ট্রান্সপারেন্সি
ইন্টারন্যাশনাল
বাংলাদেশ
দুর্নীতিবিরোধী সামাজিক আন্দোলন

ডেটা
সাংবাদিকতা সহায়িকা

ডেটা

সাংবাদিকতা
সহায়িকা

ডেটা সাংবাদিকতা সহায়িকা

প্রথম প্রকাশ : ফেব্রুয়ারি ২০২৩

পুর্নমুদ্রণ : ফেব্রুয়ারি ২০২৫

© ট্রান্সপারেন্সি ইন্টারন্যাশনাল বাংলাদেশ (টিআইবি)

সহায়িকায় উপস্থাপিত তথ্য-উপাত্ত, মন্তব্য ও ব্যাখ্যার সম্পূর্ণ দায়দায়িত্ব প্রণয়নে সংশ্লিষ্টদের ও টিআইবির।

পরিকল্পনা ও প্রণয়ন সহযোগিতা

ডিজিটালি রাইট লিমিটেড

সম্পাদনা ও প্রকাশনা সহায়তা

আউটরিচ অ্যান্ড কমিউনিকেশন বিভাগ, টিআইবি

প্রচ্ছদ অলংকরণ ও সজ্জাবিন্যাস

সামছুদ্দোহা সাফায়েত

মূল্য : ৪০০ টাকা

ট্রান্সপারেন্সি ইন্টারন্যাশনাল বাংলাদেশ
মাইডাস সেন্টার (৫ম ও ৬ষ্ঠ তলা)
বাড়ি - ০৫, সড়ক - ১৬ (নতুন) ২৭ (পুরোনো)
ধানমন্ডি, ঢাকা ১২০৯
ফোন : ৮৮০২ ৪১০২১২৬৭-৭০
ফ্যাক্স : ৮৮০২ ৪১০২১২৭২
ই-মেইল : info@ti-bangladesh.org
ওয়েবসাইট : www.ti-bangladesh.org
ফেসবুক : www.facebook.com/TIBangladesh

ISBN : 978-984-35-3808-6

মুখবন্ধ

ট্রান্সপারেন্সি ইন্টারন্যাশনাল বাংলাদেশ (টিআইবি) বাংলাদেশে দুর্নীতির বিরুদ্ধে একটি কার্যকর ও টেকসই সামাজিক আন্দোলন গড়ে তুলতে কাজ করছে। এ উদ্দেশ্য অর্জনে টিআইবি জাতীয় ও স্থানীয় পর্যায়ে গবেষণা এবং অ্যাডভোকেসির পাশাপাশি অংশীজনদের সক্ষমতা ও দুর্নীতিবিরোধী সামাজিক আন্দোলনে নাগরিক সম্পৃক্ততা বৃদ্ধির লক্ষ্যে বিভিন্ন কার্যক্রম পরিচালনা করছে। এই সামগ্রিক কার্যক্রমে অন্যতম অংশীজন ও সহযাত্রী হিসেবে দুর্নীতিবিষয়ক তথ্য প্রকাশ ও অনুসন্ধানী সাংবাদিকতায় গণমাধ্যমকর্মীদের সক্ষমতা বৃদ্ধির লক্ষ্যে টিআইবি ধারাবাহিকভাবে বিভিন্ন ধরনের প্রশিক্ষণ ও ফেলোশিপ প্রদানসহ নানা কার্যক্রম পরিচালনা করছে। এ ছাড়া ১৯৯৯ সাল থেকে দুর্নীতিবিরোধী প্রতিবেদন প্রণয়নে গণমাধ্যমকর্মীদের কাজের স্বীকৃতি ও উৎসাহ প্রদানের অংশ হিসেবে টিআইবি প্রতিবছর দুর্নীতিবিরোধী অনুসন্ধানী সাংবাদিকতা পুরস্কার প্রদান করছে।

বৈশ্বিক প্রেক্ষিতে ডেটা সাংবাদিকতা অনেকটাই এগিয়ে গেছে। সেই তুলনায় বাংলাদেশ ও বাংলা ভাষাভাষী অঞ্চলে ডেটা সাংবাদিকতা ততটা এগোয়নি। কিছু পত্রিকায় ডেটাভিত্তিক প্রতিবেদন হয়েছে, তা মূলত একক ও বিচ্ছিন্ন উদ্যোগ হিসেবে এসেছে, কিন্তু সার্বিক কোনো প্রাতিষ্ঠানিক কাঠামো বা সুনির্দিষ্ট পেশাগত উৎকর্ষ উন্নয়নের উদ্যোগ দেখা যায় না। প্রযুক্তির ব্যবহার করে বড় আকারের ডেটাসেট নিয়ে বিশ্লেষণ করার মতো দক্ষতা দেশের সাংবাদিকতার ক্ষেত্রে তৈরি হয়নি, একইসাথে ডেটা বিশ্লেষণে সক্ষম ব্যক্তিদের সাংবাদিকতায় সম্পৃক্ত করার প্রয়াসও তেমন দেখা যায় না। ডিজিটাল বাংলাদেশের অন্যতম প্রত্যাশিত প্রাপ্তির ক্ষেত্র উন্মুক্ত ডেটা; এক্ষেত্রে বাংলাদেশের অগ্রগতি ধীর হলেও ক্রমবর্ধমান অগ্রগতি লক্ষণীয়। আর দেশের সেবাখাতসহ সামগ্রিক উন্নয়ন প্রক্রিয়া যত তথ্য-প্রযুক্তিনির্ভর হবে, তত বেশি ডেটা তৈরি হবে, উন্মুক্ত তথ্যের পরিসরও বাড়বে। এই ডেটা কীভাবে সাংবাদিকতায় অনুসন্ধানী ও বিশ্লেষণমূলক প্রতিবেদন তৈরিতে পেশাদারিত্বের সঙ্গে ব্যবহার নিশ্চিত করা যায়, সেটিই এখন গুরুত্ব দিয়ে ভাবার সময়।

গোটা বিশ্বেই স্বচ্ছতা ও জবাবদিহি প্রতিষ্ঠায় বিশেষ করে দুর্নীতি, অর্থপাচার, বিদেশে অবৈধ সম্পদ অর্জন এবং রাজস্ব ফাঁকির ঘটনা উদ্‌ঘাটনে ডেটা সাংবাদিকতা বড় অবদান রাখছে। কেননা, ডেটাভিত্তিক প্রতিবেদন অনেক বেশি সুনির্দিষ্ট ও প্রায় অকাট্য। একইসাথে ডেটা সাংবাদিকতা নতুন নতুন অনুসন্ধানের ক্ষেত্র তৈরি করে জবাবদিহি ও বস্তুনিষ্ঠতার সুযোগকে আরও সমৃদ্ধ করে। যার বড় উদাহরণ ২০১৫ সালে ইন্টারন্যাশনাল কনসোর্টিয়াম অব ইনভেস্টিগেটিভ জার্নালিস্টস (আইসিআইজে) প্রকাশিত পানামা পেপারস, পরবর্তী সময়ে পেগাসাস পেপারস, প্যান্ডোরা পেপারস ইত্যাদি। যার মাধ্যমে কোটি কোটি তথ্যকে বিশ্লেষণ করে দেশে দেশে গভীরতর অনুসন্ধানের মাধ্যমে ক্ষমতার অপব্যবহার ও দুর্নীতি, বিশেষকরে অর্থপাচারের গোপন এক অধ্যায়কে উদ্‌ঘাটন করা সম্ভব হয়েছে। যদি সাংবাদিকতার লক্ষ্য হয় তথ্যের অবাধ প্রকাশ, স্বচ্ছতা ও জবাবদিহি নিশ্চিত করা এবং জনগণকে বস্তুনিষ্ঠ ও অকাট্য তথ্য প্রদান করা, তাহলে ডেটা সাংবাদিকতা সেই প্রচেষ্টায় অন্যতম সময়োপযোগী হাতিয়ার।

এ প্রেক্ষিতে টিআইবি ডেটা সাংবাদিকতার বিকাশে বাংলা ভাষায় এবং স্থানীয় প্রেক্ষিত বিবেচনায় সমন্বিত, যুগোপযোগী ও তথ্যসমৃদ্ধ সহায়িকা প্রস্তুত করার কাজ হাতে নেয়। তারই অংশ হিসেবে এই সহায়িকাটি বাংলায় ডেটা সাংবাদিকতা বিষয়ে প্রয়োজনীয় শিক্ষা উপকরণের অভাবকে কিছুটা হলেও পূরণ করবে।

সহায়িকাটির পরিকল্পনা, রচনা ও সম্পাদনাসহ সার্বিক সমন্বয়ের জন্য বিশেষভাবে কৃতজ্ঞতা প্রকাশ করছি ''ডিজিটালি রাইট'' ও এর ব্যবস্থাপনা পরিচালক মিরাজ আহমেদ চৌধুরীর প্রতি। ধন্যবাদ জানাচ্ছি মিডিয়া হ্যাক কালেক্টিভের সহ-প্রতিষ্ঠাতা ও আউটলায়ার সম্পাদক অ্যালাস্টেয়ার ওটার এবং ইনভেস্টিগেটিভ রিপোর্টার্স অ্যান্ড এডিটরস (আইআরই)- এর গ্রাজুয়েট রিসার্চ অ্যাসিস্ট্যান্ট সৌরভ রহমানকে সহায়িকাটি রচনার জন্য। এ ছাড়া ডেটা সাংবাদিকতায় নিজেদের অভিজ্ঞতা জানানোর জন্য কৃতজ্ঞতা প্রকাশ করছি প্রথম আলোর পরামর্শক কুর্‌রাতুল আইন তাহমিনা এবং মোঙ্গাবের কন্ট্রিবিউটিং এডিটর আবু সিদ্দিকের প্রতি। বইটি রিভিউ করে দেওয়ায় বিশেষ ধন্যবাদ ডেটাফুলের প্রতিষ্ঠাতা পলাশ দত্তের প্রতি। এ ছাড়া সহায়িকাটি প্রণয়ন ও প্রকাশে সার্বিক উদ্যোগ গ্রহণ এবং তার বাস্তবায়নে অনন্য ভূমিকার জন্য টিআইবির আউটরিচ অ্যান্ড কমিউনিকেশন বিভাগের প্রাক্তন পরিচালক শেখ মনজুর-ই-আলম ও বর্তমান পরিচালক মোহাম্মদ তৌহিদুল ইসলামসহ অন্যান্য সহকর্মীদের প্রতি আন্তরিক অভিনন্দন জানাচ্ছি।

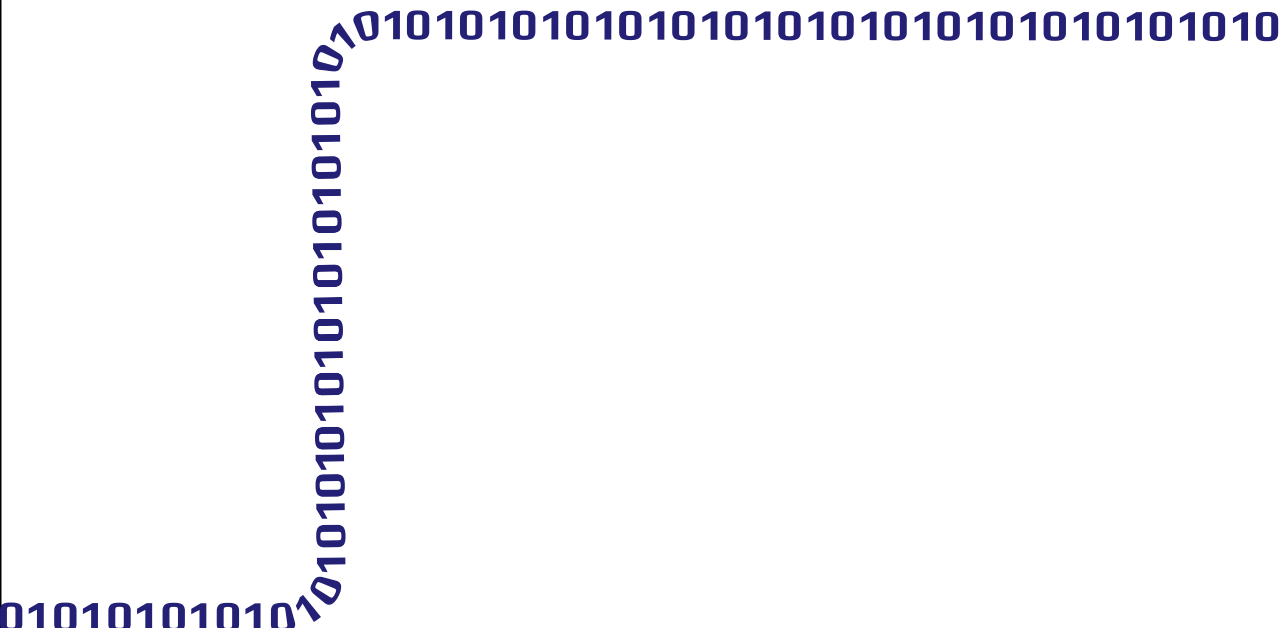
প্রথম প্রকাশের পর গনমাধ্যমকর্মী, শিক্ষার্থী ও সাধারণ পাঠকের ব্যাপক চাহিদার ভিত্তিতে সহায়িকাটি পুর্নমুদ্রণ করা হলো।

আশা করি সহায়িকাটি বাংলাদেশে ডেটা সাংবাদিকতায় সার্বিকভাবে নতুন দিগন্ত উন্মোচন করবে এবং গণমাধ্যমকর্মী তথা সংশ্লিষ্ট অংশীজনদের ডেটানির্ভর অনুসন্ধানী সাংবাদিকতা চর্চায় সক্ষমতা বৃদ্ধিতে কার্যকর ভূমিকা রাখবে। ডেটা সাংবাদিকতাবিষয়ক টিআইবির কার্যক্রম এবং এই সহায়িকা সংক্রান্ত যে কোনো সুচিন্তিত মন্তব্য ও গঠনমূলক পরামর্শকে আমরা স্বাগত জানাই।

ইফতেখারুজ্জামান
নির্বাহী পরিচালক

সূচিপত্র

অধ্যায় এক	:	পরিপ্রেক্ষিত	০৮
অধ্যায় দুই	:	মৌলিক ধারণা	১৮
অধ্যায় তিন	:	যত ধাপ	৩৪
অধ্যায় চার	:	ডেটা সংগ্রহ	৪২
অধ্যায় পাঁচ	:	তথ্য পেতে আরটিআই	৭৬
অধ্যায় ছয়	:	ক্লিনিং ও সংরক্ষণ	৮৬
অধ্যায় সাত	:	বিশ্লেষণ	১১৪
অধ্যায় আট	:	ডেটা থেকে গল্প	১৫২
অধ্যায় নয়	:	চিত্রায়ণ	১৭৬
অধ্যায় দশ	:	নৈতিকতা	১৮৮
অধ্যায় এগারো	:	কীভাবে ডেটা সাংবাদিক হবেন?	১৯৬



অধ্যায় ০১

পরিপ্রেক্ষিত

পরিপ্রেক্ষিত

> প্রতিদিন নতুন নতুন ডেটা মানে, প্রতিদিন নতুন নতুন স্টোরির সম্ভাবনা। এই ডেটা বার্তাকক্ষের আচরণকেই বদলে দেবে। এজন্য বেশি দিন অপেক্ষা করতে হবে না। এটি কোনো নির্দিষ্ট বিট নয়, বরং প্রতিটি বিটের জন্য ডেটা হয়ে উঠবে এবং অনেক দেশে এরইমধ্যে হয়ে উঠেছে সাংবাদিকতার অন্যতম উপকরণ।

প্রতিদিন বিশ্বে প্রায় ২.৫ টেরাবাইট ডেটা ইন্টারনেটে আসছে। এই হিসেব চার বছর আগের। ধরা হচ্ছে, ২০২৫ সাল নাগাদ প্রতিদিন ৪৬৩ এক্সাবাইট ডেটা অবমুক্ত হবে।[১] আপনি যদি এক্সাবাইট শব্দটির সঙ্গে অপরিচিত হন, তাহলে জেনে নিন এক এক্সাবাইটে যে পরিমাণ ডেটা আঁটে, তা ধারণ করতে ২ কোটির বেশি ডিভিডি লাগবে। আমরা খেতে খেতে, হেঁটে যেতে যেতে, খবর পড়তে পড়তে, কথা বলতে বলতে, মেসেজ পাঠাতে পাঠাতে, ছবি দেখতে দেখতে এমনকি ঘুমাতে ঘুমাতেও প্রতিনিয়ত ডেটা তৈরি করে যাচ্ছি। অনলাইনে যা অর্ডার করছেন, আপনার চলাচলের প্রতিটি জিপিএস পয়েন্ট, স্মার্টফোনের হেলথ অ্যাপে ঘুমানো বা হাঁটার যে রেকর্ড দেখছেন, ফেসবুক বা টুইটারে প্রতিটি লাইক বা পোস্ট এর সবই ডেটা। ইন্টারনেটে সংযুক্ত আমাদের প্রতিটি মুহুর্ত অবিরাম ডিজিটাল ডেটা তৈরি করে যাচ্ছে। আমরা ডেটার দুনিয়ায় বাস করছি। ডেটা আমাদের ব্যক্তি-অভ্যাস, অর্থনীতি, সমাজজীবন সব কিছুকেই প্রভাবিত করছে।

ডেটা কী এ নিয়ে পরে আরো আলোচনা হবে। কিন্তু, এখন মোটা দাগে বলা যায়, ডেটা আসলে কী নয়? সব কিছুই ডেটা দিয়ে দেখা, বোঝা এবং ব্যাখ্যা করা যায়। এ তো গেল ইন্টারনেটে পাওয়া ডিজিটাল ডেটার কথা। একবার ভাবুন, প্রতিদিন প্রতিটি অফিসে কত নথি তৈরি হচ্ছে, তাতে কত তথ্য থাকছে? প্রযুক্তি যতই এগুচ্ছে আমাদের নথিগুলোও ততোই ডিজিটাল হয়ে উঠছে। একজন সাংবাদিক হিসেবে, আগে সোর্সের কাছ থেকে একটি কাগজের কপি আনতে কত বেগ পেতে হতো। এখন ফাইলে থাকা কাগজটির ছবি তুলে নিলেই হচ্ছে। অনেক ক্ষেত্রে যেতেও হচ্ছে না; আপনার ইমেইলে, হোয়াটসঅ্যাপে বা মেসেঞ্জারেই সোর্স নথি পাঠিয়ে দিচ্ছেন। এখন এমন সাংবাদিক পাওয়া মুশকিল, যার মোবাইলে কোনও নথি বা সরকারি পরিসংখ্যানের ইমেজ বা পিডিএফ নেই। সেই নথিগুলোকেই আপনি যখন একটি ওসিআর বা অপটিক্যাল ক্যারেক্টার রিকগনিশন অ্যাপ দিয়ে টেক্সটে বদলে নিচ্ছেন, অথবা সেই নথির তথ্য নিজে কম্পিউটারের কোনো ফাইলে টুকে রাখছেন সেটি তখন ডিজিটাল ডেটায় বদলে যাচ্ছে।

---

[১] *How much data is generated each day?* (2020c, February 8). World Economic Forum. Retrieved October 12, 2022, from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

আমাদের সামগ্রিক ব্যবস্থাপনা যত প্রযুক্তিনির্ভর হচ্ছে, ততো বেশি তথ্য উন্মুক্ত হচ্ছে। আগে যেসব তথ্যের জন্য দিনভর কোনো অফিসে গিয়ে বসে থাকতে হতো, এখন সেটি হয়তো নিয়মিত ওয়েবসাইটে পাওয়া যাচ্ছে। দেশে প্রশাসনিক স্বচ্ছতা যত বাড়বে, ডেটাও ততো বেশি হারে উন্মুক্ত হতে থাকবে। এর কোনো বিকল্প নেই। এই পরিবর্তন কোথাও আগে শুরু হয়েছে, কোথাও দ্রুত হচ্ছে, কোথাও ধীর গতিতে। কিন্তু এটি অমোঘ, অবশ্যম্ভাবী। প্রশ্ন হলো, এই পরিবর্তনের সঙ্গে তাল মেলাতে একজন সাংবাদিক হিসেবে আমরা কতটা তৈরি?

প্রতিদিন নতুন নতুন ডেটা মানে, প্রতিদিন নতুন নতুন স্টোরির সম্ভাবনা। এই ডেটা বার্তাকক্ষের আচরণকেই বদলে দেবে। এ জন্য বেশি দিন অপেক্ষা করতে হবে না। এটি কোনো নির্দিষ্ট বিট নয়, বরং প্রতিটি বিটের জন্য ডেটা হয়ে উঠবে এবং অনেক দেশে এরইমধ্যে হয়ে উঠেছে সাংবাদিকতার অন্যতম উপকরণ।

ডেটা সাংবাদিকতা, বৈশ্বিক প্রেক্ষিত ও বাংলাদেশ

ডেটা সাংবাদিকতার সংজ্ঞা, ধরন ও প্রথাগত সাংবাদিকতার সঙ্গে এর পার্থক্য নিয়ে পরের অধ্যায়ে আরো আলোচনা থাকছে। তবে মোটাদাগে এর মানে হচ্ছে যেখানে ডেটা-ই আপনার প্রতিবেদনে তথ্যকে উপস্থাপনের মূল উৎস বা টুল। ২০১৭ সালে গুগল নিউজ ইনিশিয়েটিভ ডেটা সাংবাদিকতার বৈশ্বিক হালচাল নিয়ে একটি প্রতিবেদন প্রকাশ করেছিল।[২] তাতে দেখা যায়, যুক্তরাষ্ট্র ও ইউরোপের অর্ধেক বার্তাকক্ষেই অন্তত একজন ডেটা সাংবাদিক আছেন। যাদের ডেটা সাংবাদিকতার দক্ষতা যত বেশি, তারা সাংবাদিক হিসেবে ততো বেশি আয় করছেন। কোডিং থেকে শুরু করে ভিজ্যুয়ালাইজেশন সবখানেই নতুন নতুন টুল আসছে, যা ডেটা সাংবাদিকতাকে আরো সহজ এবং কার্যকর করে তুলছে। ২০২১ সালে ডেটাজার্নালিজম ডট কম[৩] একই ধরনের আরেকটি জরিপ পরিচালনা করে। তাতেও দেখা যায়, বার্তাকক্ষে সেই সাংবাদিকের আয় সবচেয়ে বেশি যার বহুমাত্রিক ডেটা দক্ষতা আছে।

> **পর্তুগালের সাংবাদিক রুই বারোস, তার দেশে ডেটা সাংবাদিকতার অন্যতম প্রবক্তা। তিনি বলেছেন,[৪] "সাংবাদিকতার জগতে এখন তারাই সোনার হরিণ, যারা কোডিং জানেন। এদের বলা হয় জার্নোকোডার। এরা সেই বিরল প্রজাতির সাংবাদিক যাদের রিপোর্টিং দক্ষতা আছে, আবার প্রোগ্রামিং এবং ডেটা বিশ্লেষণেও পটু। এই বিশেষ দক্ষতাগুলোর মিশেলে, তারা হয়ে ওঠেন যে কোনো নিউজরুমের জন্যই আরাধ্য।"**

---

২ Rogers, S. (2017c, September 18). The state of data journalism in 2017. Google. Retrieved October 12, 2022, from https://blog.google/outreach-initiatives/google-news-initiative/data-journalism-2017/

৩ DataJournalism.com. (2022d, February 16). State of Data Journalism Survey 2021: 11 surprising findings. Retrieved October 12, 2022, from https://datajournalism.com/read/blog/data-journalism-survey-2021

৪ যদি হতে চান সাংবাদিকতার "সোনার হরিণ" (2019c, December 20). Global Investigative Journalism Network. Retrieved October 12, 2022, fromhttps://gijn.org/2019/09/03/%E0%A6%B8%E0%A6%BE%E0%A6%82%E0%A6%AC%E0%A6%BE%E0%A6%A6%E0%A6%BF%E0%A6%95%E0%A6%A4%E0%A6%BE%E0%A6%B0-%E0%A6%B8%E0%A7%8B%E0%A6%A8%E0% A6%BE%E0%A6%B0-%E0%A6%B9%E0%A6%B0%E0%A6%BF%E0%A6%A3/

বিশেষকরে পশ্চিমা বিশ্বে, ডেটা সাংবাদিকতা অনেক এগিয়ে গেছে। বড় আকারের ডেটাকে বিশ্লেষণ ও চিত্রায়ণ করতে গিয়ে তারা মাইক্রোসফট এক্সেল বা গুগল স্প্রেডশিটের মতো টুলের গণ্ডি পেরিয়ে এখন হরহামেশাই পাইথন ও "আর" এর মতো প্রোগ্রামিং ল্যাংগুয়েজ ব্যবহার করছেন; ফ্লোরিশ বা ট্যাবিলিউর মতো টুল কাজে লাগাচ্ছেন। মজার বিষয় হলো, ২০২১ সালের জরিপটিতে ৭০ শতাংশ সাংবাদিকরাই বলেছেন, তারা ডেটা সাংবাদিকতার কৌশলগুলো শিখেছেন নিজের চেষ্টায়, অনলাইন বা অফলাইন রিসোর্স পড়ে বা দেখে। তাদের কারোরই ডেটা সাংবাদিকতায় প্রাতিষ্ঠানিক শিক্ষা নেই।

ডেটা সাংবাদিকতা যে বিটের গণ্ডি ছাড়িয়ে কতটা অভিনব হয়ে উঠেছে তার নজির মিলবে এই উদাহরণে

২০২১ সালে কোভিড মহামারির সময় সাংবাদিক সাইমন হিউওয়াইলার, তার নিজ দেশ সুইজারল্যান্ডে এই মহামারির প্রতিটি দিন এবং একেক দিনে যত মানুষ মারা গেছেন তার প্রতিটি ডেটাকে একটি মিউজিক বক্স ও চার-মিটার পাঞ্চ কার্ডে সাজিয়ে একটি মিউজিক তৈরি করেন। লেখার পরের অংশে যাবার আগে কিউআর কোডটি স্ক্যান করে তার সেই সাউন্ড ভিজ্যুয়ালাইজেশন শুনে নিতে পারেন।

ইউটিউবে হিউওয়াইলারের "সং অন ক্রাউনস অ্যান্ড টিয়ার্স" শুনতে স্ক্যান করুন

ডেটা সাংবাদিকতার ব্যাপকতা সম্পর্কে আরো পরিচিত উদাহরণ হতে পারে, ইন্টারন্যাশনাল কনসোর্টিয়াম অব ইনভেস্টিগেটিভ জার্নালিস্টসের (আইসিআইজে) নেতৃত্বে পরিচালিত প্যান্ডোরা পেপারস অনুসন্ধান।[৫]

> **১৫০টি গণমাধ্যম প্রতিষ্ঠানের ৬০০ জনের বেশি সাংবাদিক মিলে ফাঁস হওয়া ৬৪ লাখ নথি, লাখ লাখ ছবি ও ইমেইল এবং প্রায় পাঁচ লাখ স্প্রেডশিট ব্যবহার করে অফশোর দুনিয়ায় মুদ্রাপাচার ও দুর্নীতির এক বৈশ্বিক চিত্র উন্মোচন করেন।**

[৫] Monteiro, C. (2022c, October 3). Pandora Papers. ICIJ. Retrieved October 12, 2022, from https://www.icij.org/investigations/pandora-papers/

বাংলাদেশের গণমাধ্যমে ডেটা সাংবাদিকতা

সেই তুলনায় বাংলাদেশ ও বাংলা ভাষাভাষী অঞ্চলে ডেটা সাংবাদিকতা ততোটা এগোয়নি। মূলধারার গণমাধ্যমগুলোতে ডেটা সাংবাদিক বলে কিছু নেই। বিচ্ছিন্নভাবে কিছু পত্রিকায় ডেটাভিত্তিক প্রতিবেদন হয়েছে (কিছু কেস স্টাডি পরবর্তী অধ্যায়গুলোতে থাকছে), ডেটাফুলের মতো একক ক্ষুদ্র উদ্যোগ এসেছে, কিন্তু সার্বিক কোনো কাঠামো দাঁড়ায়নি।

বাংলাদেশে ডেটা প্রাপ্তিতে সীমাবদ্ধতা আছে, বাংলা ভাষা উপযোগী প্রযুক্তিগত টুলের ঘাটতি আছে–কিন্তু এই বাস্তবতা আরো অনেক দেশের ক্ষেত্রেই প্রযোজ্য। এর চেয়ে বড় চ্যালেঞ্জ হলো, বড় আকারের একটি ডেটাসেট নিয়ে তাকে এক্সেল বা অন্য কোনো টুলে ফেলে, নিজস্ব বিশ্লেষণ দাঁড় করানোর মতো দক্ষতা দেশের বেশিরভাগ সাংবাদিকেরই গড়ে ওঠেনি।

উন্মুক্ত ডেটা প্রাপ্তির ক্ষেত্রে বাংলাদেশে অগ্রগতি হচ্ছে। চিন্তা করে দেখুন, আগের যে কোনো সময়ের চেয়ে সরকারি ও বেসরকারি ওয়েবসাইটে এখন কত বেশি তথ্য ও নথি পাওয়া যাচ্ছে। কিন্তু একটি বিষয় ধরে উন্মুক্ত প্রতিটি নথি সংগ্রহ করে তার ডেটাভিত্তিক বিশ্লেষণের চর্চাটি এখনো সমৃদ্ধ হয়নি। একটি উদাহরণ দিলে বিষয়টি আরও পরিষ্কার হবে। গত কয়েক বছরে সরকারি কর্মকর্তাদের বিদেশ সফর নিয়ে অসংখ্য স্টোরি হয়েছে গণমাধ্যমে। এখন এ সংক্রান্ত সরকারি আদেশের কপিগুলো বেশিরভাগ মন্ত্রণালয়ের ওয়েবসাইটেই পাওয়া যায়। কিন্তু এমন সাংবাদিকতা কী চোখে পড়েছে, যেখানে কোনো গণমাধ্যম প্রতিটি মন্ত্রণালয়ের তিন বা চার বছরের বিদেশ সফর-সংক্রান্ত সরকারি আদেশ সংগ্রহ করে তাকে স্প্রেডশিটে সাজিয়ে মন্ত্রণালয়ের তুলনা বা খরচের একটি সার্বিক চিত্র তুলে ধরে প্রতিবেদন করেছে? ডেটাকে নিয়ে বড় পরিসরের এই ভাবনাটাই এখনো অনুপস্থিত।

এখানে পরিসংখ্যান নিয়ে সাংবাদিকতা ও বিশ্লেষণ অনেক হয়, কিন্তু বড় ধরনের ডেটাকে প্রযুক্তির সাহায্যে বোধগম্য বিশ্লেষণ ও আকর্ষণীয় চিত্রায়ণের বিষয়টি এখনো দেখা যায় না। ডেটা বিশ্লেষণ করে চার্ট বা গ্রাফের মাধ্যমে তাকে উপস্থাপনের চর্চা কিছু গণমাধ্যমে চোখে পড়ে। কিন্তু ওয়েবসাইটে এসব বিশ্লেষণের ইন্টারঅ্যাকটিভ চিত্রায়ণ, মানচিত্রায়ণ বা এক্সপ্লেইনারের মতো গতিশীল উপস্থাপন কতটা দেখেছেন? চট করে মনে করার মতো উদাহরণ নেই। ডেটার ইন্টারঅ্যাকটিভ উপস্থাপনার ক্ষেত্রে অনলাইন যে আশীর্বাদ বয়ে এনেছে সেটি কাজে লাগানোর সুযোগ রয়ে গেছে।

একটি বিষয় মনে রাখা দরকার

ডেটা সাংবাদিকতা কোনো বিট নয়, বরং এটি কিছু কারিগরি ও বিশ্লেষণ দক্ষতার সমষ্টি। প্রত্যেক বিটের সাংবাদিকই ডেটা সাংবাদিক হয়ে উঠতে পারেন। একটি বার্তাকক্ষের সব সাংবাদিক ডেটা সাক্ষর জ্ঞানসম্পন্ন (ডেটা লিটারেট) হবেন এমনটা নয়। কিন্তু এই দক্ষতাগুলো সাংবাদিকতার নতুন দরজা খুলে দিতে পারে প্রতিটি গণমাধ্যমের জন্য, যা প্রযুক্তি ও ইন্টারনেটের এই যুগে টিকে থাকার জন্য অপরিহার্যও বটে।

স্বচ্ছতা ও জবাবদিহি প্রতিষ্ঠায় ডেটা সাংবাদিকতার প্রভাব

গোটা বিশ্বেই স্বচ্ছতা ও জবাবদিহি প্রতিষ্ঠায় ডেটা সাংবাদিকতা বড় ধরনের অবদান রেখে চলেছে। পশ্চিমে যাবার দরকার নেই, বাংলাদেশের আশপাশে তাকালেই ডেটা সাংবাদিকতার এমন অনেক নজির মিলবে। রয়টার্স ইনস্টিটিউটের প্রকাশ করা এক সাম্প্রতিক গবেষণায়[৬] মালিহা সিদ্দিকী দেখিয়েছেন, ডেটা সাংবাদিকতা, তার দেশ পাকিস্তানে কতটা গভীর প্রভাব ফেলছে। পানামা পেপার্স প্রকল্পে উমর চিমার ডেটাভিত্তিক অনুসন্ধানের জেরে দুর্নীতির অভিযোগে দেশটির তৎকালীন প্রধানমন্ত্রী নওয়াজ শরীফের পতন হয়েছে।

> **পেশাওয়ারের একটি স্কুলে সন্ত্রাসী হামলায় ১৪৪ জন শিশুর মৃত্যুর তিন বছর পর দেশটির অন্যতম পত্রিকা ডন, "ওয়ান ফর্টিফোর স্টোরিজ" নামের এক ডেটা ভিজ্যুয়ালাইজেশনে দেখিয়েছে যে পাকিস্তানের সন্ত্রাস বিরোধী সরকারি ও সামরিক নীতি কীভাবে ব্যর্থ হয়েছে। সোশ্যাল মিডিয়ায় নিষিদ্ধ ঘোষিত সন্ত্রাসীদের তৎপরতা, কর্মক্ষেত্রে অন্তসত্ত্বা নারীদের চাকরি হারানো, এমন অনেক বিষয় ডেটা সাংবাদিকতার মাধ্যমে সামনে এনেছে দেশটির গণমাধ্যমগুলো।**

ভারতেও রাজ্যসভার সদস্যদের ব্যক্তিগত কর্মীদের তথ্য বিশ্লেষণ করে ইন্ডিয়ান এক্সপ্রেস দেখিয়েছে, ১৪৬ জন এমপি তাদের ব্যক্তিগত সহকারী হিসেবে আত্মীয়দের নিয়োগ দিয়েছেন আর এভাবে স্বার্থের সংঘাত আমলে না এনে তারা ব্যক্তিগত সহকারীর জন্য বরাদ্দ মাসিক ৩০ হাজার রুপি নিজ পরিবারের সদস্যদের দিয়েছেন। এই প্রতিবেদন প্রকাশের পর এ ধরনের চর্চা বন্ধ করতে উদ্যোগ নেয়া হয়। আমলাদের বিদেশভ্রমণের বিলাসিতা নিয়ে একই পত্রিকার আরেকটি প্রতিবেদনের জের ধরে এর লাগাম টানতে বাধ্য হয়েছিল দেশটির সরকার।

ছবি: রয়টার্স

[৬] Reuters Institute for the Study of Journalism. (n.d.-d). Reuters Institute. Retrieved October 12, 2022, from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Showing+the+way_0.pdf

২০১৯ সালে ভারতের নির্বাচনকে সামনে রেখে ওপরের ডেটা প্রতিবেদনটি[৭] তৈরি করেছিল রয়টার্স

রিপোর্টটি স্ক্রল করে নিচে যেতে যেতে দেখা যাবে নির্বাচনের যত প্রার্থী দাঁড়িয়েছেন তাদের ছবি, বৈচিত্র্য, সম্পদ ও অপরাধের চিত্র এবং এমন আরও অনেক কিছু। মাত্র একটি পেইজে, ভারতের মতো বড় একটি দেশের নির্বাচনের সামগ্রিক চিত্র কোনো দীর্ঘ বর্ণনা ছাড়াই ভোটারদের সামনে তুলে ধরা গেছে ডেটা সাংবাদিকতার কল্যাণে। আর এখানেই ডেটার চিত্রায়ণ ডেটা সাংবাদিকতার একটি অবিচ্ছেদ্য অংশ হয়ে ওঠে।

You had to scroll through 7,320 candidates to get a woman. Only 723 (8.9 percent) of all the contestants are female.

ছবি: রয়টার্স

সাংবাদিকতার লক্ষ্য হচ্ছে, ক্ষমতাকে জবাবদিহি করা এবং জনগণকে বস্তুনিষ্ঠ তথ্য দিয়ে যাওয়া যেন তারা সজ্ঞান-সিদ্ধান্ত নিতে পারেন। ডেটা সাংবাদিকতা নতুন নতুন অনুসন্ধানের ক্ষেত্র তৈরি করেছে, জবাবদিহি ও বস্তুনিষ্ঠতার সুযোগকে আরও সমৃদ্ধ করেছে এবং আগের যেকোনো সময়ের চেয়ে বেশি হারে।

ডেটা সাংবাদিকতা ও জনস্বার্থ

সব প্রতিবেদনেই যে দুর্নীতি, অনিয়ম, রাজনীতি, নায়ক-ভিলেন থাকতে হয় এমন নয়। অনেক সাংবাদিকতা শুধু মাত্র সঠিক তথ্যটি জনগণকে পৌঁছে দিয়ে জনস্বার্থের দাবিটিকে প্রতিষ্ঠিত করা যায়। আপনাকে শুধু ভাবতে হবে–যে বিষয়টি নিয়ে স্টোরি করতে যাচ্ছেন তাকে ডেটায় রূপান্তরিত কীভাবে করবেন। আবার একটি উদাহরণ দেওয়া যাক।

ভারতের হিন্দুস্তান টাইমস ১৯৫০ সাল থেকে বলিউডের ব্যবসা সফল ছবিগুলোর ২৪ হাজার গান বিশ্লেষণ করে দেখতে পেয়েছে,[৮] সেখানে একক গানে নারী শিল্পীদের কণ্ঠ দেওয়া দিনে দিনে কমে আসছে। নিচের ছবিটি দেখুন, শুধু পুরুষ শিল্পীর গাওয়া গান ও নারী শিল্পীর গাওয়া গানের শিরোনাম দিয়ে তৈরি করা একটি গ্রাফিক্স, যেখানে নারীদের তুলনায় পুরুষ শিল্পীরা দ্বিগুণ গানে কণ্ঠ দেওয়ার সুযোগ পেয়েছেন। বলিউডের গানের জগতে ক্রমবর্ধমান জেন্ডার বৈষম্য বোঝার জন্য সবুজ ও কালো রঙের পার্থক্যটাই একটি শক্তিশালী বার্তা পৌঁছে দিচ্ছে।

[৭] The figures behind the faces. (n.d.-c). Reuters. Retrieved October 12, 2022, from https://graphics.reuters.com/INDIA-ELECTION-CRIMINAL-CANDIDATES/0100925031T/index.html

[৮] Women are fading out from Bollywood music. (n.d.-b). Hindustan Times. Retrieved October 12, 2022, from https://www.hindustantimes.com/interactives/bollywood-songs-gender/

ছবি: হিন্দুস্তান টাইমস

> **বাংলাদেশের প্রেক্ষিতে একটি হাইপোথিসিসে আসা যাক। ধরুন, আপনি ঢাকা জেলার প্রতিটি থানা থেকে সেখানকার অপরাধের তথ্য সংগ্রহ করলেন। প্রথমেই মাথায় আসবে কোথায় কত অপরাধ হচ্ছে, কোথায় কোন ধরনের অপরাধ বেশি হচ্ছে, ইত্যাদি।**

ভেবে দেখুন, মামলার নথি থেকে আপনি যদি অপরাধের নির্দিষ্ট স্পটগুলোকে একটি স্প্রেডশিটে সংকলন করতে পারেন এবং সেই স্পটের জিপিএস অবস্থান দিয়ে একটি মানচিত্র তৈরি করতে পারেন, তাহলে কেমন দাঁড়াবে। এতে দুটি কাজ হবে– প্রথমত নাগরিকেরা বুঝতে পারবেন কোন কোন স্পটে বেশি অপরাধ হয়, কখন বেশি অপরাধ হয় এবং এই তথ্য তাদের সতর্ক হতে সাহায্য করবে। অন্যদিকে সেই নির্দিষ্ট স্পটগুলোতে অপরাধ নিয়ন্ত্রণে দ্রুত ব্যবস্থা নেওয়ার জন্য আইনশৃঙ্খলা রক্ষাকারী বাহিনীর ওপরও চাপ তৈরি হবে। এখানে অনেক বড় টেক্সটনির্ভর প্রতিবেদন হয়তো তৈরি হচ্ছে না, কিন্তু একটি মাত্র ইন্টারঅ্যাকটিভ মানচিত্র (যেখানে ক্লিক করলেই প্রতিটি স্পটে অপরাধের প্রবণতা সম্পর্কে তথ্য পাওয়া যায়) জনস্বার্থকে ঠিকই সামনে নিয়ে আসছে। এ ধরনের ডেটা স্টোরির আবেদন এবং প্রাসঙ্গিকতাও অনেক দিন থাকে যা স্টোরির শেলফ-ভ্যালুও বাড়িয়ে দেয়। বস্তুনিষ্ঠতা এবং একটি বিষয়কে ওপর দেখে সামগ্রিকভাবে দেখার, নতুন দৃষ্টিভঙ্গি তুলে ধরার এই ক্ষমতা ডেটা সাংবাদিকতাকে প্রভাবের বিচারে আরও কার্যকর করে তোলে।

এই সহায়িকায় যা থাকছে

আগেই বলা হয়েছে, ডেটা সাংবাদিকতা কোনো বিট নয়, বরং কিছু দক্ষতার সমষ্টি। এর চর্চা শুরু করতে যে প্রাথমিক দক্ষতা ও বোঝাপড়ার প্রয়োজন তার সঙ্গেই সাংবাদিকদের পরিচয় করিয়ে দেবে এই সহায়িকা। বলে নেয়া ভালো, সহায়িকাটি তাদের কথা মাথায় রেখে তৈরি যারা এই বিষয়টি নিয়ে আগ্রহী, অথবা নিজে থেকে কিছুটা চেষ্টা করছেন, অথবা আপনার পুরোনো দক্ষতাগুলোকে ঝালিয়ে নিতে চান। আপনি হতে পারেন সাংবাদিক, শিক্ষার্থী, সাংবাদিকতার শিক্ষক বা স্বচ্ছতা প্রতিষ্ঠায় নিয়োজিত একনিষ্ঠকর্মী। এটি পেশাদার ডেটা সাংবাদিকদের জন্য নয়। সহায়িকাটির উদ্দেশ্যই হলো নতুনদের বিষয়টিতে আগ্রহী করে তোলা।

বিশ্বে ডেটা সাংবাদিকতা অনেক দূর এগিয়েছে। বড় বা মাঝারি গণমাধ্যমগুলোতে ডেটা সাংবাদিকতার দলও গড়ে উঠছে। বিশেষ করে পশ্চিমের স্কুল-কলেজ পাঠ্যসূচিতে প্রোগ্রামিং অন্তর্ভুক্ত করার পর থেকে যে প্রজন্ম পড়ালেখা শেষ করে সাংবাদিকতায় এসেছে, তারা সেসব দেশগুলোতে ডেটা সাংবাদিকতার বাঁকবদল করে দিয়েছে। এখন অনেক রিপোর্টারই মিলবে যারা ডেটা সাংবাদিকতায় নিয়মিত প্রোগ্রামিং দক্ষতা কাজে লাগাচ্ছেন; ছবি, ভিডিও, গান বা স্যাটেলাইট ইমেজারিকে ডেটা বানিয়ে নিত্য নতুন উদ্ভাবনী রিপোর্ট করছেন। আগে কঠিন মনে হতো, বা যেসব কাজের জন্য বিশেষজ্ঞদের কাছে যেতে হতো প্রযুক্তির কল্যাণে সেসব কাজ এখন সাংবাদিকরাই করছেন।

কিন্তু ডেটা সাংবাদিকতার সেই অ্যাডভান্সড বা অগ্রসর বিষয়গুলো এই সহায়িকায় থাকছে না। বরং এখানে মনোযোগ দেওয়া হয়েছে সংজ্ঞা থেকে শুরু করে পকিল্পনা, ডেটা সংগ্রহ, বিশ্লেষণ ও চিত্রায়ণসহ মৌলিক ধারণা, পদ্ধতি ও টুলগুলোর ওপর। যেমন– পাইথন বা এসকিউএলের মতো প্রোগ্রামিং শিক্ষায় না গিয়ে এখানে নজর দেওয়া হয়েছে স্প্রেডশিটে। কারণ এখান থেকেই ডেটা সাংবাদিকতার শুরু।

আশা হলো, যারা মৌলিক বিষয়গুলোকে চর্চা করে দক্ষতার একটি পর্যায়ে পৌঁছাবেন, তারা ধীরে ধীরে উন্নততর পদ্ধতি বা কৌশলগুলো শিখতে থাকবেন ও ডেটা সাংবাদিকতাকে এগিয়ে নেবেন।

এই সহায়িকায় অনেক উদাহরণ আছে, কেস স্টাডি আছে এবং কারিগরি দক্ষতার জায়গাটিকে প্রায়োগিকভাবে তুলে ধরার চেষ্টা আছে। আপনি পড়তে পড়তে বাইরের অনেক রিসোর্স থেকে ঘুরে আসতে পারবেন এবং পড়ার পাশাপাশি চর্চাও করতে পারবেন। সহায়িকাটির উদ্দেশ্য কিছু মৌলিক দক্ষতার সঙ্গে পরিচয় করানো, যা ডেটা সাংবাদিক হতে সাহায্য করবে। বাদবাকি পুরোটাই আপনার চেষ্টা।

010101010101010101010101010101010101010101010101010101010101010101010101010101010101010101

অধ্যায় ০২

মৌলিক ধারণা

মৌলিক ধারণা

> আমাদের চারপাশের সবকিছুকেই ডেটা দিয়ে ব্যাখ্যা করা যায়। প্রশ্ন হলো আপনি তাকে কীভাবে ডিজিটাইজড তথ্যে রূপান্তরিত করছেন। এর আগে আমরা দেখিয়েছি, কোভিডের প্রতিটি দিন ও মৃত্যুকে কীভাবে ডটে রূপান্তরিত করে একটি সাউন্ড ভিজ্যুয়ালাইজেশন তৈরি করা হয়েছে।

ডেটা সাংবাদিকতার সংজ্ঞা এক অর্থে খুব সাদামাটা, যেখানে ডেটাকে সংবাদের প্রধান সোর্স হিসেবে ব্যবহার করা হয়। কিন্তু ডেটা আসলে কী, প্রতিবেদনে কী পরিমাণ ডেটার কথা বলা হচ্ছে, সেই ডেটা বিশ্লেষণে কোন প্রযুক্তি ব্যবহার হচ্ছে এবং সেটিকে কীভাবে উপস্থাপন করা হচ্ছে– এমন প্রশ্নগুলো সেই সহজ সংজ্ঞাটিকেই কিছুটা অস্পষ্ট ও জটিল করে তোলে। কারণ যুগ যুগ ধরে সাংবাদিকেরা তাদের প্রতিবেদনে পরিসংখ্যান ব্যবহার করছেন, চার্ট বা গ্রাফও অনেক প্রতিবেদনেই দেখা যায়। তাহলে ডেটা সাংবাদিকতা আর নতুন কী? এই অস্পষ্টতার জবাব পেতে হলে কিছু বিষয়ে মৌলিক ধারণা প্রয়োজন। শুরু করা যাক ডেটার সংজ্ঞা থেকে।

ডেটা কী?

প্রথম অধ্যায়েই কিছুটা ধারণা পেয়েছেন যে ডেটা আসলে কত ব্যাপক একটি বিষয়। ডেটা নিয়ে যুক্তরাজ্যভিত্তিক দ্য ব্যুরো অব ইনভেস্টিগেটিভ জার্নালিজম (টিবিআইজে)[৯] যা মনে করে সেটি অস্পষ্টতা কাটাতে সহায়ক হতে পারে: "ডেটা মানে নিছক স্প্রেডশিটে সাজানো সংখ্যা নয়। আমাদের কাছে ডেটা মানে ডিজিটাইজড তথ্য। একেকটি জনগোষ্ঠীর বয়ান।"

অর্থ্যাৎ, পরিসংখ্যানও ডেটা, তবে কেবল পরিসংখ্যানই ডেটা নয়। ডেটা মানে মানুষও, যখন আপনি তার অভিজ্ঞতা ও বয়ানকে ডিজিটাইজ করে বিশ্লেষণ উপযোগী করে তুলছেন। একটা উদাহরণ দেওয়া যাক। আপনার কাছে মানুষের নামের বিশাল একটি তালিকা আছে। আপাতদৃষ্টিতে সেগুলো শুধুই নাম। মনে হতে পারে, তা ডেটা নয়। কিন্তু আপনি যখন গুনবেন, সেখানে কতগুলো মানুষের নাম আছে, অথবা সেখানে সৈয়দ, চক্রবর্তী, বড়ুয়া, চৌধুরী বা কোন বংশের নাম কত বার এসেছে, বেশিরভাগ নাম শুরু হচ্ছে কী দিয়ে, বেশিরভাগ নাম শেষ হচ্ছে কী দিয়ে, বাংলা নাম কত, আরবি থেকে আসা নাম কতগুলো–এখান থেকে আপনি বাংলাদেশের একটি নামের মানচিত্র তৈরি করতে পারবেন। ধরা যাক সেখানে ৫০ হাজার নামের একটি তালিকা আছে।

[৯] What is data journalism? (n.d.-b). The Bureau of Investigative Journalism (en-GB). Retrieved October 12, 2022, from https://www.thebureauinvestigates.com/explainers/what-is-data-journalism

বিশ্লেষণের জন্য আপনি যখন তাদের বংশ, নামের শুরু, শেষ বা ভাষাগত উৎস বিচার করতে চাইবেন সেই নামগুলোই তখন হয়ে উঠবে ডেটা।

ধরা যাক, জাতীয় সংসদের একটি অধিবেশনে সংসদ সদস্যরা যত কথা বলেছেন তার একটি ট্রান্সক্রিপ্ট আপনার হাতে এসেছে। আপনি যদি প্রযুক্তি ব্যবহার করে খুঁজে বের করেন সেই বক্তব্যে কোন শব্দটি সবচেয়ে বেশিবার এসেছে, সংসদ সদস্যরা কোন ইস্যুতে কত মিনিট কথা বলেছেন–তখন দেখবেন অসংখ্য বক্তব্য আসলে ডেটার পরিণত হচ্ছে।

বিভিন্ন উৎস থেকে রূপান্তরিত ডেটাকে মূলত দুই ভাগে ভাগ করা হয়ে থাকে

১ **গুণগত ডেটা :** যে ডেটা কোনো কিছুর গুণ ও মানকে বোঝায়। যেমন, রঙের বর্ণনা, অনুভূতি, অভিজ্ঞতা, সাক্ষাৎকার, ইত্যাদি।

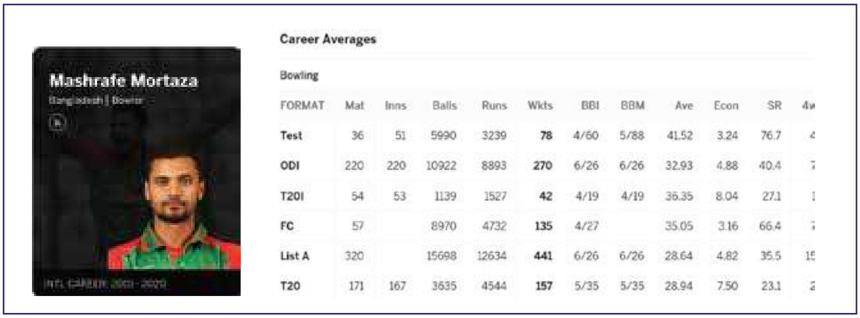
২ **পরিমাণগত ডেটা :** যে ডেটা পরিমাপ করা যায়। কোনো কিছুর সংখ্যা, আকার, দাম, ইত্যাদি।

আমাদের চারপাশের সবকিছুকেই ডেটা দিয়ে ব্যাখ্যা করা যায়। প্রশ্ন হলো আপনি তাকে কীভাবে ডিজিটাইজড তথ্যে রূপান্তরিত করছেন। এর আগে আমরা দেখিয়েছি, কোভিডের প্রতিটি দিন ও মৃত্যুকে কীভাবে ডটে রূপান্তরিত করে একটি সাউন্ড ভিজ্যুয়ালাইজেশন তৈরি করা হয়েছে। ২৪ হাজার গান থেকে শুধু কণ্ঠদাতাদের নাম নিয়ে একটি ডেটা সাংবাদিকতা হয়েছে। তাই ডেটাকে নিছক সংখ্যা হিসেবে না ভেবে, আমাদের চারপাশের প্রতিটি অভিজ্ঞতাকে বিশ্লেষণযোগ্য তথ্য হিসেবে ভাবুন। ডেটা সাংবাদিকতার শুরুটা এখান থেকেই।

ডেটা, পরিসংখ্যান ও তথ্য : কোনটি কী?

বাংলায় ডেটা অর্থ হলো, উপাত্ত; কিন্তু ডেটা শব্দটিও বহুল প্রচলিত। অবশ্য ডেটার প্রসঙ্গ এলেই আমরা অনেক সময় পরিসংখ্যান বা তথ্যের ধারণার সঙ্গে একে গুলিয়ে ফেলি। ডেটা নতুন জ্ঞান বা তথ্য তৈরির জন্য ব্যবহৃত হয়। এটি মূলত বিচ্ছিন্ন, একক ও পরস্পর সম্পর্কহীন সংখ্যা, ছবি, কোড, চিহ্ন, বা এমন আরও অনেক কিছুই হতে পারে, যার কোনো সুনির্দিষ্ট তাৎপর্য নাও থাকতে পারে। আর ডেটা সংগ্রহ ও বিশ্লেষণের যে প্রক্রিয়া তা হলো পরিসংখ্যান।

একটি উদাহরণ দিলে স্পষ্ট হবে। মাশরাফি বিন মোর্ত্তজা, তাঁর ক্রিকেট ক্যারিয়ারে অসংখ্য ম্যাচ খেলেছেন। এখানে মাশরাফি নামটি একটি ডেটা। ভেন্যুর নাম একটি ডেটা। তিনি সেই ম্যাচগুলোতে যত ওভার বল করেছেন, তার প্রত্যেকটি ডেলিভারি একেকটি ডেটা। কিন্তু শুধু নাম মাশরাফি বা তার শুধু একটি ডেলিভারি কোনো তথ্য বা পরিসংখ্যান নয়। আপনি যখন তার প্রতিটি বল, ম্যাচের ডেটা বিশ্লেষণ করে বলবেন তিনি মোট কত ওভার বল করেছেন, কয়টি ম্যাচ খেলেছেন, কয়টি উইকেট পেয়েছেন, বা তার বোলিং গড় কত, তখন ওই বিশ্লেষণের প্রক্রিয়াটি হলো পরিসংখ্যান।



Career Averages

Bowling

FORMAT	Mat	Inns	Balls	Runs	Wkts	BBI	BBM	Ave	Econ	SR	4w
Test	36	51	5990	3239	**78**	4/60	5/88	41.52	3.24	76.7	[illegible]
ODI	220	220	10922	8893	**270**	6/26	6/26	32.93	4.88	40.4	[illegible]
T20I	54	53	1139	1527	**42**	4/19	4/19	36.35	8.04	27.1	[illegible]
FC	57		8970	4732	**135**	4/27		35.05	3.16	66.4	[illegible]
List A	320		15698	12634	**441**	6/26	6/26	28.64	4.82	35.5	[illegible]
T20	171	167	3635	4544	**157**	5/35	5/35	28.94	7.50	23.1	[illegible]

ছবি: ক্রিকইনফো

অর্থ্যাৎ পরিসংখ্যান হলো ডেটার ইন্টারপ্রিটেশন বা ব্যাখ্যা, যাকে একটি নির্দিষ্ট পদ্ধতিতে সাজানো ও বিশ্লেষণ করা হয়েছে। ডেটা যদি হয় বিছিন্ন কিছু "ফ্যাক্ট" এর সংগ্রহ, তাহলে তথ্য হচ্ছে সেই ফ্যাক্টগুলোকে একটি নির্দিষ্ট প্রেক্ষাপটে বিশ্লেষণ ও বর্ণনা করা। ডেটা কেন ও কীভাবে জাতীয় প্রশ্নের উত্তর দিতে পারে না, কিন্তু তথ্য সেটি দিতে পারে। ডেটা বা পরিসংখ্যানকে যেখানে উপস্থাপন করা হয় গ্রাফ, সংখ্যা, চার্ট বা চিত্রের মাধ্যমে, সেখানে তথ্যকে সাধারণত তুলে ধরা হয় শব্দে, ভাষায়, বর্ণনায়, চিন্তায় এমনকি ধারণার মাধ্যমেও। অর্থ্যাৎ ডেটা বা পরিসংখ্যানে আপনি যখন প্রেক্ষিত জুড়ে দেবেন তখন সেটি তথ্য হয়ে উঠবে।

আরেকটি উদাহরণ : ধরুন, আপনি একটি সরকারি নথি পেয়েছেন যেখানে বলা আছে বরিশাল জেলায় কয়টি সরকারি অ্যাম্বুলেন্স আছে, তাদের কয়টি নষ্ট, এবং কয়টি সচল। সেই নথিটি আপনার জন্য কিছু তথ্যের যোগান দিচ্ছে, কারণ তাতে একটি নির্দিষ্ট সময়ে, নির্দিষ্ট জেলায় সচল ও অচল অ্যাম্বুলেন্সের বিবরণ রয়েছে, তা একটি নির্দিষ্ট অফিস থেকে এসেছে; অর্থ্যাৎ সংখ্যার সঙ্গে প্রেক্ষিত যুক্ত হয়ে সেটি একটি অর্থবহ তথ্য হয়ে উঠেছে।

এখন ধরুন, আপনি গোটা দেশের ৬৪ জেলা থেকে একই তথ্য পেলেন। তারপর সেখান থেকে জেলার নাম, প্রতিজেলায় সচল ও অচল অ্যাম্বুলেন্সের সংখ্যা নিয়ে একটি স্প্রেডশিটে বসালেন। তখন প্রতিটি নাম ও সংখ্যা একেকটি ডেটা, আর সেই ডেটা বিশ্লেষণ করে আপনি যখন গোটা দেশের চিত্র তুলে আনলেন এবং প্রেক্ষাপট ও বর্ণনা দিয়ে ব্যাখ্যা করলেন, সেটি তখন একটি নতুন তথ্য জন্ম দিলো। তাই এক কথায়, ডেটা সাংবাদিকতা হলো অসংখ্য একক "ফ্যাক্ট" কে বিশ্লেষণের মাধ্যমে তথ্যে রূপ দেয়া, যা মানুষকে সচেতনভাবে সিদ্ধান্ত নিতে সাহায্য করে।

ডেটা ও তথ্যের মধ্যে পার্থক্য সম্পর্কে আরও জানতে চাইলে পাশের কিউআর কোডটি স্ক্যান করুন

ডেটা সাংবাদিকতা কী?

ডেটাকে ভিত্তি ধরে যে সাংবাদিকতা করা হয়, সেটিই ডেটা সাংবাদিকতা। সহজ করে ডেটা সাংবাদিকতার এ সংজ্ঞা দেয়া হলেও, তা দিয়ে নতুন ধরনের এই সাংবাদিকতার ব্যাপ্তি বা বিশালতাকে বোঝানো যায় না।

> ১০ বছর আগেও ডেটা সাংবাদিকতা কেবলই সরাসরি সংখ্যার বিশ্লেষণে মনোযোগী ছিল। কিন্তু প্রযুক্তির এই বিশ্বে এখন সব কিছুই ডেটা। কারণ সবকিছুই এখন সংখ্যা দিয়ে বর্ণনা করা সম্ভব। হোক তা ছবি, ভিডিও, মানচিত্র, ইতিহাস, এমনকি মানুষের কথপোকথন। শূন্য (০), এক (১)–এই দুই অঙ্কের মাধ্যমে বর্ণনামূলক সবকিছুরই এখন প্রযুক্তির সহায়তায় সংখ্যায় রূপান্তর করা যায়। সেই সংখ্যাকে বিশ্লেষণ করে তৈরি হয় প্রতিবেদন। ফলে বিশেষ করে পশ্চিমা উন্নত প্রযুক্তির অভিগম্যতা সম্পন্ন দেশগুলিতে ডেটা সাংবাদিকতা এখন কেবল সরাসরি সংখ্যা হিসেবে পাওয়া সংখ্যার বিশ্লেষণে থেমে নেই।

ডেটা সাংবাদিকতা নিয়ে যখন ভাববেন, মনে রাখবেন যে তাতে

০১ ডেটা হলো গল্পের অন্যতম উৎস

০২ বিশ্লেষণ বা উপস্থাপনে প্রযুক্তির ব্যবহার আছে

০৩ একটি প্রবণতা বা প্যাটার্ন মিলছে

০৪ ডেটাচিত্র হচ্ছে গল্পের অংশ

০৫ ডেটা দিয়ে বলা গল্পটা আসলে মানুষের।

ডেটা সাংবাদিকতা কি তাহলে প্রথাগত সাংবাদিকতা থেকে আলাদা? মোটেও তা নয়। বরং এটি হলো সংবাদ তৈরির পৃথক একটি পদ্ধতি, যে পদ্ধতিতে সংবাদ তৈরি বা সংগ্রহের প্রথাগত পদ্ধতি ব্যবহৃত হয় না। দৈনন্দিন ঘটনার খবর জানানো ডেটা সাংবাদিকতার কাজ নয়। বলকান ইনভেস্টিগেটিভ রিপোর্টিং সেন্টারের ক্রিনা বোরোস ও লরেন্স মারজুক[১০] মনে করেন, ডেটা সাংবাদিকতা এসেছে মূলত প্রথাগত অনুসন্ধানী সাংবাদিকতা থেকে, তবে এক্ষেত্রে সেই অনুসন্ধানী সাংবাদিকের সংখ্যাতাত্ত্বিক ও ডিজিটাল দক্ষতা থাকতে হয়।

---

[১০] *Boroş, C., & Marzouk, L. (2018).* GETTING STARTED IN DATA JOURNALISM. Balkan Investigative Reporting Network in Albania. https://birn.eu.com/wp-content/uploads/2018/08/Data-journalism-single- page.pdf

তাদের মতে, ডেটা সাংবাদিকতা হলো সংখ্যা নিয়ে খেলা; অথবা টেক্সট, ছবি বা মেটাডেটাকে বিশ্লেষণের মাধ্যমে এমন কোনো প্যাটার্ন বা আউটলায়ার খুঁজে বের করা যা কোনো স্টোরির ভিত্তি দেয়, অথবা নতুন বৈশিষ্ট্য ও ইঙ্গিতের সন্ধান দেয়।

> ফিলিস্তিনে আজ বা গত কয়েকদিনে কী পরিমাণ জায়গা দখল হলো, সেটি একটি খবর। কিন্তু সেটি একটি বড় ধরনের ডেটা সাংবাদিকতা হয়ে উঠেছে যখন আলজাজিরা স্যাটেলাইট থেকে সংগ্রহ করা ৬০ বছরের ছবি বিশ্লেষণ করে মানচিত্র ও গ্রাফিক্সের মাধ্যমে তুলে ধরেছে ১৯৪০ এর দশক থেকে ইসরায়েল কীভাবে ফিলিস্তিনের ভুখণ্ড দখল করেছে এবং এই সংঘাতে কত মানুষ আহত বা নিহত হয়েছেন। দেখুন–

- এখানে একটি প্রবণতা পাওয়া যাচ্ছে;
- জানা একটি বিষয়কে নতুনভাবে উপস্থাপনের চেষ্টা রয়েছে;
- ডেটা প্রাপ্তিতে স্যাটেলাইট প্রযুক্তি ব্যবহার হয়েছে;
- বিশ্লেষণেও তারা প্রযুক্তিগত টুল ব্যবহার করেছে;
- সর্বোপরি, এই মানচিত্র আসলে সেই ভূখণ্ডের মানুষদের কথা বলছে।

ইনফোগ্রাফিক: আল-জাজিরা

অনেক সময় ডেটা নিজেই স্টোরি। যেমন, উইকিলিকস বা পানামা পেপার্স। সেই ডেটার প্রভাব বা প্রতিক্রিয়া এত বড়, এত ব্যাপক যে সেখানে ডেটাই আপনার স্টোরিকে মহিমান্বিত করে তুলবে। কিন্তু দিন শেষে ডেটা আপনাকে বলবে কী হচ্ছে; কিন্তু কেন আর কীভাবে–এসব প্রশ্নের উত্তর পেতে আপনাকে প্রথাগত সাংবাদিকতার পথেই হাঁটতে হবে। মানুষ সোর্সের কাছে যেতে হবে, সরেজমিনে পর্যবেক্ষণ করতে হবে। তবেই সেটি একটি স্টোরি হয়ে উঠবে। তাই, নিছক সংখ্যার উল্লেখ নয়, সংখ্যাকে প্রশ্ন করার মধ্য দিয়ে ডেটা সাংবাদিকতা শুরু হয়।

সংখ্যাকে প্রশ্ন করা

> বাংলোদেশে প্রতি মাসে মূল্যস্ফীতির তথ্য প্রকাশ করা হয়। সাংবাদিকরা এটি নিয়ে নিয়মিত রিপোর্টও করে থাকেন। কেউ কেউ মূল্যস্ফীতির তুলনামূলক চিত্র দেখিয়ে চার্ট বা গ্রাফও উপস্থাপন করে থাকেন। কিন্তু সেটি ডেটা সাংবাদিকতার মানদণ্ডে উৎরাবে কী? আপনি যদি গত ৩০ বছরের মুল্যস্ফীতির মাসভিত্তিক ও খাতভিত্তিক ডেটাকে একটি স্প্রেডশিটে নিয়ে সেখানে কোনো প্রবণতা খোঁজার চেষ্টা করেন, তখন সেটি ডেটা সাংবাদিকতা হয়ে উঠতে পারে। আর এই কাজটি করা যায় মূলত সংখ্যাকে প্রশ্ন করার মাধ্যমে।

গত ৩০ বছরে
মূল্যস্ফীতি কোন কোন সময়ে অস্বাভাবিক পর্যায়ে গিয়ে ঠেকেছে
এর সঙ্গে বিশ্ব বাজারে দাম বৃদ্ধির সম্পর্ক কী
যেসব বছর দেশে বন্যা বা খরা হয়েছে সেসময় মূল্যস্ফীতি কেমন ছিলো
কোন কোন জাতীয় বা আন্তর্জাতিক সংকটে মূল্যস্ফীতি বেড়েছে
সময়ভেদে খাত বিচারে মূল্যস্ফীতির হার বৃদ্ধির প্রবণতায় পরিবর্তন কেমন

বড় আকারের এই ডেটাসেটকে প্রশ্ন করে আপনি হয়তো জানতে পারেন, কোন ধরনের সংকট কোন খাতে মূল্যস্ফীতি বাড়ায়, জলবায়ু পরিবর্তনের প্রভাবের সঙ্গে এর সম্পর্ক কী। আপনি হয়তো মানুষের আয় বৃদ্ধির হারের সঙ্গে মূল্যস্ফীতির হারের তুলনা করে নতুন কোনো অ্যাঙ্গেলও খুঁজে দেখতে চাইতে পারেন।

আগেও খবর বা প্রতিবেদনে সংখ্যা বা পরিসংখ্যান ব্যবহৃত হতো। সেই পরিসংখ্যানের যোগান দিতেন পরিসংখ্যানবিদরা। সেখানে তাদের সংখ্যাতাত্ত্বিক বিশ্লেষণ থাকতো। কিন্তু ডেটা সাংবাদিকতায়, সাংবাদিক নিজেই ডেটাকে প্রশ্ন করছেন এবং বিশ্লেষণ করে রিপোর্টের একটি নতুন অ্যাঙ্গেল দাঁড় করাচ্ছেন। বিগ ডেটার এই যুগে তথ্যের স্বচ্ছতা আর সত্য অনুসন্ধানে ডেটা সাংবাদিকতা ও এর অনুশীলন শুধু গুরুত্বপূর্ণই নয়, প্রয়োজনীয় হয়ে উঠেছে। কেননা এখন সাংবাদিকদের আগের চেয়ে বেশি তথ্য সংগ্রহ, যাচাই এবং বিশ্লেষণ করতে হয়।

সাংবাদিকতার বিশ্ব বিখ্যাত সব শিক্ষক আর জ্যেষ্ঠ সাংবাদিকেরা, নতুন ধারার এই সাংবাদিকতাকে বিভিন্ন সময়ে সংজ্ঞায়িত করেছেন। এমন তিনটি সংজ্ঞা এখানে তুলে ধরা হলো।

অবাধ তথ্য প্রবাহ

যখন তথ্যের অভাব ছিল, তখন আমাদের বেশিরভাগ প্রচেষ্টা ছিল তথ্য সংগ্রহের জন্য। এখন তথ্যের সরবরাহ যখন অবাধ, তখন তার প্রক্রিয়াকরণ আরও বেশি গুরুত্বপূর্ণ। আমরা দুটি স্তরে এই প্রক্রিয়া সম্পন্ন করি: ১. ডেটার অফুরন্তপ্রবাহ থেকে সংবাদ উপজীব্য ডেটা চিহ্নিত করা ও তার কাঠামো আনতে বিশ্লেষণ এবং ২. পাঠকের জন্য কী গুরুত্বপূর্ণ তা নির্ধারণ ও প্রাসঙ্গিকতা উপস্থাপন করা। ডেটা সাংবাদিকতা, বিজ্ঞানের মতো করে সংশ্লিষ্ট পদ্ধতিগুলো অনুসরণ করে এবং তার ফলাফলগুলোকে এমনভাবে উপস্থাপন করে, যা প্রভাব যাচাই করা যায়।

– ফিলিপ মায়ার
প্রফেসর এমেরিটাস
উত্তর ক্যারোলিনা বিশ্ববিদ্যালয়
চ্যাপেল হিল

গল্প বলার নতুনপদ্ধতি

ডেটা সাংবাদিকতা একটি ছাতার মতো। যা গল্প বলার জন্য প্রয়োজনীয় সকল সরঞ্জাম, কৌশল এবং পদ্ধতিকে একটি প্রক্রিয়ায় অন্তর্ভুক্ত করে। এটি ঐতিহ্যগত কম্পিউটার-সহায়ক রিপোর্টিং থেকে আরো উন্নত ধরনের। যেখানে সবচেয়ে আধুনিক ডেটা ভিজ্যুয়ালাইজেশন এবং সংবাদ অ্যাপ্লিকেশনগুলির সবকিছু অন্তর্ভুক্ত রয়েছে। আর এই সব প্রযুক্তি আর কৌশল একত্রিত করার লক্ষ্য হলো একটি, সাংবাদিকতা–আজকের গুরুত্বপূর্ণ বিষয় সম্পর্কে সকলকে অবহিত করতে তথ্য এবং বিশ্লেষণ প্রদান করা।

- অ্যারন পিলহোফার
নিউ ইয়র্ক টাইমস

তথ্য আমাদের জীবন

ভালো তথ্য সাংবাদিকতা কঠিন, কারণ ভালো সাংবাদিকতা কঠিন। এর অর্থ হচ্ছে কীভাবে ডেটার সন্ধান পাওয়া যায়, সংগ্রহ করা যায়, কীভাবে এটি বোঝা যায় এবং কীভাবে গল্পটি খুঁজে পাওয়া যায় তা নির্ধারণ করা, মূলত কঠিন কাজ। কখনও এই অনুসন্ধান সফল ভাবে শেষ হয়। কখনও বা আবার কোনো মহান গল্পের সৃষ্টি হয় না। সর্বোপরি, যদি এটি কেবলি একটি সুইচ চাপার বিষয় হতো, তবে এটি হয়তো সাংবাদিকতা হতো না। কিন্তু একটি মুক্ত ও ন্যায্য সমাজের জন্য আমাদের জীবনে ডেটা এখন অপরিহার্য।

–ক্রিস ট্যাগার্ট
ওপেন কর্পোরেটস

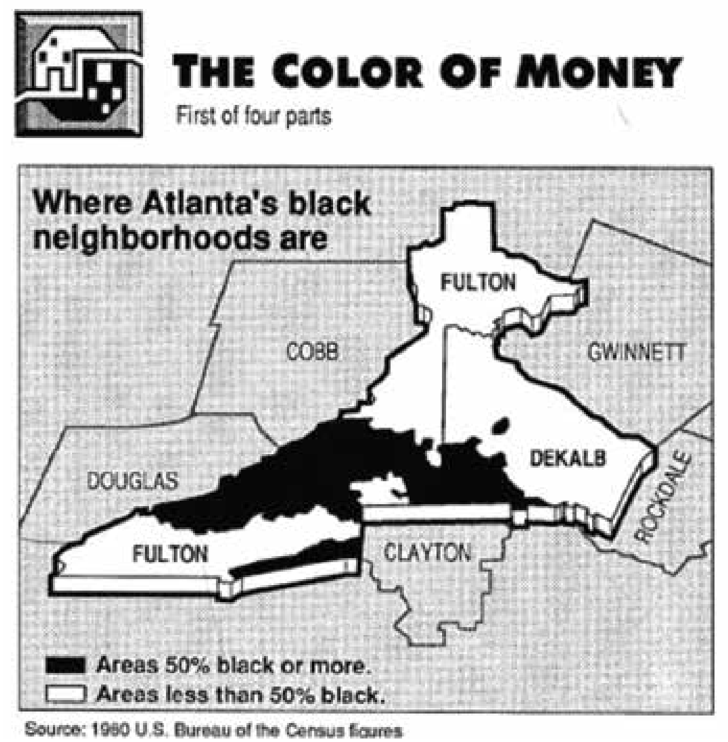
বিবর্তন

প্রতিবেদনের উৎকর্ষতার জন্য সাংবাদিকতায় পরিসংখ্যান ব্যবহারের ইতিহাস দীর্ঘ। এখন আমরা যে ডেটা সাংবাদিকতা করি, সেটির শুরু সেই পরিসংখ্যানভিত্তিক প্রতিবেদন তৈরির প্রক্রিয়া থেকে। যার পুরোটাই নির্ভরশীল ছিলো কম্পিউটারের ওপর, বিশেষ করে মাইক্রোসফট এক্সেলে। ডেটা সংগ্রহ ও বিশ্লেষণ করার জন্য কম্পিউটার ব্যবহার করার প্রথম কাঠামো ও পদ্ধতিগত এই প্রক্রিয়া “কম্পিউটার-সহায়তা রিপোর্টিং” বা "CAR" হিসেবে পরিচিত।

ডেটা জার্নালিজম হ্যান্ডবুকের তথ্য অনুযায়ী,[১১] মার্কিন গণমাধ্যম কলাম্বিয়া ব্রডকাস্টিং সিস্টেম - সিবিএস ১৯৫২ সালে দেশটির রাষ্ট্রপতি নির্বাচনের ফলাফলের পূর্বাভাস দেয়ার মধ্য দিয়ে সাংবাদিকতায় “কম্পিউটার-সহায়তা রিপোর্টিং” বা "CAR" এর ব্যবহার শুরু করেছিল। এরপর ১৯৬০ এর দশকে মার্কিন অনুসন্ধানী সাংবাদিকেরা এই প্রযুক্তিসহায়ক কৌশল ব্যবহার করতে শুরু করেন। সে সময় মূলত সরকারি তথ্য ভাণ্ডারের ডেটা বিশ্লেষণ করা হতো CAR এর মাধ্যমে।

[১১] DataJournalism.com. (2020b, October 26). Data Journalism in Perspective. Retrieved October 12, 2022, from https://datajournalism.com/read/handbook/one/introduction/data-journalism-in-perspective

১৯৮৬ সালে মার্কিন সাংবাদিক বিল ডেডম্যান আটলান্টা রাজ্যের আর্থিক প্রতিষ্ঠানগুলোর ৫ বছরের ডেটা বিশ্লেষণ করে, দ্যা কালার অব মানি[১২] প্রতিবেদনে ঋণ অনুমোদনে বর্ণবাদ আর জাতিগত পক্ষপাতের চিত্র তুলে ধরেন।

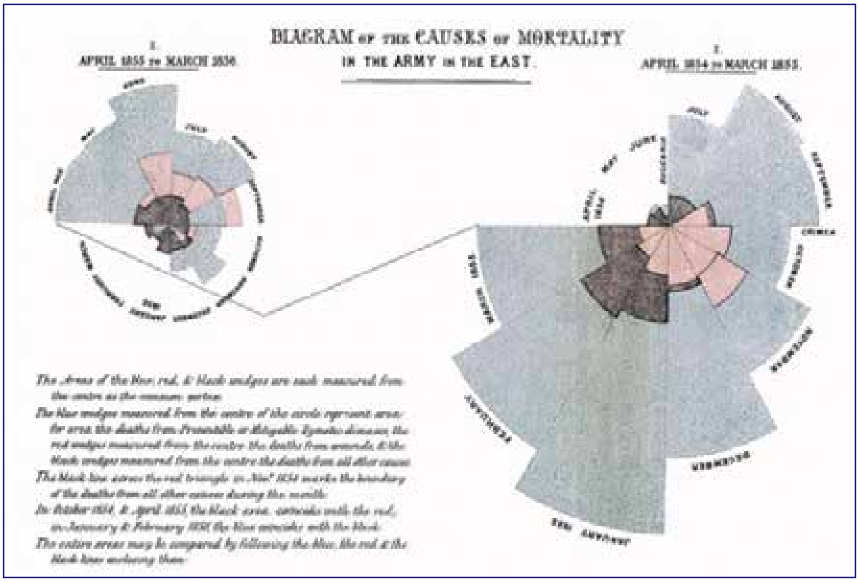


ইনফোগ্রাফিক: বিল ডেডম্যান

তবে এরও প্রায় ১০০ বছর আগে যুক্তরাজ্যের পার্লামেন্টে সেনাবাহিনীর স্বাস্থ্যসেবার উন্নতি বিষয়ে এক প্রতিবেদনে ডেটাভিত্তিক বিশ্লেষণ হাজির করেছিলেন বিখ্যাত ব্রিটিশ পরিসংখ্যানবিদ ও সেবিকা ফ্লোরেন্স নাইটঅ্যাঙ্গেল। 'মরটালিটি অব দ্য ব্রিটিশ আর্মি'[১৩] নামের সেই প্রতিবেদনে চার্টের মাধ্যমে তিনি দেখিয়েছিলেন যে, বুলেট নয়, বরং যথাযথ স্বাস্থ্সেবার অভাব হলো যুদ্ধক্ষেত্রে বৃটিশ সেনাদের মৃত্যুর বড় কারণ।

[১২] *The Color of Money.* (n.d.-b). Retrieved October 12, 2022, from http://powerreporting.com/color/

[১৩] *Mortality of the British army: at home and abroad, and during the Russian war, as compared with the mortality of the civil population in England;* illustrated by tables and diagrams : Free Download, Borrow, and Streaming: (n.d.-b). Internet Archive. Retrieved October 12, 2022, from https://archive.org/details/mortalityofbriti00lond/page/n31/mode/2up?view=theaterRetrieved October 12, 2022, from https://archive.org/details/mortalityofbriti00lond/page/n31/mode/2up?view=theater



ইনফোগ্রাফিক: ফ্লোরেন্স নাইটিঙ্গেল

গবেষক ও সাংবাদিকদের উৎকর্ষতার চেষ্টা ও দীর্ঘ অণুশীলনের ফল আজকের ডেটা সাংবাদিকতা। তবে ডেটা সাংবাদিকতা, CAR এর উন্নত ধরন নাকি পুরোপুরি নতুন এবং আলাদা একটি কৌশল এ নিয়ে বিতর্ক আছে। লন্ডনের কিংস কলেজের শিক্ষক ও পাবলিক ডেটা ল্যাবের সহ-প্রতিষ্ঠাতা লিনিয়ানা বুনেগ্রু[১৪] এ নিয়ে গবেষণায় দুই পক্ষের মতামত তুলে ধরেছেন। একপক্ষ মনে করে, CAR শুধু ডেটা সংগ্রহ ও তার বিশ্লেষণের একটি কৌশল। যা ব্যবহার করে একজন প্রতিবেদক, তথ্যের ভিত্তিতে প্রতিবেদনের পূর্বানুমানকে প্রমাণ করতে পারে। তাদের মতে CAR ডেটা অনুসরণ করে বা ডেটার ওপর নির্ভর করে প্রতিবেদন তৈরি করতে হয়। পক্ষান্তরে ডেটা সাংবাদিকতা, প্রয়োজনে নিজে ডেটা তৈরি করে। এই কৌশলে ডেটার পেছনের গল্প অনুসন্ধানের পাশাপাশি ডেটা নিজে একটি সংবাদ হয়ে ওঠে।

> তবে কেউ কেউ যুক্তি দেন যে, ডেটা সাংবাদিকতা এবং CAR এর মধ্যে কোনো অর্থপূর্ণ পার্থক্য নেই। তাদের মতে CAR এর ধারাবাহিক প্রযুক্তিগত উৎকর্ষতার রূপ আজকের ডেটা সাংবাদিকতা, যা আগামী দিনে হয়তো আরও এগিয়ে যাবে। CAR ও ডেটা সাংবাদিকতার বড় পার্থক্যটি করে দিয়েছে প্রযুক্তি। বিশেষ করে, ডেটা সংগ্রহ এবং চিত্রায়ণে। পাইথনের মত সফটওয়ার এসে যে ডেটা বিশ্লেষণের মাত্রা কে অন্য পর্যায়ে নিয়ে গেছে। তাই এটি অস্বীকারের উপায় নাই যে, সাংবাদিকতায় ডিজিটাল প্রযুক্তির ব্যবহার এবং ওয়েব তথ্য প্রকাশের পদ্ধতিকে মৌলিকভাবে পরিবর্তন করছে ডেটা সাংবাদিকতা। এর অনুশীলনের কৌশল এবং প্রভাব পাঠক বা দর্শকের কাছে খুব কার্যকর ভাবে পৌঁছেছে।

[১৪] Bounegru, L. (2021, December 6). SUPERB to promote forest restoration and adaptation across Europe. Retrieved October 12, 2022, from https://lilianabounegru.org/

ডেটা সাংবাদিকতা কেন করবেন?

ডেটা সাংবাদিকতা শুধু প্রতিবেদন তৈরি বা গল্প উপস্থাপনের ধরনকেই বদলে দেয়নি, দৈনন্দিন জীবনে সাংবাদিকতার প্রভাবও বাড়িয়েছে বহুগুণ। সংখ্যার এই তুলনামূলক উপস্থাপন ও সহজ ব্যাখ্যা, পাঠক ও দর্শকের সঙ্গে গণমাধ্যমের সংযোগও দৃঢ় করেছে। মানুষের ব্যক্তি জীবন, সমাজ, রাষ্ট্র এবং সাংবাদিকতায় ডেটা সাংবাদিকতার গুরুত্ব ও সুবিধাগুলো চলুন জেনে নেই।

০১ ডেটা সাংবাদিকতা একজন প্রতিবেদককে আকর্ষণীয় ইনফোগ্রাফিকের মাধ্যমে একটি জটিল গল্প সহজে বলতে সাহায্য করে। ফলে প্রতিবেদনটি একদিকে যেমন সহজ পাঠযোগ্য হয়, তেমনি বেশি সংখ্যক পাঠক বা দর্শকের কাছে পৌঁছায়।

০২ একটি ঘটনা একটি বৃহৎ জনগোষ্ঠীকে কীভাবে প্রভাবিত করে, কিংবা সরকারের একটি নীতি সিদ্ধান্ত কতটা একজন নির্দিষ্ট ব্যক্তির জীবন সম্পর্কিত, সেটি ব্যখ্যা করার সবচেয়ে কার্যকর কৌশল হচ্ছে ডেটা সাংবাদিকতা।

০৩ ডেটার বিশ্লেষণ সংবাদ বা প্রতিবেদনকে নতুন রূপ দেয় ডেটা সাংবাদিকতা। খালি চোখে আমরা সংবাদকে যেভাবে দেখি বা বিশ্লেষণ করি, ডেটা সেই ধারণাকে বদলে দেয়। ফলে ডেটাভিত্তিক প্রতিবেদনটি বাকি পাঁচটি প্রতিবেদন থেকে ভিন্ন হয়। আর সে জন্য ডেটা সাংবাদিকতাকে প্রতিবেদকের নতুন চোখ বা ক্যামেরাও বলা হয়। যা ব্যবহার করে একজন সাংবাদিক সংবাদকে একটি সম্পূর্ণ ভিন্ন পেক্ষাপটে ব্যাখ্যা করতে পারে। ডেটাকে বিভিন্নভাবে বিশ্লেষণ করার এই ক্ষমতা, সাংবাদিককে অধিক সংখ্যক প্রতিবেদন তৈরি করার সুযোগ করে দেয়। প্রতিবেদনের ক্ষেত্র বড় হয়।

০৪ ডেটার বিশ্লেষণ মানুষকে ব্যক্তি জীবনে সঠিক সিদ্ধান্ত নিতে সহায়তা করে। শুধু সাধারণ মানুষ নয়, সরকারও এমন বিশ্লেষণ দ্বারা উপকৃত হয়। ধরুন, কোনো এলাকায় দাঙ্গা বা রাজনৈতিক সংঘাত দেখা দিলো। সে সময় ওই এলাকার বসবাসকারীদের ধর্ম, বয়স, পেশা, ভোটের ফলাফল কিংবা পূর্বের এমন জটিল পরিস্থিতির গতিশীলতা বিশ্লেষণ, সরকারকে সমস্যার সম্ভাব্য সমাধানে সাহায্য করতে পারে। আবার, জেলাভিত্তিক কর্মসংস্থানের ডেটা কিংবা চাকরি বাজারের গতিশীলতা বিশ্লেষণ, তরুণ বেকারদের সিদ্ধান্ত গ্রহণে সাহায্য করতে পারে।

০৫ সংখ্যা বা পরিসংখ্যান যে কোনোও তথ্যকে আরও সুনির্দিষ্ট করে। সংখ্যার এই শক্তিকে কাজে লাগিয়ে ডেটা প্রতিবেদন, সাংবাদিকতার বিশ্বাসযোগ্যতা বাড়ায় এবং বিভ্রান্তি দূর করে। সংশ্লিষ্টদের সাক্ষাৎকার, কিংবা প্রত্যক্ষদর্শীর বয়ানের পরিবর্তে, একজন সাংবাদিকের তার কাছে থাকা পরিসংখ্যান ও তার ভিত্তিতে তৈরি হওয়া বিশ্লেষণ প্রতিবেদনকে একটি শক্তিশালী রূপ দেয়। তথ্যভিত্তিক এইসব প্রতিবেদন সাংবাদিকতার অবস্থান ও প্রভাবকেও বাড়ায়।

০৬ চাকরির বাজারে ডেটা সাংবাদিকদের চাহিদা বাড়ছে। এমনকি বিশ্বজুড়ে যখন নিউজরুমগুলোর আকার ছোট হয়ে আসছে, অনেক সাংবাদিক জনসংযোগে যাওয়ার পরিকল্পনা করছেন, তখন ডেটা সাংবাদিক বা ডেটা বিজ্ঞানীরা নতুন করে সাংবাদিকতায় যোগ দিচ্ছেন। গণমাধ্যমগুলোই তাদের খুঁজে নিচ্ছেন। তাই চাকরির বিজ্ঞাপনেও আলাদা করে উল্লেখ থাকছে ডেটা সাংবাদিকতা, অন্তত পক্ষে মাইক্রোসফট এক্সেল কিংবা গুগল শিটের দক্ষতার কথা। ডেটা সাংবাদিকদের বেতনও বাকিদের থেকে বেশি বই কম নয়। ফ্রিল্যান্সারদের জন্যও, ডেটার দক্ষতা নতুন কাজের সুযোগ এবং স্থিতিশীল আয়ের পথ দেখায়।

০৭ ডেটা সাংবাদিকতা নাগরিকের ওপর থেকে তথ্যের চাপ কমায়। অবাধ তথ্যের এই যুগে অনেক সময়ই সাধারণ মানুষের সামনে অনেক বেশি তথ্য একবারে চলে আসে। যা যাচাই করে সেগুলো থেকে নিজের প্রয়োজনীয় তথ্য সংগ্রহ করা যেমন কঠিন হয়ে যায়, তেমনি একাধিক তথ্য মানুষকে বিভ্রান্ত করে। ডেটা সাংবাদিকরা এক্ষেত্রে সহায়ক হিসেবে কাজ করে। তারা সেই বিশাল তথ্য ভাণ্ডার থেকে ডেটার শ্রেণিবিন্যাস করে। ক্ষেত্রভেদে পৃথক প্রতিবেদন তৈরি করে এবং সংশ্লিষ্ট ইনফোগ্রাফিক বা তথ্যচিত্র প্রকাশ করে, যাতে পাঠক বা দর্শক সহজে তার প্রয়োজনীয় তথ্যটি খুঁজে ও বুঝে নিতে পারে।

০৮ ডেটা সাংবাদিকতা তথ্যের অসামঞ্জস্যতার বিরুদ্ধে লড়াই করতে সহায়তা করে। এর মাধ্যমে ভুল কিংবা মিথ্যা তথ্য সহজে চিহ্নিত করা যায়। রাজনীতিবিদ হোক, সরকার অথবা বেসরকারি খাতের প্রতিষ্ঠান, সবাই নিজ নিজ তথ্য নিয়ে জনগণের কাছে হাজির হয়; হোক তা প্রতিষ্ঠানের লাভ ক্ষতির হিসাব, দেশের বেকারত্বের সংখ্যা এবং জিডিপির প্রবৃদ্ধি। এসব তথ্য সংশ্লিষ্টরা প্রায়শই নিজেদের সুবিধামত কম-বেশি করে কিংবা ভুল ব্যাখ্যা উপস্থাপন করে। একজন ডেটা সাংবাদিক তথ্যের সেই অসংগতিটি খুব সহজে চিহ্নিত করতে পারে।

০৯ ডেটা প্রতিবেদন প্রকাশের মাধ্যমে সমাজে এবং পাঠক-দর্শকদের মাঝে এক ধরনের সচেতনতা তৈরি হয়। তারা সংখ্যা ও পরিসংখ্যানের গুরুত্ব অনুধাবন করতে শুরু করে। ফলে সমাজে সংখ্যাতাত্ত্বিক তথ্যের চাহিদা তৈরি হয়। সরকার ও সংশ্লিষ্ট কর্তৃপক্ষের ওপর হালনাগাদ পরিসংখ্যান প্রকাশের এক ধরনের চাপ তৈরি হয় জনগণের পক্ষ থেকে। অর্থাৎ সমাজে অবাধ তথ্য ও পরিসংখ্যানের এমন একটি চাহিদা তৈরি করতে ডেটা সাংবাদিকতা বড় অবদান রাখে।

১০ ডেটা সাংবাদিকতা প্রতিবেদকের সময় এবং শ্রম দুটোই কমাতে সহায়তা করে। কাগুজে দলিল, পিডিএফ ফাইল কিংবা ওয়েবসাইট থেকে ডেটা বের করা কিংবা কয়েক হাজার ডেটা বা পরিসংখ্যানের হিসাব ডেটা সাংবাদিকতার কৌশল ব্যব্যহার করে মুহূর্তে করা সম্ভব। সেই হিসেবে নতুন ধারার এই সাংবাদিকতা, সাংবাদিকতার গতিশীলতা ও কর্মক্ষমতা বাড়িয়েছে।

0101010101010101010101010101010101010101010101010101

1010101010101010101

অধ্যায় ০৩

যত ধাপ

যত ধাপ

ডেটা সাংবাদিকতা একটি দীর্ঘ যাত্রা। এই যাত্রায় একজন প্রতিবেদক কিংবা ডেটা দলকে অন্তত আটটি ধাপ পার করতে হয়। প্রতিটি ধাপ বা পর্যায়ই সমানভাবে গুরুত্বপূর্ণ এবং একটি অন্যটির সঙ্গে সম্পর্কযুক্ত। একটি পর্যায় বাদ রেখে পরবর্তী পর্যায়ের কাজ করাও সম্ভব নয়। প্রতিটি ধাপেই নির্দিষ্ট কিছু লক্ষ্য বা কাজ থাকবে, যা আপনি নিজেই নির্ধারণ করবেন এবং সেগুলো সফলভাবে সম্পন্ন করার মধ্য দিয়ে একটি শক্তিশালী ডেটা প্রতিবেদন তৈরির দিকে এগিয়ে যাবেন। আর এই পুরো প্রক্রিয়াটি লিপিবদ্ধ থাকবে একটি ফাইলে। এই ফাইলটি মাস্টারফাইল নামে পরিচিত। প্রতিটি ডেটা প্রতিবেদনের জন্যে পৃথক একটি মাস্টারফাইল থাকে। যেটি প্রতিবেদক তৈরি করেন প্রতিবেদনটির পরিকল্পনার শুরু থেকে।

একটি ডেটা প্রতিবেদন তৈরিতে ধারাবাহিকভাবে যেসব ধাপ সম্পন্ন করতে হয়

পরিকল্পনা

যে কোনো কাজ শুরুর আগেই সে বিষয়ে পরিকল্পনার দরকার। সাংবাদিকতার ক্ষেত্রে এর প্রয়োজনীয়তা বেশি। ডেটা সাংবাদিকতার ক্ষেত্রে যেটির গুরুত্ব আরও বেশি। কেননা প্রাথমিক পর্যায়ের এই কার্যক্রমের ওপরই নির্ভর করে পরের ধাপগুলোর কার্যকারিতা ও সফলতা। পরিকল্পনার এই ধাপটি মূলত প্রতিবেদন তৈরির প্রস্তুতি পর্ব। এখানে একজন ডেটা প্রতিবেদক কিংবা তার দলকে বেশ কয়েকটি বিষয়ে প্রস্তুতি এবং সিদ্ধান্ত নিতে হয়।

> এই পর্বে সবার আগে আপনাকে প্রতিবেদনের বিষয় নির্ধারণ করতে হবে। তৈরি করতে হবে প্রতিবেদনের হাইপোথিসিস বা পূর্বানুমান। যেখানে এই প্রতিবেদনের মাধ্যমে সম্ভাব্য কোন বিষয়ের বিশ্লেষণ পাঠক ও দর্শকের সামনে আপনি তুলে ধরতে চান, তা নির্ধারণ করবেন। ডেটা সাংবাদিকতার সবচেয়ে গুরুত্বপূর্ণ পর্যায়গুলোর একটি এই পূর্বানুমান। পরীক্ষাগারে একজন গবেষক যেভাবে একটি ধারণার ওপর নির্ভর করে, বিভিন্ন উপকরণের ওপর পরীক্ষা চালিয়ে নতুন কোনো সূত্র বা পণ্য আবিষ্কার কিংবা সিদ্ধান্তে উপনীত হন, ঠিক একইভাবে একজন ডেটা সাংবাদিক, তার সাংবাদিকতার অভিজ্ঞতা দিয়ে যে অনুমান করেন, পরবর্তীতে ডেটার বিশ্লেষণের মাধ্যমে তা প্রমাণ করেন।

পূর্বানুমান তৈরি করতে আপনাকে প্রতিবেদনের পটভূমি এবং এ সংক্রান্ত পুরোনো ডেটা জানতে হবে। একইবিষয়ে তৈরি অন্য গণমাধ্যমে প্রকাশিত প্রতিবেদন পড়তে হবে। একইসঙ্গে পরিকল্পনার এ পর্যায়ে আপনাকে জানতে হবে, পূর্বানুমানটি প্রতিবেদনের মাধ্যমে প্রমাণ করতে চাইলে আপনার কী ধরনের ডেটার প্রয়োজন হবে। সেসব ডেটার সম্ভাব্য উৎস কী হবে। আবার ডেটা বিশ্লেষণে পাওয়া তথ্য-বক্তব্যের সঙ্গে ইতিমধ্যে প্রকাশিত তথ্য-প্রতিবেদনের সামঞ্জস্য বা অসামঞ্জস্য তুলনা করা যেতে পারে। সেজন্য ওই বিষয় সম্পর্কিত পুরনো প্রতিবেদন পড়া প্রয়োজন হবে।

> ধরুন আপনি দেশের বেকারত্ব নিয়ে একটি ডেটা প্রতিবেদন তৈরি করবেন বলে ঠিক করেছেন। পরিকল্পনার এই পর্যায়েই আপনাকে জানতে হবে যে, পরিসংখ্যান ব্যুরোর শ্রমশক্তি জরিপে আপনি এই সংক্রান্ত ডেটা পাবেন। তবে শুধু ডেটা কোথায় পাবেন জানলেই হবে না। সেই ডেটা উন্মুক্ত নাকি আপনাকে আলাদা করে আবেদন করে সংগ্রহ করতে হবে, সেটিও জানা থাকতে হবে প্রতিবেদন তৈরির এই পর্যায়েই। এমন কি ডেটার ধরন সম্পর্কেও পুরোপুরি ধারণা থাকতে হবে। কেননা এসব তথ্য জানলেই কেবল আপনি প্রতিবেদনটির জন্য একটি শক্তিশালী পূর্বানুমান তৈরি করতে পারবেন, যা একটি ভালো প্রতিবেদনের পূর্বশর্ত।

একইসঙ্গে প্রতিবেদনের সম্ভাব্য সংবাদ সূচনা বা ইন্ট্রো কী হতে পারে তার ধারণা থাকতে হবে। নিজেকে প্রশ্ন করতে হবে কেন এই প্রতিবেদনটি তৈরি করা গুরুত্বপূর্ণ কিংবা কেন এই সংবাদটি দর্শক ও পাঠকের জানা উচিত? এসব বিষয়ে উত্তর আর সিদ্ধান্ত ইতিবাচক হলে আপনি প্রতিবেদন তৈরির প্রাথমিক কাজ শুরু করবেন। যাবেন পরের ধাপে। যেখানে আপনার কাজ হবে পরিকল্পনা অনুযায়ী ডেটা সংগ্রহ করা।

ডেটা সংগ্রহ

ডেটা সংগ্রহের এই পর্যায়ে একজন প্রতিবেদককে কিংবা ডেটা দলকে অনেক বেশি তৎপর থাকতে হয়। বাংলাদেশের মত দেশগুলোতে কাজটি অনেক বেশি পরিশ্রমসাধ্য এবং সময়সাপেক্ষ। কেননা এদেশে উন্মুক্ত ডেটা খুব একটা পাওয়া যায় না। বেশিরভাগ প্রতিষ্ঠান ডেটা গুছিয়ে রাখে না। সরকারি বিভিন্ন সংস্থার প্রকাশিত ডেটা অবশ্য একটি বড় উৎস হতে পারে। সেই ডেটার সমস্যা হলো সেগুলো উন্মুক্ত থাকলেও প্রয়োজনমতো ব্যবহারের উপযোগী অবস্থায় থাকে না।

> আপনি যখন ডেটা সাংবাদিকতা শুরু করবেন, অনেক ডেটাই আপনাকে নিজস্ব উদ্যোগে সংগ্রহ করতে হবে এবং তা লিপিবদ্ধ করতে হবে। তবে ডেটা সংগ্রহের ক্ষেত্রে সরকারি ও বেসরকারি সংস্থা বা প্রতিষ্ঠানের প্রকাশিত তথ্যের বাইরে, তথ্য অধিকার আইনের মাধ্যমে আবেদন করে ডেটা সংগ্রহ করতে পারবেন। এটি ডেটা সাংবাদিকতার ক্ষেত্রে বিশ্বজুড়ে একটি কার্যকর পদক্ষেপ। এর বাইরে প্রতিবেদনের স্বার্থে ডেটা তৈরি করার প্রয়োজন হলে সেটি এই ধাপেই শেষ করতে হবে। ডেটার ধরন সম্পর্কেও পুরোপুরি ধারণা থাকতে হবে। কেননা এসব তথ্য জানলেই কেবল আপনি প্রতিবেদনটির জন্য একটি শক্তিশালী পূর্বানুমান তৈরি করতে পারবেন, যা একটি ভাল প্রতিবেদনের পূর্বশর্ত।

ডেটা সংরক্ষণ

প্রতিবেদন কিংবা অনুসন্ধানের মূল কাজ শুরুর আগে সংগ্রহ করা ডেটা সঠিক এবং সুশৃঙ্খল ভাবে সংরক্ষণ করতে হয়। কেননা দীর্ঘ অনুসন্ধানের বিভিন্ন পর্যায়ে নানান ধরনের ডেটা হাতে আসবে, সেসব ডেটা নিঃসন্দেহে প্রতিবেদনে কার্যকর অবদান রাখবে, তবে সেজন্য ডেটাকে বিশ্লেষণ ও যাচাই উপযোগী করতে হয়। সঠিকভাবে সাজিয়ে-গুছিয়ে রাখতে হয়। প্রাথমিকভাবে সংগৃহীত সেই ডেটা একটি পৃথক ফাইলে সংরক্ষণ করতে হয়।

> সংখ্যাতাত্ত্বিক ডেটা সাধারণত মাইক্রোসফট এক্সেল অথবা গুগল শিটে সংরক্ষণ করা হয়। কেননা বর্তমানে বিশ্বের প্রায় সব ডেটা সাংবাদিকই মাইক্রোসফট এক্সেল অথবা গুগল শিট ব্যবহার করেই ডেটা সাংবাদিকতার প্রাথমিক কাজগুলো করছেন। ডেটা সংরক্ষণের ক্ষেত্রে আপনাকে ডেটার নিরাপত্তার বিষয়েও সচেতন হতে হবে। একইসঙ্গে ডেটা দলের বাকি সদস্যরা যেন সেই ডেটাসেট ব্যবহার করতে পারেন, সেটিও নিশ্চিত করতে হবে। আর এ কারণেই ডেটা সাংবাদিকেরা গুগল শিট ব্যবহার করেন, যাতে দলের সব সদস্য একসঙ্গে ডেটাসেটের বিশ্লেষণের কাজ করতে পারেন। তবে বাড়তি নিরাপত্তার স্বার্থে ডেটার মূল কপিটি একাধিক কম্পিউটারের হার্ডড্রাইভেও সংরক্ষণ করতে হবে। বিশ্বের বড় সব গণমাধ্যম ও অনুসন্ধানী সংবাদমাধ্যমগুলো, ডেটা সংরক্ষণের ক্ষেত্রে নিজস্ব ক্লাউড সার্ভার ব্যবহার করে থাকে। ডেটার নিরাপত্তার জন্যে এমন ক্লাউড সার্ভার খুবই প্রয়োজনীয় এবং গুরুত্বপূর্ণ। তবে ছোট পরিসরের ডেটা সাংবাদিকতার জন্যে মাইক্রোসফট এক্সেল অথবা গুগল শিট সবচেয়ে কার্যকরী টুলস।

ডেটা ক্লিনিং

বিভিন্ন উৎস থেকে যে ডেটা আপনি সংগ্রহ করবেন, ধরে নেয়া যায় প্রাথমিক পর্যায়ে তার বেশিরভাগই বিশ্লেষণ উপযোগী থাকবে না। বেশিরভাগ ডেটাই এলোমেলো থাকবে এবং এটাই স্বাভাবিক। কেননা সংশ্লিষ্ট প্রতিষ্ঠানগুলো তাদের কাজের ধরনের আলোকে ডেটা সংরক্ষণ করেন, সেটি বিশ্লেষণ উপযোগী কি-না সেটি নিয়ে তাদের মাথাব্যথা থাকে না। যে সব ডেটার ক্ষেত্রে এমনটি থাকবে না, সেগুলোকে পরিষ্কার বা ব্যবহার উপযোগী করতে হয়।

> ডেটা ক্লিনিং এর আরেকটি কাজ ডেটার স্ট্যান্ডার্ডাইজেশন। ভিন্ন ভিন্ন উৎস থেকে পাওয়া ডেটা, ভিন্ন ভিন্ন মানদণ্ডে থাকতে পারে। বিশ্লেষণের জন্যে এইসব ডেটাকে একটি সমান মানদণ্ডে আনতে হবে। ডেটা স্ট্যান্ডার্ডাইজেশনের এই কাজটি করা যাবে মাইক্রোসফট এক্সেল অথবা গুগল শিট ব্যবহার করে।

ধরুন আপনি চালের দাম, চাহিদা, উৎপাদন, মজুদ আর আমদানি পরিস্থিতি নিয়ে প্রতিবেদন তৈরি করবেন। এ জন্য কৃষি সম্প্রসারণ অধিদপ্তর থেকে উৎপাদন, খাদ্য মন্ত্রণালয় থেকে মজুদ, রাজস্ব বিভাগ থেকে আমদানি আর বাণিজ্য মন্ত্রণালয় থেকে দামের তথ্য সংগ্রহ করে একটি ডেটাসেট তৈরি করলেন। কিন্তু দেখা গেল, কোনো সংস্থা চালের ডেটা লিপিবদ্ধ করছে টন হিসেবে। কেউ কেজি হিসেবে। তো কেউ মণ হিসেবে। স্প্রেডশিটের বিশ্লেষণে সঠিক ফলাফল পেতে, সবগুলো ডেটাকে আপনার একটি নির্দিষ্ট মানদণ্ডে আনতে হবে।

> তবে শুধু স্ট্যান্ডার্ডাইজেশনের মধ্য দিয়ে ডেটা ক্লিনিং–এর কাজ শেষ হবে না। সংগ্রহ করা ডেটা সঠিক কি-না তাও যাচাই করতে হবে ডেটা সাংবাদিকতার এই ধাপে।

ডেটা বিশ্লেষণ

ডেটা সাংবাদিকতার মূল কাজটি শুরু হবে এই পর্বে। এখান থেকেই শুরু বিশ্লেষণের। এই ধাপে আপনি ডেটাকে প্রশ্ন করবেন, উন্মোচন করবেন ডেটার আড়ালে থাকা সংবাদ আর তার বিশ্লেষণ। ডেটা সাংবাদিকতার প্রথম ধাপে করা পনূর্বানুমানকে যাচাই ও প্রমাণ করার উদ্দেশ্যে আপনি প্রশ্নগুলো তৈরি করবেন। এই বিশ্লেষণের জন্য প্রাথমিক পর্যায়ে আপনাকে মূলত মাইক্রোসফট এক্সেল বা গুগল শিট ব্যবহার করতে হবে। এই টুলস কীভাবে ব্যবহার করবেন সেই বিষয়ে পরবর্তী অধ্যায়ে আলোচনা করা হবে। এই বিশ্লেষণের মাধ্যমে ডেটা থেকে যে তথ্য আপনি পাবেন সেটির ভিত্তিতেই তৈরি হবে আপনার ডেটা প্রতিবেদন।

> এই ধাপে মাইক্রোসফট এক্সেল ও গুগল শিটের বাইরেও অন্য টুলস বা বিশেষজ্ঞের পরামর্শের দরকার হতে পারে। এ ক্ষেত্রে বিশেষজ্ঞের পরামর্শ প্রতিবেদনকে বুলেটপ্রুফ বা নির্ভুল করতে সহায়তা করবে।

ডেটা চিত্রায়ণ

চিত্রায়ণ বা ভিজ্যুয়ালইজেশন ডেটা সাংবাদিকতার অন্যতম গুরুত্বপূর্ণ অংশ। ডেটার সঠিক চিত্রায়ণের ওপর নির্ভর করে পুরো প্রতিবেদনের সাফল্য। চিত্রায়ণ ভাল না হলে, ডেটা সংগ্রহ থেকে শুরু করে বিশেণ্ঢষণ প্রত্যেকটি ধাপের পরিশ্রমই বৃথা যাবে। তাই চিত্রায়ণের এই ধাপে প্রতিবেদক এবং ডেটা দলকে একদিকে যেমন সৃজনশীল হতে হয়। অন্যদিকে পাঠক ও দর্শক সর্ম্পকেও পরিষ্কার ধারণা রাখতে হয়। অনেক বড় অনুসন্ধানের ক্ষেত্রেই এমনটি দেখা গেছে যে, সঠিক চিত্রায়ণের অভাবে সেটি বেশি সংখ্যক পাঠক বা দর্শকের কাছে পৌঁছায়নি। তাই চিত্রায়ণের মূল উদ্দেশ্য হচ্ছে গ্রাফিক ও ডিজাইনের মাধ্যমে কত সহজে পাঠক বা দর্শকের কাছে তথ্যটি পৌঁছে দেয়া যায়।

> কোন ধরনের চিত্রায়ণ বা গ্রাফিক পাঠক বা দর্শক সহজে বুঝতে পারবে, কিংবা নির্দিষ্ট প্রতিবেদন বা ডেটা চিত্রায়ণের জন্য কোন ধরনের গ্রাফিক্স ব্যবহার করা হবে, সেটি নির্ধারণও এই পর্যায়ের বড় কাজ। এই কাজে সাংবাদিক ও গ্রাফিক ডিজাইনারদের বেশ কয়েকটি ধাপ পাড়ি দিতে হয়। মাইক্রোসফট এক্সেল ও গুগল শিটে নানান ধরন ও ডিজাইনের চার্ট রয়েছে ডেটার চিত্রায়ণের জন্য। এর বাইরে অনলাইন এবং অফলাইনে বেশ কিছু কার্যকর টুল আছে, যেগুলো ব্যবহার করে সহজেই ডেটার চিত্রায়ণ বা ভিজুয়ালাইজেশন করা হয়।

স্টোরি লেখা

প্রতিবেদন লেখা ডেটা সাংবাদিকতার চ্যালেঞ্জিং কাজগুলোর একটি। প্রতিবেদন তৈরির এই পর্যায়ে আপনাকে সংখ্যাতাত্ত্বিক হিসাব থেকে জীবনের গল্প বের করতে হবে এবং সেটিকে পাঠক বা দর্শকের সামনে উপস্থাপন করতে হবে। তবে সেই প্রতিবেদন লেখার জন্যে প্রথমে ডেটাকে প্রশ্ন করতে হবে। ডেটার বিশ্লেষণ থেকে জানতে হবে সংশ্লিষ্ট বিষয়ের সমস্যা, প্রভাব ও কারণ। একইসাথে তুলে ধরতে হবে তার সমাধান।

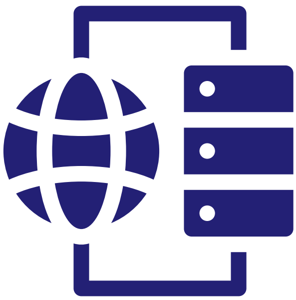
> ডেটা সাংবাদিকতায় গল্প বলার বেশ কিছু ধরন রয়েছে। এর কয়েকটি প্রচলিত সাংবাদিকতার মতোই। কিছু ধরন শুধু ডেটা সাংবাদিকতাতেই ব্যবহৃত হয়। গল্প বলার জন্য আপনি কোন ধরনটি বেছে নেবেন, তা অনেকটাই নির্ভর করবে ডেটা বিশ্লেষণ থেকে পাওয়া তথ্যের ওপর। অনেক ক্ষেত্রে প্রকাশ বা প্রচারের মাধ্যম এবং পাঠক বিবেচনায় নিয়েও গল্প বলার ধরন নির্বাচন করা হয়। একইসঙ্গে প্রতিবেদনের বাক্য গঠনেও বাড়তি সচেতনতা প্রয়োজন। শুধু সংখ্যার আধিক্য পাঠকের মাঝে যেন বিরক্তির সৃষ্টি না করে, সেটিও খেয়াল রাখতে হবে।

প্রভাব মূল্যায়ন

প্রভাব সাংবাদিকতার সবচেয়ে বড় স্বীকৃতি। কোনো প্রতিবেদন প্রকাশের পর সেটি সমাজ কিংবা রাষ্ট্রে কীভাবে প্রভাব ফেললো, মানুষকে কতটা ভাবালো কিংবা নীতি পরিবর্তনে কতৃপক্ষের ওপর কতটা চাপ সৃষ্টি করতে পারলো তার ওপর নির্ভর করে সেই সংবাদ বা প্রতিবেদনের সার্থকতা।

> প্রতিবেদন প্রকাশের পর সমস্যার সমাধান কিংবা দোষীদের শাস্তির মাধ্যমেই কেবল এর প্রভাব শেষ হয় না। সমাজে সচেতনতা সৃষ্টিও বেশ গুরুত্বপূর্ণ। তাই ডেটা সাংবাদিকতার শেষ ধাপ এটি। যার মাধ্যমে প্রতিবেদনটি সার্থক হয়। তবে এই প্রভাব মূলত দুই ধরনের। প্রথমটি বাণিজ্যিক, দ্বিতীয়টি সামাজিক। বাণিজ্যিক দিকটি গণমাধ্যমের ব্যবসার দিকটি নির্দেশ করে। অর্থাৎ কত সংখ্যক পাঠক বা দর্শক প্রতিবেদনটি পড়েছে বা দেখেছে। একইসাথে এই প্রতিবেদনকে ঘিরে প্রতিষ্ঠান কী পরিমাণ বিজ্ঞাপন পেয়েছে বা আর্থিক আয়ের পরিমাণ কী–এই সবকিছুই বাণিজ্যিক প্রভাবের অন্তর্গত। প্রতিবেদনের লক্ষ্য পূরণের পাশাপাশি সামাজিক প্রভাব গণমাধ্যম ও সাংবাদিকের মর্যাদা বৃদ্ধিকে নির্দেশ করে।

তবে এই প্রভাব নিশ্চিত করতে পাঠক বা দর্শকের সঙ্গে গণমাধ্যমের সম্পর্ক বা যোগাযোগ বাড়ানোর উদ্যোগ নিতে হয়। কোন মাধ্যমে কীভাবে পাঠক বা দর্শক সংবাদের সঙ্গে বেশি সম্পৃক্ত হবে সেটি নির্ধারণ করে, তা বাস্তবায়ন করাই ডেটা সাংবাদিকতার এই শেষ ধাপের কাজ। ডেটা সাংবাদিকতার বাকি সব পর্যায়ের মতো এটিও সমান গুরুত্বপূর্ণ। কেননা এতো সময় ও শ্রম দিয়ে তৈরি করা প্রতিবেদন পাঠক বা দর্শকের কাছে পৌঁছাতে না পারলে প্রতিবেদনের সার্থকতা মিলবে না।



অধ্যায় ০৪

ডেটা সংগ্রহ

ডেটা সংগ্রহ

> অনেক ক্ষেত্রে সোর্স নিজে এসে সাংবাদিককে ডেটা সরবরাহ করে, কিংবা ডেটার সম্ভাব্য উৎস সম্পর্কে ধারণা দেয়। পরবর্তীতে যেখানে অনুসন্ধান করে সাংবাদিক ডেটা সংগ্রহ করতে পারেন। আবার অনেক ক্ষেত্রে সাংবাদিক নিজের অনুসন্ধানী দৃষ্টিকোণ থেকে বুঝতে পারেন কোথায় খোঁজ করলে ডেটা পাওয়া যাবে। এভাবে ডেটা প্রাপ্তির নজির কম নেই নতুন ধারার এই সাংবাদিকতায়।

ডেটা কোথায় পাবেন? ডেটা সাংবাদিকতার প্রথমেই একজন প্রতিবেদককে এই প্রশ্নের মুখোমুখি হতে হয়। যার উত্তরের জন্য ডেটা সাংবাদিক বা ডেটা দলের সামনে বেশ কিছু পথ থাকে। এখানে পথ বলতে ডেটার উৎস বোঝানো হয়েছে। সাংবাদিক কিংবা ডেটা দল, বিভিন্ন উৎস থেকে তাদের প্রয়োজনীয় ডেটাসেট বা তথ্যভাণ্ডার সংগ্রহ করে পারেন।

অনেক ক্ষেত্রে সোর্স নিজে এসে সাংবাদিককে ডেটা সরবরাহ করে, কিংবা ডেটার সম্ভাব্য উৎস সম্পর্কে ধারণা দেয়। পরবর্তীতে যেখানে অনুসন্ধান করে সাংবাদিক ডেটা সংগ্রহ করতে পারেন। আবার অনেক ক্ষেত্রে সাংবাদিক নিজের অনুসন্ধানী দৃষ্টিকোণ থেকে বুঝতে পারেন কোথায় খোঁজ করলে ডেটা পাওয়া যাবে। এভাবে ডেটা প্রাপ্তির নজির কম নেই নতুন ধারার এই সাংবাদিকতায়। তাই একজন ডেটা সাংবাদিককে জানতে হবে বা ধারণা থাকতে হবে কথা গেলে ডেটা পাওয়া যাবে, কারা ডেটা সংরক্ষণ করেন এবং কেন। জেনে নেওয়া যাক, একটি রাষ্ট্রে সাধারণত কোথা থেকে ডেটা সংগ্রহ করা যায়, বা ডেটার সাধারণ উৎসগুলো কী কী।

ডেটার সাধারণ উৎস

নিয়ন্ত্রক সংস্থা

নিয়ন্ত্রক সংস্থাগুলো সংশ্লিষ্ট খাতের ডেটা সংরক্ষণ করে নিজেদের কাজের প্রয়োজনে। নিয়ন্ত্রক সংস্থার কাছে থাকা ডেটা সাংবাদিকতার ক্ষেত্রে বেশ কার্যকর, কেননা এসব সংস্থা তার আওতাধীন প্রতিষ্ঠানগুলোর কার্যক্রম তদারকির জন্যে এসব ডেটা লিপিবদ্ধ করে। ফলে এই ডেটাসেট থেকে জানা যায় কোন প্রতিষ্ঠানের কাজের মান কেমন, কোন প্রতিষ্ঠান কতটা নিয়ম মেনে চলছে বা কী ধরনের অনিয়ম করছে।

ফলে নিয়ন্ত্রক সংস্থাগুলো ডেটা সাংবাদিকতার জন্যে একটি বড় উৎস। যেমন ব্যাংকিং-সংক্রান্ত ডেটার জন্য বাংলাদেশ ব্যাংক, টেলিযোগাযোগ-সংক্রান্ত ডেটার জন্য বাংলাদেশ টেলিযোগাযোগ নিয়ন্ত্রণ সংস্থা (বিটিআরসি) ইত্যাদি।

পরিসংখ্যান বিভাগ

রাষ্ট্রের যে সংস্থা বা প্রতিষ্ঠান, জরিপ, সমীক্ষা বা শুমারি পরিচালনা করে, তারা ডেটার একটি বড় উৎস। অর্থাৎ একজন সাধারণ নাগরিক হিসেবে আপনি যখনই একটি জরিপে অংশ নিচ্ছেন বা সমীক্ষার প্রশ্নপত্র পূরণ করছেন, আপনি জেনে যাচ্ছেন ডেটা কারা সংগ্রহ করছে এবং পরবর্তীতে এ সংক্রান্ত ডেটা কোথায় পাওয়া যাবে। বাংলাদেশে সাধারণ মানুষের কাছ থেকে এমন ডেটা সংগ্রহ করে বাংলাদেশ পরিসংখ্যান ব্যুরো। তবে এর বাইরে সরকারি ও বেসরকারি অনেক প্রতিষ্ঠান-সংশ্লিষ্ট বিষয়ে নির্দিষ্ট জনগোষ্ঠীর ওপরে জরিপ করে থাকে।

কর্তৃপক্ষ

সরকারি বা রাষ্ট্রায়ত্ত সব প্রতিষ্ঠানেরই ডেটা সংগ্রহের নিয়ম বা বিধান রয়েছে। এসব প্রতিষ্ঠান নিজস্ব নিয়মে ডেটা সংগ্রহ এবং সংরক্ষণ করে। সেসব ডেটা হয়তো পরিসংখ্যান আকারে লিপিবদ্ধ থাকে না। তবে সেই ডেটা সহজেই সংখ্যায় রূপান্তর সম্ভব।

বার্ষিক প্রতিবেদন

যেকোনো সংস্থা সারা বছর কী কাজ করলো, তা তাদের বার্ষিক প্রতিবেদনে উল্লেখ করে। ফলে সংশ্লিষ্ট সংস্থার প্রকাশযোগ্য বেশিরভাগ ডেটাই বার্ষিক প্রতিবেদনে পাওয়া যায়। তবে আমাদের দেশে অনেক ক্ষেত্রেই এমন বার্ষিক প্রতিবেদনে ধীরগতি দেখা যায়। অর্থাৎ বছর শেষ হবার অনেক পরে ওই বছরের বার্ষিক প্রতিবেদন প্রকাশ পায়।

মানবসম্পদ বিভাগ

ডেটার আরেকটি বড় উৎস মানব সম্পদবিভাগ। তবে এখান থেকে শুধু প্রতিষ্ঠানের কর্মীদের ডেটা পাওয়া যাবে। সরকারের পাশাপাশি বেসরকারি প্রতিষ্ঠানগুলো থেকে এমন ডেটা সংগ্রহ করার সুযোগ আছে।

নথিপত্র

নথি, দলিল-দস্তাবেজ ডেটা সাংবাদিকতার অন্যতম বড় উৎস। সেই নথি হতে পারে আদালতের, হাসপাতালের, কিংবা সরকারি ও বেসরকারি দপ্তরের। এসব কাগজ থেকে ডেটা সংগ্রহ করে এবং তা সংখ্যায় রূপান্তর করে দেশে বিদেশে শক্তিশালী সব ডেটা প্রতিবেদন তৈরি হয়েছে, হচ্ছে।

উন্মুক্ত ডেটা

যে ডেটা এরইমধ্যে প্রকাশিত হয়েছে এবং যা ব্যবহারে কোনো ধরনের বিধি নিষেধ নেই, এমন ডেটাকে উন্মুক্ত ডেটা হিসেবে চিহ্নিত করা হয়। এই ডেটার বিশ্লেষণ ও তা প্রচারেও কোনো বাধা থাকে না। এমন উন্মুক্ত ডেটা শুধু সাংবাদিকতাকেই সহজ করে না, সাধারণ মানুষের জ্ঞানের পরিধিকে বিস্তৃত ও শক্তিশালী করে। একইসঙ্গে সমাজে তথ্য প্রাপ্তির ক্ষেত্রে সাম্যতা নিশ্চিত হয়।

> সরকারি ও বেসরকারি সব ধরনের সংস্থা বা প্রতিষ্ঠানই উন্মুক্ত ডেটা প্রকাশ করে। এই ডেটা যেমন প্রতিষ্ঠানের ওয়েবসাইটে প্রকাশিত হয়, তেমনি ছাপা কাগজেও পুস্তক আকারে প্রকাশ করা হয়। বিশেষ কিছু ব্যতিক্রম ছাড়া, একবার উন্মুক্ত করা ডেটা তুলে নেয়া হয় না। তবে তা নিয়মিত পরিমার্জন ও পরিবর্ধন করা হয়। যে ব্যক্তি বা প্রতিষ্ঠান ওই ডেটা প্রকাশ করেন, তারা এটি করে থাকেন।

এমন সব ডেটা মূলত সাধারণ মানুষের তথ্য জানার অধিকারের আওতায় প্রকাশিত হয়ে থাকে। যে দেশের সরকার তার জনগণের কাছে যত বেশি স্বচ্ছ ও জবাবদিহিমূলক, সেখানে উন্মুক্ত ডেটার পরিমাণ ততো বেশি। সবমিলে উন্মুক্ত ডেটা বৃহত্তর জনসচেতনতা তৈরি এবং সরকারকে কার্যকর, দায়িত্বশীল ও জবাবদিহি করতে সহায়তা করে। তবে সাধারণ মানুষ উন্মুক্ত এমন উৎস থেকে প্রয়োজনীয় ডেটা সংগ্রহ করে ঠিকই, কিন্তু ওই ডেটার ব্যাখ্যা বা তার পেছনের খবর জানতে গণমাধ্যমের ওপর নির্ভর করে। কেননা মূল সেসব ডেটাতে শুধু সংখ্যা বা পরিসংখ্যান থাকে, যা সাধারণ পাঠকের জন্যে বোঝা বা বিশ্লেষণ করা অনেক ক্ষেত্রেই কঠিন হয়ে পড়ে। গণমাধ্যম সেসব পরিসংখ্যানের সহজ ব্যাখ্যা তুলে ধরে এবং জনসাধারণকে নিজস্ব ব্যাখ্যা বা দৃষ্টিভঙ্গি নির্ধারণে সহায়তা করে।

> উন্মুক্ত ডেটা সাংবাদিকতার জন্য অত্যন্ত গুরুত্বপূর্ণ। এমন ডেটা, একজন প্রতিবেদককে প্রাসঙ্গিক ডেটা সনাক্ত করতে সহায়তা করে এবং দীর্ঘমেয়াদি অনুসন্ধানের দিকে পরিচালিত করে। বিশেষকরে বিশ্বজুড়ে ডেটা সাংবাদিকতা বিকাশের ক্ষেত্রে উন্মুক্ত ডেটা বড় একটি ভূমিকা রাখছে। ডেটা সাংবাদিকেরা এই ডেটার ওপর তাদের বিশ্লেষণ দক্ষতা ও প্রযুক্তির ব্যবহারের মাধ্যমে নতুন তথ্য বা সংবাদের সৃষ্টি করছে। যা শুধু ওই ডেটা প্রতিবেদনকেই শক্তিশালী করছে না, সাংবাদিকতার ওপর মানুষের আস্থা ও বিশ্বাস বাড়াতে বড় আবদান রাখছে।

সাংবাদিকরা শুধুমাত্র নিজের দেশের সরকার প্রকাশিত উন্মুক্ত ডেটা ব্যবহার করেই থেমে থাকছে না। যেহেতু এসব ডেটা ইন্টারনেটে প্রকাশিত, তাই যে কোনো সাংবাদিক বিশ্বের যে কোনো প্রান্তে বসে অন্য দেশ কিংবা জাতিসংঘ, বিশ্বব্যাংক, ইউরোপীয় ইউনিয়নের মতো বহুজাতিক সংস্থা প্রকাশিত তথ্য সংগ্রহ এবং ব্যবহার করছে। প্রতিটি দেশ এবং সংস্থাই নিয়মিত বিরতিতে তাদের প্রকাশিত উন্মুক্ত ডেটা হালনাগাদ করে।

বৈশ্বিক ডেটাবেস খুঁজে পেতে পারেন ওয়ার্ল্ড ব্যাংক ডেটা,[১৫] জাতিসংঘ,[১৬] ইন্টারন্যাশনাল মনিটারি ফান্ড (আইএমএফ)[১৭]–এর মতো সংস্থার কাছ থেকে। বাংলাদেশ, দক্ষিণ আফ্রিকা[১৮] ও যুক্তরাষ্ট্রের মতো জাতীয় পর্যায়ের বিভিন্ন ডেটাবেসও আপনি খুঁজে পেতে পারেন। বিশ্ববিদ্যালয়গুলোও তৈরি করে আওয়ার ওয়ার্ল্ড ইন ডেটা-র মতো ডেটাসেট। প্রোপাবলিকার[১৯] মতো সংবাদমাধ্যম, স্ট্যাটিসটা[২০] ও গ্লোবাল পেট্রোল প্রাইসেস[২১]–এর মতো ব্যক্তিমালিকানার প্রতিষ্ঠানও বিভিন্ন ডেটাসেট তৈরি ও ব্যবস্থাপনা করে।

> উন্মুক্ত ডেটার ক্ষেত্রে সাংবাদিক বা প্রতিবেদককে কিছুটা সচেতন থাকতে হয়। কেননা ইন্টারনেট বা অনলাইনে নতুন ও পুরাতন সব ধরনের ডেটাই থাকে। সেক্ষেত্রে প্রতিবেদককে হালনাগাদ তথ্য খুঁজতে হয়। কোন তথ্য খুজছি সে বিষয় পরিষ্কার ধারণা থাকা খুব গুরুত্বপূর্ণ। একইসঙ্গে উন্মুক্ত ডেটার সত্যতা এবং বিশ্বাসযোগ্যতা যাচাই করে নেয়া প্রয়োজন।

সেইসাথে অনলাইন থেকে প্রয়োজনীয় ডেটাটি খুঁজে বের করাও একটি দক্ষতার বিষয়। প্রযুক্তির এই যুগে অনলাইনে উন্মুক্ত ডেটার সংখ্যা প্রতি মুহূর্তে বাড়ছে। মাঝে মাঝে মনে হবে এত তথ্যের মাঝে আপনি নিজেই তলিয়ে যাচ্ছেন। ফলে যে ডেটা খুঁজতে আমরা অনলাইনে বসি, অনেক সময়ই সেটা খুঁজে পাওয়া কষ্টকর হয়ে ওঠে। তবে সহজে ডেটা খুঁজে পাবার কিছু কৌশলও রয়েছে। যেগুলা অনুসরণ করলে, দ্রুত কাঙ্ক্ষিত তথ্য বা ডেটাটি খুঁজে পাওয়া যাবে বলে আশা করা যায়।

- বিষয়ের ওপর কোন কোন প্রতিষ্ঠান ডেটা সংগ্রহ করে, সে সম্পর্কে ধারণা নিন।
- আগে জানুন, গুগল অনুসন্ধানে কোন কি-ওয়ার্ড লিখলে কাঙ্ক্ষিত তথ্য মিলবে।
- তথ্যের ফুটনোট, রেফারেন্স এবং সংজ্ঞা দেখুন।
- পুরোনো খবর পড়ে সেখানে কোন কোন উৎসের উল্লেখ রয়েছে তা নোট করুন।
- প্রয়োজনে সরাসরি সংশ্লিষ্ট সংস্থা, ব্যক্তি বা প্রতিষ্ঠানের সঙ্গে যোগাযোগ করুন।

[১৫] World Bank Open Data | Data. (n.d.-b). Retrieved October 12, 2022, from https://data.worldbank.org/

[১৬] UNdata. (n.d.-b). Retrieved October 12, 2022, from https://data.un.org/

[১৭] IMF Data. (n.d.-b). IMF. Retrieved October 12, 2022, from https://www.imf.org/en/Data

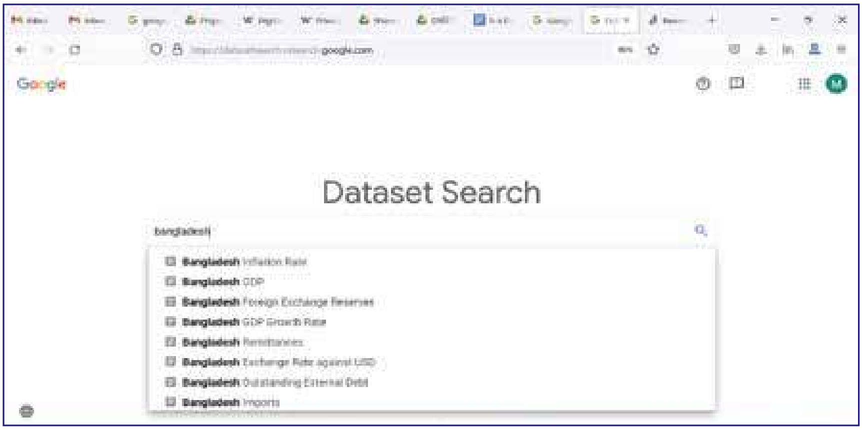
[১৮] Africa, S. S. (n.d.). Statistics South Africa | The South Africa I Know, The Home I Understand. Retrieved October 12, 2022, from https://www.statssa.gov.za/

[১৯] ProPublica Datastore. (2022b, February 2). ProPublica Data Store. Retrieved October 12, 2022, from https://www.propublica.org/datastore/

[২০] Statista. (n.d.-b). Statista - The Statistics Portal. Retrieved October 12, 2022, from https://www.statista.com/

[২১] Download gasoline and diesel price data. (n.d.-b). GlobalPetrolPrices.com. Retrieved October 12, 2022, from https://www.globalpetrolprices.com/data/

গুগল থেকে ডেটা



ছবি: গুগল ডেটাসেট সার্চ

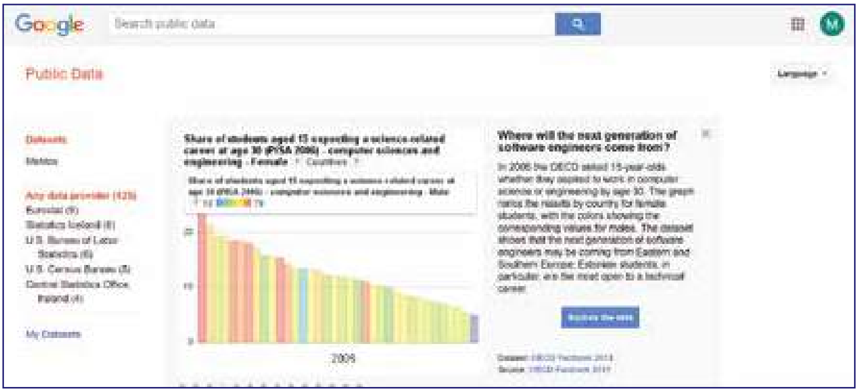
অনলাইনে ডেটার একটি বড় উৎস গুগল ডেটাসেট সার্চ।[২২] এটি গুগলের একটি পৃথক সেবা। এই সেবাটি মূলত গবেষনা ও সাংবাদিকতাকে সহায়তা করার উদ্দেশ্যে তৈরি। এর মাধ্যমে আপনি উন্মুক্ত ডেটা খুব সহজে সংগ্রহ করতে পারবেন। এই সার্চ ইঞ্জিনে আপনি নিদির্ষ্ট শর্ত জুড়ে দিয়ে অর্থাৎ ফিল্টারের মাধ্যমে ডেটা খুঁজতে পারবেন।

> প্রথমেই সাধারণ গুগল সার্চের মতো করে আপনার কাঙ্ক্ষিত ডেটা সংশ্লিষ্ট শব্দটি/গুলো লিখুন। এরপর ফলাফলের সংখ্যাটি দেখে নিন। ডেটাসেট সম্পর্কে বিস্তারিত দেখতে প্রতিটি এন্ট্রিতে ক্লিক করুন। একইসঙ্গে ডেটাসেটটি আপনার ব্যবহারের ক্ষেত্রে উপযুক্ত কি-না তা দেখতে বিবরণ পড়ুন।

প্রয়োজনীয় ডেটাটি সহজে পেতে, কিংবা কিছু নির্দিষ্ট উৎস থেকে তথ্য পেতে site: সংযুক্ত করে সংশ্লিষ্ট শব্দ লিখুন। এর বাইরে অনুসন্ধানকে আরো নির্দিষ্ট করতে গুগল ডেটাসেটও সময়ভিত্তিক বেশকিছু অপশন বা ফিল্টার আপনার সামনে তুলে ধরবে। প্রয়োজনে সেগুলো ব্যবহার করুন।

> কিছু কার্যকর ফিল্টারের মধ্যে একটি Last Updated বা সর্বশেষ/হালনাগাদ তথ্য। আপনি কি ধরনের ডেটা চান, অর্থাৎ লেখা, টেবিল না-কি ছবি সেটিও নির্দিষ্ট করে দেয়ার সুযোগ আছে এই অনুসন্ধানে। চাইলে শুধু পিডিএফ ফাইল বা অন্য কোনো ফরম্যাটের ফাইলের জন্যও আলাদা করে অনুসন্ধান করা যাবে। তখন ওই ফরম্যাট ছাড়া অন্য কোনো ফরম্যাটের ডেটা আপনার সামনে আসবে না। এই সার্চ ইঞ্জিনের সবচেয়ে ইতিবাচক দিক হচ্ছে আপনি ডেটা ব্যবহারের মেধাসত্ত্ব যাচাই করে নিতে পারবেন, ডেটা সংগ্রহের আগেই। যদি কোনো ডেটার বাণিজ্যিক ব্যবহারে নিষেধাজ্ঞা থাকে, তবে আপনি তা ফিল্টারের মাধ্যমে আলাদা করতে পারবেন। এমন একাধিক ফিল্টার এই সার্চ ইঞ্জিনে আপনি এক সাথেও ব্যবহার করতে পারবেন।

[২২] Dataset Search. (n.d.-b). Google Dataset Search. Retrieved October 12, 2022, from https://datasetsearch.research.google.com



ছবি: গুগল ডেটাসেট সার্চ

এর বাইরে গুগল পাবলিক ডেটা এক্সপ্লোরারের[২৩] মাধ্যমে আপনি আর্ন্তজাতিক সংস্থা ও প্রতিষ্ঠানগুলোর উন্মুক্ত ডেটা সহজেই খুঁজে বের করতে পারবেন।

ডেটার বিকল্প উৎস

ডেটা যখন উন্মুক্ত উৎস হতে পাওয়া যায় না, তখন প্রতিবেদক আর ডেটা দলকে বিকল্প ভাবতে হয়, নতুন উৎস খুঁজতে হয়। উৎস বা ডেটা তৈরি করতে হয়। নিজস্ব ডেটা তৈরির জন্য সাংবাদিকেরা যেসব পদ্ধতি অনুসরণ করে এবার তা আলোচনা করব। সেই সঙ্গে এইসব উৎস থেকে ডেটা সংগ্রহের ক্ষেত্রে যেসব সতর্কতা অনুসরণ করা দরকার সেগুলোও জানব।

তথ্য চেয়ে আবেদন

উন্মুক্ত উৎস থেকে প্রয়োজনীয় ডেটা না পেলে, সংশ্লিষ্ট কর্তৃপক্ষের কাছে তথ্য চেয়ে আবেদন করার অধিকার সাংবাদিকের আছে। বাংলাদেশসহ যে সব দেশে তথ্য অধিকার আইন আছে, সেসব দেশে তথ্যের জন্য যে কেউ আবেদন করতে পারবেন। তবে সেজন্য আপনাকে জানতে হবে সরকারের কোন দপ্তর কি কাজ করে এবং আপনার প্রয়োজনীয় তথ্যটি সরকারের কোন সংস্থা বা প্রতিষ্ঠানের কাছে জমা রয়েছে। কীভাবে তথ্য চেয়ে আবেদন করতে হবে অর্থাৎ আবেদন ফরম কোথায় পাওয়া যায়, কী লিখতে হয়, এসব বিষয়ে বিস্তারিত প্রস্তুতি রাখতে হবে। এ বিষয়ে বিস্তারিত আছে, এই গ্রন্থের 'তথ্য পেতে আরটিআই' অধ্যায়ে।

জরিপ পরিচালনা

ডেটা সাংবাদিকতায় বেশিরভাগ ক্ষেত্রেই বড় ডেটাসেটের প্রয়োজন হয়। অনেক প্রতিবেদন তৈরি করতে গিয়ে যখন উন্মুক্ত উৎস হতে ডেটা পাওয়া যায় না, তখন বিকল্প উৎস হিসেবে জরিপ সামনে আসে। অর্থাৎ ডেটা দল বা সংবাদ মাধ্যম নিজেরাই জরিপ পরিচালনা করে ডেটা সংগ্রহ করে। জরিপ সঠিকভাবে সম্পন্ন করা গেলে অনেক ধরনের ডেটা পাওয়া যায়। ফলে প্রতিবেদন আরও টেকসই হয়।

[২৩] Google Public Data Explorer. (n.d.-c). Retrieved October 12, 2022, from https://www.google.com/publicdata/directory

সাংবাদিকতায় জরিপের একটি ভালো উদাহরণ হতে পারে, যশোরভিত্তিক দৈনিক গ্রামের কাগজের "মাতৃত্বকালীন ভাতা নিয়ে অনুসন্ধানটি। কিউআর কোড স্ক্যান করে পড়ে নিতে পারেন সেই কেস স্টাডি।	

ডেটা জার্নালিজম হ্যান্ডবুক ২.০ বইয়ে ক্রিনা গ্যাব্রিয়েলা বোরোস জরিপ পরিচালনার আগে কয়েকটি বিষয় মাথায় রাখার পরামর্শ দিয়েছেন[২৪]–

- জরিপের অংশগ্রহণকারী হিসেবে শুধু আপনার পরিচিতদের বেছে নিলে হবে না। "প্রতিনিধিত্ব" নিশ্চিত করতে হলে তাতে বিভিন্ন শ্রেণী, বয়স, লিঙ্গ, ও ভৌগলিক বৈচিত্রের উপস্থিতি লাগবে।

- অংশগ্রহণকারীদের বাছাই করতে হবে দৈবচয়নের ভিত্তিতে। অংশগ্রহণকারী বাছাইয়ের একটি সুনির্দিষ্ট পদ্ধতি থাকতে হবে। তা না হলে আপনার জরিপের ফলাফল বিভ্রান্তিকর হতে পারে।

- প্রতিনিধিত্বশীল করতে হলে জরিপে অংশগ্রহণকারীর সংখ্যার একটি ন্যুনতম মানে পৌঁছাতে হবে। রাওসফট, সার্ভে মাংকি[২৫] এবং সার্ভে সিস্টেমসের[২৬] মতো প্রতিষ্ঠানগুলোর অনলাইন ক্যলকুলেটর আছে। সেখান থেকে দেখে নিন আপনি যে বিষয় বা জনগোষ্ঠী নিয়ে কাজ করছেন, তার জন্য ন্যুনতম কী পরিমাণ নমুনা লাগবে।

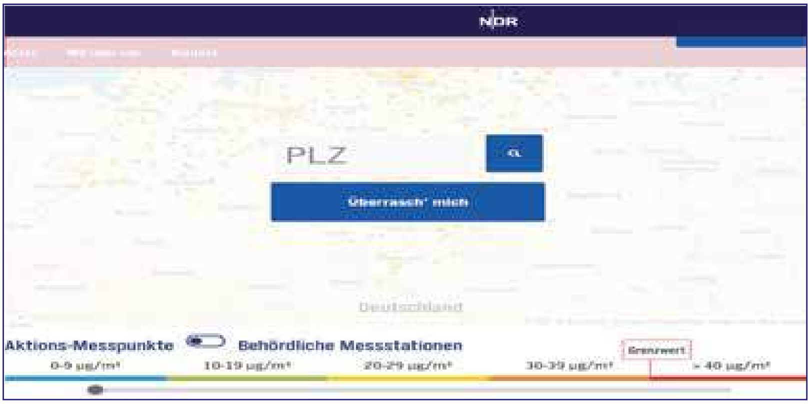
- জরিপ নিয়ে উচ্চাভিলাষী হওয়ার আগে ভাবুন। কারণ, আকার ও পরিধি বেশি বড় হয়ে গেলে সেটি সামাল দেওয়া সাংবাদিক হিসেবে আপনার জন্য কঠিন হয়ে দাঁড়াতে পারে।

- আপনার জরিপপদ্ধতি প্রতিবেদনে সঠিকভাবে ব্যাখ্যা করে বলুন। কাদের বাছাই করেছেন, কীভাবে এবং কেন এই জবাবগুলো পাঠককে জানাতে হবে। অনেকগুলো সাক্ষাৎকার নেওয়া মানেই জরিপ নয়। স্বচ্ছ থাকুন।

জরিপের পাশাপাশি সংশ্লিষ্ট কোনো নথি পাওয়া যায় কি-না সে বিষয়ে নজর রাখুন। জরিপের মাধ্যমে সংগ্রহকৃত ডেটা সঠিক কি-না, সেটি প্রশ্ন করে যাচাই করতে হবে। যদি পদ্ধতি নিয়ে আত্মবিশ্বাসী না হন, তাহলে বিশেষজ্ঞের পরামর্শ নিয়ে জরিপপদ্ধতি দাঁড় করান এবং সেটি মেনে চলুন।

[২৪] Data Journalism.com. (2021, August 18). Running Surveys for Investigations. Retrieved October 12, 2022, from https://datajournalism.com/read/handbook/two/assembling-data/mapping-pollution-in-indian-cities
[২৫] Sample Size Calculator by Raosoft, Inc. (n.d.-b). Retrieved October 12, 2022, from http://www.raosoft.com/samplesize.html
[২৬] Sample Size Calculator - Confidence Level, Confidence Interval, Sample Size, Population Size, Relevant Population - Creative Research Systems. (n.d.-b). Retrieved October 12, 2022, from https://www.surveysystem.com/sscalc.html

ক্রাউডসোর্সিং এই পদ্ধতিতে ডেটা সাংবাদিক বা গণমাধ্যম সাধারণ মানুষের কাছ থেকে ডেটা সংগ্রহ করে। এখানে সাধারণ মানুষ বলতে একজন দুইজন ব্যক্তি নয়। এই প্রক্রিয়ার মাধ্যমে একটি বড় জনগোষ্ঠীর কাছ থেকে ডেটা সংগ্রহ করা হয়। তাদের কাছে নির্দিষ্ট একটি বিষয়ে ডেটা আহ্বান করা হয়। বিপরীতে সাধারণ মানুষ, ওই কার্যক্রমে অংশ নিয়ে তাদের কাছে থাকা ডেটা, এমনকি অভিজ্ঞতা জানান। পুরো প্রক্রিয়া বা কার্যক্রমটি পরিচালিত হয় অনলাইনে। সোশ্যাল মিডিয়া ব্যবহার করে ক্রাউডসোর্সিং এখন ডেটা সাংবাদিকতায় ডেটা সংগ্রহের একটি বহুল ব্যবহৃত এবং কার্যকর উদ্যোগ। এর মাধ্যমে গণমাধ্যমের সাথে সাধারণ মানুষের সম্পর্ক আরও গভীর হয়। গণমাধ্যমের ওপর মানুষের আস্থা বাড়ে। শুধু ডেটা সংগ্রহের জন্য নয়, ডেটা যাচাই এবং সেখান থেকে গুরুত্বপূর্ণ সংবাদ উপকরণ খুঁজে বের করতেও বিশ্বখ্যাত অনেক গণমাধ্যম এখন ক্রাউডসোর্সিং করে থাকে।



ইনফোগ্রাফিক: এনডিআর

জার্মান গণমাধ্যম এনডিআর দেশটির উত্তরাঞ্চলে বাতাসে ক্ষতিকর নাইট্রোজেন ডাইঅক্সাইডের পরিমাণ জানতে সাধারণ মানুষের প্রতি আহ্বান জানিয়েছিল।[২৭] ৫ হাজার মানুষ তাতে সাড়া দেয় এবং গণমাধ্যমটি তাদের মধ্য থেকে ১ হাজার জনের কাছে নমুনা সংগ্রহের উপকরণ পাঠিয়ে দেয়। তারা নমুনাগুলো ল্যাবে পাঠিয়ে বাতাসে ক্ষতিকর নাইট্রোজেন ডাইঅক্সাইডের পরিমাণ নির্ধারণ করে এবং সেই ডেটা দিয়ে একটি মানচিত্র তৈরি করে। সেই মানচিত্রে প্রবেশ করে জার্মানির উত্তরাঞ্চলের যে কেউ জানতে পারবেন, তার এলাকার বাতাস কতটা দূষিত।

তবে একইসাথে ক্রাউডসোর্সিং ব্যবহার করে সাংবাদিকতার ঝুঁকিও রয়েছে। তাই কিছু বিষয়ে ডেটা সাংবাদিক এবং গণমাধ্যমকে আগে থেকেই সতর্ক থাকতে হবে। প্রথমেই এমন পদ্ধতি খুঁজে বের করতে হবে, যার মাধ্যমে আপনি জানতে পারবেন ক্রাউডসোর্সিংয়ে অংশ নেওয়া মানুষেরা প্রকৃত কি-না। কেননা এমন কার্যক্রমে প্রায়শই ভুয়া মানুষ বা বিষয়ের সাথে সম্পৃক্ত নন এমন মানুষকে অংশ নিতে দেখা যায়। যেহেতু পুরো প্রক্রিয়াটি অনলাইনে পরিচালিত হয়, তাই আগে থেকেই যে কোনো ধরনের

[২৭] *Was atmest du? Das Ergebnis der großen Luftmessaktion.* (2019, February 24). NDR. Retrieved October 12, 2022, from https://www.ndr.de/fernsehen/sendungen/45_min/Die-groesste-Luft-Messaktion-im-Norden,luft132.html

সাইবার হামলা থেকে তথ্যের নিরাপত্তা নিশ্চিত করতে হবে। পিছিয়ে পড়া বা প্রান্তিক জনগোষ্ঠী সকল শ্রেণি পেশার মানুষের অংশগ্রহণ নিশ্চিত করতে হবে। ক্রাউডসোর্সিংয়ের মাধ্যমে সংগ্রহ করা তথ্যের বিশ্বাসযোগ্যতা কীভাবে নিশ্চিত হবে, সেটিও আগে নির্ধারণ করে রাখতে হবে সাধারণ মানুষের ব্যাক্তিগত তথ্য যেন অন্য কারো হাতে চলে না যায়, সে বিষয়ে নজরদারি বাড়াতে হবে কিংবা বাড়তি সতর্কতা অবলম্বন করতে হবে।

সামাজিক যোগাযোগ মাধ্যম

সোশ্যাল মিডিয়া এখন সাংবাদিকতার বড় সোর্স বা তথ্যের উৎসে পরিণত হয়েছে। ফেসবুক থেকে শুরু করে ইন্সটাগ্রাম, টুইটার, স্নাপচ্যাট ব্যবহারকারিদের স্ট্যাটাস, ছবি, আর ভৌগলিক অবস্থান অনুসরণ করে ডেটা সংগ্রহ করা যায়। এর বাইরে লিঙ্কড-ইনও, অনলাইন অনুসন্ধান এবং ডেটা সংগ্রহের বড় উৎস। বিশ্বের অনেক ভাল এবং শক্তিশালী অনুসন্ধানী ডেটা প্রতিবেদন তৈরি হয়েছে শুধু সামাজিক যোগাযোগ মাধ্যমে থাকা ডেটার ওপর ভিত্তি করে।

সোশ্যাল মিডিয়া থেকে ডেটা সংগ্রহের সময় যা মাথায় রাখবেন

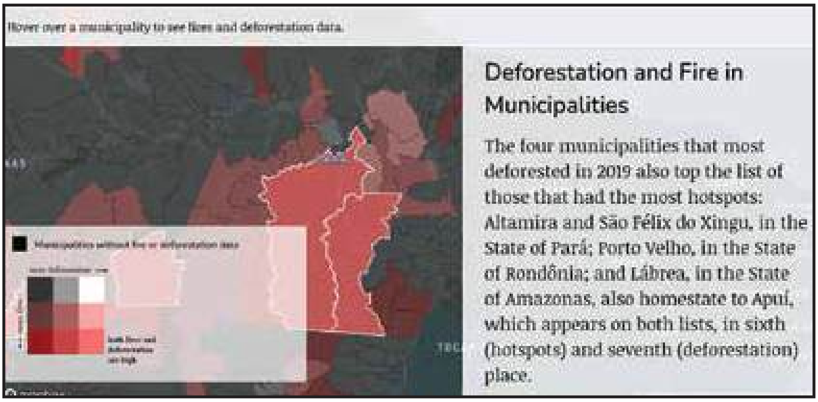
- যে বিষয়টি খুঁজছেন, সেটি সংশ্লিষ্ট কি-ওয়ার্ডগুলো বাছাই করুন এবং টুকে রাখুন;
- কোন কোন প্ল্যাটফর্মে খুঁজবেন (ফেসবুক, টুইটার না অন্যকিছু) আগে ঠিক করুন;
- কোন প্ল্যাটফর্মে সার্চের জন্য কোন টুলটি ভালো তা আগে জেনে নিন এবং চর্চা করুন। (যেমন: ফেসবুকের জন্য হতে পারে হু পোস্টেড হোয়াট,[২৮] টুইটারের জন্য টুইপসম্যাপ[২৯] বা ফলোয়ারওঙ্ক[৩০]। প্রতিনিয়ত এসব টুল বদলাচ্ছে। তাই আগে খোঁজ করুন;)
- কন্টেন্ট কীভাবে ডেটাবেজে সংরক্ষণ করবেন আগে ঠিক করে নিন;
- যদি বিশ্লেষণটি কন্টেন্টনির্ভর হয় তাহলে প্রতিটি কন্টেন্টের সত্যতা যাচাই করুন;
- যাদের তথ্য সংগ্রহ করছেন তাদের প্রাইভেসি সম্পর্কে সতর্ক ও শ্রদ্ধাশীল থাকুন;
- ছবির ক্ষেত্রে অবশ্যই উৎস নিশ্চিত হয়ে নিন। কারণ প্রচুর ভুয়া ছবি বের হয়;

এই মাধ্যম থেকে ডেটা সংগ্রহের সবচেয়ে বড় চ্যালেঞ্জ হল কৌশল জানা। নির্দিষ্ট ব্যক্তির বিষয়ে অনুসন্ধানের বাইরে যখন একটি বড় জনগোষ্ঠীর কর্মকাণ্ড আপনি সোশ্যাল মিডিয়াতে অনুসরণ করতে চাইবেন, তখন প্রথমেই নির্দিষ্ট কিছু লক্ষ্য নির্ধারণ করতে হবে। যেমন– এলাকা, বয়সসীমা, পেশা ইত্যাদি। এভাবে অনুসন্ধানের ক্ষেত্র আরও ছোট ও নির্দিষ্ট করে আনা যায়। সামাজিক যোগাযোগ মাধ্যমে আপনি অনেক তথ্য পাবেন যার সবগুলো হয়তো আপনার প্রতিবেদনের সঙ্গে সঙ্গতিপূর্ণ নয়। সেগুলো বাদ দেয়াটাও একটি দক্ষতার কাজ।

[২৮] *Who posted what?* (n.d.-b). Retrieved October 12, 2022, from https://whopostedwhat.com
[২৯] Tweepsmap. (n.d.-b). *Tweepsmap - Deep Twitter Analytics and Publishing.* Retrieved October 12, 2022, from https://tweepsmap.com/
[৩০] *Followerwonk: Tools for Twitter Analytics, Bio Search and More.* (n.d.-b). Retrieved October 12, 2022, from https://followerwonk.com/

ড্রোন এবং স্যাটেলাইট

ছবি থেকে ডেটা সংগ্রহের প্রচলন শুরু হয়েছে অনেক আগে থেকেই। তবে গেল দশ বছরে বৈশ্বিক সাংবাদিকতায় ডেটার জন্য সেই ছবি সংগ্রহের প্রযুক্তিতে বৈপ্লবিক পরিবর্তন এসেছে। এখন ছবি সংগ্রহ করা হচ্ছে স্যাটেলাইট আর ড্রোন ক্যামেরার মাধ্যমে। পাখির চোখের সেই ছবিতে এখন অনেক কিছুই তুলে আনা সম্ভব হচ্ছে, যা সমতলে তোলা ছবিতে সম্ভব না। আবার অনেক স্থানে ছবি তোলার সুযোগই থাকে না, যা ড্রোন এবং স্যাটেলাইটের মাধ্যমে সহজেই করা সম্ভব। স্যাটেলাইট থেকে তথ্যও সংগ্রহের সবচেয়ে বড় সুবিধা আপনি রিয়েলটাইম তথ্য পাবেন এবং আপনি চলমান ঘটনার পরিবর্তনটি দেখতে পারবেন। ফলে যেসব স্থানে তথ্য সংগ্রহের জন্যে আপনি সরাসরি পৌঁছাতে পারবেন না, স্যাটেলাইট আপনাকে সেখানে নিয়ে যাবে। সময়ের সাথে সাথে সংশ্লিষ্ট বিষয়ের পরিবর্তনের ডেটা দেবে।



ইনফোগ্রাফিক: অ্যাম্বিয়েন্টাল মিডিয়া

অ্যাম্বিয়েন্টাল মিডিয়া তাদের স্মোক স্ক্রিন প্রজেক্টে স্যাটেলাইট ডেটার সতর্ক বিশ্লেষণ করে দেখিয়েছে, আমাজন বনে আগুন লাগার মূল কারণ মূলত বন দখল করে মানুষের বসতি গড়া। প্রতিবেদন অনুযায়ী, ২০১৯ সালে বনটির যে চারটি এলাকায় সবচেয়ে বেশি আগুন লাগার ঘটনা দেখা গেছে তার ৭২ শতাংশই ব্যক্তিমালিকানাধীন। তারা আরো দেখিয়েছেন যে আদ্রতা, ঘন কুয়াশা ও মসজাতীয় উদ্ভিদের উপস্থিতির কারণে সেখানে আগুন লাগার কথা নয়। কিন্তু বন উজাড়ের কারণে সেই বৈশিষ্ট্য থাকছে না এবং আগুন লাগছে।

ড্রোন থেকেও একই ধরনের সুবিধা পাওয়া যাবে, সীমিত পরিসরে। তবে ড্রোনের মাধ্যমে ছবি তোলার আগে অন্য কোনো ভাবে ঘটনা বা বিষয়বস্তুর ছবি তোলা যায় কি-না সেটা যাচাই করে দেখতে হবে। কেননা, সব দেশেই ড্রোন পরিচালনার নির্দিষ্ট কিছু বিধিমালা আছে, সেগুলো আপনি ভাঙছেন কি-না সে বিষয়ে সতর্ক থাকতে হবে। এতো আয়োজন এবং অর্থ ব্যয় করে আপনি যে ছবি তুলছেন তার সংবাদমূল্য কতটা, সেই ছবি বা ডেটা আপনার প্রতিবেদনকে কতটা সমৃদ্ধ করে সেটিও বুঝতে হবে। একইসঙ্গে ড্রোন থেকে তোলা ওই ছবি কীভাবে বিশ্লেষণ এবং ডেটায় পরিমাপ করা হবে, সেই পদ্ধতি আগে চূড়ান্ত করতে হবে।

ডেটা ইমপোর্ট

ডেটা সংগ্রহ করে মাইক্রোসফট এক্সেল বা গুগল শিটে সংরক্ষণ করা হয়। কেননা, এই দুই টুলস এর যে কোনো একটি ব্যবহার করেই সাধারণত ডেটার বিশ্লেষণ করা হয়ে থাকে। দুইটি টুলস প্রায় একই রকম। এই অনুশীলনে গুগল শিট ব্যবহার করা হবে।

BangladeshCSV.txt নামের একটি টেক্স ফাইল ডেটা সাংবাদিকতা সহায়িকার গুগল ড্রাইভে পাওয়া যাবে। কিউআর কোডটি স্ক্যান করে টেক্সট ফাইলটি পাবেন।

চলুন, গুগল শিটে ডেটা যুক্ত করার মধ্য দিয়ে ডেটা সাংবাদিকতার প্রথম পর্যায়ের অনুশীলন শুরু করি। এ জন্য দেখতে হবে আপনি কোন মাধ্যমে ডেটাটি সংগ্রহ করেছেন বা পেয়েছেন। আপনি যদি কাগুজে নথি বা হাতে লেখা ডেটা সংগ্রহ করে থাকেন, তবে সেগুলোকে টাইপ করে গুগল শিটে যুক্ত করতে হবে। যা কিছুটা কষ্টসাধ্য এবং সময়সাপেক্ষ। তবে ডেটার পরিমাণ যদি কম হয়, আপনি সেটি সহজেই করতে পারবেন।

> যদি আপনার সোর্স গুগল শিট বা মাইক্রোসফট এক্সেলে ডেটা সংরক্ষণ করে এবং তা আপনাকে সরবরাহ করেন, তাহলে আপনি ভাগ্যবানদের একজন। আপনাকে তাহলে গুগল শিটে ডেটা যুক্ত করতে বাড়তি কিছু করতে হবে না। ফাইলটি গুগল ড্রাইভ আপলোড করলেই হবে।

তবে বেশিরভাগ ডেটা সাংবাদিকতার ক্ষেত্রে এতো সৌভাগ্যবান হওয়া যায় না। আপনাকে হয়তো বিভিন্ন ফরম্যাটের ডেটা নিয়ে কাজ করতে হবে। কখনও সেটি টেক্সট, সিএসভি কখনো বা ট্যাব। ভাল খবর হচ্ছে টেক্সট ফাইলগুলো প্রোগ্রামের মাধ্যমে কনভার্ট বা রূপান্তর করা যায় এবং গুগল শিট বা এক্সেলে যুক্ত করা যায়।

> এবার BangladeshCSV.txt নামের টেক্সটটি আপনার গুগল শিটে যুক্ত করতে ফাইল মেন্যু থেকে ইম্পোর্ট অপশনে ক্লিক করুন। তবে টেক্সট ফাইলগুলো দেখতে আপনার কম্পিউটারে একটি টেক্সট এডিটর টুলস বা সফটওয়্যারের প্রয়োজন হবে।

টেক্সট এডিটর

টেক্সট এডিটরের মাধ্যমে যেকোনো ধরনের ফাইল দেখা যায়। যেহেতু ডেটা সাংবাদিকতায় বিভিন্ন উৎস থেকে সংগ্রহ করা ডেটা নিয়ে কাজ করা হয়, তাই কম্পিউটারে এক বা একাধিক টেক্সট এডিটর থাকা প্রয়োজন।

সব কম্পিউটারেই একটি করে টেক্সট এডিটর থাকে। উইন্ডোজ কম্পিউটারের ক্ষেত্রে নোটপ্যাড আর ম্যাকে থাকে টেক্সট এডিট। দুটোই প্রাথমিক বা শুরুর দিককার কাজের জন্যে উপযোগী এবং কার্যকর। কিন্তু ডেটা সাংবাদিকতার বিভিন্ন ক্ষেত্রে যখন আপনি আরও ভিন্ন ভিন্ন ফরম্যাট বা ধরনের টেক্সট ডেটা পাবেন, তখন কিছুটা উন্নত মানের টেক্সট এডিটরের প্রয়োজন হবে। যেখানে আপনি ফাইলটি সম্পর্কে বিস্তারিত জানতে পারবেন। এমনকি কিছু ক্ষেত্রে ক্লিনিং এবং ফরম্যাট বদলের কাজও করার সুযোগ থাকবে।

অনলাইন থেকে বিনামূল্যে ডাউনলোড করা যাবে এমন কিছু টেক্সট এডিটর

- সাবমাইল টেক্সট (উইন্ডোজ এবং ম্যাক)
- নোটপ্যাড++ (উইন্ডোজ)
- বিবিএডিট (ম্যাক)

টেক্সট ডিলিমিটারস এবং কোয়ালিফায়ারস

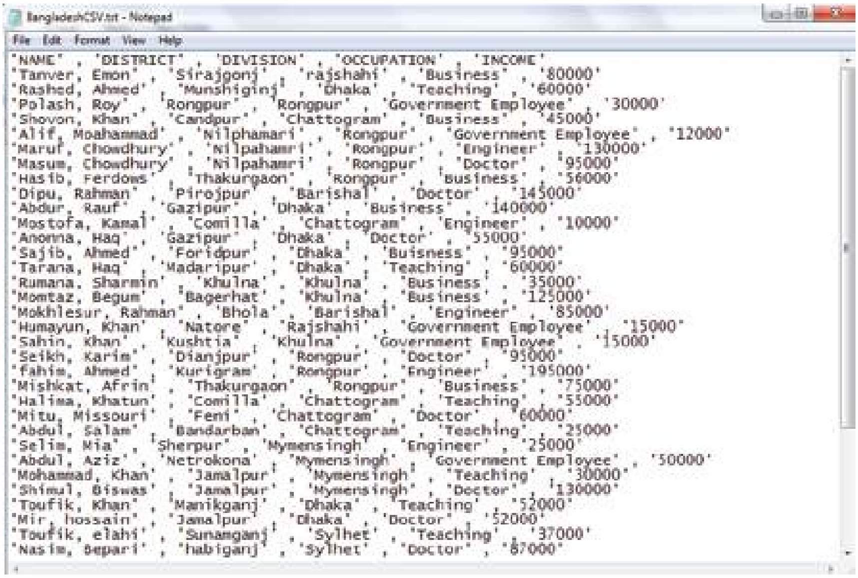
টেক্সট ফাইল থেকে ডেটা গুগল শিটে স্থানান্তরের প্রথম পর্যায়ে আপনাকে টেক্সট এডিটর খুলে ডেটা বা তথ্যগুলো দেখে নিতে হবে। যখন প্রথমবারের মতো ডেটাগুলো দেখবেন, তখন আপনার কাছে সেটিকে তথ্যের জঙ্গল মনে হতে পারে। কিন্তু তার মাঝেও একটি কাঠামো থাকবে। শুধু জানতে হবে, আপনি কোন ডেটাটি খুঁজছেন।

প্রথমেই আপনাকে দেখতে হবে ফাইলে থাকা ডেটাগুলোর কাঠামো কোথায়। কিভাবে সাজানো হয়েছে। কোন যতিচিহ্ন দিয়ে একটি ডেটাকে অন্যটির থেকে আলাদা করেছে। বেশিরভাগ ক্ষেত্রেই ফাইলে থাকা ডেটা কোনো না কোনো কাঠামোতে ভাগ করা থাকে। একটি ডেটাকে আরেকটি ডেটা থেকে আলাদা করতে সবচেয়ে বেশি ব্যবহৃত হয় কমা (,)। এই টেক্সট ফাইলগুলোর নামও তাই, কমা সেপারেটেড ভ্যালুস ফাইলস বা সিএসভি। সিএসভি ফাইল টেক্সট বা গুগল শিটে যুক্ত করাটাও সহজ। কমার (,) বাইরে অন্য যতিচিহ্ন দিয়েও ফাইলে ডেটা আলাদা করা থাকতে পারে।

ডেটাকে আলাদা করতে যেসব যতিচিহ্ন বহুল ব্যবহৃত

- কমা
- ট্যাব
- স্পেস
- সেমিকোলন
- পাইপ (।)

এবার একটি সিএসভি ফাইল নিয়ে অনুশীলন করে দেখা যাক। গুগল ড্রাইভ থেকে BangladeshCSV.txt টি কম্পিউটারে ডাউনলোড করুন এবং নোটপ্যাডের মাধ্যমে সেটি ওপেন করুন। আপনি যদি কলামের শিরোনামগুলো দেখেন তবে আপনি একটি কাঠামো দেখতে পাবেন। প্রতিটি কলামের তথ্য কমা (,) যতিচিহ্ন দিয়ে আলাদা করা।


BangladeshCSV.txt - Notepad
File Edit Format View Help

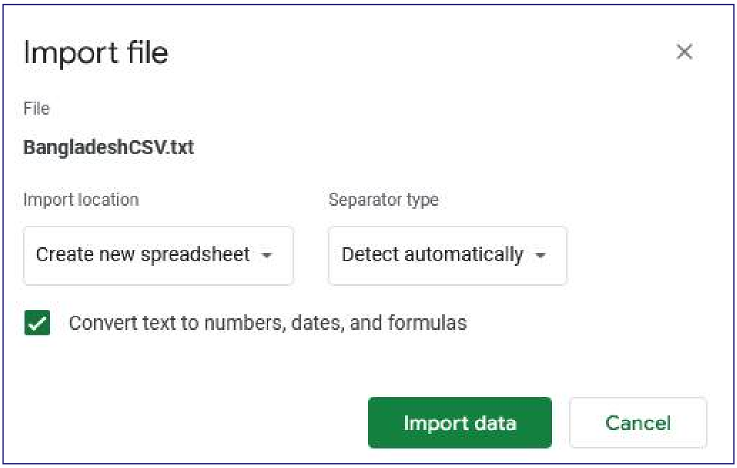
```
'NAME' , 'DISTRICT' , 'DIVISION' , 'OCCUPATION' , 'INCOME'
'Tanver, Emon' , 'Sirajgonj' , 'rajshahi' , 'Business' , '80000'
'Rashed, Ahmed' , 'Munshiginj' , 'Dhaka' , 'Teaching' , '60000'
'Polash, Roy' , 'Rongpur' , 'Rongpur' , 'Government Employee' , '30000'
'Shovon, Khan' , 'Candpur' , 'Chattogram' , 'Business' , '45000'
'Alif, Moahammad' , 'Nilphamari' , 'Rongpur' , 'Government Employee' , '12000'
'Maruf, Chowdhury' , 'Nilpahamri' , 'Rongpur' , 'Engineer' , '130000'
'Masum, Chowdhury' , 'Nilpahamri' , 'Rongpur' , 'Doctor' , '95000'
'Hasib, Ferdows' , 'Thakurgaon' , 'Rongpur' , 'Business' , '56000'
'Dipu, Rahman' , 'Pirojpur' , 'Barishal' , 'Doctor' , '145000'
'Abdur, Rauf' , 'Gazipur' , 'Dhaka' , 'Business' , '140000'
'Mostofa, Kamal' , 'Comilla' , 'Chattogram' , 'Engineer' , '10000'
'Anonna, Haq' , 'Gazipur' , 'Dhaka' , 'Doctor' , '55000'
'Sajib, Ahmed' , 'Foridpur' , 'Dhaka' , 'Buisness' , '95000'
'Tarana, Haq' , 'Madaripur' , 'Dhaka' , 'Teaching' , '60000'
'Rumana, Sharmin' , 'Khulna' , 'Khulna' , 'Business' , '35000'
'Momtaz, Begum' , 'Bagerhat' , 'Khulna' , 'Business' , '125000'
'Mokhlesur, Rahman' , 'Bhola' , 'Barishal' , 'Engineer' , '85000'
'Humayun, Khan' , 'Natore' , 'Rajshahi' , 'Government Employee' , '15000'
'Sahin, Khan' , 'Kushtia' , 'Khulna' , 'Government Employee' , '15000'
'Seikh, Karim' , 'Dianjpur' , 'Rongpur' , 'Doctor' , '95000'
'fahim, Ahmed' , 'Kurigram' , 'Rongpur' , 'Engineer' , '195000'
'Mishkat, Afrin' , 'Thakurgaon' , 'Rongpur' , 'Business' , '75000'
'Halima, Khatun' , 'Comilla' , 'Chattogram' , 'Teaching' , '55000'
'Mitu, Missouri' , 'Feni' , 'Chattogram' , 'Doctor' , '60000'
'Abdul, Salam' , 'Bandarban' , 'Chattogram' , 'Teaching' , '25000'
'Selim, Mia' , 'Sherpur' , 'Mymensingh' , 'Engineer' , '25000'
'Abdul, Aziz' , 'Netrokona' , 'Mymensingh' , 'Government Employee' , '50000'
'Mohammad, Khan' , 'Jamalpur' , 'Mymensingh' , 'Teaching' , '30000'
'Shimul, Biswas' , 'Jamalpur' , 'Mymensingh' , 'Doctor' , '130000'
'Toufik, Khan' , 'Manikganj' , 'Dhaka' , 'Teaching' , '52000'
'Mir, hossain' , 'Jamalpur' , 'Dhaka' , 'Doctor' , '52000'
'Toufik, elahi' , 'Sunamganj' , 'Sylhet' , 'Teaching' , '37000'
'Nasim, Bepari' , 'habiganj' , 'Sylhet' , 'Doctor' , '87000'
```

সিএসভি ফাইলটিতে আরও একটি যতিচিহ্ন আপনি খুঁজে পাবেন, সেটি হচ্ছে উদ্ধৃতি চিহ্ন বা কোটেশন (‘)। প্রতিটি শব্দকে এই চিহ্ন দিয়ে আলাদা করা হয়েছে। কেননা প্রথম কলাম কমা দিয়ে আলাদা করা হয়েছে। এখন কলামের মধ্যে আবার দুই শব্দ আছে। সেটিকে আলাদা করতেই উদ্বৃতি চিহ্ন (‘) ব্যবহার করা হয়েছে। এর ফলে এক্সেল বা গুগল শিট জানতে পারবে, কমার মধ্যেই কোটেশন আছে এবং সেটি একই কলামের অন্তর্গত। ডেটা পৃথক হয়েছে শুধু কমার মাধ্যমে।

সিএসভি ফাইল ইম্পোর্ট

এরইমধ্যে জেনে গেছেন যে, কমা (,) দিয়ে প্রতিটি কলামের ডেটা আলাদা করা হয়েছে আর উদ্বৃতি চিহ্ন দিয়ে কলামের মধ্যকার ডেটা আলাদা হয়েছে। এখন এসব ডেটা গুগল শিটে যুক্ত করা যাক।

ডেটা সাংবাদিকতা সহায়িকার গুগল ড্রাইভে থাকা গুগল শিটটিতে নতুন একটি শিট যুক্ত করুন এবং ফাইল মেন্যুতে গিয়ে ইম্পোর্ট অপশনটি ক্লিক বা নির্বাচন করুন। এরপর আপনার সামনে অনেকগুলো উৎস থেকে ফাইল যুক্ত করার অপশন আসবে। আপলোডে ক্লিক করুন, এবং আগে থেকে ডাউনলোড করা BangladeshCSV.txt ফাইলটি নির্বাচন করুন।



এরপর আপনার সামনে ফাইলটি কীভাবে ইম্পোর্ট করবেন, এ বিষয়ে বিভিন্ন অপশন আসবে।

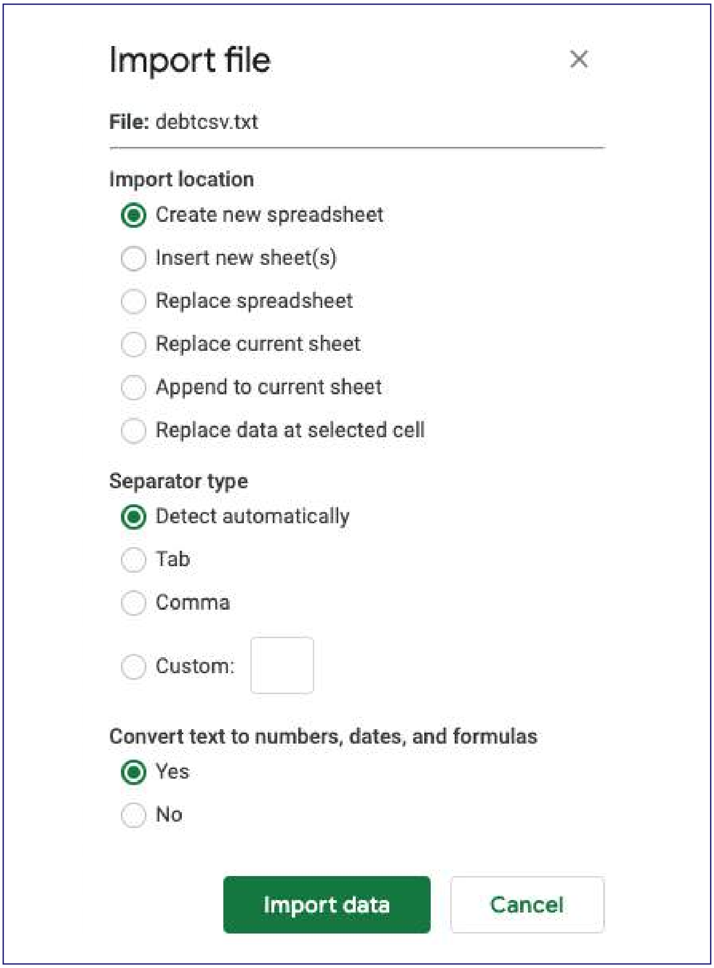
এই ধাপে তিনটি অপশন আপনার সামনে আসবে। প্রতিটি ক্ষেত্রেই এমন অপশন আপনাকে নির্বাচন করতে হবে, যার মাধ্যমে কাজের সুবিধা হয় এবং যে ভাবে আপনি ডেটাগুলোকে সাজাতে চান।

ধাপ-১ ইম্পোর্ট লোকেশন

গুগল শিট ছয়টি পৃথক অপশন দেখিয়ে আপনার কাছে জানতে চাইবে ডেটাগুলো কোথায় যুক্ত করবেন। আপনি একটি নতুন শিট তৈরি বা বিদ্যমান ফাইলে একটি নতুন শিট যুক্ত করার অপশনটি বেছে নিবেন। কারণ এতে পরবর্তী সময়ে কাজ অনেক সহজ হবে। তবে তার আগে কোন অপশনটি কি কাজ করবে, তা জেনে নেয়া যাক।

- Create new spreadsheet: যে গুগল শিটসটিতে কাজ করছেন, তার বাইরে সম্পূর্ণ নতুন একটি গুগল শিট তৈরি হবে।
- Insert new sheet(s): যে গুগল শিটটিতে কাজ করছেন, সেখানে নতুন একটি শিট বা ট্যাব যুক্ত হবে।
- Replace spreadsheet: যে গুগল শিটসটিতে কাজ করছেন, সেটি মুছে দিয়ে সম্পূর্ণ নতুন একটি গুগল শিটস তৈরি হবে।

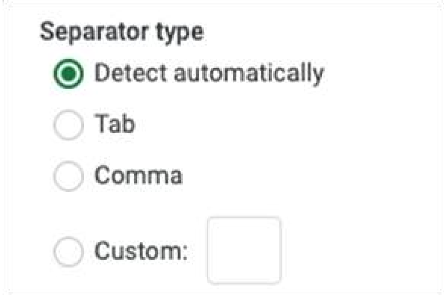
- Replace current sheet: সম্পূর্ণ গুগল শিটসটি পরিবর্তন হবে না, তবে যে ট্যাব বা শিটসে কাজ করছেন সেটি মুছে টেক্সট ফাইল থেকে নতুন তথ্য যুক্ত হবে।
- Append to current sheet: যে ট্যাব বা শিটসে কাজ করছেন, সেখানে মাউসের কার্সরটি শিটসের যেখানে রয়েছে, সেখানে নতুন তথ্য যুক্ত হবে।
- Replace data at selected cell: যে ট্যাব বা শিটে কাজ করছেন, সেখানে মাউসের কার্সরটি শিটসের যে সেলে রয়েছে সেই সেলের সব ডেটা মুছে নতুন তথ্য যুক্ত হবে।



এই অনুশীলনের জন্যে দ্বিতীয় অর্থাৎ Insert new sheet(s) *বেছে নিন।*

ধাপ-২ ডেটা পৃথকীকরণ চিহ্ন

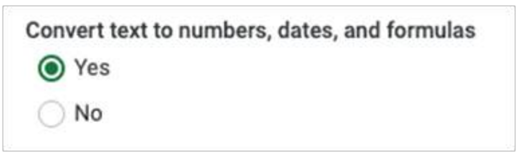
গুগল শিটস স্বয়ংক্রিয়ভাবে ডিলিমিটার বা ডেটা পৃথককরণ চিহ্নকে চিহ্নিত করবে। তবে আপনি যদি পৃথককরণ চিহ্ন উল্লেখ করে দেন, তবে ভুলের আশঙ্কা কম থাকবে। এখানে দুইটি বহুল ব্যবহৃত যতি চিহ্ন কমা এবং ট্যাব উল্লেখ করা আছে। একইসঙ্গে অন্য কোনো ডিলিমিটার বা ডেটা পৃথককরণ চিহ্ন উল্লেখের সুযোগ আছে কাস্টম অপশনে। তবে কলামের মধ্যে থাকা যতি-চিহ্নের মাধ্যমে শব্দ আলাদা করা হলেও, তা শিটসে উল্লেখ করার সুযোগ নেই।



ধাপ-৩ ডেটার ধরন

টেক্সট থেকে গুগল শিটসে ডেটা যুক্ত করার এই প্রক্রিয়াটি খুব শক্তিশালী না। এখানে স্বয়ংক্রিয়ভাবে ডেটা স্থানান্তরের সুযোগ আছে। যা আপনার ডেটাকে কলাম অনুযায়ী ভাগ করে শিটে যুক্ত করবে। তবে এক্ষেত্রে ডেটা যুক্ত হবার পর আপনাকে সবগুলো ডেটায় একবার চোখ বুলিয়ে নিতে হবে। এর বাইরে আপনি চাইলে নিজেও ডেটার কলাম ধরে তথ্যের ফরম্যাট ঠিক করে নিতে পারেন।

এক্ষেত্রে স্বয়ংক্রিয় পদ্ধতির ওপরই আস্থা রাখুন।



এবার আপনি টেক্সট ফাইলটি থেকে শিটসে ডেটা যুক্ত করতে Import Data তে ক্লিক করতে পারেন।

আশা করা যায়, আপনার গুগল শিটসে এরইমধ্যে টেক্সট ফাইলে থাকা সব ডেটা যুক্ত হয়েছে। তবে একটি ত্রুটি রয়ে গেছে। খেয়াল করে দেখবেন, ঠিকানার কলামে বাক্যটির নামের দ্বিতীয় অংশ বসেছে। এর কারণ আগেই উল্লেখ করা হয়েছে, ডিলিমিটারস অর্থাৎ কমার মধ্যে থাকা Qualifier অর্থাৎ উদ্ধৃতি চিহ্ন গুগল শিটস চিহ্নিত করতে পারে না। ফলে সেটি পরবর্তী কলাম অর্থাৎ ঠিকানায় যুক্ত হয়েছে। এর ফলে শেষ কলামে থাকা ডেটা শিরোনামহীন রয়ে গেছে।

ডেটা সাংবাদিকতা সহায়িকা

File Edit View Insert Format Data Tools Extensions Help Last edit was seco-

	A	B	C	D	E	F
1	NAME'	DISTRICT'	DIVISION'	OCCUPATION'	INCOME'	
2	Tanver	Emon'	Sirajgonj'	rajshahi'	Business'	80000'
3	Rashed	Ahmed'	Munshiginj'	Dhaka'	Teaching'	60000'
4	Polash	Roy'	Rongpur'	Rongpur'	Government Emp	30000'
5	Shovon	Khan'	Candpur'	Chattogram'	Business'	45000'
6	Alif	Moahammad'	Nilphamari'	Rongpur'	Government Emp	12000'
7	Maruf	Chowdhury'	Nilpahamri'	Rongpur'	Engineer'	130000'
8	Masum	Chowdhury'	Nilpahamri'	Rongpur'	Doctor'	95000'
9	Hasib	Ferdows'	Thakurgaon'	Rongpur'	Business'	56000'
10	Dipu	Rahman'	Pirojpur'	Barishal'	Doctor'	145000'
11	Abdur	Rauf'	Gazipur'	Dhaka'	Business'	140000'
12	Mostofa	Kamal'	Comilla'	Chattogram'	Engineer'	10000'
13	Anonna	Haq'	Gazipur'	Dhaka'	Doctor'	55000'
14	Sajib	Ahmed'	Foridpur'	Dhaka'	Buisness'	95000'
15	Tarana	Haq'	Madaripur'	Dhaka'	Teaching'	60000'
16	Rumana	Sharmin'	Khulna'	Khulna'	Business'	35000'
17	Momtaz	Begum'	Bagerhat'	Khulna'	Business'	125000'
18	Mokhlesur	Rahman'	Bhola'	Barishal'	Engineer'	85000'

এই জটিলতার সমাধান বেশ সহজ। আপনি কলামের শিরোনামগুলো একধাপ পিছিয়ে দিতে পারেন এবং নামের শেষ ভাগের ক্ষেত্রে নতুন শিরোনাম যুক্ত করতে পারেন। একইসঙ্গে ডেটার মধ্যে যে উদ্ধৃতি চিহ্ন রয়েছে সেটি মুছে ফেলতে পুরো শিটে থাকা সব ডেটা সিলেক্ট করুন, এরপর Edit > Find and Replace, অপশনে গিয়ে সেগুলো মুছে দেন।

	A	B	C	D	E	F
1	FIRST NAME	LAST NAME	DISTRICT	DIVISION	OCCUPATION	INCOME
2	Tanver	Emon	Sirajgonj	rajshahi	Business	80000
3	Rashed	Ahmed	Munshiginj	Dhaka	Teaching	60000
4	Polash	Roy	Rongpur	Rongpur	Government Emp	30000
5	Shovon	Khan	Candpur	Chattogram	Business	45000
6	Alif	Moahammad	Nilphamari	Rongpur	Government Emp	12000
7	Maruf	Chowdhury	Nilpahamri	Rongpur	Engineer	130000
8	Masum	Chowdhury	Nilpahamri	Rongpur	Doctor	95000
9	Hasib	Ferdows	Thakurgaon	Rongpur	Business	56000
10	Dipu	Rahman	Pirojpur	Barishal	Doctor	145000
11	Abdur	Rauf	Gazipur	Dhaka	Business	140000
12	Mostofa	Kamal	Comilla	Chattogram	Engineer	10000
13	Anonna	Haq	Gazipur	Dhaka	Doctor	55000
14	Sajib	Ahmed	Foridpur	Dhaka	Buisness	95000
15	Tarana	Haq	Madaripur	Dhaka	Teaching	60000
16	Rumana	Sharmin	Khulna	Khulna	Business	35000
17	Momtaz	Begum	Bagerhat	Khulna	Business	125000
18	Mokhlesur	Rahman	Bhola	Barishal	Engineer	85000
19	Humayun	Khan	Natore	Rajshahi	Government Emp	15000

ওয়েবসাইট থেকে ডেটা সংগ্রহ বা স্ক্র্যাপিং

সাংবাদিকদের প্রায়শই ওয়েবসাইট থেকে ডেটা সংগ্রহ করে তা মাইক্রোসফট এক্সেল বা গুগল শিটসে বিশ্লেষণ করতে হয়। এ কারণে ওয়েবসাইটের সেই ডেটা শিটে যুক্ত করতে হয়। যা ওয়েব স্ক্র্যাপিং নামে পরিচিত।

> ওয়েবসাইটে কাঠামো ও ডেটা বিন্যাসের ওপর নির্ভর করে কোন পদ্ধতিতে স্ক্র্যাপিং করে ডেটা সংগ্রহ করা হবে। একটি জটিল ওয়েবসাইটের ক্ষেত্রে আপনাকে হয়তো পৃথক স্ক্র্যাপিং প্রোগ্রাম তৈরি করতে হতে পারে। তবে পাইথন, জাভাস্ক্রিপ্ট এবং আর নামের বেশকিছু কার্যকর স্ক্র্যাপিং প্রোগ্রাম রয়েছে, যা দিয়ে জটিল কাঠামোর ওয়েবসাইট থেকে ডেটা সংগ্রহ করা যায়।

তবে এই অনুশীলনে কেনো ধরনের কোডিং বা প্রোগ্রামিং ছাড়া ওয়েবসাইট থেকে সহজে ডেটা সংগ্রহ করে তা গুগল শিটস ও এক্সেলে যুক্ত করা হবে।

> যেহেতু গুগল শিটস অনলাইননির্ভর অ্যাপ্লিকেশন, তাই ওয়েবসাইটের টেবিল স্ক্র্যাপ করতে এটি সবচেয়ে কার্যকর। অ্যাপ্লিকেশনটি এমনভাবে তৈরি করা যেখানে আপনি গণিতের সূত্র ব্যবহার করে অনলাইন থেকে ডেটা সংগ্রহ করতে পারবেন।

সরাসরি অনুশীলনে যাওয়া যাক। এই টেবিলটি বাংলাদেশ পুলিশ ওয়েবসাইটের। যেখানে দেশের সকল মহানগর পুলিশ ও সারাদেশের পুলিশের সকল রেঞ্জে ২০১৯ সালে দায়ের করা মামলা বা সংঘঠিত অপরাধের হিসাব আছে।

CRIME STATISTICS 2019

Unit Name	Dacoity	Robbery	Murder	Speedy Trial	Riot	Woman & Child Repression	Kidnapping	Police Assault	Burglary	Theft	Other Cases	Recovery Cases: Arms Act	Explosive	Narcotics	Smuggling	Total	Total Cases
DMP	1	15	15	10	0	118	8	6	48	127	413	6	3	1276	14	1393	2133
CMP	0	3	9	5	0	22	0	5	6	15	123	36	0	449	1	486	678
KMP	1	3	4	1	0	8	0	0	6	4	19	2	0	272	3	277	317
RMP	0	4	1	1	1	11	0	0	4	6	52	3	0	283	35	321	431
BMP	0	8	0	0	0	14	1	0	4	6	40	1	0	91	9	92	166
SMP	0	2	5	3	0	15	0	13	0	2	55	0	0	44	2	46	143
Dhaka Range	7	11	75	10	0	218	12	13	28	79	888	18	3	1721	6	1758	3080
Mymensingh Range	1	1	23	1	0	87	2	6	5	24	435	5	2	430	7	448	1004
Chittagong Range	4	11	74	4	0	147	8	10	25	61	809	54	1	1108	10	1173	2300
Sylhet Range	4	5	23	4	0	67	1	8	7	53	412	1	3	139	21	164	725
Khulna Range	4	4	37	8	0	88	2	2	7	36	544	13	3	814	20	850	1585
Barisal Range	2	3	18	2	0	88	3	2	16	23	378	1	0	302	8	364	855
Rajshahi Range	5	8	29	6	0	99	8	2	5	49	458	23	13	1186	134	1340	2011
Rangpur Range	1	2	33	8	0	118	0	0	8	38	625	4	0	551	64	649	1477
Railway Range	0	1	2	0	0	0	0	0	0	5	9	0	0	58	12	67	84

স্ক্রিনশট: বাংলাদেশ পুলিশ

HTML লিংক জানার পাশাপাশি আপনাকে এটাও জানতে হবে যে ওয়েবসাইট থেকে কোন তথ্য আপনি স্ক্র্যাপ করতে চান এবং সেটি কোন ফরম্যাটে আছে।

- পুরো টেবিল না-কি শুধু একটি সারি?
- কোন টেবিল অথবা কোন সারি (প্রথম, দ্বিতীয়, তৃতীয়, ইত্যাদি)?

এবার চলুন স্ক্র্যাপিং শুরু করা যাক

গুগল শিটসে, ওয়েবসাইট থেকে ডেটা স্ক্র্যাপের পদ্ধতিকে বলা হয় importHTML

=importHTML(“url”, “table”, 1)

এবার সূত্রটির ব্যাখ্যা জেনে নেয়া যাক

url = যে ওয়েবপেইজটি থেকে টেবিলের ডেটা স্ক্র্যাপ করবেন, তার url কপি করুন এবং উপরে উল্লেখ করা ফর্মুলা বা সূত্রের “url” এর url শব্দের বদলে, সংশ্লিষ্ট ওয়েবপেইজটির কপি url টি পেস্ট করতে হবে।

table = পরের উদ্ধৃতি চিহ্নের মধ্যে দুইটি অপশন ব্যবহার করা যায়। “table” অথবা “list”। যেহেতু ওয়েবপেইজটি থেকে টেবিলের ডেটা স্ক্র্যাপ করা হবে, তাই দ্বিতীয় উদ্ধৃতি চিহ্নের মধ্যে “table” উল্লেখ করতে হবে।

১ = ওয়েবপেইজটিতে কয়টি টেবিল আছে। তার ওপর নির্ভর করবে এই সংখ্যা। যদি পেইজে একটি টেবিল থাকে তবে নিঃসন্দেহে এক্ষেত্রে ১ বসবে। যদি দুই বা তিনটি টেবিল থাকে তবে আপনি কোন টেবিলটির ডেটা স্ক্র্যাপ করতে চান, সংখ্যাটি তার ওপর নির্ভর করবে। যদি আপনি তৃতীয় টেবিলটি নিতে চান সেক্ষেত্রে এখানে ৩ বসবে। তবে যেহেতু এই ওয়েবপেইজটিতে ১টি টেবিল, এখানে ১ লিখতে হবে।

এবার ডেটা স্ক্র্যাপ করতে গুগল শিটসে নতুন একটি শিটস বা পেইজ তৈরি করুন এবং A1-এ ফর্মুলাটি লিখুন।

=importHTML("https://www.police.gov.bd/en/crime_statistic/year/2019", "table", 1)

Unit Name	Dacoity	Robbery	Murder	Speedy Trial	Riot	Woman & Child Repression	Kidnapping	Police Assault	Burglary	Theft	Other Cases	Recovery Cases Arms Act	Explosive	Narcotics	Smuggling	Total
DMP	1	10	18	10	0	718	9	9	49	127	471	9	3	1370	14	1390
CMP	0	3	9	5	0	22	6	6	6	15	125	38	0	449	1	488
KMP	1	0	4	1	0	8	0	0	6	4	56	2	0	272	3	277
RMP	0	4	1	1	1	11	0	0	4	6	52	3	0	263	35	321
BMP	0	0	0	0	0	14	1	0	4	8	48	1	0	91	0	92
SMP	0	2	8	3	0	15	0	13	2	2	55	0	0	44	2	46
Dhaka Range	2	11	75	10	0	219	12	13	29	70	868	10	3	1731	6	1734
Mymensingh Range	1	1	23	1	0	87	2	8	8	24	405	8	2	435	7	444
Chittagong Range	4	11	74	4	0	147	8	10	25	61	801	54	1	1108	10	1173
Sylhet Range	4	5	20	4	0	67	1	8	7	33	412	1	3	139	21	164
Khulna Range	4	4	37	0	0	90	2	2	7	36	544	13	3	814	20	850
Barisal Range	2	3	16	2	0	86	3	2	15	23	378	1	0	363	0	364
Rajshahi Range	5	8	29	6	0	95	8	2	6	40	469	25	15	1166	134	1340
Rangpur Range	1	2	33	0	0	116	0	0	8	38	625	4	0	551	94	649
Railway Range	0	1	2	0	0	0	0	0	0	5	9	0	0	55	12	67
GMP	2	3	3	1	0	22	1	2	2	8	65	3	0	130	2	135
RRMP	0	0	1	0	0	12	1	0	0	6	33	0	0	58	0	58
ATU	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	9
Total	32	68	391	48	1	1139	46	69	174	484	5428	174	30	9050	361	9634

মনে রাখবেন, এই পুরো ডেটাসেট এখানে যুক্ত হয়েছে A1 সেলে বসানো ফর্মুলা বা সূত্রের কারণে। পরবর্তীতে অন্য কোনো সূত্রের সাহায্যে এই ডেটা সেটকে বিশ্লেষণ করতে চাইলে, ডেটাকে আলাদা করে কপি করে নিতে হবে।

> যেহেতু অনলাইন প্রোগ্রামিং-এর মাধ্যমে গুগল শিট এই ডেটা স্ক্র্যাপ করছে, তাই মূল ওয়েবসাইটে ডেটার পরিবর্তন হলে, গুগল শিটসেও তা সক্রিয় তা পরিবর্তন হয়ে যাবে। তাই যদি পুরোনো ডেটা সংরক্ষণে রাখতে চান, গুগল শিটসের একটি স্ক্রিনশট নিয়ে রাখুন।

ওয়েবসাইট থেকে ডেটা সংগ্রহের জন্যে অনুশীলনে বাংলাদেশ পুলিশের ডেটার যে লিঙ্কটি ব্যবহার করা হয়েছে, সেখানে যেতে কিউআর কোডটি স্ক্যান করুন।

পিডিএফ থেকে ডেটা সংগ্রহ

ডেটা সাংবাদিকতার বিড়ম্বনাগুলোর একটি পিডিএফ থেকে ডেটা সংগ্রহ। অনেক প্রতিষ্ঠান তাদের ডেটা এক্সেল অথবা টেক্সট ফাইল হিসেবে সরবরাহ করে। যদি আপনি এই ফরম্যাটের ডেটার বদলে পিডিএফ পান, প্রথমেই তা বদলে কর্তৃপক্ষের কাছ থেকে এক্সেল অথবা টেক্সট ফাইলে ডেটা সংগ্রহের চেষ্টা করুন। তা না হলে আপনাকে পিডিএফ ফাইলটিকে কনভার্ট করতে হবে। তবে দুঃখজনক খবর হচ্ছে, সব ধরনের পিডিএফ থেকে ডেটা সংগ্রহের জন্যে কোনো একক সমাধান নেই। এটি পুরোপুরি নির্ভর করবে আপনার পিডিএফ ফাইলের ধরনের ওপর।

> তবে আপনি যে টুলস বা অ্যাপ্লিকেশনই ব্যবহার করেন না কোনো, পিডিএফ থেকে ডেটা সিএসভি ফাইল কিংবা শিটসে রূপান্তরের পর প্রথম কাজ হবে নতুন ডেটা ভালভাবে যাচাই করা। কেননা আপনি যে পদ্ধতিতেই এই রূপান্তর করুন না, এখানে কিছু না কিছু ভুল হবার আশঙ্কা থাকবেই।

রূপান্তর বা কনভার্সনের প্রথম ধাপে আপনাকে চিহ্নিত করতে হবে, যে পিডিএফটি আপনি পেয়েছেন সেটি কি কম্পিউটারে টাইপ করা টেক্সট ফাইল নাকি কোনো ছবি থেকে তৈরি হওয়া পিডিএফ। ছবি বলতে কোন দৃশ্য নয়, অনেক সময় কম্পিউটারে টাইপ করা টেক্সটের ছবি তুলে সেই ছবির পিডিএফ তৈরি করা হয়। এখানে সেই ছবি পিডিএফের কথা বলা হয়েছে।

টেক্সট পিডিএফ

এটি বুঝতে আপনাকে একটি ছোট পরীক্ষা করতে হবে। মাউসের কার্সর দিয়ে পিডিএফে থাকা টেক্সটগুলো কে সিলেক্ট করার চেষ্টা করুন। যদি আলাদা করে নির্দিষ্ট কিছু শব্দ বা বাক্য সিলেক্ট করতে পারেন, তাহলে আপনাকে অভিনন্দন। কেননা আপনার ফাইল রূপান্তরের প্রক্রিয়া অনেকটা সহজ হয়ে গেছে। কারণ এটি একটি টেক্সট পিডিএফ। এই পরীক্ষার আরেকটি পদ্ধতিও আছে। CTRL+F চেপে কোনো শব্দ সার্চ দিয়ে দেখতে পারেন। যদি টেক্সট পিডিএফ হয়, সার্চে আপনার কাঙ্ক্ষিত শব্দ খুঁজে পাওয়া যাবে।

যদি আপনি শব্দ বা বাক্য সিলেক্ট করতে না পারেন অথবা সার্চ কাজ না করে তবে সেটি ইমেজ বা ছবি পিডিএফ। এই অনুশীলনে দুই ধরনের পিডিএফ কনভার্সন বা রূপান্তরের কৌশল নিয়েই আলোচনা করা হবে।

টেক্সট পিডিএফ থেকে ডেটা সংগ্রহ অপেক্ষাকৃত সহজ। বেশ কিছু সফটওয়্যার ও অনলাইন অ্যাপ্লিকেশনের মাধ্যমে বিনামূল্যে এই রূপান্তরের কাজটি করা যায়। তবে ডেটা সংগ্রহের এই কাজ কতটা সফল ভাবে হবে, তা নির্ভর করে পিডিএফটি কোন কাঠামোতে গঠিত তার ওপর।

নিচে উল্লেখ করা টুলস বা অ্যাপ্লিকেশনগুলো ব্যবহার করে আপনি পিডিএফ থেকে তথ্য সংগ্রহের কাজটি করতে পারবেন।

ট্যাবুলা[৩১]

এটি বিনামূল্যে ব্যবহার করা যায়, ওপেন সোর্স ডেস্কটপ অ্যাপ্লিকেশন। এটি তৈরিই করা হয়েছে সাংবাদিকতার কাজ করার জন্য, সাংবাদিকদের জন্য। আর তৈরিও করেছেন সাংবাদিকদের একটি দল।

সুবিধা: ক্লাউডে অর্থাৎ কোনো সার্ভারে রূপান্তর করতে যাওয়া পিডিএফটি যুক্ত করতে হয় না। এটি স্পর্শকাতর ও গোপনীয় নথির জন্যে ইতিবাচক। আপনার যতটুকু অংশের ডেটা দরকার, পিডিএফ থেকে শুধু সেই অংশটুকু নির্দিষ্ট বা সিলেক্ট করে রূপান্তর করা সম্ভব।

অসুবিধা: শুধু সিএসভি অথবা প্রোগ্রামিং ফরম্যাটেই রূপান্তর সম্ভব।

কমেট ডকস[৩২]

এই অ্যাপ্লিকেশনটিও অনলাইনভিত্তিক। পিডিএফ থেকে ডেটা রূপান্তরের প্রাথমিক কিছু সেবা এখানে বিনামূল্যে পাওয়া যায়। তবে আরও উন্নত প্রযুক্তির কিছু সেবার জন্যে টাকা দিতে হবে।

সুবিধা: পিডিএফ থেকে ডেটা নিয়ে এক্সেল, টেক্সট, পাওয়ারপয়েন্টসহ বিভিন্ন ধরনের ফাইলে রূপান্তর করা যায়।

অসুবিধা: পুরো পিডিএফ একবারে কনভার্ট বা রূপান্তর করতে হয়। নির্দিষ্ট অংশের ডেটা আলাদা করে সংগ্রহ করা যায় না।

---

[৩১] Tabula: Extract Tables from PDFs

[৩২] Cometdocs File Converter. Convert Excel to Word and more.

আরও আছে

- অ্যাডোবি পিডিএফ অনলাইন[৩৩] (১৫ ডলার/মাসে)
- স্মলপিডিএফ[৩৪] (৫ ডলার/মাসে)

ছবির পিডিএফ

ছবি থেকে তৈরি করা পিডিএফ বা Image pdf থেকে ডেটা সংগ্রহ বা রূপান্তর কিছুটা কঠিন। যেহেতু এই ফাইলগুলো স্ক্যান করা অথবা ক্যামেরায় তোলা ছবি দিয়ে তৈরি তাই সেখান থেকে অক্ষর চিহ্নিত করতে ওসিআর (অপ্টিকাল ক্যারেক্টার রিকগনিশন) সফটওয়্যারের প্রয়োজন হয়। ইংরেজি ভাষায় বেশ ভালো মানের ওসিআর পাওয়া যায়। কিন্তু বাংলায় যেসব ওসিআর আছে, ডেটা স্ক্র্যাপিংয়ের জন্য এখনো সেগুলো যথেষ্ট উপযুক্ত নয়।

গুগল পিন পয়েন্ট

সাংবাদিক এবং গবেষকদের ডেটা সংগ্রহ এবং তা বিশ্লেষণে একটি কার্যকর টুলস হল পিনপয়েন্ট। পিনপয়েন্ট[৩৫] ব্যবহার করে আপনি কয়েক হাজার নথি, ছবি, ইমেল, হাতে লেখা নোট, এবং নির্দিষ্ট শব্দ বা বাক্যাংশ, অবস্থান, সংস্থা এবং লোকেদের জন্য অডিও ফাইল আপলোড এবং অনুসন্ধান করতে পারবেন। পিনপয়েন্টের সবচেয়ে বড় সুবিধা হচ্ছে, আপনার কাঙ্ক্ষিত ডেটাটি যদি হাতে লিখে কেউ পিডিএফ বা তার ছবি আপলোড করে, পিনপয়েন্ট তা আপনাকে খুঁজে দেবে।

অনলাইন ওসিআর ডট নেট[৩৬]

এটি একটি অনলাইনে ক্লাউডভিত্তিক অ্যাপ্লিকেশন।

সুবিধা: বিনামূল্যে ব্যবহার করা যায়।

অসুবিধা: বড় আকারের ফাইল রূপান্তর করা যায় না।

---

[৩৩] How to convert a PDF to Excel | Adobe Acrobat
[৩৪] Smallpdf.com - A Free Solution to all your PDF Problems
[৩৫] About Pinpoint - Pinpoint Help (google.com)
[৩৬] Free Online OCR - Image to text and PDF to Doc converter

অ্যাবি ফাইনরিডার[৩৭]

ওসিআর সফটওয়্যারগুলোর মধ্যে অন্যতম সেরা ও কার্যকর।

সুবিধা: টেক্সট ও ওয়ার্ড ফাইলে ডেটা রূপান্তর করা যায়। নির্ভুলভাবে ডেটা সংগ্রহ করা যায়।

অসুবিধা: বেশ ব্যয়বহুল।

এসব টুলস ব্যবহার করে সহজে ডেটা সংগ্রহ করা সম্ভব। তবে এর বাইরে আরও অনেক কৌশল রয়েছে যেগুলো ব্যবহার করে আপনি ওয়েবসাইট থেকে ডেটা স্প্রেডশিটে সংযুক্ত করতে পারবেন।

ওয়েবসাইট থেকে ডেটা সংগ্রহের আরও কৌশল জানতে ডেটা সাংবাদিকতার এই সহায়িকাটি[৩৮] অনুসরণ করতে পারেন। বিশেষ করে এই অংশে, এখানে স্ক্র্যাপিং এর কোডিং এর বাইরে, এ সংক্রান্ত বেশ কিছু উদাহরণও উল্লেখ করা হয়েছে। কিউআর কোডটি স্ক্যান করেও ডেটা জার্নালিজম হ্যান্ডবুকটি পড়ে নিতে পারবেন।

[৩৭] How https://pdf.abbyy.com/
[৩৮] https://datajournalism.com/read/handbook/one

কেস স্ট্যাডি

ছবি: ঢাকা ট্রিবিউন

এই স্টোরিটি ডেটা চিত্রায়ণের জন্য সাউথ এশিয়া জার্নালিজম অ্যাওয়ার্ড পেয়েছে। রিপোর্টের বিষয়বস্তু ছিলো গঙ্গার উৎস থেকে বাংলাদেশে প্রবেশ করা পর্যন্ত উজানে পানি অপসারণ ও তার প্রভাব। আর এর মূল চ্যালেঞ্জ ছিল, ডেটার সীমাবদ্ধতা। নানা প্রতিবন্ধকতার পরও বিভিন্ন উৎস থেকে পানি সরিয়ে নেওয়ার ডেটা সংগ্রহ করে, সেটিকে যাচাই ও বিশ্লেষণ করে, ডেটাকে একটি ক্রসবর্ডার অনুসন্ধানের কেন্দ্রবিন্দুতে পরিণত করেছে স্টোরিটি। এই প্রতিবেদনে কীভাবে ডেটা ব্যবহার করা হয়েছে, তা নিয়ে লিখেছেন এর প্রতিবেদক আবু সিদ্দিক।

ঢাকা ট্রিবিউনে প্রকাশিত আবু সিদ্দিক-এর দ্য হাংরি চিলড্রেন অব মাদার গ্যাঞ্জেস প্রতিবেদনটি পড়তে কিউআর কোডটি স্ক্যান করুন

প্রতিবেদনের লক্ষ্য

বাংলাদেশ ও ভারতের মধ্যে প্রবাহিত আন্তর্জাতিক নদী গঙ্গার পানি বণ্টন নিয়ে আলোচনা দীর্ঘদিনের পুরোনো। প্রায় ২৫০০ কিলোমিটার লম্বা এই নদীটি ভারতের উত্তরখাণ্ড রাজ্যের গঙ্গোত্রী হিমবাহ থেকে উৎপত্তি হয়ে দেশটির বিভিন্ন রাজ্য হয়ে বাংলাদেশে প্রবেশ করেছে আমাদের চাঁপাইনবাবগঞ্জ জেলা দিয়ে। বাংলাদেশে এসে এই নদীটি পদ্মা নামে পরিচিতি পেয়েছে। আবার পদ্মা নদী কুষ্টিয়া জেলায় গড়াই নদী হয়ে বাংলাদেশের দক্ষিণ-পশ্চিম অঞ্চলের নদীগুলার সাথে যোগ হয়েছে। এইভাবে মিশে গেছে বঙ্গোপসাগরের সাথে, সুন্দরবন হয়ে। আমরা জানি যে, আমাদের দক্ষিণের নদীগুলোতে প্রতিদিন দুইবার জোয়ার-ভাটা হয়। জোয়ারের সময় এইসব নদী সাগর থেকে লোনা পানি নিয়ে উপরের দিকে আসে, আবার ভাটার সময় নীচে নেমে যায়। যেহেতু, পদ্মা নদীতে পানি আগের চেয়ে অনেক কম, কাজেই পানি সাগরের দিকে যাবার ক্ষমতাও কমে এসেছে। এই কারণে, জোয়ারের সময় যে লোনা পানি উপরে উঠে আসে, ভাটার সময় তা আর নামতে পারে না। স্বাভাবিক ভাবেই, লোনা পানি ক্রমাগত দক্ষিণ থেকে উত্তরে উঠে আসছে, কৃষিজমি লবণাক্ততায় আক্রান্ত হচ্ছে। কৃষি উৎপাদন কমে যাছে, সেই সব এলাকায়। আবার নদীতে পানি কমে যাওয়াতে পানিতে বাস করা বিভিন্ন জলজ প্রাণীর বিচরণক্ষেত্র সংকুচিত হয়ে আসছে। কমে আসছে তাদের সংখ্যা ও বৈচিত্র্য।

১৯৭০ এর দশকে ভারতের তৈরি করা ফারাক্কা ব্যারাজ চালু হবার পর থেকেই এই সমস্যা দেখা দিয়েছে। আন্তর্জাতিক এই নদীটি থেকে, শুকনো মৌসুমে, দুই দেশ যাতে সমান ভাগ পায় এই কারণে দুই দেশের মধ্যে একটি পানি বণ্টন চুক্তি হয়েছে ১৯৯৭ সালে, যা "বাংলাদেশ-ভারত গঙ্গা পানি বণ্টন চুক্তি" নামে পরিচিত। চুক্তির আওতায় প্রাকাশিত তথ্য দেখলে বোঝা যায়, বাংলাদেশ চুক্তির শর্ত অনুযায়ী পানি পাচ্ছে ঠিকই, কিন্তু পানির পরিমাণ ক্রমাগত কমে আসছে। এই নিয়ে বাংলাদেশের কিছু বলার নেই, কারণ চুক্তিতে বলা আছে– ফারাক্কা ব্যারাজে যে পরিমাণ পানি থাকবে সেটা ভাগ হবে। পুরো গঙ্গা নদীতে ভারত কি করবে তা বলা নেই।

> এই প্রতিবেদনের উদ্দেশ্য ছিলো, পুরো গঙ্গা থেকে ভারত কীভাবে পানি উত্তোলন করে আসছে এবং আর কী কী পরিকল্পনা করছে তার একটা হিসাব বের করা। উজানের দেশ ভারতের এইভাবে এককভাবে গঙ্গার পানি তুলে নেওয়াতে বাংলাদেশের ক্ষতি কীভাবে বাড়ছে, তা দেখানো।

> পুরো প্রতিবেদনটি তৈরি করতে বিস্তর গবেষণার পাশাপাশি, বাংলাদেশ ছাড়াও ভারতের গঙ্গা অংশের বিভিন্ন জায়গায় প্রায় এক মাস ফিল্ডওয়ার্ক, বিভিন্ন সোর্স থেকে ডেটা সংগ্রহ এবং পর্যালোচনা করতে হয়েছে। একেবারে শেষ পর্যায়ে এসে, পুরো প্রতিবেদনটি লেখা হয়েছে। সাথে প্রমাণ হিসেবে দেওয়া হয়েছে ছবি, ভিডিও ক্লিপ, গ্রাফ, চার্টসহ বিভিন্ন অনুষঙ্গ।

দরকারি ডেটা

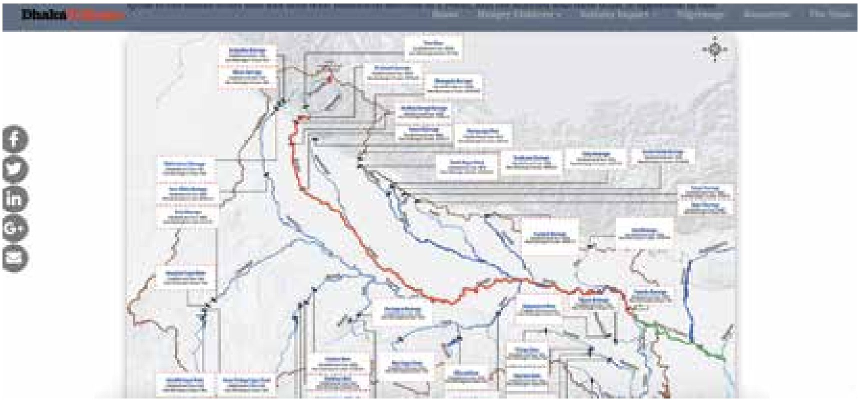
গঙ্গার পানি বন্টনের এই প্রতিবেদনে অনেক ডেটার ব্যবহার করা হয়েছে। এই প্রতিবেদনটির মূল উদ্দেশ্য ছিল, আন্তর্জাতিক নদী গঙ্গা থেকে ভারত এবং বাংলাদেশ, কে কতটুকু ভাগ পাচ্ছে তা বের করা? ভাটিরদেশ হিসেবে বাংলাদেশ যদি কম পানি পেয়ে থাকে, তাতে কী ধরনের ক্ষতি হচ্ছে? এ কারণে আমাকে প্রথমেই খুঁজতে হয়েছে কোন দেশ কতটুকু পানি পাচ্ছে? খরার মৌসুমে পানির অভাবে বাংলাদেশে কী কী ধরনের সমস্যা হচ্ছে? এইভাবে পানির সংকট চলতে থাকলে সামনের দিনে কী কী সমস্যা হতে পারে? এইভাবে ধাপে ধাপে অনেক প্রাসঙ্গিক প্রশ্ন এবং তার উত্তর পেতে অনেক তথ্যের প্রয়োজন পড়েছে।

কোথায় পেলাম ডেটা?

গঙ্গার পানি বন্টনের এই প্রতিবেদনে অনেক ডেটার ব্যবহার করা হয়েছে। এই প্রতিবেদনটির মূল উদ্দেশ্য ছিল, আন্তর্জাতিক নদী গঙ্গা থেকে ভারত এবং বাংলাদেশ, কে কতটুকু ভাগ পাচ্ছে তা বের করা? ভাটিরদেশ হিসেবে বাংলাদেশ যদি কম পানি পেয়ে থাকে, তাতে কী ধরনের ক্ষতি হচ্ছে? এ কারণে আমাকে প্রথমেই খুঁজতে হয়েছে কোন দেশ কতটুকু পানি পাচ্ছে? খরার মৌসুমে পানির অভাবে বাংলাদেশে কী কী ধরনের সমস্যা হচ্ছে? এইভাবে পানির সংকট চলতে থাকলে সামনের দিনে কী কী সমস্যা হতে পারে? এইভাবে ধাপে ধাপে অনেক প্রাসঙ্গিক প্রশ্ন এবং তার উত্তর পেতে অনেক তথ্যের প্রয়োজন পড়েছে।

গঙ্গার স্টোরির বেলায় আমি সবচেয়ে বড় ঝামেলায় পড়েছিলাম শুকনো মৌসুমের পানি প্রবাহের তথ্য জোগাড় করতে গিয়ে। যৌথ নদী কমিশনের ওয়েবসাইটে শুধু দেওয়া আছে, জানুয়ারি থেকে মে মাস পর্যন্ত কোন দেশ কতটুকু পানি পেয়েছে। সেটাও ২০০৮ সাল থেকে ২০১৭ সাল পর্যন্ত। এর বাইরে আর কোনো তথ্য কোথাও নেই। গঙ্গা নদীর উপরে ভারতের তৈরি করা ফারাক্কা ব্যারাজ, যেটা বাংলাদেশ সীমান্ত থেকে ১৫ কিলোমিটার উজানে। দুই দেশের মধ্যে ১৯৯৭ সালে হওয়া চুক্তি অনুযায়ী ফারাক্কা ব্যারাজে বছরের জানুয়ারি থেকে মে মাস পর্যন্ত যে পরিমাণ পানি থাকবে, তার ওপর ভাগাভাগি হবে।

> এর বাইরে পুরো নদীর উজানে কতটুকু পানি ভারত নিচ্ছে তা জানার কোনো রাস্তা বাংলাদেশের নেই। কারণ চুক্তিতে এই বিষয়ে কিছু বলা নেই। তবে, ২০০৮ সাল থেকে ২০১৭ সাল পর্যন্ত যে পানিবণ্টনের হিসাব আমরা পাই, তাতে এটা পরিষ্কার যে, ফারাক্কার উজানে পানির ব্যবহার বাড়ছে। কতটুকু বাড়ছে, কীভাবে বাড়ছে? আমরা জানি না। যেহেতু বাংলাদেশে এই তথ্য নেই, আমি বিকল্প হিসেবে গুগল সার্চ ইঞ্জিনকে কাজে লাগাই। অনলাইনে সার্চ করে খুঁজতে থাকি, গঙ্গা এবং যেসব নদী গঙ্গা নদীতে পানি সরবরাহ করে তাদের উপরে কী কী ধরনের অবকাঠামো আছে। সেগুলো কেন বানানো হয়েছে, কখন বানানো হয়েছে। কী কাজ করে। আমি এই ধরনের অবকাঠামো সম্পর্কে বিভিন্ন ওয়েবসাইট (বেশির ভাগই ভারত সরকারের বিভিন্ন মন্ত্রণালয় আর প্রকল্প সংক্রান্ত) গভীরভাবে পড়তে শুরু করি। আর যেখানে যতটুকু তথ্য পাই, সব কপি করতে থাকি সোর্সসহ। এইভাবে খুঁজতে খুঁজতে প্রায় ৫০টির মতো ছোটবড় ব্যারাজ ও পানি বিদ্যুৎ প্রকল্পের সন্ধান পাই।



ইনফোগ্রাফিক: ঢাকা ট্রিবিউন

এর পরের ধাপে আমি আমার পরিচিত ভারতীয় সাংবাদিক এবং বিভিন্ন এনজিও তে কাজ করে এমন কিছু বন্ধুদের অনুরোধ করি, আমার হয়ে ভারতের বিভিন্ন সরকারি অফিসে তথ্য অধিকার আইনে আবেদন করার জন্য। প্রতিটি ক্ষেত্রে আমার চাওয়া ছিলো সেইসব অবকাঠামো দিয়ে শুকনো মৌসুমে কী পরিমাণ পানি বিভিন্ন কাজে সরানো হয়, তা জানা। কিন্তু সকল ক্ষেত্রেই–ডেটা ক্লাসিফাইড, দেওয়া যাবে না বলে উত্তর আসে। যেটা ছিলো খুবই হতাশার। অনুসন্ধান থামিয়ে দেবার জন্য যথেষ্ট। কারণ এই বিশেষ তথ্যের হিসাব না দিতে পারলে পুরো কাজটাই আটকে যাবে।

ইনফোগ্রাফিক: ঢাকা ট্রিবিউন

এই পর্যায়ে, আমি অনলাইনে যেসব ভারতীয় প্রকল্প খুঁজে পেয়েছিলাম, সেগুলোর প্রজেক্ট ডিটেইলস খুঁজতে থাকি, প্রজেক্টগুলোর সাথে সম্পর্কিত বিভিন্ন মন্ত্রণালয়, প্রজেক্টের নিজস্ব ওয়েবসাইট, সেসব সম্পর্কিত বিভিন্ন জার্নাল এবং পুরোনো খবরও।

এইভাবে আমি কমবেশি ৩০ টির মতো অবকাঠামোতে নদী থেকে ক্যানেল বা খালের মাধ্যমে পানি সরানোর সক্ষমতা বের করে ফেলি। পরবর্তীতে এই সংখ্যাগুলো যোগ দিয়ে বের করা হয়, প্রায় ৩ লাখ কিউসেক পানি ফারাক্কা ব্যারাজে আসার আগেই ভারত উজানে প্রত্যাহার করে নিতে এবং শুকনো মৌসুমে বিভিন্ন কাজে লাগাতে পারে। তবে আমি শুধু মাত্র সেসব ডেটা ব্যবহার করেছি, সেগুলো ভারতীয় বিভিন্ন সরকারি ডকুমেন্ট, ওয়েবসাইট বা ডেটাবেইজ এ উল্লেখ আছে। কিছু কিছু ক্ষেত্রে একইবিষয়ে একাধিক তথ্য সামনে এসেছে। সেক্ষেত্রে সরকারি হিসাবটি ব্যবহার করেছি। আবার বেসরকারি হিসাবে উল্লেখ আছে, কিন্তু সরকারি হিসাবে নেই, এই ধরনের ডেটা বাদ দিয়ে দিয়েছি। কারণ আমরা এমন কোন তথ্য দিতে চাইনি, যেটা নিয়ে কোনো মতবিরোধ আছে। সবচেয়ে বড় বিষয়, এই প্রতিবেদনের শেষে আমরা, কোনো তথ্য কোথা থেকে নিয়েছি, তার একটি হিসেব এবং সোর্স লিস্ট দিয়ে দিয়েছি যাতে করে যে কেউ দেখে নিতে পারে।

ইনফোগ্রাফিক: ঢাকা ট্রিবিউন

ডেটা কোথায় রেখেছি

এই প্রতিবেদন করার জন্য প্রাথমিক ফোকাস ছিল, ভারত ফারাক্কা ব্যারাজের উপরে কতটুকু পানি এককভাবে ব্যবহার করে তার হিসাব বের করা। এবং সেই হিসাব থেকে বের করা যে বাংলাদেশ কতটুকু কম পানি পায়, সেই অংশ পেলে বাংলাদেশের কতটুকু ক্ষতি কমতো। আবার এটা দেখানো প্রয়োজন ছিল যে, পানি কম পাওয়াতে বাংলাদেশের কী কী ক্ষতি হচ্ছে। যেমন– মাছের ওপর কেমন নেতিবাচক প্রভাব পড়েছে, দেশের দক্ষিণ-পশ্চিম অঞ্চলে লবণ বেড়ে যাওয়াতে ধানের উৎপাদনে কেমন হেরফের হয়েছে, গত বছরগুলাতে বিশেষকরে ফারাক্কা ব্যারাজ হবার পর থেকে। এ সবের প্রয়োজনে সম্ভাব্য সকল ধরনের তথ্য জোগাড় করতে হয়েছে। যেহেতু অনেক ধরনের তথ্য ও সেগুলো লম্বা সময়কালের। উদাহরণ হিসেবে বলি, এই প্রতিবেদনে চিত্রা এবং নবগঙ্গা নদীর কয়েকটি পয়েন্টের লবনাক্ততার পরিমাণ দেখানো হয়েছে ১৯৬১ সাল থেকে। আবার ধান উৎপাদনের হিসাব দেওয়া হয়েছে গত ৪৫

বছরের। ডেটা যাতে করে সহজে ব্যবহার করা যায়, প্রয়োজনে খুঁজে পাওয়া যায় এবং যে কোনো সময় দেখা যায় এ কারণে কাজের শুরুতেই আমি গুগল ড্রাইভে আলাদা ফোল্ডার তৈরি করে ফেলি। প্রতিটি বিষয়ের জন্যেও আলাদা আলাদা ফোল্ডারে সকল তথ্য জমা করতে থাকি।

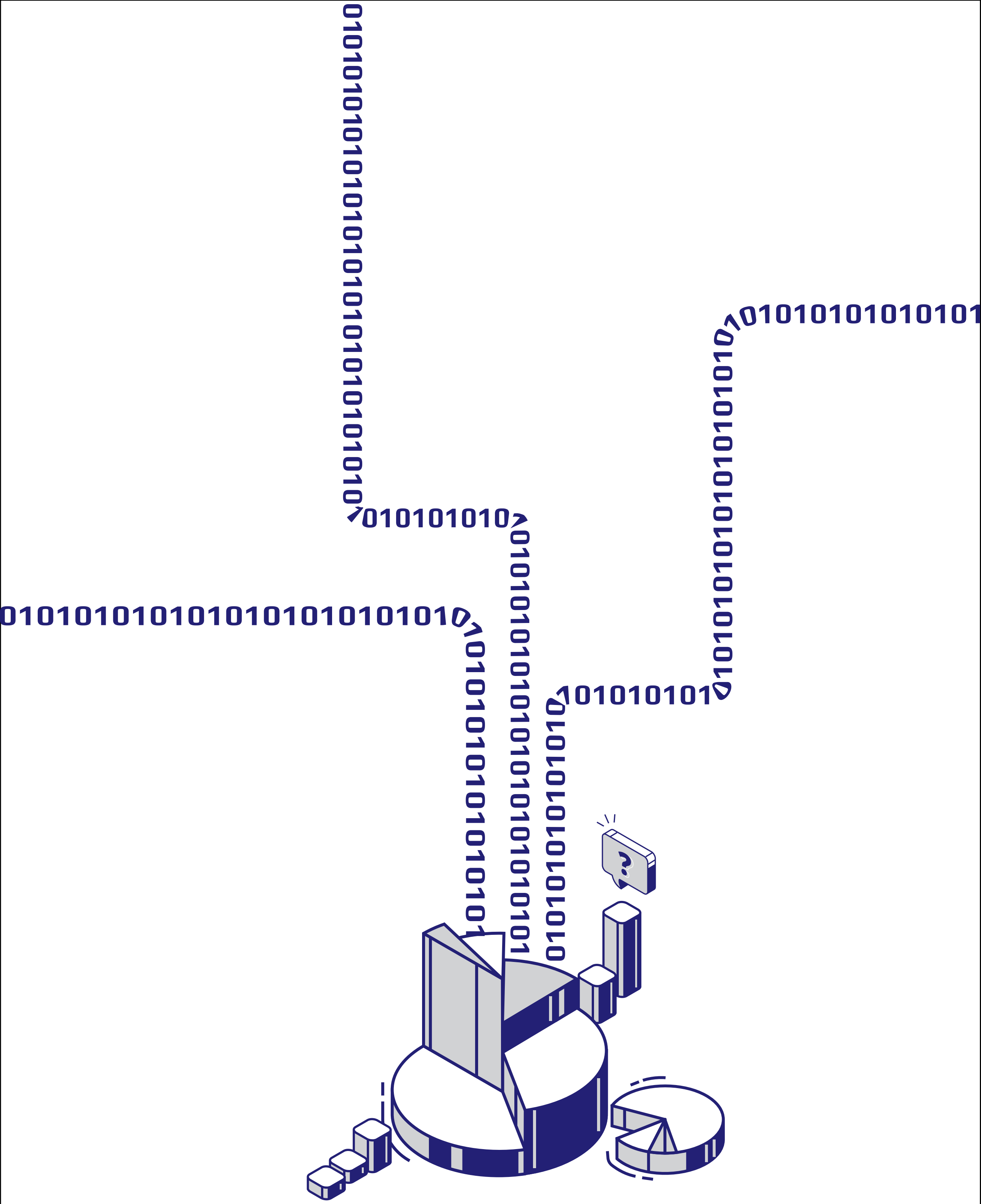
> এরপর ডেটা সংগ্রহের সকল কাজ শেষ হয়ে গেলে, ডেটা বিশ্লেষণ করে দেখি যে আমরা আসলে কোন ডেটা কতটুকু ব্যবহার করতে পারবো। কতটুকু গল্পের প্রয়োজনে দরকার। সব তথ্য আপনার সেই প্রতিবেদনে দিতে হবে এটা ভাবার কোনো কারণ নেই। পরবর্তীতে কখনো কাজে লাগবে।

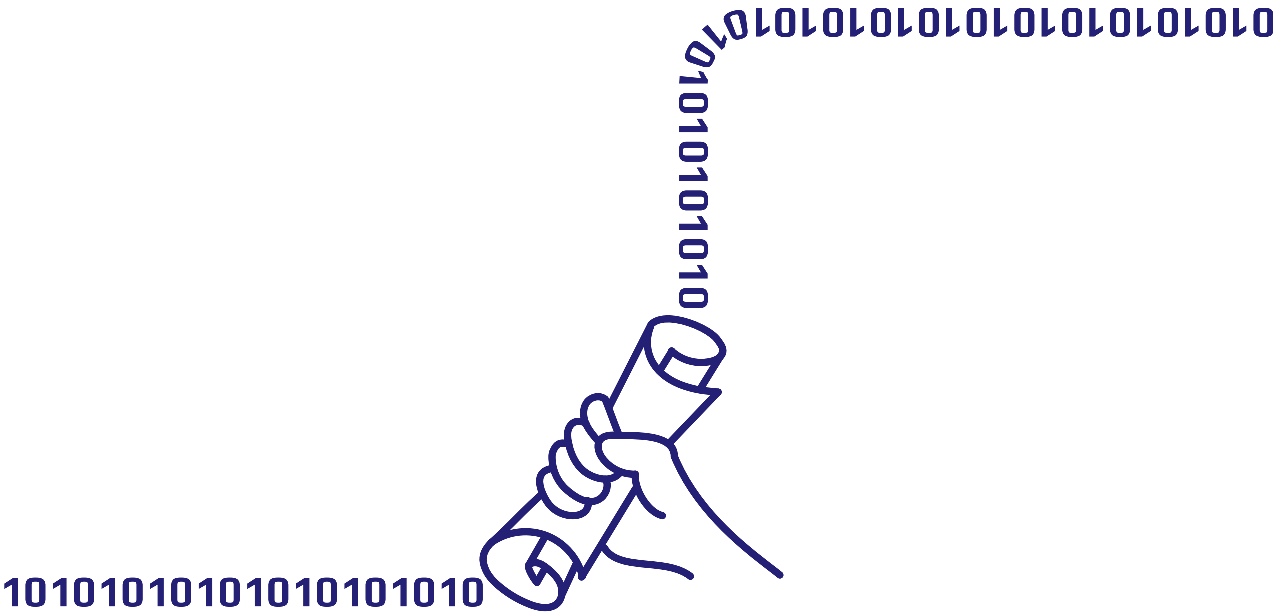
ডেটার উপস্থাপনা

গঙ্গার প্রতিবেদনে আমি যতটুকু সম্ভব গল্প বলার চেষ্টা করেছি। পাশাপাশি গল্পের প্রয়োজনে সত্যতা হিসেবে ডেটাগুলাকে চার্ট, গ্রাফ বা টেবিল ফরম্যাটে উপস্থাপন করেছি। আবার কোথাও কোথাও অনলাইনভিত্তিক ম্যাপের ব্যবহারও করেছি। যেহেতু এটি গঙ্গার পানি ইস্যুতে একটি বিস্তারিত প্রতিবেদন এবং সেখানে অনেক ধরনের তথ্যের সন্নিবেশ ছিল, আমরা প্রতবেদনটিকে একটি আলাদা স্পেস দেওয়ার জন্য ঢাকা ট্রিবিউনের মূল ওয়েবসাইটের সাথে লিঙ্ক করে একটি আলাদা সাব-ডোমেইন এ উপস্থাপন করা হয়েছে। ডেটা প্রেজেন্টেশান এর কিছু নমুনা এখানে স্ক্রিনশট হিসেবে দেখানো হলো–

ইনফোগ্রাফিক: ঢাকা ট্রিবিউন

সবশেষে একটি গুরুত্বপূর্ণ বিষয় মাথায় রাখতে হবে, যে বিষয়ে প্রতিবেদন তৈরি করবেন, সে বিষয় সম্পর্কে ভালভাবে জানতে হবে। এ কারণে প্রচুর বিষয়ভিত্তিক পড়ালেখার কোনো বিকল্প নেই। কারণ বিষয়টি সম্পর্কে ভালভাবে না জানলে আপনি কী করতে চান, কীভাবে এগুতে হবে, কী কী তথ্যের প্রয়োজন, সেগুলো কোথায় পাবেন এইসব বুঝতে পারবেন না। পাশাপশি সংশ্লিষ্ট সেক্টরের বিভিন্ন মানুষের সাথে যোগাযোগ বাড়াতে হবে, কারণ তাদের সাথে কথা বলেই আপনি অনেক বিষয়ে পরিষ্কার ধারণা পাবেন। পাশাপাশি, তাদের অনেকেই কখনো কখনো কাজ করবে আপনার তথ্যের যোগানদার হিসেবে।





অধ্যায় ০৫

তথ্য পেতে আরটিআই

তথ্য পেতে আরটিআই

> ২০০৮ সালের কথা। তখন ভারতের প্রধানমন্ত্রী ছিলেন মনমোহন সিং। কাজের প্রয়োজনে তাঁর মন্ত্রিসভার সদস্যরা নিয়মিত বিদেশ যেতেন। এ নিয়ে কারো মাথাব্যাথাও ছিলো না, কারণ মন্ত্রীরা বিদেশে যাবেন সেটাই সবাই দেখে আসছেন। কিন্তু তাদের এই সফর নিয়ে প্রশ্ন উঠলো ইন্ডিয়ান এক্সপ্রেস পত্রিকার সাংবাদিক শ্যামলাল যাদবের মনে।

তিনি তথ্য অধিকার আইনে সরকারের বিভিন্ন দফতরে আবেদন পাঠিয়ে জানতে চাইলেন, কেন্দ্রীয় সরকারের কোন মন্ত্রী কতবার বিদেশ সফরে গিয়েছেন, কেনো, কোথায় এবং তাতে সরকারের কত টাকা খরচ হয়েছে। তিনি ৭১টি মন্ত্রণালয় থেকে তথ্য পান, যার ফলাফল হরো একটি পুরস্কারজয়ী ডেটাভত্তিক অনুসন্ধান। “ফ্রিকোয়েন্ট ফ্লায়ার্স” শিরোনামের প্রতিবেদনটিতে, যাদব বিশ্লেষণ করে দেখান, ভারতের মন্ত্রীরা সাড়ে তিন বছর সময়ে বিদেশযাত্রায় যত মাইল দূরত্ব ভ্রমণ করেছেন, তাতে ২৫৬ বার গোটা পৃথিবীকে আবর্তণ করা যায়। মাত্র ৪৭ জন মন্ত্রী সেই সময়ের মানে ভ্রমণে ২৭ কোটি রুপি খরচ করেছেন।

ছবি: ইন্ডিয়ান এক্সপ্রেস

এই প্রতিবেদনের পর মন্ত্রণালয়গুলোকে বিদেশভ্রমণের খরচ কমাতে নির্দেশ দেন দেশটির তৎকালীন প্রধানমন্ত্রী মনমোহন সিং। পরবর্তীতে একই পত্রিকা ফের বিদেশ ভ্রমণের[৩৯] তথ্য চেয়ে আবেদন করে দেখতে পায়, এই খাতে সরকারের খরচ কমেছে। এটি মাত্র একটি উদাহরণ। যাদব এরপর তার অনেক অনুসন্ধানী প্রতিবেদনে তথ্য অধিকার আইন (আরটিআই) ব্যবহার করেছেন, নথি থেকে তথ্য নিয়ে স্প্রেডশিটে সাজিয়ে সেটিকে ডেটা সাংবাদিকতায় রূপ দিয়েছেন। সাধারণ সাংবাদিকতা, অনুসন্ধানী সাংবাদিকতা এবং ডেটা সাংবাদিকতায় তথ্য অধিকার আইন কীভাবে একটি বড় টুল হয়ে উঠলো দেশে দেশে তারই উদাহরণ হয়ে উঠেছে প্রতিবেদনগুলো। এ নিয়ে তিনি একটি বইও লিখেছেন, "জার্নালিজম থ্রু আরটিআই: ইনফর্মেশন, ইনভেস্টিগেশন, ইমপ্যাক্ট" শিরোনামে।

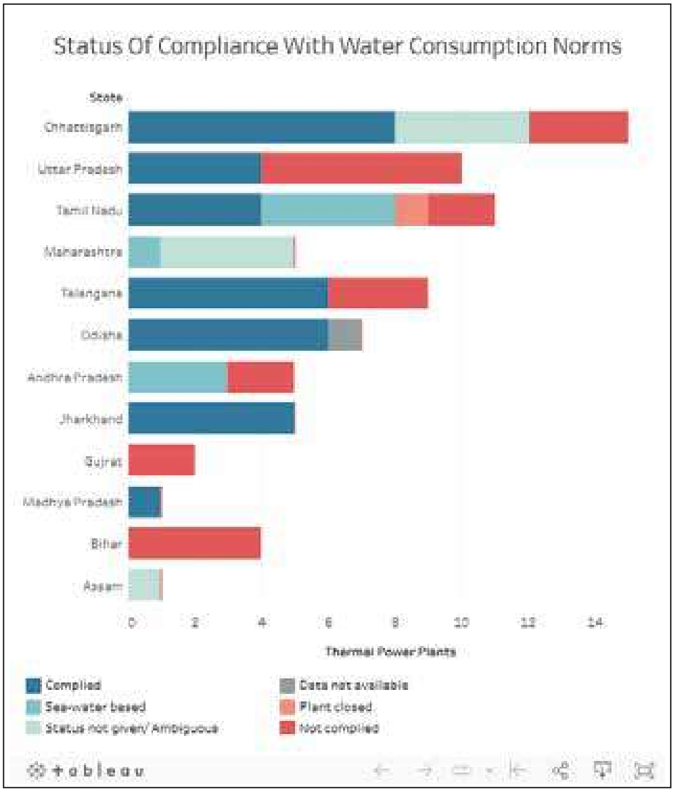
> যাদব বা ইন্ডিয়ান এক্সপ্রেসের মতো নিয়মিত না হলেও তথ্য অধিকার আইন ও ডেটা ব্যবহার করে সরকারি তহবিল ব্যবস্থাপনায় অস্বচ্ছতা ও দুর্নীতি উন্মোচনের নজির বাংলাদেশেই আছে। ২০১৯ সালে দক্ষিণাঞ্চলের জনপ্রিয় পত্রিকা গ্রামের কাগজ[৪০] একটি অনুসন্ধান করেছিল। তাতে বেরিয়ে আসে, সরকার যশোরের দরিদ্র ও দুস্থ মায়েদের জন্য যে মাতৃত্বকালীন ভাতা প্রদান করেছে, তার সুবিধাভোগীদের বড় অংশই দরিদ্র নয়। অনুসন্ধানটি শুরু হয়েছিল তথ্য অধিকার আইনে সুবিধাভোগীদের থানাভিত্তিক তালিকা চেয়ে আবেদন করে। এভাবে তারা প্রায় সাড়ে সাত হাজার সুবিধাভোগীর নাম-ঠিকানা সংগ্রহ করেন। সেই তালিকা ধরে অন্তত ৪০০ জন ভাতাভোগী মায়ের সাক্ষাৎকার নেন। সেই ডেটাকে বিশ্লেষণ করে তারা যে ধারাবাহিকটি প্রকাশ করে, সেটি টিআইবির অনুসন্ধানী সাংবাদিকতা পুরস্কারসহ বেশ ক'টি পুরস্কার লাভ করে।

[৩৯] Yadav, S. (2021b, January 5). 10 ways in which RTI has changed the functioning of govt, officials. The Indian Express. Retrieved October 12, 2022, from https://indianexpress.com/article/explained/10-ways-in-which-rti-has-changed-the-functioning-of-govt-officials/

[৪০] Gramerkagoj || Daily Newspaper in Bangladesh. (n.d.). https://www.gramerkagoj.com. Retrieved October 12, 2022, from https://www.gramerkagoj.com/

নাগরিক সংগঠনের সহায়তা নিন

গোটা বিশ্বেই তথ্য অধিকার আইন এখন ডেটা সংগ্রহের একটি বড় টুল হিসেবে বিবেচিত হয়। এটি সাংবাদিকেরা যেমন ব্যবহার করছেন তেমনি স্বচ্ছতা ও জবাবদিহিতা নিয়ে কাজ করা অলাভজনক নাগরিক সংগঠনগুলোও এই আইনের সুবিধাকে কাজে লাগিয়ে অনেক অজানা বিষয়কে সামনে নিয়ে আসছে। যেমন: ভারতের মন্থণ অধ্যয়ন কেন্দ্র নামের একটি গবেষণা প্রতিষ্ঠান প্রতিটি রাজ্যে আরটিআই আবেদন করে জানতে চেয়েছিলো সেখানকার তাপবিদ্যুৎ কেন্দ্রগুলো কী পরিমাণ ভূগর্ভস্থ পানি টেনে তুলছে।[৪১] তারা তথ্য পেয়েছিল ১৫৬টি বিদ্যুৎকেন্দ্রের। সেই তথ্য বিশ্লেষণ করে প্রতিষ্ঠানটি দেখতে পায়, কেবল ৫১ শতাংশ বলেছে যে তারা নির্ধারিত সীমা মেনে পানি তুলছে, এবং ১৯ শতাংশ সরাসরি বলেছে যে তারা সীমার বেশি পানি তুলেছে। এটি পরে অনেক গণমাধ্যমের জন্য খবরের একটি বড় উৎস হয়ে ওঠে।



ইনফোগ্রাফিক: আল-জাজিরা

[৪১] Pardikar, R. (2019b). Barely Regulated Thermal Power Plants Use Up More Water Than Permitted, RTI Data Show. India Spend. https://www.indiaspend.com/barely-regulated-thermal-power-plants-use-up-more-water-than-permitted-rti-data-show/

অনেক ক্ষেত্রে এসব সংগঠন ও গণমাধ্যম, নির্দিষ্ট বিষয় নিয়ে জোট বেঁধে কাজ করছে এবং একে অপরের জন্য তথ্য ও বিশ্লেষণের উৎস হিসেবে আবির্ভূত হয়েছে। প্রাসঙ্গিকতা এবং স্থানীয় প্রেক্ষাপটের কারণে এখানে ভারতের উদাহরণ দেওয়া হচ্ছে, কিন্তু পশ্চিমের দিকে তাকালে এমন অসংখ্য উদাহরণ মিলবে। যুক্তরাষ্ট্রে তথ্য অধিকার আইনের নাম হলো "ফ্রিডম অব ইনফর্মেশন অ্যাক্ট" যা সংক্ষেপে ফোয়া নামে পরিচিত। সেখানে মাকরক (muckrock.com) নামে একটি সংগঠনই আছে যারা ফোয়ার মাধ্যমে তথ্যপ্রাপ্তির জন্য সবাইকে সহায়তা করে, প্রাপ্ত তথ্য ওয়েবসাইটে উন্মুক্ত করে এবং সেই তথ্য ব্যবহারের মাধ্যমে সাংবাদিকতাকেও উৎসাহিত করে।

> বাংলাদেশে এই আইনের বিকাশ ও ব্যবহার বাড়ানোর ক্ষেত্রে গণমাধ্যম ও নাগরিক সংগঠন উভয়েরই দায়িত্ব আছে। এক্ষেত্রে একজন সাংবাদিক হিসেবে আপনি যেমন এগিয়ে আসতে পারেন, তেমনি ইস্যুভিত্তিক তথ্য সংগ্রহে স্থানীয় ও জাতীয় পর্যায়ের নাগরিক সংগঠনগুলোও তাদের কাজে আরটিআইয়ের ব্যবহার বাড়াতে পারে। যেখানেই এ ধরনের আইন কার্যকর হয়েছে সেখানেই এই দুই ধরনের প্রতিষ্ঠানের যৌথ এবং একক প্রয়াস দেখা গেছে। বাংলাদেশে টিআইবি ও এমআরডিআইসহ বিভিন্ন প্রতিষ্ঠান সাংবাদিকদের তথ্য অধিকার আইনে তথ্যের জন্য আবেদন করার ক্ষেত্রে সহায়তা দিয়ে থাকে। তাদের সঙ্গে যোগাযোগ করলে, আপনি এ বিষয়ে সহায়তা পেতে পারেন।

বাংলাদেশে তথ্য অধিকার আইন

বাংলাদেশে তথ্য অধিকার আইন অত্যন্ত শক্তিশালী হলেও এর প্রয়োগ কম এবং সরকারি কর্মকর্তাদের বিরুদ্ধে তথ্য প্রদানে অনীহারও অভিযোগ রয়েছে।[৪২] কিন্তু সাংবাদিক এবং নাগরিক সংগঠনগুলো এই আইনের ব্যবহার যতবেশি করবেন, ততোই তথ্য প্রদানের নতুন সুযোগ তৈরি হবে এবং সরকারি কর্মকর্তাদের মধ্যে তথ্য দেওয়ার ক্ষেত্রেও অভ্যস্ততা তৈরি হবে। ভারতের তথ্য অধিকার আইনের সঙ্গে বাংলাদেশের আইন ও সংস্কৃতিতে অনেক মিল আছে। দেশটিতে সাংবাদিক এবং বিশেষ করে নাগরিক সংগঠনগুলো এই আইন ব্যবহার করে ডেটা সংগ্রহ, বিশ্লেষণ এবং সর্বোপরি জবাবদিহি প্রতিষ্ঠায় বড় ধরনের অবদান রেখে চলেছেন।

> বাংলাদেশে তথ্য অধিকার আইনে তথ্য পেতে হলে আপনাকে লিখিতভাবে আবেদন করতে হবে। এজন্য তথ্য কমিশনের সাইটে ফর্ম আছে, অথবা আপনি সেটি দেখে নিজে হাতে লিখে, সশরীরে নির্দিষ্ট অফিসে গিয়ে বা ইমেইলে ফর্মটি পাঠিয়েও তথ্যের জন্য আবেদন করতে পারেন। তথ্যের জন্য অনুরোধ পাঠাতে হয় তথ্য অধিকার কর্মকর্তা বরাবর। তারপর তিনি নির্দিষ্ট সময়ের মধ্যে আপনাকে তথ্য দেবেন, বা তথ্যটি দিতে অপরাগতা জানাবেন। যদি তার জবাব আপনার পছন্দ না হয়, তাহলে আপনি ঊর্ধতন অফিসে আপিল করতে পারেন, সেই আপিলেও কাজ না হলে তথ্য কমিশনে অভিযোগ দায়ের করতে পারেন।

[৪২] (2014, June 26). তথ্য প্রদানে অনীহা Prothomalo. Retrieved October 12, 2022, from https://www.prothomalo.com/opinion/%E0%A6%A4%E0%A6%A5%E0%A7%8D%E0%A6%AF-%E0%A6%AA%E0%A7%8D%E0%A6%B0%E0%A6%A6%E0%A6%BE%E0%A6%A8%E0%A7%87-%E0%A6%85%E0%A6%A8%E0%A7%80%E0%A6%B9%E0%A6%BE

বাংলাদেশে সাংবাদিকতায় তথ্য অধিকার আইনের ব্যবহার নিয়ে গণমাধ্যম উন্নয়ন প্রতিষ্ঠান ম্যানেজমেন্ট অ্যান্ড রিসোর্সেস ডেভেলপমেন্টে ইনিশিয়েটিভ (এমআরডিআই) এর একটি গাইড রয়েছে, যেখানে বিস্তারিতভাবে তুলে ধরা হয়েছে, একজন সাংবাদিক কীভাবে আইনটি ব্যবহার করে তথ্য পেতে ও সেটি কাজে লাগাতে পারেন। নিচে দেয়া কিউআর কোডটি স্ক্যান করে আপনিও চাইলে গাইডটি ডাউনলোড করে নিতে পারেন এবং এটি অনুসরণ করে তথ্য চাইতে পারেন।

কিউআর কোডটি স্ক্যান করে আপনিও চাইলে গাইডটি ডাউনলোড করে নিতে পারেন এবং এটি অনুসরণ করে তথ্য চাইতে পারেন

তথ্য অধিকার আইন ব্যবহারের বেশ কিছু সুবিধা আছে। প্রথমত, এই আইন ব্যবহার করে আপনি যে তথ্য বা নথি পাচ্ছেন সেটি আনুষ্ঠানিক বা অফিসিয়াল। এটি নিয়ে বিতর্কের সুযোগ কম। দ্বিতীয়ত, আপনার আবেদন যদি স্পষ্ট হয়, তাহলে তথ্যটি হবে আপনার স্টোরির জন্য প্রাসঙ্গিক, হালনাগাদ এবং তাতে বাহুল্য কম থাকবে। তৃতীয়ত, এ ধরনের তথ্য ব্যবহারে আইনি ঝুঁকিতে পড়ার আশঙ্কা কম বা নেই বললেই চলে। বেনামি সূত্রের কাছ থেকে পাওয়া নথি আপনাকে ডিজিটাল নিরাপত্তা আইনসহ নানা রকম আইনি ও মামলার ঝুঁকিতে ফেলতে পারে। এ জন্য কোনো বেনামি বা স্বল্পপরিচিত সোর্সের কাছ থেকে স্পর্শকাতর নথি পাওয়ার পরও অনেকেই তথ্য অধিকার আইনে একই নথির জন্য আবেদন জানিয়ে সোর্সের কাছ থেকে পাওয়া তথ্য যাচাই করে নেন।

ধারণা

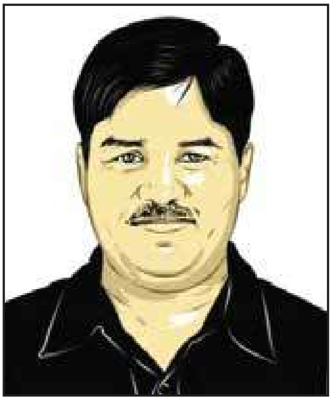
তথ্য অধিকার আইনে আপনি যে কত রকমের তথ্য পেতে পারেন, তা বলে শেষ করা সম্ভব নয়। এটি প্রতিটি বিটের জন্য ডেটা সাংবাদিকতা চর্চার অন্যতম হাতিয়ার হয়ে উঠতে পারে। এখানে শুধু একটি উদাহরণ দেয়া হলো, ধারণাটিকে স্পষ্ট করার জন্য। ধরুন, আপনি সরকারি অ্যাম্বুলেন্স সেবা নিয়ে একটি স্টোরি করতে চান। এজন্য কোন জেলায় কতটি সরকারি অ্যাম্বুলেন্স আছে, এর কতটি সচল, কতটি অচল এবং কত ড্রাইভার আছে জানতে চেয়ে একই আবেদন পাঠিয়ে দিলেন স্বাস্থ্য মন্ত্রণালয়ে, স্বাস্থ্য অধিদপ্তরে এবং প্রতিটি জেলার সিভিল সার্জন অফিসে। একবার ভাবুন, এই একটি আবেদন থেকে কত রকমের প্রতিবেদন সম্ভব। প্রথমত, অ্যাম্বুলেন্সের সংখ্যায় জেলাভেদে কোনও বৈষম্য আছে কি-না; কোন এলাকায় মাথাপিছু অ্যাম্বুলেন্সের সংখ্যা কত এবং সেটি চাহিদার তুলনায় কতটা পর্যাপ্ত বা অপর্যাপ্ত; কোথায় সচল অ্যাম্বুলেন্স বেশি, কোথায় কোথায় অ্যাম্বুলেন্স অচল থাকার কারণে মানুষের বিড়ম্বনা হচ্ছে বেশি। ডেটাকে যত প্রশ্ন করবেন, তত বেশি নতুন নতুন আইডিয়া আপনার মাথায় আসবে।

প্রতিবেদনের কথা না হয় বাদ থাকল। একবার ভেবে দেখুন এটি প্রতিটি এলাকার মানুষের জন্য কত গুরুত্বপূর্ণ হয়ে উঠতে পারে। আপনি না হয় প্রতিবেদন করলেনই না। নিছক ডেটাগুলো দিয়ে একটি ইন্টারঅ্যাকটিভ মানচিত্র তৈরি করলেন, যেখানে জেলায় অ্যাম্বুলেন্সের সংখ্যা, যোগাযোগের নম্বর এবং ভাড়ার একটি তালিকা দিয়ে দিলেন। এই একটি মানচিত্রই তখন আপনার ডেটা সাংবাদিকতার আউটপুট যেটি প্রতিটি জেলার মানুষকে গুরুত্বপূর্ণ স্বাস্থ্যতথ্য যুগিয়ে যেতে পারে অনেক দিন ধরে।

এমন নয় যে, আপনি তথ্য চাইবেন আর সবই পেয়ে যাবেন। হয়তো অনেক জেলার তথ্য পাবেন, অনেকগুলো পাবেন না। ভারতের বিদ্যুৎ কেন্দ্রের উদাহরণটির কথাই ভাবুন। তারা সব অঙ্গরাজ্যে তথ্য চেয়ে আবেদন জানিয়েছিল, কিন্তু ৪০০টি কেন্দ্রের মাত্র ১৫৬টির তথ্য পেয়েছে। কিন্তু সেটি তাদের প্রতিবেদন তৈরিতে কোনও বাধা হয়ে দাঁড়ায়নি। আপনার শুধু দরকার একটি আইডিয়া এবং একটু গবেষণা, যে এই স্টোরির জন্য কোন কোন ডেটা লাগবে, সেটি কোথায় কী অবস্থায় আছে, কীভাবে আবেদন করলে তথ্যটি পাওয়া যাবে। তারপর আবেদনটি পাঠিয়ে দিন।

পরামর্শ

ফিরে আসা যাক শ্যামলাল যাদবের কথায়। তিনি এখন ইন্ডিয়ান এক্সপ্রেসের ইনভেস্টিগে-টিভ ব্যুরোর সিনিয়র এডিটর হিসেবে কাজ করছেন।



ছবি: শ্যামলাল যাদব

২০২০ সাল নাগাদ তিনি প্রায় ১০ হাজার আবেদন করেছেন[৪৩] তথ্য অধিকার আইনে এবং এখনো করে যাচ্ছেন। ২০১৭ সালে প্রকাশিত "জার্নালিজম থ্রু আরটিআই" বইয়ের শেষ অধ্যায়ে তিনি সাংবাদিকদের জন্য কয়েকটি গুরুত্বপূর্ণ পরামর্শ দিয়েছেন। পরামর্শগুলো[৪৪] আপনারও কাজে আসবে–

০১	প্রথমে চিন্তাভাবনা গুছিয়ে নিন।
০২	আরটিআই ছাড়াই যে তথ্য পাওয়া যায় সেগুলো দিয়ে শুরু করুন।
০৩	প্রশ্ন করবেন না, শুধু তথ্যের কথা জানতে চান।
০৪	আবেদন সহজ ও স্পষ্ট রাখুন। ফাঁকফোকর সম্পর্কে সচেতন থাকুন।
০৫	অধিকার প্রয়োগ করার আগে জেনে নিন, কোন তথ্য পাওয়া যায় না।
০৬	একই তথ্যের জন্য সংশ্লিষ্ট সব সরকারি কর্তৃপক্ষকে ব্যবহার করুন।
০৭	দায়িত্বপ্রাপ্ত কর্মকর্তার সঙ্গে পরিচিত হয়ে নিন।

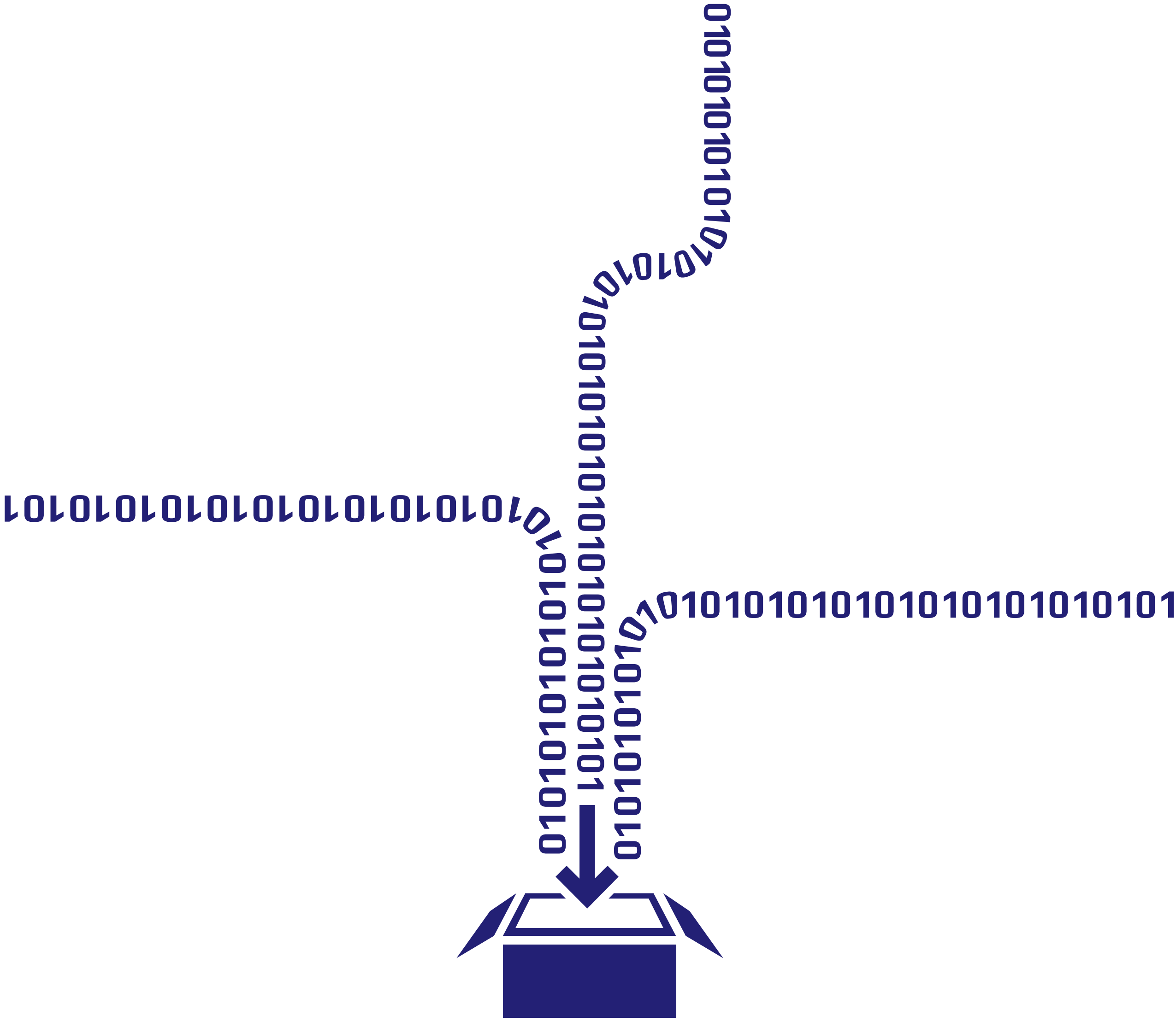
[৪৩] Yadav, Shyamlal. (2020, February 25). Using RTI should be every journalist's routine: Bennett University. Retrieved October 12, 2022, from https://www.bennett.edu.in/using-rti-should-be-every-journalists-routine-shyamlal-yadav/

[৪৪] Global Investigation Journalism Network. (2019b, October 14). তথ্য অধিকার আইন প্রয়োগের যত রকম কৌশল. Global Investigative Journalism Network. Retrieved October 12, 2022, from https://gijn.org/%E0%A6%A4%E0%A6%A5%E0%A7%8D%E0%A6%AF-%E0%A6%85%E0%A6%A7%E0%A6%BF%E0%A6%95%E0%A6%BE%E0%A6%B0-%E0%A6%86%E0%A6%87%E0%A6%A8-%E0%A6%AA%E0%A7%8D%E0%A6%B0%E0%A7%9F%E0%A7%8B%E0%A6%97%E0%A7%87%E0%A6%B0-...%E0%A7%9F%E0%A7%8B%E0%A6%97/

০৮	প্রয়োজন বুঝে লক্ষ্য পরিবর্তন করুন।
০৯	সাংবাদিক পরিচয় প্রকাশ করার কোনো প্রয়োজন নেই।
১০	যে কোনো ধরনের প্রতিক্রিয়া পাওয়ার জন্য প্রস্তুত থাকুন।
১১	মনে রাখবেন, “না” এর অর্থ সবসময়ই না-ই হবে, এমন কোনো কথা নেই।
১২	সময় নষ্ট করবেন না, আবেদন অনলাইনে জমা দিন অথবা ডাকে।
১৩	ধৈর্য্য ধরুন, হাল ছাড়বেন না।
১৪	তথ্য পাওয়ার জন্য বাড়তি কাজ করুন।
১৫	আরটিআই কর্মকর্তাদের প্রতি সংবেদনশীল হোন।
১৬	তথ্য কমিশনারদের সাহায্য নিন।
১৭	কাগজপত্র পরিদর্শন করুন।
১৮	ফলোআপ চালিয়ে যান।

আরটিআই নিয়ে আরও জানতে চান, কিউআর কোডটি স্ক্যান করে জিআইজেএনের এই গাইডটি পড়ে নিতে পারেন।

0101010101010101010101010101010101010101010101010101



অধ্যায় ০৬

ক্লিনিং ও সংরক্ষণ

ক্লিনিং ও সংরক্ষণ

> সংগ্রহ করা ডেটা গুগল শিটসে লিপিবদ্ধ করার সময় ভুল বা সংখ্যার তারতম্য হতে পারে। একই তথ্য একাধিকবার ব্যবহার হতে পারে। আবার কিছু ডেটা হারিয়ে যেতে পারে। বিশ্লেষণের আগে এই সব অনাকাঙ্ক্ষিত ভুল শুধরে নিতে আপনাকে ডেটা যাচাই করতে হবে।

প্রতিবেদনের জন্যে সাংবাদিকদের বিভিন্ন উৎস থেকে ডেটা সংগ্রহ করতে হয়। সেই ডেটা যে সব সময় শতভাগ নির্ভুল বা ব্যবহার উপযোগী থাকে, সেটি বলা যাবে না। সংগ্রহ করা ডেটা গুগল শিটসে লিপিবদ্ধ করার সময় ভুল বা সংখ্যার তারতম্য হতে পারে। একই তথ্য একাধিকবার ব্যবহার হতে পারে। আবার কিছু ডেটা হারিয়ে যেতে পারে। বিশ্লেষণের আগে এই সব অনাকাঙ্ক্ষিত ভুল শুধরে নিতে আপনাকে ডেটা যাচাই করতে হবে। চলুন জেনে নেই কীভাবে ডেটাকে নির্ভুল বা বুলেটপ্রুফ করা যায়।

ধাপ ১

সংগ্রহ এবং পর্যালোচনা

অনলাইন থেকে হোক কিংবা অন্য কোনো উৎস থেকে, আপনি যখন কোনো ডেটা সংগ্রহ করবেন প্রথমেই দেখে নিতে হবে সেখানে পূর্ণাঙ্গ ডেটা আছে কি-না এবং আপনি সেই ডেটা বুঝতে পারছেন কি-না অর্থাৎ ডেটা বোধগম্য কি-না।

- পুরো ডেটাতে একবার চোখ বুলিয়ে নিন। দেখে নিন আপনি যা চেয়েছেন তার পুরোটা আছে কি-না। সব ডেটার ব্যাখ্যা আছে কি-না। যদি কোনো তথ্যের ব্যাখ্যা বা বিস্তারিত না থাকে, তবে ডেটা সরবরহকারি প্রতিষ্ঠানের দৃষ্টি আকর্ষণ করুন এবং অনুপস্থিত সেই ব্যাখ্যা সংগ্রহ করুন।
- ডেটার একাধিক কপি করুন এবং একাধিক পৃথক স্থানে সেটি সংরক্ষণ করুন।
- পূর্ণাঙ্গ তথ্য পড়ুন এবং নিশ্চিত করুন যে আপনার সংগ্রহ করা সব ডেটার মানে বা অর্থ বুঝতে পেরেছেন। না পারলে ডেটা সংগ্রহকারী বা সরবরাহকারীর সহযোগিতা নিন।
- কাজের সুবিধার্থে একটি ডেটা ডায়েরি তৈরি করুন এবং সেখানে কাজের তালিকা লিপিবদ্ধ করুন। যাতে কাজের দীর্ঘ এই যাত্রায় কোনো কিছু ভুলে না যান।
- গুগল শিটসে ডেটা যুক্ত করতে সঠিকপদ্ধতি অনুসরণ করুন।

ধাপ ২

ডেটার কাঠামো যাচাই

বিশ্লেষণের আগে ডেটাকে প্রাথমিক কিছু যাচাইয়ের মধ্য দিয়ে নিতে হবে। যাতে নিশ্চিত হতে পারেন ব্যবহার করা ডেটা নির্ভুল। এক্ষেত্রে কারিগরি কৌশলের থেকে পর্যবেক্ষণ ক্ষমতার ব্যবহার বেশি গুরুত্বপূর্ণ। যদি কোনো কিছু অদ্ভূত বা খটকা লাগে তবে আরও এক দফা ভাল করে যাচাই করুন। বিভ্রান্তির মূল কারণ অনুসন্ধান করুন। বাস্তবে সেখানে কোনো ভুল থাকতে পারে, আবার নাও থাকতে পারে। তবে যাচাই করে নেয়া সর্বোত্তম।

- প্রতিটি টেবিল বা শিটসের জন্য ডেটা পরীক্ষা করতে হবে। গুনে দেখুন সরবরাহকারী সংস্থা যে পরিমাণ ডেটা দিয়েছে বা যত ডেটা সংগ্রহ করা হয়েছে, তার সঙ্গে মোট সংখ্যার মিল আছে কি-না। কোনো ডেটা যেন হারিয়ে না যায়, তা নিশ্চিত করুন।
- ডেটা টেবিলের উপরের বাম থেকে উপরের ডানদিকে, নীচের ডানে এবং নীচের বাম দিকে চলে যান। নিশ্চিত করুন যে ডেটা টেবিলে কোনো শূন্যস্থান বা ফাঁকা জায়গা নেই। সব ডেটা তাঁর নির্দিষ্ট জায়গায় আছে।
- ডেটা টেবিল বা গুগলশিটের কাঠামো পরীক্ষা করতে হবে। প্রতিটি টেবিলের ক্ষেত্রে কাঠামো একই আছে কি-না তা যাচাই করুন। সংখ্যার স্থনে সংখ্যা, তারিখের স্থানে তারিখ বসেছে কি-না, তা নিশ্চিত করুন।
- যদি একাধিক টেবিলে ডেটা থাকে তবে নিশ্চিত করুন সব টেবিলের ডেটা একই ধরনের বা ফরম্যাটের। ভিন্ন ভিন্ন ফরম্যাটের ডেটা থাকলে সেটি এক শিটসে বিশ্লেষণ করা যাবে না।

ধাপ ৩

অপরিষ্কার/অপরিচ্ছন্ন ডেটা খুঁজুন

ডেটাশিটে নামের বানান ও স্পেস বেশ গুরুত্বপূর্ণ। ক্লিনিংয়ের সময় এটা মাথায় রাখতে হয়। আপনি সব সময় একেবারে তৈরি, পরিচ্ছন্ন ডেটাসেট পাবেন না। কোনো কোনো সময় র' ডেটা ফাইল নিয়েও কাজ করতে হতে পারে। সেখানে দেখা যায়, ম্যানুয়াল ডেটা ইনপুটের কারণে সারিতে একই শব্দের নাম বিভিন্ন হয়। আবার অনেক ক্ষেত্রে আপনি যে সোর্স থেকে ডেটা সংগ্রহ করেছেন তারাও এমন ভুল করতে পারে। যেমন: আপনার ডেটাসেটে হয় তো ১০০ বার ইংরেজিতে Dhaka শব্দটি রয়েছে। কিন্তু ডেটা এন্ট্রির সময় কোথাও ঢাকা শব্দটির আগে একটি স্পেস (খালি জায়গা) পড়েছে, কোথাও হয়তো তাড়াহুড়োয় একটি a বাদ পড়ে গিয়ে Dhka লেখা হয়েছে। এমন ক্ষেত্রে এই একটি স্পেস থাকা বা অক্ষর বাদ পড়ার কারণে, আপনার বিশ্লেষণে পুরোপুরি ভুল ফলাফল আসবে। তাই যে কোনো ডেটাসেট বিশ্লেষণের

আগে তাকে পরিষ্কার বা ক্লিন করে নিতে হয়। বলা হয়, কোনো ডেটা বিশ্লেষণের কাজে ৯০ ভাগ সময়ই ব্যয় হয় ডেটা পরিষ্কারের কাজে।

মাইক্রোসফট এক্সেল ও গুগল শিটসের পিভট টেবিল বা ফিল্টার টুল ব্যবহার করে ভুল ডেটা চিহ্নিত করা সহজ।

> – পিভট টেবিল ব্যবহার করলে, আপনি যে কলামটির ডেটা যাচাই করছেন সেটি সিলেক্ট করে, পিভট টেবিল তৈরি করুন। এরপর Pivot Table Editor এর ROW এবং Value তে যে কলামের তথ্য যাচাই করবেন সেটি যুক্ত করুন। এরপর গুগল শিটস স্বয়ংক্রিয়ভাবে একটি টেবিল তৈরি করবে। যেখানে প্রতিটি ডেটা আলাদা করে প্রদর্শিত হবে এবং ডেটাশিটে সেটি কতবার আছে তাও উল্লেখ থাকবে। একইসাথে কোন ডেটাটি সবচেয়ে বেশি বার উল্লেখ আছে এবং কোনটি কম। অর্থাৎ ব্যতিক্রম বা চোখে পড়ার মতো যে কোনো অসঙ্গতি পিভট টেবিল আপনাকে খুঁজে দেবে। অনেক প্রতিবেদনে ডেটা এতো বেশি থাকে যে একবারে পুরোটা পিভট টেবিলের মাধ্যমে যাচাই করা সম্ভব হয় না। এসব ক্ষেত্রে দুই-তৃতীয়াংশ ডেটার ক্ষেত্রে এটি করা যেতে পারে। অথবা, ডেটার যে অংশটি আপনার কাছে সন্দেহজনক মনে হবে সেটুকুও করতে পারেন।

– যদি কোনো ডেটা কতবার আছে তা জানতে না চান, তবে গুগল শিটসের ফিল্টার অপশন ব্যবহার করতে পারেন। যেখানে আপনি খুব দ্রুত প্রতিটি কলামে থাকা একই ধরনের কতগুলো ডাটা আছে, তা জানতে পারবেন। তবে ডেটার পরিমাণ যদি দীর্ঘ হয়, তবে ফিল্টার অপশন ব্যবহার করে অসামঞ্জস্যপূর্ণ ডেটা বের করা কঠিন।

যেখানে সমস্যা পাবেন

যে সব ডেটা আপনার বিশ্লেষণে সমস্যা করতে পারে, সেগুলো চিহ্নিত করে আলাদা ভাবে টুকে রাখুন। যেন বিশ্লেষণের ফলাফলে সমস্যা দেখা দিলে আপনি সহজেই বুঝতে পারেন সমস্যা কোথায়। নিচে ভুল ডেটার কিছু প্রচলিত ধরন তুলে ধরা হলো–

তারিখ

ডেটা শিটসে থাকা তারিখগুলো ঠিক আছে কি-না তা জানতে Ascending এবং Descending দুইভাবেই যাচাই করুন। শুরুর দিন থেকে শেষ দিন পর্যন্ত অর্থাৎ ডেটার সময়সীমাটি আলাদা করে নোট করে রাখুন। কেননা প্রতিবেদন লেখার সময় এই তথ্যটি উল্লেখ করার প্রয়োজন হতে পারে। একইসঙ্গে তারিখের ভিত্তিতে ঘটনার বা ডেটার ধারাবাহিকতায় কোনো অসংগতি আছে কি-না সেটাও যাচাই করুন। ত্রুটি থাকলে ডেটা সংগ্রহকারীদের দৃষ্টি আকর্ষণ করুন।

অজানা অক্ষর

অনেক ক্ষেত্রে একই ডেটাসেটা ভিন্ন ভাষার ডেটা বা তথ্য উল্লেখ থাকতে পারে। আবার ফন্টের ভিন্নতার কারনেও অনেক ক্ষেত্রে মাইক্রোসফট এক্সেল ও গুগল শিটের বিশ্লেষণে সমস্যা হতে পারে। তাই ডেটাসেটের সব তথ্যের অক্ষর একই ভাষার হওয়া প্রয়োজন।

শূন্যস্থান

ডেটাসেটের কোনো সেল যদি ফাঁকা থাকে তবে সেটিও আপনার বিশ্লেষণে সমস্যা তৈরি করতে পারে। শুধু ডেটা না থাকার কারণেই যে সেলে ডেটা নেই, সবসময় বিষয়টি এমন নাও হতে পারে। কাজের মাঝে কখনো কোনো সেল হতে তথ্য মুছে যেতে পারে। তাই ডেটা সেটের সব সেলে তথ্য আছে কি-না, না থাকলে কেনো নেই সে বিষয়ে পরিষ্কার ধারণা থাকতে হবে।

বিরতি চিহ্ন

ডেটার মাঝে, কিংবা আগে পরে অনাকাঙ্ক্ষিতভাবে স্পেস বা বিরতি চিহ্ন থাকলে তথ্য বদলে যায়, তখন স্প্রেডশিটের বিশ্লেষণেও ভিন্নতা আসে।

কোড

কর্তৃপক্ষ সব সময় দুই ধরনের তথ্য সংরক্ষণ ও সরবরাহ করে। কোডেড এবং নন-কোডেড বা কোড ছাড়া। কোডেড অর্থ কোনো সংকেত ব্যবহার করে তথ্য লিপিবদ্ধ করা হয়েছে। এই ধরনের ডেটা বিশ্লেষণের ক্ষেত্রে কোডের অর্থ জানতে হবে। যা আপনাকে ডেটা সংগ্রহ বা সংরক্ষণকারি সংস্থা সরবরাহ করতে পারবে।

আইডি

ডেটাশিটের একটি কলামে শুধুমাত্র অনন্য তথ্য যেমন, আইডি কার্ডের নম্বর থাকতে পারে। সেটি সঠিকভাবে আছে কি-না, তা পরীক্ষা করতে কলামটির তথ্য পিভট টেবিলের মাধ্যমে যাচাই করা যেতে পারে। এরপর সর্বনিম্ন থেকে সর্বোচ্চের ভিত্তিতে ডেটাকে সাজান। এরপর যদি সেই তালিকায় অসংগতি দেখতে পান, বুঝতে হবে ডেটাতে ত্রুটি আছে। কারিগরি এসব কৌশলের বাইরে ডেটার ত্রুটি চিহ্নিত করে আপনাকে সাধারণ জ্ঞানেরও প্রয়োগ করতে হবে। সেটি ডেটা সাংবাদিকতার এ পর্যায়ে খুবই গুরুত্বপূর্ণ।

সূত্রে সমাধান

গুগল শিটসের বেশিরভাগ কার্যক্রমই গণিতনির্ভর। গণিতের সাহায্যে শিটস সকল ধরনের হিসাব করে থাকে। যোগ, বিয়োগ, গুণ, ভাগ থেকে শুরু গড় গণিতের সকল সূত্রেরই ব্যবহার আছে এখানে। সংখ্যার বাইরেও ডেটার যে কোনো ধরন যেমন, টেক্সট বা অক্ষরনির্ভর তথ্যও গুগল শিটস বিশ্লেষণ করতে পারে নিজস্ব কার্যক্রম বা সূত্রের মাধ্যমে। সংখ্যার বাইরে অক্ষর নির্ভর তথ্যের এই কার্যক্রমকে স্ট্রিং ফাংশন্স বলা হয়। কারণ কম্পিউটার বিজ্ঞানে টেক্সটকে স্ট্রিংস হিসেবে চিহ্নিত করা হয়।

> স্ট্রিং ফাংশন্স গুগল শিটসে থাকা ডেটার মধ্যে প্যাটার্ন খুঁজতে সহায়তা করে। এর মাধ্যমে টেক্সটকে একটি থেকে আরেকটিকে আলাদা করা যায়। আবার জোড়াও দেয়া যায়। সেই সঙ্গে শিটসে বড় হাতের বা ছোট হাতের অক্ষর লেখা যায়, এই ফাংশন্সের মাধ্যমে।

এর বাইরে গুগল শিটসে আরও দুইটি ফাংশন্স কাজ করে: লজিক্যাল ফাংশন্স এবং ডেট ফাংসন্স। তারিখ-সংক্রান্ত তথ্যের পুনর্বিন্যাস করতে ডেট ফাংশন্স ব্যবহার করা হয়। আর লজিক্যাল ফাংশন্স তথ্যের সত্যতা খুঁজতে সহায়তা করে।

> তবে আপনাকে এই সব ফাংশন্স বা সূত্র মনে রাখতে হবে না। সেটি সম্ভবও না। কাজের সময়ে তাৎক্ষনিক কোনো সূত্র বা ফাংশন্স মনে না পড়লে, Google Sheets help online এর সহায়তা নিতে পারেন কিংবা গুগলে সমাধান খুঁজতে পারেন।

অনুশীলন করুন

ডেটা ক্লিনিং এর এই অনুশীলনে যে ডেটা নিয়ে কাজ করা হবে, তা ইংরেজি। এর কারণ মাইক্রোসফট এক্সেল বা গুগল শিটসের সূত্রগুলো এমনভাবে সাজানো যে, ইংরেজি ছাড়া অন্য কোনো ভাষায় হিসাব করতে গেলে তা ভুল হবার আশঙ্কা তৈরি হয়, বিশেষ করে বাংলার ক্ষেত্রে অনেক যুক্তাক্ষর থাকায় সূত্র মেলানো কঠিন হয়ে যায়। তাই এই পাঠ্যক্রম ও অনুশীলনে বাংলাদেশের ইংরেজি ডেটা ব্যবহার করা হবে। এ ছাড়া আরেকটি বড় কারণ হচ্ছে, দেশের সরকারি বেসরকারি যে সব সংস্থা ডেটা সংরক্ষণ বা সরবরাহ করে তারা সবাই ইংরেজিতেই ডেটা সংরক্ষণ করে।

ডেটা ক্লিনিং ও সংরক্ষণের অনুশীলনের জন্য ডেটা সাংবাদিকতা সহায়িকার গুগল ড্রাইভে থাকা Data Cleaning স্প্রেডশিটটি ব্যবহার করতে হবে। গুগল ড্রাইভ থেকে কপি করে আপনার নিজের গুগল ড্রাইভে রাখুন। কিউআর কোডটি স্ক্যান করলে স্প্রেডশিটটি পাবেন।

অনুশীলন ১ : TRIM/ট্রিম

এই অনুশীলনের জন্য ডেটা সাংবাদিকতা সহায়িকার গুগল শিটস ক্লিনিং ফাইলটি ব্যবহার করা হবে। তাই কাজের শুরুতে ফাইলটি খুঁজে নিন।

> ডেটা ক্লিনিং অনুশীলনের প্রথম এই ধাপে প্রথমে অপরিচ্ছন্ন বা এলোমেলো ডেটা খুঁজে বের করতে হয়। এই কাজের জন্যে পিভট টেবিল কার্যকর একটি ফাংশন, যা আগেই বলা হয়েছে। এর মাধ্যমে জানা যায়, গুগল শিটসে ডেটা গ্রুপভিত্তিক ভাবে সাজানো আছে না কি বিক্ষিপ্তভাবে আছে। ডেটার বানানে কোনো ভুল আছে কি-না সেটিও যাচাই করা যাবে এই ফাংশন্সের মাধ্যমে।

অনুশীলনের প্রথমেই চলুন দেখে নেই একই ধরনের তথ্য শিটসে কতগুলো পৃথক জায়গায় রয়েছে। অর্থাৎ একই ধরনের তথ্য ভিন্ন ভিন্ন সেলে উল্লেখ করা হয়েছে। এটা জানতে পিভট টেবিলের সহায়তা নিতে হবে। Trim নামের শিটসটিতে থাকা অবস্থায় Insert মেন্যু থেকে Pivot Table-এ ক্লিক করুন এবং একটি Pivot Table তৈরি করুন। এজন্য New Sheets অপশনে ক্লিক করুন এবং ডেটা রেঞ্জের ক্ষেত্রে প্রথম থেকে শেষ পর্যন্ত সব সেল উল্লেখ করুন। এরপর Row বা সারি অপশনে। যে কলামের ডেটা সম্পর্কে জানতে চাচ্ছেন সেটি বসান। এই অনুশীলনের ক্ষেত্রে বিভাগ বসাতে হবে। বিভাগগুলোর নামে তারতম্য আছে কি-না তা যাচাই করতে Value অপশনেও বিভাগ যুক্ত করুন।

	A	B
1	*DIVISION*	COUNTA of DIVISION
2	DHK	1
3	RNG	1
4	BRL	2
5	BRL	1
6	CTG	6
7	DHK	6
8	KHL	2
9	KHL	1
10	MNS	2
11	MNS	2
12	rajsahi	1
13	RJS	3
14	RJT	1
15	RNG	6
16	ROG	1
17	SLT	3

পিভট টেবিলের বিশ্লেষণে দেখা যাচ্ছে, একই নাম বা কাছাকাছি নামের অনেক তথ্য আছে এই গুগল শিটসটিতে। কোনো কোনো বানানের ক্ষেত্রে একটি অক্ষরের ভিন্নতা রয়েছে। আবার কোথাও একটি স্পেসের কারণে গুগল শিটস সেটিকে আলাদা হিসেবে গণনা করেছে। ভাল খবর হলো গুগল শিটসের এমন কিছু স্ট্রিং ফাংশন্স আছে, যেগুলো ব্যবহার করে ডেটা বা টেক্সটের মধ্যে ভুলবশত থাকা স্পেস মুছে, একই ধরনের ডেটাকে এক গ্রুপের আওতায় নিয়ে আসা যায়। গুগল শিটসের এই পদ্ধতিটির নাম TRIM।

=TRIM(cell)

গুগল শিটস বা মাইক্রোসফট এক্সেল যেখানেই ডেটা ক্লিনিং–এর কাজ করুন না কোনো, কখনো ভুল বা অপরিচ্ছন্ন ডেটা মুছে, সেই একই সেলে নতুন পরিষ্কার ডেটা বসানো যাবে না। পরিষ্কার ডেটার জন্যে নতুন কলাম তৈরি করুন।

এই শিটে যেহেতু কলাম A এবং B তে দুইটি ডেটাসেট আছে, তাই ক্লিন করা ডেটা C কলামে রাখুন। নতুন কলামটিকে নতুন নাম দিন। এরপর C2 সেলে TRIM এর সূত্র বা ফাংশনটি লিখুন।

এক্ষেত্রে সূত্রটি **=TRIM(B2)**

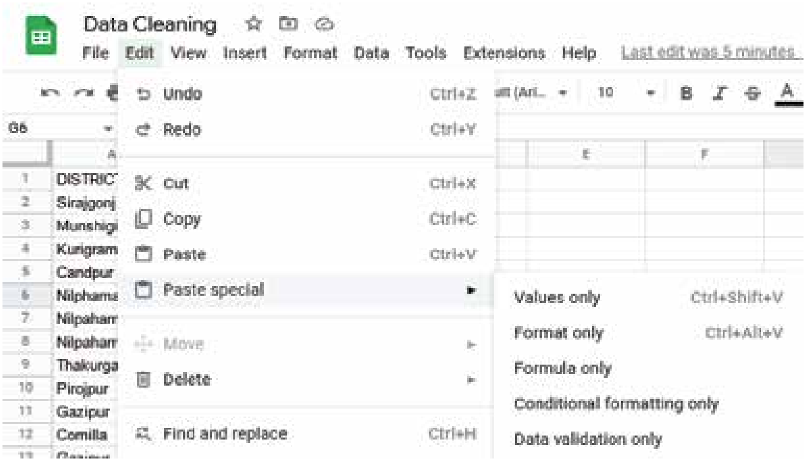
C2 | *fx* | =TRIM(B2)

	A	B	C
1	DISTRICT	DIVISION	RNG × R DIVISION
2	Nilpahamri	RNG	=TRIM(B2)
3	Manikganj	DHK	DHK
4	Pirojpur	BRL	BRL
5	Habiganj	SLT	SLT
6	Bhola	BRL	BRL
7	Sunamganj	SLT	SLT
8	Comilla	CTG	CTG
9	Jamalpur	DHK	DHK
10	Bhola	BRL	BRL
11	Candpur	CTG	CTG
12	Feni	CTG	CTG
13	Noakhali	CTG	CTG
14	Munshiginj	DHK	DHK

এই ফাংশন বা সূত্র ডেটাসেটের অনেক অসংগতি দূর করবে। তবে এরপর কিছু টেক্সটের ক্ষেত্রে ত্রুটি হয়তো থেকে যাবে, যেগুলো আপনি হাতে ধরে সংশোধন করতে পারবেন।

তবে আপনি সরাসরি C কলামের টেক্সট পরিবর্তন বা এডিট করতে পারবেন না। কেননা, C কলামে থাকা টেক্সটগুলো একটি ফাংশন বা সূত্রের সাহায্যে সেখানে বসানো হয়েছে। এ কারণে আপনাকে C কলামের সব ডেটাকে CUT করে একই কলাম অথবা পাশের ভিন্ন কোনো কলামে Special Paste করতে হবে। এক্ষেত্রে Special Paste টি করতে হবে Value-তে, Formula-তে নয়।

Edit > Paste special > Paste values only.



এর ফলে কলাম C তে থাকা ডেটা সূত্রের বাইরে চলে আসছে। এবার আপনি স্বাচ্ছন্দ্যে টেক্সট পরিবর্তন বা পরিমার্জন করতে পারবেন। আর পুরো কলামের ডেটা একবারে সিলেক্ট করতে কলামের উপরে A, B, C তে ক্লিক করুন।

এবার নতুন কলামে থাকা ক্লিনিং করা ডেটা আরও একবার পিভট টেবিলে তুলে দেখুন, এখন কোন কোন টেক্সটে অক্ষরগত ত্রুটি আছে।

	A	B
1	CLEAR DIVISION	COUNTA of CLEAR DIVISION
2	BRL	3
3	CTG	6
4	DHK	7
5	KHL	3
6	MNS	4
7	rajsahi	1
8	RJS	3
9	RJT	1
10	RNG	7
11	ROG	1
12	SLT	3
13	Grand Total	39

এসব অসংগতি সূত্র ছাড়া টাইপ করেই সংশোধন করতে হবে আপনাকে।

অনুশীলন ২ : স্প্লিট ১

এবার গুগল শিটসে থাকা পরবর্তী শিট বা ট্যাবটি-তে যান। শিটটির নাম দেয়া হয়েছে "Split Dates" কারণ অনুশীলনের এই পর্যায়ে তারিখের ত্রুটি দূর করা হবে।

"Split Dates" নামের শিটটি ব্যবহারের জন্যে কিউআর কোডটি স্ক্যান করুন।

এই ফাইলে থাকা জন্ম তারিখগুলো টেক্সট আকারে যুক্ত করা হয়েছে। ফলে সংখ্যাগুলো দেখে সেটি কোনো তারিখ নির্দেশ করে, তা বোঝা কিছুটা কষ্টকর। তবে এমন টেক্সট থেকে তারিখ বের করে আনার সূত্র গুগল শিটসে আছে।

স্প্রেডশিটে টেক্সট ডেটাটি আছে ১৯৮২০৩০১ এভাবে। কিন্তু তারিখটি মূলত ০১/০৩/১৯৮২। অর্থাৎ টেক্সটের ফরম্যাটের কারণে পুরো ডেটাটি কোনো অর্থ বহন করছে না।

	A	B	C	D	E
1	Date of Birth	Year	Month	Day	New Date of Birth
2	19820301	1982	03	01	01-03-1982
3	19820519	1982	05	19	19-05-1982
4	19800106	1980	01	06	06-01-1980
5	19880103	1988	01	03	03-01-1988
6	19870101	1987	01	01	01-01-1987
7	19850714	1985	07	14	14-07-1985
8	19850512	1985	05	12	12-05-1985
9	19880307	1988	03	07	07-03-1988
10	19820322	1982	03	22	22-03-1982

এই অনুশীলনে ডেটার বিদ্যমান ফরম্যাটটি পরিবর্তন করে, বাংলাদেশে যেভাবে তারিখ লেখা হয়, সে ফরম্যাটে রূপান্তর করা হবে। এ জন্যে প্রথমে বিদ্যমান ডেটা থেকে একটি তারিখের প্রত্যেকটি অংশ অর্থাৎ বছর, মাস ও তারিখ আলাদা করতে হবে। তারপর আবার সবগুলো ডেটাকে যুক্ত করতে হবে।

মনে রাখবেন, বিদ্যমান কলামে ডেটার পরিবর্তন বা কোনো অনুশীলন করা উচিত না। এ কারণে খালি কলামগুলোতে এই টেক্সট থেকে ডেটা আলাদা করতে হবে এবং পরে আবার তারিখের ফরম্যাটে তা যুক্ত করা হবে।

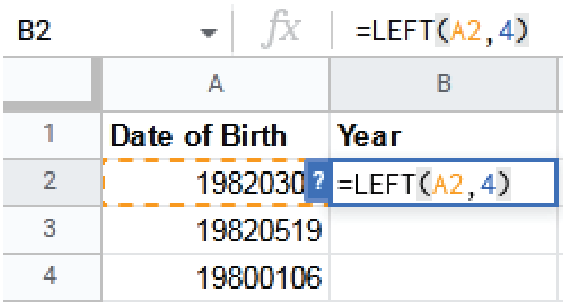
টেক্সট থেকে ডেটা আলাদা করার কাজটি আপনি ডেটার যে কোনো অংশ থেকে শুরু করতে পারেন। এখানে টেক্সটের বামদিক থেকে শুরু করা হবে। তাই এই সূত্রটিও শুরু হয় LEFT শব্দ দিয়ে। এই সূত্রে আপনি লিখবেন, আপনি কোন পাশের অক্ষর আলাদা করতে চান। এই পৃথককরণ কোনো সেল থেকে শুরু হবে এবং টেক্সটে থাকা কয়টি অক্ষর আপনি আলাদা করবেন।

=LEFT (সেল নম্বর, আলাদা করতে চাওয়া অক্ষর সংখ্যা)

টেক্সট থেকে প্রথমেই বছরের ডেটাকে আলাদা করুন। কারণ এটি সবার বামে আছে। নতুন কলাম Year এ, নতুন ডেটাটি রাখতে হবে। তাই সূত্রটি সেই কলামে বসান।

এখানে সূত্র হবে **=LEFT(A2,4)**

কারণ আপনি বাম দিকের ডেটা আলাদা করতে চাইছেন। আপনি যে টেক্সট থেকে ডেটা নিতে চাইছেন সেটি A2 তে আছে এবং বামদিক থেকে প্রথম ৪টি অক্ষর আলাদা করতে চাইছেন।



বছরের কলামে আপনি এরই মধ্যে নতুন ডেটা পেয়ে গেছেন। B2 তে বছরের ডেটা যুক্ত হয়েছে।

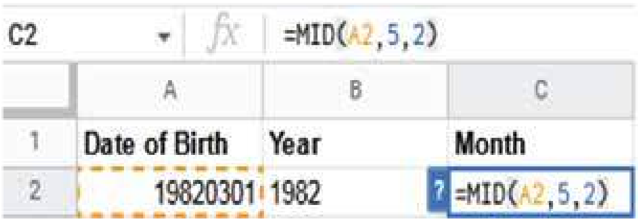
নতুন ডেটা যুক্ত হবার সাথে সাথে একইসূত্র মেনে কলামের বাকি টেক্সটগুলো থেকে বছর আলাদা করার একটি অপশন আপনার সামনে দেয়া হয়েছে। আপনি Okay অপশনে ক্লিক করতে পারেন। তাহলে পুরো কলামের বছরের তথ্যই একই সূত্রে স্বয়ংক্রিয়ভাবে আলাদা হয়ে যাবে।

বছরের পর, এবার টেক্সট থেকে মাসের ডেটা আলাদা করুন। প্রাথমিক টেক্সট দেখেই বোঝা যাচ্ছে মাঝের দুইটি অক্ষর মাসকে নির্দেশ করছে। তাই আমাদের সূত্র হবে মাঝ বা MID দিয়ে। যেহেতু এখানে আপনি টেক্সটের মাঝামাঝি স্থান থেকে অক্ষর আলাদা করতে চাইছেন, তাই এবার আপনাকে ঠিক কোন অবস্থান বা কোন অক্ষর থেকে ডেটা আলাদা করার কাজ শুরু করতে চাইছেন, তা উল্লেখ করতে হবে। তারপর আগের মতোই কয়টি অক্ষর আলাদা করতে চাইছেন তাও উল্লেখ করতে হবে।

=MID (সেল নম্বর, শুরুর অবস্থান, আলাদা করতে চাওয়া অক্ষর সংখ্যা)

এই নিয়ম অনুসারে, মাসের তথ্য আলাদা করতে আমাদের সূত্র বা ফরমুলা হবে =MID (A2,5,2)। কারণ মূল টেক্সট অর্থাৎ যে টেক্সট থেকে মাসের ডেটা আলাদা করা হবে, সেটি আছে A2 তে। যে অক্ষর আপনি আলাদা করতে চান, টেক্সটটে সেটির অবস্থান 5 এ। আর আপনি দুইটি অক্ষর আলাদা করতে চান মাসের ডেটা বের করতে চান, তাই সবশেষে 2।

কলাম **C** তে এই সূত্র বসালেই আপনি মাসের ডেটা পেয়ে যাবেন।



এবার একইপদ্ধতিতে ডান পাশের ডেটা টেক্সট থেকে পৃথক করুন। ডান পাশের ডেটাটি দেখেই মনে হচ্ছে এটি দিনকে নির্দেশ করছে। তাই এবার সূত্রের শুরু হবে RIGHT দিয়ে। টেক্সট যেহেতু A কলামের সেল 2 থেকে নিচ্ছি তাই A2। আর দিনের ডেটা বের করতে হবে, দুই অক্ষর নির্দিষ্ট করতে হবে অর্থাৎ 2।

=RIGHT (সেল নম্বর, আলাদা করতে চাওয়া অক্ষর সংখ্যা)

এর মাধ্যমে আপনি টেক্সট থেকে তারিখের সবগুলো ডেটা আলাদা আলাদা করে পেয়ে গেছেন। এবার সেই ডেটাগুলোকে এক সাথে তারিখ হিসেবে ডেটা সেট এ যুক্ত করতে হবে।

D2 | fx | =RIGHT(A2,2)

	A	B	C	D
1	Date of Birth	Year	Month	Day
2	19820301	1982	03	=RIGHT(A2,2)

এক্ষেত্রে সূত্র শুরু হবে DATE দিয়ে। তারপর দিন, মাস ও বছর উল্লেখ করতে হবে সেল নম্বরের মাধ্যমে। তবে সূত্রটি তারিখ লেখার আন্তর্জাতিক ধরন অনুসরণ করা হয়েছে। যে কারণে আমাদেরও সূত্রটি ব্যবহার করতে, আগে বছর, তারপর মাস এবং সবশেষে দিন উল্লেখ করতে হবে। এভাবে–

=DATE (বছরের সেল নম্বর, মাসের সেল নম্বর, দিনের সেল নম্বর)

এই ফর্মুলা বা সূত্র অনুসারে, তারিখের জন্যে E2 সেলে =DATE (B2,C2,D2) লিখতে হবে। এই সূত্রে যে তারিখটি আসবে সেটি আবার মার্কিন যুক্তরাষ্ট্রের তারিখ লেখার ধরন অনুসরণ করে। যেখানে মাস আগে থাকবে, তারপর দিন এবং সবশেষে বছর।

আপনাকে বাংলাদেশের তারিখে সেটি বদলে নিতে হবে। E2 তে পাওয়া ডেটাটি সিলেক্ট করুন এবং তারপর ফরম্যাট মেন্যু থেকে নাম্বার অপশনে যান এবং নিচের দিকে থাকা Custom Date And Time-ক্লিক করুন। সেখানে অনেকভাবে তারিখ লেখার অপশন আছে। আপনার যেটি প্রয়োজন সেটি সিলেক্ট করুন। বাংলাদেশে প্রচলিত দিন-মাস-বছর ফরম্যাটও আছে।

FORMAT > NUMBER > CUSTOM DATE AND TIME

এবার আপনি বাংলাদেশে ব্যবহৃত তারিখ লেখার ধরন ডেটাসেটে পেয়ে গেছেন বলে আশা করা যায়।

E2 | fx | =DATE(B2,C2,D2)

	A	B	C	D	E
1	Date of Birth	Year	Month	Day	New Date of Birth
2	19820301	1982	03	01	=DATE(B2,C2,D2)

নিচের বাকি ডেটার ক্ষেত্রে একইসূত্র অনুসরণ করে নতুন ডেটা তৈরি করুন। স্বয়ংক্রিয়ভাবে এই কাজটি করতে B2, C2, D2 and E2 সিলেক্ট করে Copy করুন এবং নিচের সেলগুলোতে Paste (Formula Only) করুন। এই কাজটি আরও সহজে এবং দ্রুত করতে B2, C2, D2 এবং E2 সিলেক্ট করে E2 এর নিচে কার্সর আনুন এবং নিচের দিকে টান দিন। একই সূত্র অবলম্বন করে সবগুলো ডেটা পরিবর্তন হয়ে যাবে।

	A	B	C	D	E
1	Date of Birth	Year	Month	Day	New Date of Birth
2	19820301	1982	03	01	01-03-1982
3	19820519	1982	05	19	19-05-1982
4	19800106	1980	01	06	06-01-1980
5	19880103	1988	01	03	03-01-1988
6	19870101	1987	01	01	01-01-1987
7	19850714	1985	07	14	14-07-1985
8	19850512	1985	05	12	12-05-1985
9	19880307	1988	03	07	07-03-1988
10	19820322	1982	03	22	22-03-1982

অনুশীলন ৩ : স্প্লিট ২

ক্লিনিং অনুশীলনের এই পর্যায়ে আমরা টেক্সট পৃথককরণের কাজ করবো। এক্ষেত্রে ডেটাসেট থেকে জেলা ও বিভাগকে আলাদা করা হবে। ফলে পরবর্তী সময়ে আমরা জেলাগুলোকে বিভাগওয়ারি ভাগও করতে পারবো। ডেটা থেকে বিভাগকে আলাদা করে একটি কলামে রাখবো, আর জেলাকে অন্য আরেকটি কলামে।

“Split Text” নামের শিটটি ব্যবহারের জন্যে কিউআর কোডটি স্ক্যান করুন।

অনুশীলন শুরু করতে গুগল শিটসে থাকা পরবর্তী শিট বা ট্যাবটি-তে যান। শিটটির নাম দেয়া হয়েছে ‘Split Text’ ।

	A
1	DV, District
2	BRL, Pirojpur
3	BRL, Bhola
4	CTG, Candpur
5	CTG, Comilla
6	CTG, Feni
7	CTG, Bandarban
8	CTG, Noakhali
9	DHK, Munshiginj
10	DHK, Gazipur
11	DHK, Foridpur
12	DHK, Madaripur
13	DHK, Manikganj

প্রথমেই সহজ কাজটি দিয়ে শুরু করা যাক। ডেটাসেটে দেখা যাচ্ছে দেশের বিভাগগুলোকে ইংরেজি তিনটি অক্ষর দিয়ে চিহ্নিত করা হয়েছে। যা টেক্সটের শুরুতে বাম দিকে আছে। তাই বিভাগকেই প্রথমে পৃথক করার কাজটি শুরু করুন। এ জন্যে বিভাগের জন্যে নতুন একটি কলাম নিন।

এবার সূত্র বা ফর্মুলাতে নজর দেয়া যাক। যেহেতু বামদিকের টেক্সট নিয়ে কাজ শুরু করতে হবে, তাই সূত্র শুরু হবে LEFT দিয়ে। টেক্সট যেহেতু A2 সেলে আছে আর প্রথম তিনটি অক্ষর বিভাগকে নির্দেশ করে, তাই সূত্রটি দাঁড়াচ্ছে এমন–

=LEFT(A2,3)

B2 fx =Lower(A2)

	A	B
1	NAME	Lower
2	Tanver emon	=Lower(A2)
3	rashed Ahmed	
4	Polash Roy	

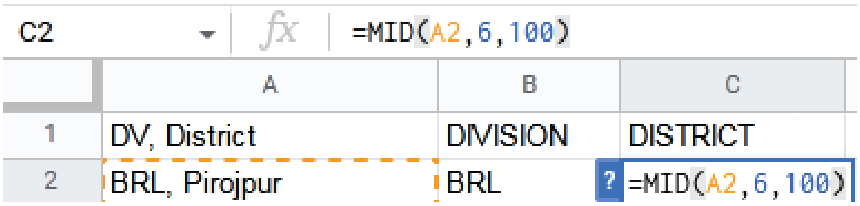
এবার জেলা। মনে হতে পারে এক্ষেত্রে ডানদিক থেকে হিসাব হবে, অর্থাৎ RIGHT সূত্র ব্যবহার হবে। কিন্তু সমস্যা হচ্ছে, ডান পাশের টেক্সট কোনো প্যাটার্ন নির্দেশ করে না। এক একটি টেক্সট বা জেলার নামের অক্ষর এক এক রকম। ফলে গুগল শিট এটিকে কোনো সূত্রে ফেলতে পারে না।

	A	B	C
1	DV, District	DIVISION	DISTRICT
2	BRL, Pirojpur	BRL	Pirojpur
3	BRL, Bhola	BRL	Bhola
4	CTG, Candpur	CTG	Candpur
5	CTG, Comilla	CTG	Comilla
6	CTG, Feni	CTG	Feni
7	CTG, Bandarban	CTG	Bandarban
8	CTG, Noakhali	CTG	Noakhali
9	DHK, Munshiginj	DHK	Munshiginj
10	DHK, Gazipur	DHK	Gazipur
11	DHK, Foridpur	DHK	Foridpur
12	DHK, Madaripur	DHK	Madaripur
13	DHK, Manikganj	DHK	Manikganj
14	DHK, Jamalpur	DHK	Jamalpur
15	KHL, Bagerhat	KHL	Bagerhat
16	KHL, Kushtia	KHL	Kushtia
17	MNS, Sherpur	MNS	Sherpur
18	MNS, Netrokona	MNS	Netrokona

আর এ কারণে এখানে MID সূত্র ব্যবহার করা হবে। তাহলে প্রশ্ন থেকে যায় MID সূত্রের ক্ষেত্রেও অক্ষরের শেষসীমা কি হবে। উত্তর যেহেতু এখানে MID এর পর আর কোনো ডেটা নাই, সেক্ষেত্রে ১০০ অক্ষর পর্যন্ত অক্ষর হিসেবে নিতে পারবে গুগল শিটস।

কলাম C তে, জেলার নাম বা ডেটা বসানোর জন্যে সূত্র C2 তে সূত্র লিখুন। MID ফরমুলা ব্যবহার হবে A2 এর ডেটা সোর্স থেকে। ডেটায় জেলার নাম শুরু হয়েছে ৬ নম্বর অক্ষর দিয়ে তাই ৬। আর অক্ষর হিসেবে নেয়া শেষ হবে ১০০ অক্ষরের ঘরে গিয়ে। তাই সূত্রটি দাঁড়াচ্ছে এমন–

=MID(A2,6,100)



আশা করা যায়, C2 তে আপনি ক্লিন ডেটা পেয়ে গেছেন।

অনুশীলন ৪ : স্প্লিট ৩

ডেটা ক্লিনিং এর এ পর্যায়ে অপেক্ষাকৃত জটিল বিষয় সমাধান করার চেষ্টা করা হবে। এবার অপরিষ্কার ডেটা থেকে ব্যক্তির নাম পৃথক করতে হবে। একজন ব্যক্তির নাম দুই ভাগে ভাগ করার চেষ্টা করবেন, নামের প্রথম ও দ্বিতীয় অংশ। অনুশীলনের জন্য একই স্প্রেডশিটের Split Text 1 ট্যাবে যান।

"Split Text" নামের শিটটি ব্যবহারের জন্যে কিউআর কোডটি স্ক্যান করুন

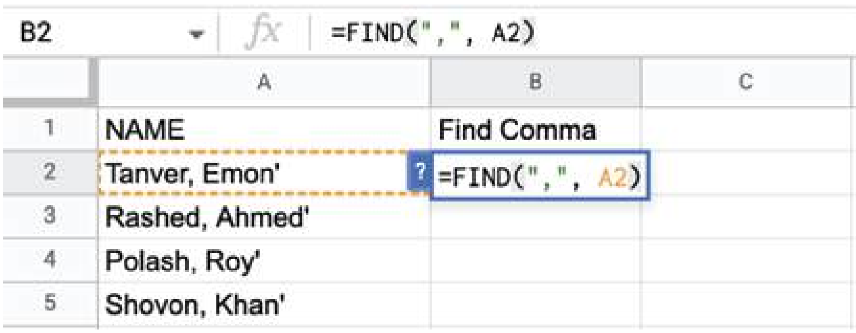
	A
1	NAME
2	Tanver, Emon'
3	Rashed, Ahmed'
4	Polash, Roy'
5	Shovon, Khan'
6	Alif, Moahammad'
7	Maruf, Chowdhury'
8	Masum, Chowdhury'
9	Hasib, Ferdows'
10	Dipu, Rahman'
11	Abdur, Rauf'
12	Mostofa, Kamal'
13	Anonna, Haq'
14	Sajib, Ahmed'
15	Tarana, Haq'
16	Rumana, Sharmin'
17	Momtaz, Begum'
18	Mokhlesur, Rahman'
19	Humayun, Khan'
20	Sahin, Khan'
21	Seikh, Karim'
22	fahim, Ahmed'

আগেই বলা হয়েছে, এই ডেটা ক্লিনিংটা একটু জটিল। কেননা এখানে আপনি কোনো প্যাটার্ন খুঁজে পাবেন না। ডাক নাম হোক অথবা আনুষ্ঠানিক–এক একটি নাম, এক এক দৈর্ঘ্যের। এক একটি নামে, অক্ষর সংখ্যা এক এক রকম। তাহলে কোন প্যাটার্ন বা কোন কাঠামো ধরে ডেটাকে ক্লিন করবে গুগল শিটস? যদি ভাল করে পুরো ডেটা সেটটি দেখেন, তবে সবগুলো ডেটায় একটি জিনিস কমন পাবেন। আর সেটি যতিচিহ্ন কমা (,)। মাঝের ওই কমা (,) ডেটা সেলে নামের প্রথম এবং দ্বিতীয় অংশকে আলাদা করেছে। গুগল সিটসের সূত্রে এই কমা ব্যবহার করে শব্দকে আলাদা করা হবে। তবে তার আগে ডেটার মাঝ থেকে কমার (,) অবস্থান চিহ্নিত করতে হবে। এ জন্যও গুগল শিটসের সূত্র ব্যবহার করতে হবে–

=FIND (“যে অক্ষর/যতিচিহ্ন/শব্দ খুঁজছেন”, সেল নম্বর)

কলাম B তে, টেক্সটে কমা (,) এর অবস্থান নির্ণয়ের জন্য ব্যবহার করুন। এক্ষেত্রে গুগল শিটসের সূত্র FIND ব্যবহৃত হবে। যে অক্ষর/যতিচিহ্ন/শব্দ খুঁজচ্ছেন সেটি প্রথমে কোট-আনকোট এর মধ্যে থাকবে। আর পরে থাকবে যে ডেটার মধ্যে খুঁজচ্ছেন তার সেল নম্বর। সেই হিসেবে B1 লেখা সূত্র বা ফরমূলাটি হবে এমন–

=FIND(",", A2)



গুগল শিটসের বাকি সব সূত্রের মতো এখানেও কমা (,) কে কোটেশন মার্কের মধ্যে রাখতে হবে। না হলে শিটস ওই কমাকে সূত্র বা ফর্মুলার অংশ হিসেব বিবেচনা করবে। ফলে উত্তর বা ফলাফলে অসঙ্গতি দেখা দেবে।

এবার বাকি টেক্সটের জন্যেও একইসূত্র প্রয়োগ করুন Copy ও Paste করে।

B3 | =FIND(",", A3)

	A	B
1	NAME	Find Comma
2	Tanver, Emon'	7
3	Rashed, Ahmed'	7
4	Polash, Roy'	7
5	Shovon, Khan'	7
6	Alif, Moahammad'	5
7	Maruf, Chowdhury'	6
8	Masum, Chowdhury'	6
9	Hasib, Ferdows'	6
10	Dipu, Rahman'	5
11	Abdur, Rauf'	6
12	Mostofa, Kamal'	8

কলাম B তে, নিশ্চয়ই এরইমধ্যে প্রত্যেকটি ডেটায় কমার (,) অবস্থান কোটায় সেটি চলে এসেছে। এবার পরবর্তী ধাপের কাজ। যার মাধ্যমে নামের প্রথম অংশ এবং দ্বিতীয় অংশ আলাদা করা হবে।

প্রথমে বাম দিক থেকে শুরু করবো। তাই ডাক নাম প্রথমে। কলাম C তে। এখানে গুগল শিটস যে কৌশলটি ব্যবহার করছে তা হলো, কমার আগে ও পরের টেক্সট-কে আলাদা করা হবে। এক্ষেত্রে বামদিক থেকে কমার এক অক্ষর আগ পর্যন্ত নির্ণয় করা হবে। সূত্র হবে B2-1। অর্থাৎ B2 থেকে 1 অক্ষর আগ পর্যন্ত। আর যেহেতু A2 এর টেক্সট থেকে ডেটা উদ্ধার করা হচ্ছে, তাই সূত্রের প্রথমেই থাকবে A2। ফলে কলাম C2 তে সূত্রটি হবে এমন–

=LEFT(A2,B2-1)

C2 | =LEFT(A2,B2-1)

	A	B	C	D
1	NAME	Find Comma	First Name	
2	Tanver, Emon'	7	=LEFT(A2,B2-1)	
3	Rashed, Ahmed'	7		
4	Polash, Roy'	7		
5	Shovon, Khan'	7		

নামের বাকি অংশের কলাম D তে একইপদ্ধতি অনুসরণ করতে হবে। তবে এবার যেহেতু টেক্সটের মধ্য ভাগ থেকে ডেটা নেয়া হবে তাই সূত্রও শুরু হবে MID দিয়ে। যেহেতু কমার (,) পরের টেক্সট নেয়া হবে তাই বিয়োগের বদলে এবার যোগ বা (+) বসবে। কমার (,) পরে একটি স্পেস আছে তাই দুই অক্ষর পর থেকে ডেটা নির্ণয় করা হবে। আর অক্ষরের উর্ধ্বসীমা ধরা হয়েছে ১০০। ফলে D2 সেলে সূত্রটি দাড়াচ্ছে এমন–

=MID(A2,B2+2,100)

D2 fx =MID(A2,B2+2,100)

	A	B	C	D	E
1	NAME	Find Comma	First Name	Last Name	
2	Tanver, Emon'	7	Tanver	=MID(A2,B2+2,100)	
3	Rashed, Ahmed'	7			
4	Polash, Roy'	7			
5	Shovon, Khan'	7			

এবার বাকি সেলগুলোর জন্যেও একইসূত্র প্রয়োগ করুন Copy ও Paste করে।

	A	B	C	D
1	NAME	Find Comma	First Name	Last Name
2	Tanver, Emon	7	Tanver	Emon
3	Rashed, Ahmed	7	Rashed	Ahmed
4	Polash, Roy	7	Polash	Roy
5	Shovon, Khan	7	Shovon	Khan
6	Alif, Moahammad	5	Alif	Moahammad
7	Maruf, Chowdhury	6	Maruf	Chowdhury

অনুশীলন ৫ : সংযুক্ত

টেক্সটকে শুধু আলাদা নয়, অনেক থেকে দুইটি আলাদা শব্দকে স্প্রেডশিটে যুক্ত করার প্রয়োজন হতে পারে। গুগলশিটে সেটাও খুব সহজে সূত্রের মাধ্যমে করা সম্ভব। এক্ষেত্রে সূত্রটির নাম “Concatenate”।

Concatenate নামের শিটটি ব্যবহারের জন্য কিউআর কোডটি স্ক্যান করুন

এই অনুশীলনের জন্যে গুগল শিটের পরবর্তী শিটটিতে যান। “Concatenate” নামের নতুন শিটটিতে আপনি দেখতে পাবেন কলাম A এবং B তে, একটি নামের প্রথম ও দ্বিতীয় অংশ আলাদা করা আছে।

	A	B	C
1	First Name	Last Name	
2	Tanver	Emon	
3	Rashed	Ahmed	
4	Polash	Roy	
5	Shovon	Khan	
6	Alif	Moahammad	
7	Maruf	Chowdhury	

এই দুই নামকে একসঙ্গে যুক্ত করতে কলাম C নির্ধারণ করুন এবং নতুন একটি Header দিন। এবার সূত্র লেখার পালা। দুইটি শব্দকে এক করার এই সূত্রে একটি Space বা বিরতি চিহ্ন দিতে হবে। কারণ দুইটি আলাদা নাম বা শব্দ যদি পাশাপাশি বসানো বা যুক্ত করা হয়, তবে তাদের মাঝে একটি বিরতি চিহ্ন বা Space দিতে হয়। আর সে জন্যেই C2 তে পূর্ণ নামের জন্যে সূত্রটি হবে এমন–

```
=CONCATENATE(A2, " ",B2 )
```

C2 =CONCATENATE(A2, " ",B2)

	A	B	C	D
1	First Name	Last Name	Full Name	
2	Tanver	Emon	=CONCATENATE(A2, " ",B2)	
3	Rashed	Ahmed		
4	Polash	Roy		
5	Shovon	Khan		
6	Alif	Moahammad		

CONCATENATE এর এই সূত্র ব্যবহার করে আপনি যে কোনো শব্দ যুক্ত করতে পারবেন। সংখ্যাকে পাশাপাশি বসাতে পারবেন এবং উদ্ধৃতি চিহ্নের মাঝে স্পেস দিয়ে দুইটি শব্দকে আলাদাও করতে পারবেন।

1	First Name	Last Name	Full Name
2	Tanver	Emon	Tanver Emon
3	Rashed	Ahmed	Rashed Ahmed
4	Polash	Roy	Polash Roy
5	Shovon	Khan	Shovon Khan
6	Alif	Moahammad	Alif Moahammad

অনুশীলন ৬ : টেক্সট ফরম্যাটিং

Formatting নামের শিটটি ব্যবহারের জন্য কিউআর কোডটি স্ক্যান করুন।	

সোর্সদের কাছ থেকে সাংবাদিকেরা বিভিন্ন ফরম্যাটে ডেটা বা তথ্য পায়। সংখ্যাভিত্তিক ডেটার বাইরে, টেক্সট ডেটার পরিমাণও নেহাত কম হয় না, একটি অনুসন্ধানী ডেটা প্রতিবেদনে। এসব টেক্সট ডেটা সব এক ফরম্যাটে থাকে না। কোনটি Uppercase অর্থাৎ বড় হাতের অক্ষরে থাকে, তো কোনটি Lowercase বা ছোট হাতের অক্ষর। যেগুলোকে একটি নির্দিষ্ট ফরম্যাট বা মানদণ্ডে নিয়ে আসা প্রয়োজন হয়। গুগল শিটসের স্ট্রিং ফাংশন্সের মাধ্যমে এটি সহজেই করা যায়। এমন একটি অনুশীলনের জন্যে আপনার স্প্রেডশিটের Text Formatting ট্যাবে যান।

	A	B	C	D
1	Full name	lower	UPPER	Proper
2	JOHN GAIL HENRY NEELY			
3	ANJUM SHARIFF			
4	EDMOND BECHIR CABBABE			
5	JERRY DALE KENNETT			
6	ANTHONY SHIHIN SHEN			
7	LENT CLIFTON JOHNSON			

যদি আপনি সোর্সের কাছ থেকে এমন একটি ডেটাসেট পান যেখানে Uppercase বা বড় হাতের অক্ষর এবং Lowercase বা ছোট হাতের অক্ষর দুটোই বিচ্ছিন্নভাবে রয়েছে এবং আপনি সবগুলোকে Lowercase বা ছোট হাতের অক্ষরে রূপান্তর করতে চান, সেক্ষেত্রে LOWER সূত্র ব্যবহার করতে হবে। সাথে যে ডেটাটি পরিবর্তন করবেন সেটি কোটেশন মার্ক ( ) বা ব্রাকেটের মধ্যে লিখতে হবে–

```
=LOWER(cell)
```

সেই হিসেবে B2 সেলে Lowercase এর ডেটা আনতে আপনাকে যা লিখতে সেটি এমন–

```
=LOWER(A2)
```

B2 | =Lower(A2)

	A	B
1	NAME	Lower
2	Tanver emon	=Lower(A2)
3	rashed Ahmed	
4	Polash Roy	

যদি আপনি সব ডেটা Uppercase এ পরিবর্তন করতে চান, তবে এক্ষেত্রেও একই সূত্র শুধু Lower এর বদলে Upper লিখতে হবে। অর্থাৎ =UPPER(cell)। সেই হিসেবে C2 সেলে সূত্রটা হবে এমন–

=UPPER(A2)

C2 | =UPPER(A2)

	A	B	C
1	NAME	Lower	Upper
2	Tanver emon	tanver emon	=UPPER(A2)
3	rashed Ahmed		
4	Polash Roy		

যদি আপনি সব ডেটা Uppercase এ পরিবর্তন করতে চান, তবে এক্ষেত্রেও একই সূত্র শুধু Lower এর বদলে Upper লিখতে হবে। অর্থাৎ =UPPER(cell)। সেই হিসেবে C2 সেলে সূত্রটা হবে এমন টেক্সট ফরম্যাটে এই দুই পরিবর্তনের বাইরেও আপনি আরেকটি পরিবর্তন করতে পারবেন, যেখানে শব্দের প্রথম অক্ষর Uppercase হবে আর বাকি অক্ষর গুলো Lower। গুগল শিটস এই ধরনকে আদর্শ বা Proper Case হিসেবে চিহ্নিত করছে। তাই সূত্রটা লেখা হয়েছে Proper দিয়ে সাথে মূল ডেটার সেল নম্বর। অর্থাৎ =PROPER(cell)। সেই হিসেবে D2 সেলে সূত্রটা হবে এমন–

=PROPER(A2)

D2 | =PROPER(A2)

	A	B	C	D
1	NAME	Lower	Upper	Proper
2	Tanver emon	tanver emon	TANVER EMON	=PROPER(A2)
3	rashed Ahmed			
4	Polash Roy			
5	shovon khan			

এবার বাকি সেলগুলোর জন্যেও একইসূত্র প্রয়োগ করুন Copy ও Paste করে।

	A	B	C	D
1	NAME	Lower	Upper	Proper
2	Tanver emon	tanver emon	TANVER EMON	Tanver Emon
3	rashed Ahmed	rashed ahmed	RASHED AHMED	Rashed Ahmed
4	Polash Roy	polash roy	POLASH ROY	Polash Roy
5	shovon khan	shovon khan	SHOVON KHAN	Shovon Khan
6	Alif Moahammad	alif moahammad	ALIF MOAHAMMAD	Alif Moahammad
7	Maruf CHowdhury	maruf chowdhury	MARUF CHOWDHURY	Maruf Chowdhury

অনুশীলন চালিয়ে যান। আপনি যত চর্চা করবেন তত বেশি ডেটার ভুলগুলো আপনার চোখে ধরা পড়বে এবং আপনি হয়তো সেগুলো ক্লিন বা পরিষ্কার করার নতুন নতুন পথও বের করতে পারবেন। টুলগুলো জানা থাকলে ভুল ডেটার ফাঁদ এড়ানো সহজ হয়।

আপনি যদি ডেটা ক্লিনিং নিয়ে আরও জানতে চান, আহলে কিউআর কোডটি স্ক্যান করে জিআইজেএনের এই টিপশিটটি অনুসরণ করতে পারেন।

010101010101010101010101010101010101010101010101010

010101010101010101010101010101010101010101010101010

অধ্যায় ০৭

বিশ্লেষণ

বিশ্লেষণ

> **ডেটা সাংবাদিকতায় একজন প্রতিবেদকের সবচেয়ে বিশ্বস্ত বন্ধু হচ্ছে স্প্রেডশিট। স্প্রেডশিটে থাকা সূত্রের সহায়তায় শুধু ডেটার নানামুখী নতুন বিশ্লেষণই শুধু মেলে না, গতিশীল হয় প্রতিবেদন তৈরির প্রক্রিয়া, সহজ হয় প্রতিবেদকের কাজ। তাই সকল সাংবাদিকেরই স্প্রেডশিট ব্যবহারের প্রাথমিক কিছু জ্ঞান থাকা প্রয়োজন।**

একইসঙ্গে প্রতিটি সংবাদকক্ষে অন্তত একজন সাংবাদিক থাকা দরকার, যার কম্পিউটারনির্ভর এই সাংবাদিকতায় বিশেষ দক্ষতা রয়েছে। কেননা, এই কৌশল আপনার প্রতিবেদনকে নিঃসন্দেহে ভিন্ন মাত্রায় পৌঁছে দিতে সক্ষম।

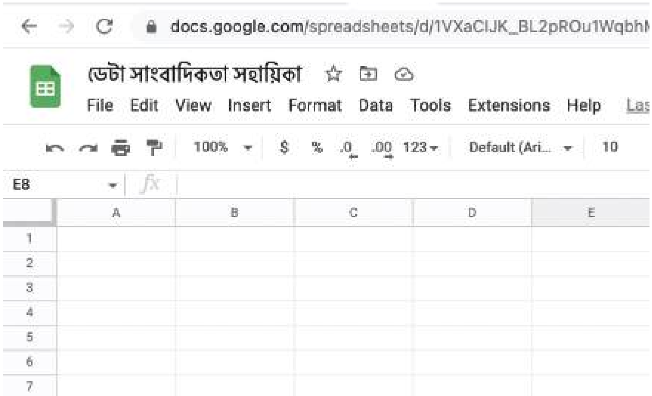
> বিশ্বে বহুল ব্যবহৃত স্প্রেডশিটটি হচ্ছে মাইক্রোসফটের এক্সেল। মূলত এক্সেলের মাধ্যমেই স্প্রেডশিটের ব্যবহার বেড়েছে বিশ্বজুড়ে। এর বাইরে আছে গুগল শিটস বা অ্যাপেলের নাম্বারস। তবে অনলাইননির্ভর হবার কারণে গুগল শিটসই এখন বহুল ব্যবহৃত। জিমেইল অ্যাকাউন্টের আওতায় গুগল শিটস বিনামূল্যে ব্যবহার করা যায়। এই স্প্রেডশিটটিতে সয়ংক্রিয়ভাবে গাণিতিক বিশ্লেষণের সুবিধা বাকিদের তুলনায় কিছুটা বেশি এবং সহজও। তবে মাইক্রোসফট এক্সেল এবং গুগল শিটস এই দুটোর প্রোগ্রামিং, সূত্র বা ফর্মুলা এবং ব্যবহারপদ্ধতি প্রায় পুরোটাই এক।

> **এই অনুশীলনে গুগল শিটস ব্যবহার করা হবে। তবে একই কৌশল মাইক্রোসফট এক্সেলের ক্ষেত্রেও প্রয়োগ করা যাবে। আর ডেটার ক্ষেত্রে বাংলাদেশ-সংশ্লিষ্ট ইংরেজি ডেটা ব্যবহার করা হবে। কেননা আগেই বলেছি, স্প্রেডশিটে বাংলায় ডেটা বিশ্লেষণে ভুল হবার আশঙ্কা থাকে এবং বাংলায় খুব বেশি ডেটাও পাওয়া যায় না।**

গুগল শিটসকে জানুন

গুগল শিটস ব্যবহার করতে আপনার জিমেইল অ্যাকাউন্টের মাধ্যমে গুগল ড্রাইভে ঢুকুন। এরপর New অপশন থেকে Google Sheets এ যান এবং নতুন Spreadsheet খুলুন।

Google Drive >New>Google Sheets>Blank spreadsheet.

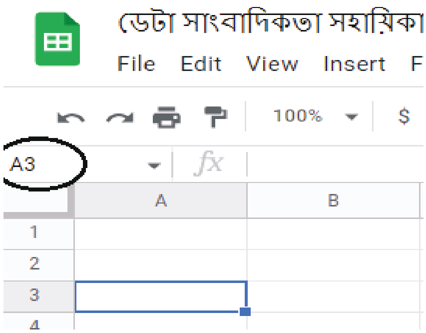


আপনার স্প্রেডশিটটি দেখতে এমন হবে। প্রথমেই Untilteld বদলে ফাইলের নতুন নাম দিন। আমরা দিয়েছি ডেটা সাংবাদিকতা সহায়িকা।

স্প্রেডশিটে থাকা A, B, C, D সারিগুলোকে বলা হয় কলাম (Column)। আর আড়াআড়ি ভাবে থাকা 1, 2, 3, 4 সংখ্যার সারিকে বলা হয় রো (Row)।

এই কলাম এবং রো–এর মাধ্যমে স্প্রেডশিটে যে চারকোণ ঘর তৈরি হয়, সেটিকে বলা হয় সেল। প্রতিটি সেলের একটি করে পৃথক চিহ্নিতকরণ নম্বর রয়েছে। যা কলাম এবং রো এর নম্বরের মিশেলে তৈরি হয়।

এখানে যেমন A3 সেলটি সিলেক্ট করা দেখানো হচ্ছে। সেলটি কলাম A এর 3 নম্বর রো তে অবস্থান করছে। এ কারণে সেল নম্বর A3।

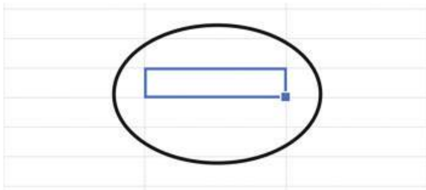


গুগল শিটসে মাউস ব্যবহার করে শটকার্ট

গুগল শিটস পরিচালনায় মাউসের কার্সরের কার্যকর ব্যবহার রয়েছে। শিটসের বেশকিছু সেবা দ্রুত পাওয়া যায় শুধু মাউস ব্যবহার করে।

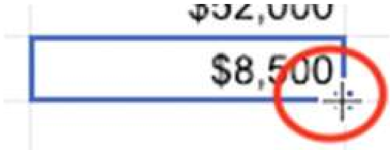
সেল নির্বাচন

সেল নির্বাচন সিলেক্ট করতে কার্সর নিয়ে ক্লিক করুন নির্দিষ্ট সেলের উপরে।



ডেটা কপি

সেলে থাকা ডেটা বা সূত্র কপি করে পাশের বা নিচের সেলে নিতে, মাউসের কার্সরকে সেলের ডানপাশের নিচের কোণ ধরে টান দিন। মূল সেলে সূত্র বা ফর্মুলা থাকলে, সে অনুযায়ী পরের সেলগুলোর তথ্যেও পরিবর্তন আসবে। আর যদি সূত্রের বদলে শুধু তথ্য থাকে, তবে বাকি সেলগুলোতেও ওই একইতথ্য COPY-PASTE হয়ে যাবে।



পুরো কলাম/ সেল নির্বাচন

স্প্রেডশিটের পুরো কলাম বা সেলে থাকা তথ্যও একইভাবে সিলেক্ট বা নির্বাচন করা যায়, কার্সরের সাহায্যে। কার্সর কলাম বা সেলের উপরে নিয়ে গেলে কার্সরটি দৃশ্যমান হবে, তখন সিলেক্ট করা যাবে। আবার গ্রুপভিত্তিক তথ্য বা স্প্রেডশিটে থাকা সব তথ্যও মাউসের মাধ্যমে একবারে সিলেক্ট বা নির্বাচন করা যায়।

	A	B	C	D	E	F
1	FIRST NAME	LAST NAME	DISTRICT	DIVISION	OCCUPATION	INCOME
2	Tanver	Emon	Sirajgonj	rajshahi	Business	80000
3	Rashed	Ahmed	Munshiginj	Dhaka	Teaching	60000
4	Polash	Roy	Rongpur	Rongpur	Government Employee	30000
5	Shovon	Khan	Candpur	Chattogram	Business	45000
6	Alif	Moahammad	Nilphamari	Rongpur	Government Employee	12000

ডেটা

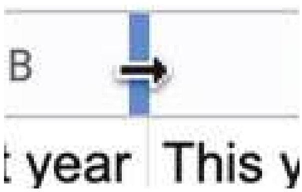
সেলে ডেটা বা তথ্য যুক্ত করতে, নির্দিষ্ট সেল নির্বাচন বা সিলেক্ট করুন এবং টাইপ করতে শুরু করুন। লেখা শেষ হলে, ENTER চেপে তথ্য যুক্ত করার প্রক্রিয়া শেষ করুন। এ ক্ষেত্রে তথ্য বা ডেটার দৈর্ঘ্যের কোনো সীমা নেই। টেক্সটের আকার দীর্ঘ হলেও সেটি সেলে থাকবে। সেলে না দেখা গেলে, ফর্মুলা বারে দেখা যাবে।

E	F
OCCUPATION	INCOME
Business	80000

সংখ্যাও একইপদ্ধতিতে সেলে যুক্ত করবেন। তবে এক্ষেত্রে সংখ্যার মাঝে কোনো কমা (,) বা মুদ্রার চিহ্ন ব্যবহার করবেন না। এগুলো পরে ফরম্যাট পরিবর্তন করে বসাতে পারবেন। তবে সংখ্যার পরিমাণ বেশি হলে এবং তা যদি কলামের বাইরে চলে যায়, তবে কলামের আকার বাড়িয়ে নিন। না হলে হিসাবে কোনো ভুল হয়তো হবে না, কিন্তু পরবর্তী সময়ে আপনি বিভ্রান্ত হতে পারেন।

B	C	D
LAST NAME	DISTRI	DIVISION
Emon	Sirajgo	rajshahi
Ahmed	Munshi	Dhaka

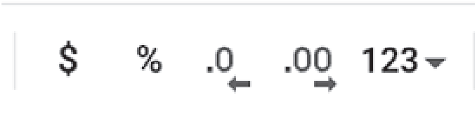
টেক্সট বা সংখ্যার আকার অনুযায়ী সেলের আকার করতে স্প্রেডশিটের উপরের সারিতে বাম কোণে অর্থাৎ A এবং 1 এর মাঝে ক্লিক করুন এবং এরপর A ও B কলামের সংযোগ বারে ক্লিক করুন। প্রতিটি কলাম ততোটুকুতে পরিণত হবে, সেলে থাকা ডেটা ধারণ করতে যতটুকু প্রয়োজন।



সংখ্যা

স্প্রেডশিটে বিভিন্ন ধরনের সংখ্যার ব্যবহার আছে। কখনও সেটি অর্থের পরিমাণ বোঝাতে ব্যবহৃত হয়, কখনও বা কোনো কিছুর তুলনা করতে। ভিন্ন ভিন্ন ব্যবহারে সংখ্যার উপস্থাপনও ভিন্ন হয় স্প্রেডশিটে।

স্প্রেডশিটে ব্যবহৃত সংখ্যার ক্ষেত্রে তিনটি ফরম্যাট সবচেয়ে বেশি ব্যবহৃত হয়। সেগুলো হল–কমা (,), মুদ্রা চিহ্ন (ডলার বা টাকার চিহ্ন) এবং শতাংশ (%)। সংখ্যার উপর এই ফরম্যাটগুলো বসাতে আপনাকে নির্দিষ্ট কলাম বা সেল সিলেক্ট করতে হবে এবং তারপর ফরম্যাট মেন্যু বার থেকে নম্বরে যেতে হবে। নম্বরের মধ্যে একাধিক অপশন পাবেন, আপনার প্রয়োজনীয়টি সিলেক্ট করুন।



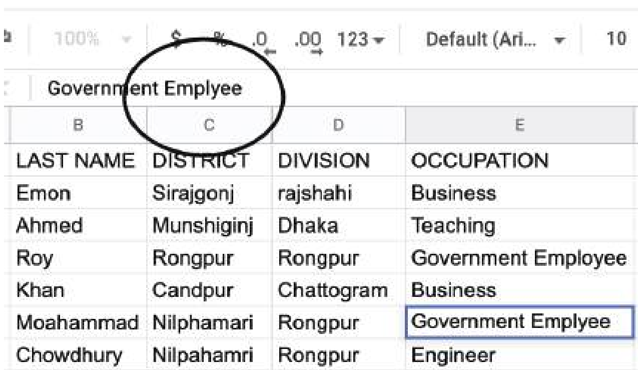
একই অপশন স্প্রেডশিটের ফর্মুলাবারের ঠিক উপরে গুগল শিটের টুলস বারে, শটকার্ট হিসেবে রয়েছে। সেখান থেকেও অপশনগুলো ব্যবহার করতে পারবেন।

ত্রুটি সংশোধন

স্প্রেডশিটের সেলে ডেটা যুক্ত করার প্রক্রিয়ায় অর্থাৎ তথ্য টাইপ করার সময়ে যদি কোনো ভুল হয়ে থাকে এবং আপনি মনে করেন যে, পুরো ডেটাটি আবার নতুন করে লিখবেন, তাহলে কি-বোর্ডের উপরের সারির বাম কোনে Esc বাটনটি চাপুন। এর মাধ্যমে আপনি ওই সেলে এতোক্ষণ যা লিখেছেন, তা মুছে যাবে, অন্য কোনো পরিবর্তন ছাড়াই।

তবে ডেটা থেকে আংশিক কিছু মুছতে Backspace বা Delete বাটন ব্যবহার করুন।

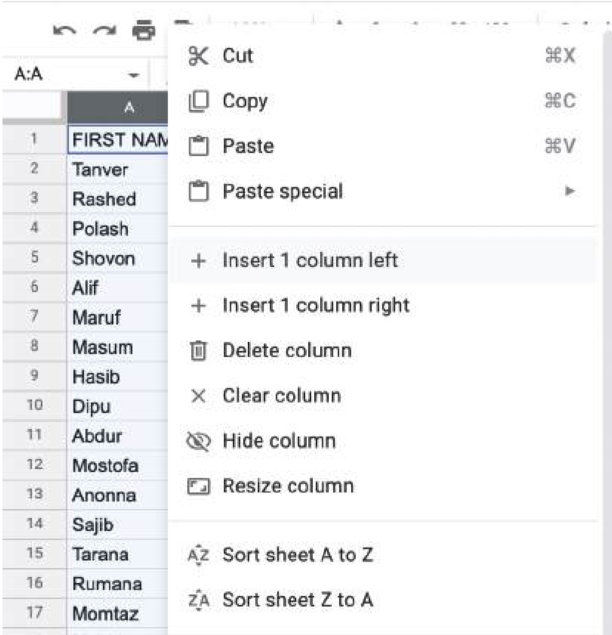
আর যদি কিছু সংশোধন করে নতুন তথ্য যুক্ত করতে হয়, তবে ওই সেলটি সিলেক্ট করুন এবং কলামের উপরে থাকা ফর্মুলা বারে ক্লিক করে, সেখানে সংযোজন বা বিয়োজন করুন।



B	C	D	E
LAST NAME	DISTRICT	DIVISION	OCCUPATION
Emon	Sirajgonj	rajshahi	Business
Ahmed	Munshiginj	Dhaka	Teaching
Roy	Rongpur	Rongpur	Government Employee
Khan	Candpur	Chattogram	Business
Moahammad	Nilphamari	Rongpur	Government Emplyee
Chowdhury	Nilpahamri	Rongpur	Engineer

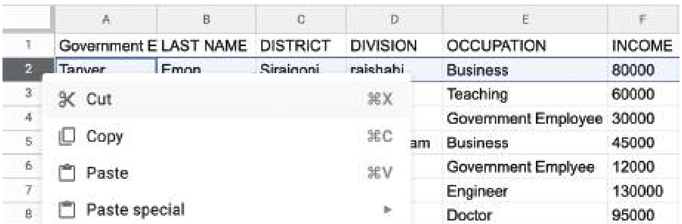
কলাম ও রো

গুগল শিটসে ডেটার বিশ্লেষণ করার সময় হটাৎ করে নতুন কিছু তথ্য আপনার কাছে চলে আসতে পারে, যা শিটে যুক্ত করা প্রয়োজন। একইসঙ্গে সেই তথ্য বিদ্যমান কোনো ডেটা কলামের পাশে বসানো দরকার, সেক্ষেত্রে নির্দিষ্ট কলামের ডানে এবং বামে দুই পাশেই নতুন একটি খালি কলাম যুক্ত করার সুযোগ আসে গুগল শিটসে। এজন্য যে কলামের ডানে বা বামে নতুন কলাম যুক্ত করবেন সেই কলামের উপরে কার্সর নিয়ে যান এবং ডাউন অ্যারো চিহ্নিত পয়েন্টে ক্লিক করুন। নতুন কলাম যুক্ত করার অপশন চলে আসবে।



এই মেন্যুবারেই পুরো কলাম ডিলিট করা কিংবা কলামের মধ্যে থাকা ডেটা মুছে ফেলার অপশন আছে। প্রয়োজনে সেগুলো ব্যবহার করতে পারেন।

রো-এর ক্ষেত্রেও নতুন রো সংযোজন বা বিয়োজনের ক্ষেত্রেও একইপদ্ধতি অনুসরণ করতে হবে। শুধু কলামের বদলে নির্দিষ্ট রো-বারে ক্লিক করতে হবে।



স্প্রেডশিটের সূত্র

ব্যবহারকারীদের কাছে এটি জাদু হিসেবেই পরিচিত, এক্সেল ম্যাজিক এবং শিটস ম্যাজিক। আর যে কারণে এই ম্যাজিক বা জাদু ঘটে সেটা হচ্ছে সূত্র বা ফর্মুলা। এক্সেল এবং শিটসের গাণিতিক সূত্র। যার ওপর ভর করে এই দুই স্প্রেডশিটের কৌশল পরিচালিত হয়। পৃথক স্প্রেডশিট হলেও, দুইটির প্রায় সব সূত্রই একইরকম। অনুশীলনের এই পর্যায়ে চলুন পরিচিত হই সেই সব জাদুকরি সূত্রের সঙ্গে, যা ব্যবহার করে স্প্রেডশিটে থাকা হাজার হাজার তথ্য মুহূর্তে বিশ্লেষণ করা সম্ভব হয়।

শুরু করা যাক একদম প্রাথমিক পর্যায় থেকে

গুগল শিটসে সব সূত্রের শুরু হয় equal (=) বা সমান সাইন দিয়ে। আপনি যে গাণিতিক হিসাব করুন না কেনো সূত্র শুরু হবে equal (=) সাইন দিয়ে। এর কারণ হচ্ছে, equal (=) লেখার মধ্য দিয়ে গুগল শিট কে বার্তা দেয়া হয়, যা লেখা হচ্ছে সেটি সূত্র বা ফর্মুলা, ডেটা নয়।

equal (=) সাইনের পর সূত্র লিখতে হবে

গুগল শিটের সবচেয়ে সহজ সূত্র বা সরল গাণিতিক হিসাবগুলো হচ্ছে, যোগ-বিয়োগ-গুণ-ভাগ। চারটি পৃথক সাইনের মাধ্যমে এইগুলো কাজ করে গুগল শিটে।

+ (যোগ)	/ (ভাগ)
- (বিয়োগ)	* (গুণ)

এর বাইরে সাংবাদিকতায় বহুল ব্যবহৃত আর কয়েকটি সূত্র হচ্ছে পার্থক্য, শতাংশের হিসাবে পার্থক্য এবং মোটের ওপর পার্থক্য। এই তিনটি সূত্রই অনুশীলন করা হবে এই আলোচনায়।

অনুশীলনের জন্য ডেটা সাংবাদিকতা সহায়িকার গুগল ড্রাইভে থাকা এই গুগল শিটটি খুলুন এবং কপি করে আপনার নিজের গুগল ড্রাইভে ফাইলটি যুক্ত করুন। কপি করা ফাইলে অনুশীলন শুরু করুন।

ডেটা বিশ্লেষণ অনুশীলনের জন্য ডেটা সাংবাদিকতা সহায়িকার গুগল ড্রাইভে থাকা এই স্প্রেডশিটটি ব্যবহার করতে হবে। গুগল ড্রাইভ থেকে কপি করে শিটটি আপনার নিজের গুগল ড্রাইভে রাখুন। কিউআর কোডটি স্ক্যান করলে স্প্রেডশিটটি পাবেন।

প্রশ্ন করে অঙ্ক করুন

ডেটা সাংবাদিকতা শুধু এক্সেল বা গুগল শিটের সূত্রের বিষয় নয়। আপনি হয়তো এই স্প্রেডশিটে সব সূত্র জানেন। একইসঙ্গে আপনার কাছে বিশাল একটি ডেটা সেটও আছে। এখন আপনি কতগুলো সূত্রও ব্যবহার করে সেই ডেটা বা তথ্য বিশ্লেষণ করবেন এবং কি তথ্যই বা বের করবেন। আর সেটি ঠিক করতে সবার আগে আপনাকে ডেটাকে প্রশ্ন করতে হবে। ডেটা থেকে আপনি কি ধরনের বিশ্লেষণ চান, সেটি আগে নির্ধারণ করতে হবে। ডেটাকে আপনার করা প্রশ্নই নির্ধারণ করে দেবে, কোন ক্ষেত্রে কি ধরনের সূত্র ব্যবহার করা হবে।

ধরুন আপনি একজন অপরাধবিষয়ক প্রতিবেদক। আপনি পুলিশ সদর দপ্তরে গেলেন সংস্থার মহাপরিচালকের সংবাদ সম্মেলন কভার করতে। সংবাদ সম্মেলনে আইজিপি ২০১৯ সালে বাংলাদেশ পুলিশের বিভিন্ন ইউনিটে হওয়া অপরাধ আর মামলার তথ্য তুলে ধরলেন।

কিউআর কোডটি স্ক্যান করলে
Police Statistics
স্প্রেডশিটটি পাবেন

	A	B	C
1	Unit Name	Cases 2018	Cases 2019
2	DMP	27150	2153
3	CMP	7731	678
4	KMP	2604	317
5	RMP	4247	401
6	BMP	1779	166
7	SMP	1716	143
8	Dhaka Range	400603	3089
9	Mymensingh Ra	13716	1004
10	Chattogram Ran	29925	2396
11	Sylhet Range	9162	725
12	Khulna Range	21253	1585
13	Barishal Range	11434	895
14	Rajshahi Range	25608	2011
15	Rangpur Range	22278	1472
16	Railway Range	989	84
17	GMP	872	244
18	RPMP	352	121
19	ATU	0	0
20	TOTAL	221419	17484

দেশের আইনশৃঙ্খলার পরিস্থিতি সম্পর্কে এই প্রাথমিক তথ্য থেকে আমরা নতুন কি তথ্য বের করবো। সেই প্রশ্ন আগে নির্ধারণ করবো। তারপর শুরু হবে বিশ্লেষণ।

এই ডেটাসেট দেখে প্রথম যে প্রশ্ন মাথায় আসে, তা হচ্ছে অপরাধ আর মামলার পরিমাণে বছরের ব্যবধানে কি পরিবর্তন হলো। অর্থাৎ মামলা কি আগের বছরের থেকে বাড়লো না-কি কমলো। এরপর আরও বিস্তারিত বিশ্লেষণ করতে হবে ধাপে ধাপে।

পরিবর্তন বা পার্থক্যের হিসাব

এই অনুশীলনের জন্য গুগল ড্রাইভে থাকা ডেটা সাংবাদিকতা সহায়িকা নামের স্প্রেডশিটটি ব্যবহার করা হবে। কিউআর কোডটি স্ক্যান করলে স্প্রেডশিটটি পাবেন।

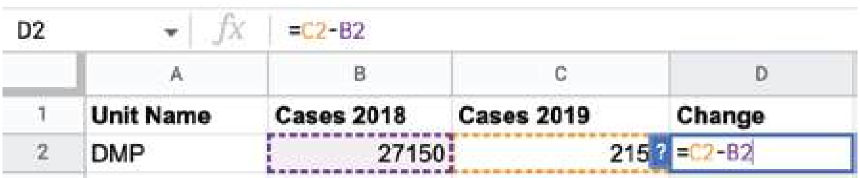
এবার কাজ শুরু করা যাক। আপনার প্রথম প্রশ্ন, অপরাধ কতটুকু বাড়ল বা কমলো? ইউনিটওয়ারি পরিবর্তন বা পার্থক্যটা কেমন? এই হিসাব বের করার সহজ গাণিতিক উপায়, নতুন বছরের মামলার থেকে আগের বছরের মামলা বিয়োগ দেয়া। গুগল শিটেও তাই করা হবে। এ জন্য বছরভেদে মামলার পার্থক্যের ডেটার জন্যে কলাম D নির্বাচন করুন এবং D1 সেলে কলামের নাম দিন। এক্ষেত্রে কলামের নাম হতে পারে 'Change'।

এবার সূত্র লিখুন। যার শুরু হবে Equal সাইন দিয়ে। এরপর নতুন বছরের বরাদ্দ বিয়োগ আগের বছরের বরাদ্দ। অর্থাৎ তা হবে এমন–

= new - old

আর D2 সেলের ক্ষেত্রে সূত্র হবে

=C2-B2



D2 | fx | =C2-B2

	A	B	C	D
1	Unit Name	Cases 2018	Cases 2019	Change
2	DMP	27150	215	=C2-B2

D2 সেলে আপনি এরইমধ্যে প্রথম ইউনিট অর্থাৎ ডিএমপি বা ঢাকা মেট্রোপলিটন পুলিশের বছরভেদে অপরাধের পার্থক্যের হিসাব পেয়ে গেছেন। যার পরিমাণ-২৪৯৯৭। এখানে মাইনাস এসেছে মানে হচ্ছে ২০১৯ এ ২০১৮ এর থেকে কম অপরাধ হয়েছে। যদি ফলাফল যোগ সংখ্যা আসতো তার মানে মামলার পরিমাণ বেড়েছে।

বাকি ইউনিটগুলোর ডেটা বের করতে কলামের বাকি সেলগুলোতে একই সূত্র অনুসরণ করুন। এজন্যে D2 সেলের নিচের ডানকোণ ধরে বাকি সেলগুলোতে টান দিয়ে নামান। এর ফলে বাকিসেল গুলোতে D2 সেলে বসানো সূত্র অনুসরণ করে নিজ নিজ ডেটা হিসাবে করবে এবং সঠিক তথ্য দেবে।

D
Change
-24997

কলাম D তে বাকি সব খাতের বছরের ব্যবধানে বরাদ্দের পার্থক্যও এরই মধ্যে চলে এসেছে।

100% $ % .0 .00 123 Default (Ari...

D10 fx =C10-B10

	A	B	C	D
1	Unit Name	Cases 2018	Cases 2019	Change
2	DMP	27150	2153	-24997
3	CMP	7731	678	-7053
4	KMP	2604	317	-2287
5	RMP	4247	401	-3846
6	BMP	1779	166	-1613
7	SMP	1716	143	-1573
8	Dhaka Range	400603	3089	-397514
9	Mymensingh Ra	13716	1004	-12712
10	Chattogram Ran	29925	2396	-27529
11	Sylhet Range	9162	725	-8437
12	Khulna Range	21253	1585	-19668
13	Barishal Range	11434	895	-10539
14	Rajshahi Range	25608	2011	-23597
15	Rangpur Range	22278	1472	-20806
16	Railway Range	989	84	-905
17	GMP	872	244	-628
18	RPMP	352	121	-231
19	ATU	0	0	0
20	TOTAL	221419	17484	-203935

পরিবর্তন বা পার্থক্যের হিসাব

এবার দেখা যাক শতকরা হিসাবে ইউনিটওয়ারি মামলায় কি পরিবর্তন আসলো। এই তথ্য হিসাব করার জন্যে E কলামটি নির্বাচন করুন এবং E1 সেলে কলামের নাম দিন "Change in Percentage"।

শতকরা হিসাবে পার্থক্য হিসাব করতে গণিতের ঐকিক নিয়ম অনুসরণ করুন। এ জন্য প্রথমেই আপনাকে নতুন বছরের থেকে আগের বছরের পার্থক্য বের করতে হবে। তারপর সেটি শতকরা হিসেবে কত, তা বের করতে হবে। শতকরা হিসাব বের করতে সংশ্লিষ্ট বছর দুইটির যে ব্যবধান, তাকে পুরোনো বছর দিয়ে ভাগ দিতে হবে।

শতকরা পরিবর্তন = (নতুন - পুরোনো)/পুরনো

সেই অনুসারে E2 সেলের ক্ষেত্রে সূত্রটা হবে এমন–

=(C2-B2)/B2

E2 | fx | =(C2-B2)/B2

	A	B	C	D	E
1	Unit Name	Cases 2018	Cases 2019	Change	Change in Percentage
2	DMP	27150	2153	-2499	=(C2-B2)/B2
3	CMP	7731	678	-7053	
4	KMP	2604	317	-2287	

যেহেতু (C2-B2) এর বিয়োগের ফলাফল D2 সেলে রয়েছে, এক্ষেত্রে আপনি চাইলে =D2/B2 সূত্রও ব্যবহার করতে পারেন।

এরপর প্রাপ্ত ফলাফলকে ১০০ দিয়ে গুণ দিন বা ফরম্যাট মেন্যুর নম্বর অপশন থেকে Percent সিলেক্ট করুন। তাহলেই পরিবর্তনের শতকরা হিসাব বের হয়ে আসবে।

Format menu > Number>Percent

E2 সেলে আপনি এইর মধ্যে প্রথম ইউনিটের বছরভেদে মামলার পার্থক্যের শতকরা হিসাব পেয়ে গেছেন। বাকি খাতগুলোর ডেটা বের করতে কলামের বাকি সেলগুলোতে একইসূত্র অনুসরণ করুন। এজন্যে E2 সেলের নিচের ডানকোন ধরে বাকি সেলগুলোতে টান দিয়ে নামান।

	A	B	C	D	E
1	Unit Name	Cases 2018	Cases 2019	Change	Change in Percentage
2	DMP	27150	2153	-24997	-92.07%
3	CMP	7731	678	-7053	-91.23%
4	KMP	2604	317	-2287	-87.83%
5	RMP	4247	401	-3846	-90.56%
6	BMP	1779	166	-1613	-90.67%
7	SMP	1716	143	-1573	-91.67%
8	Dhaka Range	400603	3089	-397514	-99.23%
9	Mymensingh Range	13716	1004	-12712	-92.68%
10	Chattogram Range	29925	2396	-27529	-91.99%
11	Sylhet Range	9162	725	-8437	-92.09%
12	Khulna Range	21253	1585	-19668	-92.54%
13	Barishal Range	11434	895	-10539	-92.17%
14	Rajshahi Range	25608	2011	-23597	-92.15%
15	Rangpur Range	22278	1472	-20806	-93.39%
16	Railway Range	989	84	-905	-91.51%
17	GMP	872	244	-628	-72.02%
18	RPMP	352	121	-231	-65.63%
19	ATU	0	0	0	

যোগ

ইউনিটওয়ারি মামলার হিসাবের পাশাপাশি প্রতিবেদনে ২০১৯ সালে সারাদেশের মোট মামলার হিসাবও উল্লেখ করা দরকার। সে কারণে অনুশীলনের এই পর্যায়ে স্প্রেডশিটে থাকা ডেটা যোগ কীভাবে করে, তা দেখা যাক। যে ডেটা শিট আপনি হাতে পেয়েছেন সেখানে এরই মধ্যে মোট মামলার একটি যোগফল আছে। তবে তারপরও সঠিক বিশ্লেষণের স্বার্থে আপনি নিজে নতুন করে আরেকবার পুরো ডেটা সেটটি যোগ করে নিন। কেননা যে প্রতিষ্ঠান ডেটাটি সরবরাহ করেছে, তাদের ভুল হতে পারে। অথবা ইচ্ছে করেই আপনাকে বিভ্রান্ত করতে পারে। তাই স্প্রেডশিটে থাকা ডেটা নতুন করে যোগ করুন। এজন্য বিদ্যমান যোগফলের নিচের আরেকটি রো-যুক্ত করুন। B21 সেল কে Total Check নাম দিন।

এরইমধ্যে জেনেছেন, গুগল শিটে থাকা ডেটা বা সংখ্যা যোগ করতে সূত্রে + (প্লাস) সাইন ব্যবহার করা হয়। এভাবে–

```
= B2 + B3 + B4 + B5… etc
```

কিন্তু যখন আপনি বড় ডেটা নিয়ে কাজ করবেন, সব সেলের নাম এক এক করে লেখা বাস্তবসম্মত নয়। আর এ কারণে স্প্রেডশিটে SUM ফর্মুলা বা সূত্র রয়েছে।

এই SUM ফর্মুলার মাধ্যমে একটি নির্দিষ্ট সীমার মধ্যে থাকা সব ডেটা কে যোগ করা যায়। আর ঐ সীমা নির্দিষ্ট করতে, প্রথম ডেটা সেল ও শেষ ডেটা সেলের মাঝে কোলন (:) চিহ্ন বসাতে। ঐ কোলন (:) চিহ্ন গুগল শিটকে ডেটা সীমার বার্তা দেয়। অর্থাৎ কলামে থাকা সব ডেটার যোগফল বের করতে সূত্র বা ফর্মুলা হবে–

=SUM (প্রথম সেল নম্বরঃ শেষ সেল নম্বর)

সেই হিসেবে এই ডেটা সেটের জন্যে B17-এ সূত্র হবে–

=SUM(B2:B19)

B21 fx =SUM(B2:B19)

	A	B
1	**Unit Name**	**Cases 2018**
2	DMP	27150
3	CMP	7731
4	KMP	2604
5	RMP	4247
6	BMP	1779
7	SMP	1716
8	Dhaka Range	400603
9	Mymensingh Range	13716
10	Chattogram Range	29925
11	Sylhet Range	9162
12	Khulna Range	21253
13	Barishal Range	11434
14	Rajshahi Range	25608
15	Rangpur Range	22278
16	Railway Range	989
17	GMP	872
18	RPMP	352
19	ATU	0
20	Total	221419
21	Total Check	=SUM(B2:B19)

কলাম C তেও আপনি একইসূত্র copy করে কিংবা, B21 সেলের নিচের ডানকোণ ধরে ডান দিকে অর্থাৎ C21 সেলে নিয়ে গেলে, কলাম C এর যোগফলও চলে আসবে।

20	Total	221419	17484
21	Total Check	581419	17480
22			

New Total এ থাকা যোগফলের সঙ্গে কলাম B অর্থাৎ আগের বছরের মামলার দেয়া তথ্যে মিল থাকলে, নতুন বছর অর্থাৎ C কলামে মোট যে হিসাব দেয়া হয়েছে সেটি ভুল ছিল। কিন্তু কেন?

ডেটায় এমন ভুল পেলে সোর্সকে জিজ্ঞেস করুন এবং সঠিক তথ্য নিয়ে প্রতিবেদনের পরিবর্তে বিশ্লেষণে এগিয়ে যান।

শতকরা হার

আপনি চাইলেই এই ডেটাসেট থেকে দেশের সবচেয়ে অবরাধ প্রবণ বা সবচেয়ে বেশি মামলা হয়েছে এমন মহানগর ও বিভাগ চিহ্নিত করতে পারবেন। এটি জানার একটি সহজ উপায় হচ্ছে ইউনিটওয়ারি মামলার সংখ্যা যাচাই করা। তবে সংখ্যার সেই হিসাব এই ধারণাকে স্পষ্ট করে না। আবার পাঠকও সহজে বুঝতে পারে না। এক্ষেত্রে শতকরা হিসাবে মামলার অনুপাত বের করা সবচেয়ে কার্যকর। এই হিসাব বের করাও বেশ সহজ।

প্রথমেই শতকরা মামলার জন্য একটি আলাদা কলাম নিয়ে নিন। এই স্প্রেডশিটে কলাম F নেয়া হয়েছে। F1 সেলে কলামের নাম "Percentage in Total Cases 2019"। যেহেতু এই কলামে শতকরা বা শতাংশের হিসাব বসবে, তাই কলামটি সিলেক্ট করে ফর্মুলা বারের Percent (%) এ ক্লিক করুন, যাতে এই কলামের হিসেবের ফলাফল শতাংশে পাওয়া যায়।

এবার সূত্রে নজর দেয়া যাক। এখানে সূত্রটি সাধারণ গাণিতিক হিসাবের। যার শতকরা হার জানতে চাইবেন, তাকে মোট দিয়ে ভাগ দিবেন। অর্থাৎ

শতাংশ = নির্দিষ্ট অংশ/মোট

সেই হিসেবে F2 সেলের জন্যে এখানে সূত্র হবে−

```
=C2/C21
```

এখানে C21 নেয়ার কারণ Total Check এ মোটের যোগফল সঠিক।

এবার F কলামের বাকি রো-গুলোতে ও শতাংশের হিসাব বের করতে একইসূত্র বা ফর্মুলা Copy করুন।

fx | =C6/C25

	B	C	D	E	F
	Cases 2018	Cases 2019	Change	Change in Percentage	Percentage in Total Cases 2019
	27150	2153	-24997	-92.07%	12.32%
	7731	678	-7053	-91.23%	#DIV/0!
	2604	317	-2287	-87.83%	#DIV/0!
	4247	401	-3846	-90.56%	#DIV/0!
	1779	166	-1613	-90.67%	#DIV/0!
	1716	143	-1573	-91.67%	#DIV/0!

সেলে Error দেখা যাচ্ছে। এর কারণ জানতে Error দেখানো সেলগুলোতে ক্লিক করুন, ফর্মুলা বারে দেখতে পাবেন, নির্দিষ্ট ঐ সেলটিকে C21 দিয়ে ভাগ করার কথা থাকলে ফর্মুলায় অন্য সেল চলে এসেছে। যেমন C6 কে C25 দিয়ে ভাগ দেয়া হয়েছে। যদিও সূত্র অনুযায়ী ভাগ দেবার কথা C21 দিয়ে।

এই সমস্যা আপনি দুইভাবে সমাধান করতে পারেন। প্রথমত, আপনি ফর্মুলা বারে সেল নাম্বার বদলে C21 দিয়ে দিতে পারেন, নির্দিষ্ট সেলে উত্তর চলে আসবে। তবে যখন ডেটা সেট বড় হবে, তখন হাতে এই ত্রুটি সংশোধন কঠিন হয়ে পড়বে। এই জন্যেই এই সমস্যার দ্বিতীয় একটি সমাধান আছে। যার নাম Anchor।

12.32%

=C3/$C$21

গুগল শিটে এই Anchor বা নোঙ্গর হিসেবে কাজ করে $ সাইন। আপনি যদি কোনো ফর্মুলায় এই $ সাইন ব্যবহার করেন অর্থাৎ Anchor বসান, তাহলে আপনার সূত্র বদলাবে না। আর তাই F2 সেলে নতুন করে ফর্মুলা বা সূত্র বসান Anchor সহ। সেক্ষেত্রে শতাংশের হিসাব বের করতে নতুন সূত্র হবে-

=C2/$C$21

	A	B	C	D	E	F
1	Unit Name	Cases 2018	Cases 2019	Change	Change in Percentage	Percentage in Total Cases 2019
2	DMP	27150	2153	-24997	-92.07%	12.32%
3	CMP	7731	678	-7053	-91.23%	3.88%
4	KMP	2604	317	-2287	-87.83%	1.81%
5	RMP	4247	401	-3846	-90.56%	2.29%
6	BMP	1779	156	-1613	-90.67%	0.95%
7	SMP	1716	143	-1573	-91.67%	0.82%
8	Dhaka Range	400603	3089	-397514	-99.23%	17.67%
9	Mymensingh Range	13716	1004	-12712	-92.68%	5.74%
10	Chattogram Range	29925	2396	-27529	-91.99%	13.71%
11	Sylhet Range	9162	725	-8437	-92.09%	4.15%
12	Khulna Range	21253	1585	-19668	-92.54%	9.07%
13	Barishal Range	11434	891	-10543	-92.21%	5.10%
14	Rajshahi Range	25608	2011	-23597	-92.15%	11.50%
15	Rangpur Range	22278	1472	-20806	-93.39%	8.42%
16	Railway Range	989	84	-905	-91.51%	0.48%
17	GMP	872	244	-628	-72.02%	1.40%
18	RPMP	352	121	-231	-65.63%	0.69%
19	ATU	0	0	0		0.00%
20	Total	221419	17484	-203935		
21	Total Check	581419	17480			100.00%

যখন কোনো একটি মোটের বিভিন্ন অংশের শতকরা হিসাব বের করবেন, তখন মনে রাখবেন সবগুলো অংশ মিলে ১০০ হবে। যদি সেটি না হয় বুঝবেন আপনার হিসাবে ভুল হয়েছে।

হার

আগের অনুশীলনে গুগল শিটস ব্যবহার করে শতকরা হার বা শতাংশের হিসাব বের করা হয়েছে। সাংবাদিকতায় অনেক ক্ষেত্রে হাজারে কতজন সেই হিসাবও বের করতে হয়। বিশেষ করে জনসংখ্যা সংশ্লিষ্ট কোনো হিসাবের ক্ষেত্রে। কেননা আপনি যখন একটি বিরাট জনগোষ্ঠীর মাঝে কোনো একটি হার বিবেচনা করবেন শতাংশে তা খুব সামান্য আসবে। যার মাধ্যমে পাঠক কিংবা দর্শক বিষয়ের গুরুত্বটা বুঝতে পারবে না। যে কারণে কোভিডে মৃত্যুর হারের হিসাব হাজারে প্রকাশ করতে দেখা গেছে। কীভাবে হাজার প্রতি হার বের করা যায়, চলুন এবার তার অনুশীলন করি। এ জন্য স্প্রেডশিটটির পরবর্তী শিট বা ট্যাবে যান। যার নাম দেয়া হয়েছে Rate।

	A	B	C	D
1	Unit Name	Metropolitan	Polpulation	Cases 2019
2	DMP	Dhaka	21,741,090	2153
3	CMP	Chattogram	4,627,000	678
4	KMP	Khulna	1,500,689	317
5	RMP	Rajshahi	1,290,600	401
6	BMP	Barishal	385,093	166
7	SMP	Sylhet	928,000	143

এই ডেটা শিটে, পুলিশের প্রতিটি ইউনিটের মামলার তথ্যের সঙ্গে সংশ্লিষ্ট মহানগরের জনসংখ্যার হিসাবও পৃথক কলামে উল্লেখ করা আছে। যেখানে দেখা যাচ্ছে যে মহানগরে জনসংখ্যা যত বেশি, সেখানে মামলার সংখ্যাও বেশি। কিন্তু এই তথ্য বাস্তবতা কে নির্দেশ করে না। মহানগরগুলোর অপরাধ প্রবণতা সর্ম্পকে বুঝতে আপনাকে জানতে হবে, কোন মহানগরে জনসংখ্যার কত শতাংশ মামলা হয়। সে কারণে এই ডেটাসেটের প্রকৃত চিত্র বুঝতে জনসংখ্যার হিসাবে মামলার হার বের করতে হবে। এখান থেকে প্রথমেই ঢাকা মেট্রোপলিটন পুলিশের হারটি বের করে নিতে পারবেন, মামলার সংখ্যাকে মোট জনসংখ্যা দিয়ে ভাগ করলে।

কলাম E তে মামলার হার হিসাব করুন। E1 সেলের নাম হতে পারে "Cases Rate"। আর সূত্র হবে ভাগের। এক্ষেত্রে-

```
=D2/C2
```

সূত্র কপি করে বাকি সেলগুলোতেও বসিয়ে দিন। এরইমধ্যে ফলাফল পেয়ে গেছেন।

C	D	E
Polpulation	Cases 2019	Cases Rate
21,741,090	2153	0.00009902907352
4,627,000	678	0.0001465312297
1,500,689	317	0.0002112363055
1,290,600	401	0.0003107081977
385,093	166	0.0004310647039
928,000	143	0.0001540948276

E কলামে আপনি যে ফলাফল বের করেছেন সেটি গাণিতিক হিসাবে সঠিক। কিন্তু প্রশ্ন হচ্ছে এই সংখ্যাকে কীভাবে আপনি পাঠকের কাছে উপস্থাপন করবেন। এতো ছোট ভগ্নাংশে কী পাঠক, ঘটনার প্রকটতা সম্পর্কে কিছু বুঝতে পারবে? পারবে না। আর এ কারণে এমন বড় জনগোষ্ঠী থেকে যখন কিছু মানুষের সংখ্যা কে নির্দিষ্ট করতে হয়, তখন তা এক হাজার বা দশ হাজারের সঙ্গে তুলনা দেখাতে হয়। অর্থাৎ প্রতি দশ হাজারে কতটি মামলা হয়েছে। এমন তথ্য আপনার প্রতিবেদন লেখার জন্যেও যেমন সহজ। পাঠকের বুঝতে তেমন সহজ হয়। আর তাই এই অনুশীলনে আমরা মহানগরে প্রতি ১০ হাজারে মামলার হার বের করবো। সেটি করতে বিদ্যমান সূত্রটি সামান্য বদলে নিতে হবে। শতাংশের হিসাবে যেমন ভাগফলকে আপনি ১০০ দিয়ে গুণ করছেন। এখানে ১০০০০ দিয়ে গুণ করতে হবে। সেক্ষেত্রে E2 তে নতুন সূত্রটি হবে–

```
=D2/C2*10000
```

C	D	E
Polpulation	Cases 2019	Cases Rate
21,741,090	215 ?	=D2/C2*10000

নতুন ফর্মুলা বা সূত্রটি বাকি সেলগুলোতে কপি করে নিন।

E7 fx =D7/C7*10000

	A	B	C	D	E
1	Unit Name	Metropolitan	Polpulation	Cases 2019	Cases Rate
2	DMP	Dhaka	21,741,090	2153	0.9902907352
3	CMP	Chattogram	4,627,000	678	1.465312297
4	KMP	Khulna	1,500,689	317	2.112363055
5	RMP	Rajshahi	1,290,600	401	3.107081977
6	BMP	Barishal	385,093	166	4.310647039
7	SMP	Sylhet	928,000	143	1.540948276

অনুপাত

অনুশীলনের এই পর্যায়ে অনুপাতের হিসাব করা শেখা যাক। এ জন্য স্প্রেডশিটের পরবর্তী শিট বা ট্যাবটিতে যান। এই শিটটির নাম দেয়া হয়েছে Ratio।

আপনার সামনে এখন যে ডেটা সেটটি আছে, সেখানে প্রতিটি মহানগরের জনসংখ্যা ও সংশ্লিষ্ট পুলিশ ইউনিটের আওতায় কয়টি করে স্টেশন আছে, তা উল্লেখ করা আছে। যা থেকে আপনাকে বের করতে হবে মহানগরগুলোতে এক একটি স্টেশন কত মানুষের নিরাপত্তায় কাজ করছে ? কোন মহানগরে জনসংখ্যার অনুপাতে স্টেশনের সংখ্যা তুলনামূলক কম।

	A	B	C	D
1	Unit Name	Metropolitan	Station	Polpulation
2	DMP	Dhaka	50	21,741,090
3	CMP	Chattogram	16	4,627,000
4	KMP	Khulna	8	1,500,689
5	RMP	Rajshahi	4	1,290,600
6	BMP	Barishal	4	385,093
7	SMP	Sylhet	6	928,000

এই হিসাব বের করা যাবে স্টেশন ও জনসংখ্যার অনুপাতের মাধ্যমে। অনুপাত বের করতে হবে, ভাগ (/) ব্যবহার করে। তবে প্রশ্ন হচ্ছে, কাকে কি দিয়ে ভাগ দিবেন? এখানে যেহেতু একটি স্টেশন, কতজন মানুষকে নিরাপত্তা দিচ্ছে এই হিসাব বের করা হবে, তাই জনসংখ্যাকে স্টেশনের সংখ্যা দিয়ে ভাগ দিতে হবে।

প্রথমে ঢাকা মেট্রোপলিটন পুলিশের হিসাব করা যাক। এ জন্য নতুন কলাম নিতে হবে। E কলামকে Ratio এর হিসাব করার জন্য নির্ধারণ করুন এবং হেডলাইন লিখুন। সেইসাথে E2 তে অনুপাত বের করার সূত্রটি লিখুন। এক্ষেত্রে সূত্র হবে জনসংখ্যা/স্টেশন সংখ্যা। অর্থাৎ- =D2/C2

E2 fx =D2/C2

	A	B	C	D	E
1	Unit Name	Metropolitan	Station	Polpulation	Ratio
2	DMP	Dhaka	50	21,741,09	=D2/C2
3	CMP	Chattogram	16	4,627,000	
4	KMP	Khulna	8	1,500,689	
5	RMP	Rajshahi	4	1,290,600	
6	BMP	Barishal	4	385,093	
7	SMP	Sylhet	6	928,000	

D2 তে পাওয়া ফলাফলটি নির্দেশ করছে যে, প্রতি চার লাখ ৩৪ হাজার ৮২ জন মানুষের জন্য একটি করে পুলিশ স্টেশন রয়েছে ঢাকা মহানগরে। বাকি মহানগরগগুলোর তথ্য বের করতে একইসূত্র অনুসরণ করুন বাকি সেলগুলোতে। সবগুলো সেলের অনুপাতের ডেটা চলে এলে, এবার নজর দেয়া যাক কোন মহানগরে স্টেশন অনুপাতে জনসংখ্যা বেশি। সেই হিসাব বের করতে সংশ্লিষ্ট কলাম অর্থাৎ এক্ষেত্রে সব ডেটা সিলেক্ট করুন। এরপর মেন্যু বার থেকে Data তে যান এবং Sort range এ ক্লিক করুন।

এরপর নিম্ন ক্রমানুসার অর্থ Descending Order (Z→ A) এ বেছে নিন। যে মহানগরে স্টেশনওয়ারি জনসংখ্যা সবচেয়ে বেশি, সেই মহানগরের নাম সবার উপড়ে উঠে আসবে।

A	B	C	D	E
Unit Name	**Metropolitan**	**Station**	**Polpulation**	**Ratio**
DMP	Dhaka	50	21,741,090	434,822
RMP	Rajshahi	4	1,290,600	322,650
CMP	Chattogram	16	4,627,000	289,188
KMP	Khulna	8	1,500,689	187,586
SMP	Sylhet	6	928,000	154,667
BMP	Barishal	4	385,093	96,273

গড়

পাশের ডেটা সেটটি আপনি আগে দেখেছেন। এটির কিছু বিশ্লেষণ আগেও করেছেন। অনুশীলনের এই পর্যায়ে চলুন দেশের সকল মহানগর পুলিশ ও রেঞ্জগুলোর গড় মামলার হিসাব বের করা যাক। অনুশীলনের জন্য স্প্রেডশিটটির Average নামের শিটটিতে যান।

গাণিতিক নিয়ম অনুসারে, সব ইউনিটের মামলা যোগ করে সেই যোগফলকে মোট ইউনিট সংখ্যা দিয়ে ভাগ করলেই গড় মামলা বের হবে। গুগল শিটস এই প্রক্রিয়া আরও সহজ করেছে Average ফর্মুলা বা সূত্রের মাধ্যমে।

	A	B
1	**Unit Name**	**Cases 2019**
2	DMP	2153
3	CMP	678
4	KMP	317
5	RMP	401
6	BMP	166
7	SMP	143
8	Dhaka Range	3089
9	Mymensingh Range	1004
10	Chattogram Range	2396
11	Sylhet Range	725
12	Khulna Range	1585
13	Barishal Range	891
14	Rajshahi Range	2011
15	Rangpur Range	1472
16	Railway Range	84
17	GMP	244
18	RPMP	121
19	ATU	0
20	Total	17484

স্প্রেডশিটের Average সূত্রটি Sum বা যোগের মত কাজ করে। যোগের ক্ষেত্রে যেমন প্রথম ও শেষ সেলের মধ্যে কোলন (:) সাইন বসিয়ে পুরো কলামের ডেটাকে যোগ করা যায়, একইভাবে প্রথম ও শেষ সেলের মধ্যে কোলন (:) চিহ্ন বসিয়ে গড় হিসাব করা যায়। তবে সূত্রের শুরুতে Average লিখতে হবে, এভাবে-

এই ডেটা সেটের ক্ষেত্রে সূত্র হবে–

```
=AVERAGE(B2:B19)
```

	A	B
1	**Unit Name**	**Cases 2019**
2	DMP	2153
3	CMP	678
4	KMP	317
5	RMP	401
6	BMP	166
7	SMP	143
8	Dhaka Range	3089
9	Mymensingh Range	1004
10	Chattogram Range	2396
11	Sylhet Range	725
12	Khulna Range	1585
13	Barishal Range	891
14	Rajshahi Range	2011
15	Rangpur Range	1472
16	Railway Range	84
17	GMP	244
18	RPMP	121
19	ATU	0
20	Total	971.1111111 ×
21	**Average**	=AVERAGE(B2:B19)

আপনি এরইমধ্যে সারাদেশের মহানগর পুলিশ ও পুলিশের রেঞ্জগুলোর গড় মামলার ডেটা পেয়ে গেছেন, গুগল শিটের সূত্রের মাধ্যমে। তবে এই সংখ্যায় কি আসলেই অপরাধ প্রবণ এলাকার সঠিক চিত্র প্রতিফলিত হয়। উত্তর না। কেননা এই হিসাবে সব মামলা আর ইউনিটকে এক সারিতে নিয়ে আসে। ফলে অনেক ক্ষেত্রেই বাস্তবচিত্র উঠে আসে না।

মধ্যমা

গড় হিসাবে যখন অনেক ক্ষেত্রেই বাস্তবচিত্র উঠে আসে না, তখন পরিসংখ্যানে মধ্যমা হিসাব বিবেচনায় নেয়া হয়। এই ফর্মুলা বা সূত্র কলামের মাঝের ডেটাকে বিবেচনায় নেয়, যোগফলকে নয়। তবে সূত্রটা প্রায় একইরকম। শুধু Average স্থানে Median লিখতে হবে।

=AVERAGE (প্রথম সেল নম্বর:শেষ সেল নম্বর)

এই সূত্র প্রথমে কলামে থাকা ডেটাকে ছোট থেকে বড়, এই ক্রমানুসারে সাজায়। এরপর পুরো ডেটার মধ্যবর্তী সংখ্যাটিকে বিবেচনায় নেয়। অনুশীলনের জন্য স্প্রেডশিটের পরবর্তী শিট বা ট্যাবে যান। যার নাম দেয়া হয়েছে Median।

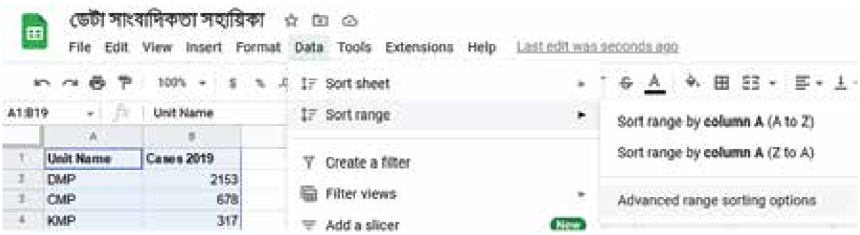
তবে গড় না-কি মধ্যমার সূত্র ব্যবহার করবেন, তা নির্ভর করবে আপনি কেমন ডেটা বিশ্লেষণ করছেন এবং কী ধরনের নতুন ডেটা বিশ্লেষণ থেকে বের করতে চাচ্ছেন তার ওপর।

সর্টিং বা বাছাই

আপনি পুলিশের পক্ষ থেকে যে ডেটাসেট পেয়েছেন সেখানে দেশের সব মহানগর পুলিশ ও রেঞ্জে করা মামলার ডেটা আছে। ইউনিটওয়ারি ওই সেটে সব ডেটা থাকলেও, তথ্য উল্লেখের ক্ষেত্রে কোনো ধারাবাহিকতা নেই। কিন্তু আপনার প্রতিবেদনের জন্য সবচেয়ে বেশি মামলা হওয়া ইউনিট বা মহানগরগুলো চিহ্নিত করতে হবে। গুগল শিটের Sort Range অপশনের মাধ্যমে এই কাজটি আপনি সহজেই করতে পারবেন। ডেটাকে ঊর্ধ্ব কিংবা নিম্ন ক্রমানুসারে সাজানো যায়।

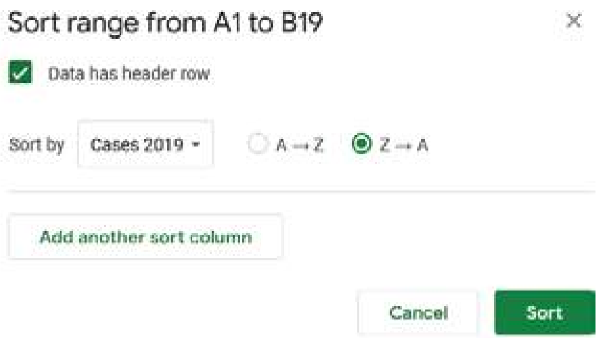
অনুশীলনের জন্য পরবর্তী শিট বা ট্যাবে যান। যার নাম দেয়া হয়েছে Sorting।

প্রয়োজন অনুযায়ী ডেটা সাজাতে, স্প্রেডশিটে থাকা সব ডেটাকে সিলেক্ট করুন। কাঠামো পরিবর্তনের আগে নিশ্চিত করুন, সব কলামের হেডার বা শিরোনাম আছে। কেননা শিরোনামের নাম ব্যবহার গুগল শিটে ডেটা নতুন করে সাজানো হবে।



এবার Data মেন্যু থেকে Sort Range অপশনে যান এবং Advance Range Sorting Option ক্লিক করুন।

এরপর আপনার সামনে নতুন একটি অপশন বক্সটি আসবে। সেখানে Data has header now চেকবক্সে ক্লিক করুন। যাতে কলামের হেডারের নামের ওপর ভিত্তি করে ডেটার ক্রমানুসার করা যায়। যেহেতু মামলার তথ্যের ওপর ভিত্তি করে কলামটি সাজাতে হবে, তাই Sort By তে Cases 2019 বা মামলা নির্বাচন করুন। এবং যেহেতু বেশি মামলা হওয়া ইউনিট চিহ্নিত করা হবে, তাই বেশি মামলা কলামের উপরে রাখতে Z→A(largest to smallest) অপশন নির্বাচন করুন।



আপনার নির্ধারিত অপশনগুলোর ওপর ভিত্তি করে এরইমধ্যে স্প্রেডশিটের ক্রমানুসারে পরিবর্তন এসেছে। যদি প্রতিবেদনে সবচেয়ে কম মামলা হওয়া ইউনিটের কথাও উল্লেখ করতে চান তবে A→Z (smallest to largest) অপশনে ক্লিক করুন। তবে সেই ডেটাসেটের জন্যে নতুন একটি শিটস তৈরি করে নিন।

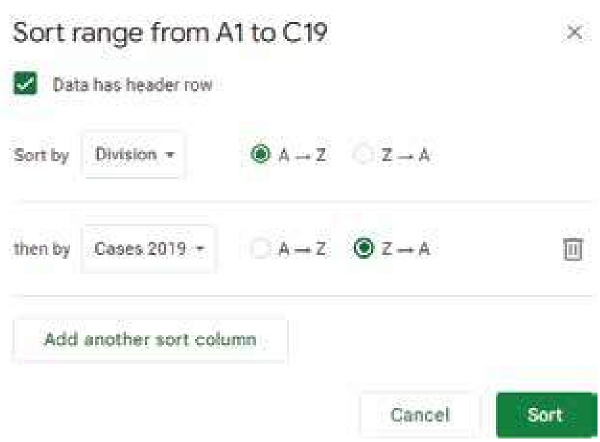
	A	B
1	**Unit Name**	**Cases 2019**
2	Dhaka Range	3089
3	Chattogram Range	2396
4	DMP	2153
5	Rajshahi Range	2011
6	Khulna Range	1585
7	Rangpur Range	1472
8	Mymensingh Range	1004
9	Barishal Range	891
10	Sylhet Range	725
11	CMP	678
12	RMP	401
13	KMP	317
14	GMP	244
15	BMP	166
16	SMP	143
17	RPMP	121
18	Railway Range	84
19	ATU	0

একসঙ্গে বিভিন্ন পর্যায়ে সর্টিং করা

যে ডেটাসেট নিয়ে সবচেয়ে বেশি মামলা হওয়া ইউনিটের ক্রমানুসার নতুন করে সাজালেন। এবারও সেই একই ডেটাসেট নিয়ে কাজ করুন, পরিবর্তন শুধু সেখানে প্রতিটি ইউনিট দেশের কোন বিভাগে কাজ করে সেটিও উল্লেখ করা আছে। যার ওপর ভিত্তি করে আপনি বিভাগওয়ারি ইউনিটের মামলার ক্রমানুসারও সাজাতে পারবেন। যার মাধ্যমে কোন বিভাগের পুলিশের ইউনিটে মামলার পরিমাণ কেমন, কোনটি কম এবং কোনটিতে মামলা কেমন হচ্ছে, তা উল্লেখ করতে পারবেন।

	A	B	C
1	**Unit Name**	**Division**	**Cases 2019**
2	DMP	Dhaka	2153
3	CMP	Chattogram	678
4	KMP	Khulna	317
5	RMP	Rajshahi	401
6	BMP	Barishal	166
7	SMP	Sylhet	143
8	Dhaka Range	Dhaka	3089
9	Mymensingh Range	Mymensingh	1004
10	Chattogram Range	Chattogram	2396

যদি আপনি পুলিশের ইউনিটগুলোকে বিভাগওয়ারি দেখতে চান, তাহলে প্রথমে দল বা গ্রুপ অনুসারে সাজাতে হবে। তারপর, যদি প্রতিটি বিভাগের মধ্যে, সর্বোচ্চ মামলা হওয়া ইউনিট দেখতে চান, তাহলে আমাদের দ্বিতীয় পর্যায়ে মামলা অনুসারে সাজাতে হবে।



এ জন্য প্রথমে পুরো ডেটাসেট সিলেক্ট করতে হবে। এরপর আগের অনুশীলনের মতো Sort B অপশনে যেতে হবে। তবে এবার দুইটি কলাম সিলেক্ট করতে হবে। প্রথমে বিভাগ এবং তারপরে মামলা। এবারও আগের মত Data মেন্যু থেকে Sort range অপশনে যান এবং Advance Range Sorting Option ক্লিক করুন। নতুন অপশন বক্সটিতে Sort By Division সিলেক্ট করুন। এরপর Add another sort column সিলেক্ট করে Sort By Casess 2019 বেছে নিন। বেশি মামলা হওয়া ইউনিটগুলোকে কলামের উপরে রাখতে Z → A (largest to smallest) অপশন নির্বাচন করুন।

ফিল্টারিং বা ছেঁকে বের করা

গুগল শিটসের Filter অপশনের সাহায্যেও ডেটার ক্রমানুসার বদলানো যায়। একইসঙ্গে Filter এ আরও বেশকিছু কার্যক্রম আছে যা ব্যবহার করে আপনার বিশ্লেষণ প্রক্রিয়া আরও সহজ করা যায়। তথ্যের যে ধরনের উপস্থাপন আপনি ডেটাশিটে দেখতে চান Filter আপনাকে তা করতে সহায়তা করবে।

> এই অনুশীলনের জন্য নতুন একটি ডেটাসেট নিয়ে কাজ করি চলুন। তবে স্প্রেডশিট একই থাকবে। অনুশীলনের জন্য Filtering নামের শিট বা ট্যাবটিতে যান। যে ডেটা সেটটি দেখতে পাচ্ছেন। সেটি আপনার পরিচিত। কারণ এই ডেটা সেটটিকে সিএসভি ফাইল হিসেবে গুগল শিটে যুক্ত করা হয়েছিলো।

> Filtering এর এই কার্যক্রম শুরু করতে পুরো ডেটা সেটটি সিলেক্ট করুন অথবা কার্সর দিয়ে A1 সেল সিলেক্ট করুন। এরপর মেন্যু বার থেকে Data তে যান এবং Create a filter-এ ক্লিক করুন। এই অপশনে ক্লিক করার পরে আপনার ডেটা সেটের প্রত্যেকটি কলামে শিরোনাম বা হেডের পাশে একটি করে ত্রিভুজ আকৃতির অপশন আসবে। যার অর্থ ডেটা সেটটি Filtering কার্যক্রম পরিচালনার জন্য প্রস্তুত।

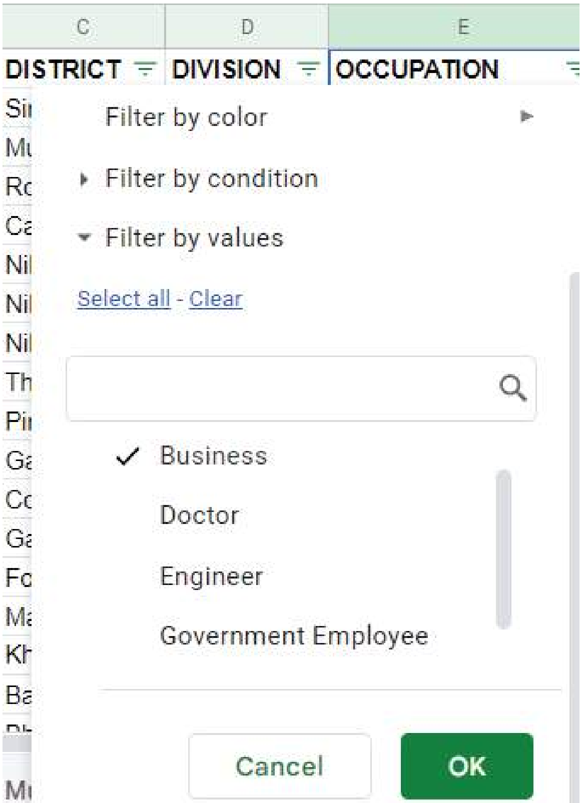
ত্রিভুজ আকৃতির অপশনটিতে ক্লিক করলে নতুন একটি মেনুবার আপনার সামনে আসবে, যার মাধ্যমে আপনি কলামে বা ডেটাসেটে কি তথ্য দৃশ্যমান করতে চান, তা নির্ধারণ করতে পারবেন।

গুগল শিটে তিনটি পৃথক পদ্ধতিতে ফিল্টার করা যায়
Value বা মান অনুসারে
Condition বা শর্ত অনুসারে
Color বা রং অনুসারে

এই অধ্যায়ে রং অনুসারে ফিল্টার পদ্ধতি নিয়ে আলোচনা করা হয়নি।

মান অনুসারে

টেক্সট ডেটার ক্ষেত্রে Value বা মান অনুসারে ফিল্টার বেশি প্রচলিত। এর মাধ্যমে আপনি আপনার প্রয়োজনীয় টেক্সট ডেটা শিটে আলাদা করে বের করতে পারবেন।



অনুশীলনে ব্যবহৃত ডেটাসেটটিতে ৩৯ জন ব্যক্তির মাসওয়ারি আয়, তাদের পেশা, জেলা ও বিভাগের ডেটা রয়েছে। বিশ্লেষণের প্রয়োজনে আপনার হয়তো শুধু পেশায় ব্যবসায়ী এমন ব্যক্তির তথ্য দরকার। আলাদা করে সেই তথ্য পেতে বিভাগ কলামের হেডারে থাকা ত্রিভুজ অপশনটিতে ক্লিক করুন। নতুন আসা মেন্যু বার থেকে Filter by values এর মধ্যে শুধু Business সিলেক্ট করুন। আলাদা করে Business সিলেক্ট করতে বাকি সব পেশার নাম Clear করে, নিচের বক্সে আপনার প্রয়োজনীয় বিভাগের নামটি লিখুন এবং সিলেক্ট করে, Ok বাটনে ক্লিক করুন।

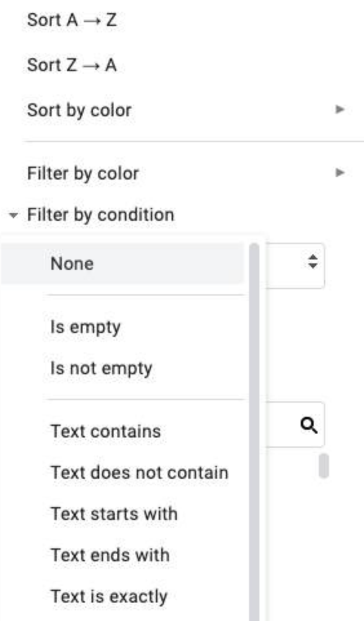
এর সাথে সাথে আপনার ডেটা শিটে শুধু ঐ নির্দিষ্ট পেশার ব্যক্তিদের তথ্যই কেবল থাকবে। তবে বাকিগুলো মুছে যাবে না। সাময়িক সময়ের জন্যে দৃশ্যমান থাকবে না। ফিল্টার তুলে নিলে, আবার সব ডেটা আগের স্থানে চলে আসবে। তাই ফিল্টার করা ডেটা পড়ে প্রয়োজন হলে, নতুন একটি শিট খুলে কপি করে রাখুন।

	A	B	C	D	E
1	NAME	DISTRIC	DIVISION	OCCUPATION	INCOI
2	Tanver Emon	Sirajgonj	Rajshahi	Business	80000
5	Shovon Khan	Candpur	Chattogram	Business	45000
9	Hasib Ferdows	Thakurgaon	Rongpur	Business	56000
11	Abdur Rauf	Gazipur	Dhaka	Business	140000
16	Rumana Sharmin	Khulna	Khulna	Business	35000
17	Momtaz Begum	Bagerhat	Khulna	Business	125000
23	Mishkat Afrin	Thakurgaon	Rongpur	Business	75000

শর্ত অনুসারে

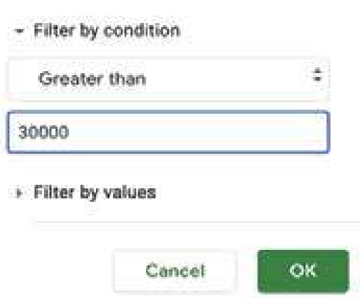
এই পদ্ধতিতে টেক্সট এবং সংখ্যা দুটোই ফিল্টার করা যায়। বিশ্লেষণের অনেক পর্যায়ে আপনার হয়তো এমন কিছু ডেটার দরকার, যার নাম বা বিশেষণ হয়তো নির্দিষ্ট করা যায় না। একটি সীমা বেধে দেয়া যায় সর্বোচ্চ। এমন সব ক্ষেত্রে Condition ফিল্টারিং ব্যবহার করা হয়।

এই অপশনে আপনি টেক্সট বা কলামগুলির জন্য আলাদা আলাদা পরিস্থিতি উল্লেখ করা যাবে। আপনি কিছু নির্দিষ্ট শব্দও উল্লেখ করতে পারবেন, যেগুলো টেক্সটের শুরুতে বা শেষে রয়েছে। এই অপশন বা বিকল্পগুলো খুঁজে পেতে, আপনি যে কলামটি ফিল্টার করতে চান তার ফিল্টার ড্রপডাউনে ক্লিক করুন। এরপর Filter by Condition থেকে প্রয়োজনীয় পরিস্থিতি নির্বাচন করুন।



ধরুন আপনি ডেটা থেকে সরকারি চাকুরিজীবীদের বাছাই করবেন। সেক্ষেত্রে আপনি Government শব্দটিকে সূত্র হিসেবে বেছে নিলেন, যে ফিল্টার শুধু তাদের বের করে আনবে যাদের পেশায় Government শব্দটি আছে। তা সে যে জেলা বা বিভাগের হোক না কেনো।

Filter by Condition> Text starts with



একইভাবে আপনি বেতনের ক্ষেত্রেও এটি করতে পারেন। ধরুন আপনি জানতে চাইলেন কতজন কর্মী এবং কারা ৩০ হাজার টাকার বেশি বেতন পান। তখনও একইভাবে Filter by Condition থেকে Greater than-৩০০০০ লিখুন।

Filter by Condition > Greater than

মুহূর্তেই আপনি আপনার কাঙ্ক্ষিত ডেটাগুলো আলাদা করে পেয়ে যাবেন। এবার এই ডেটা কপি করে নতুন একটি সিটে রাখুন।

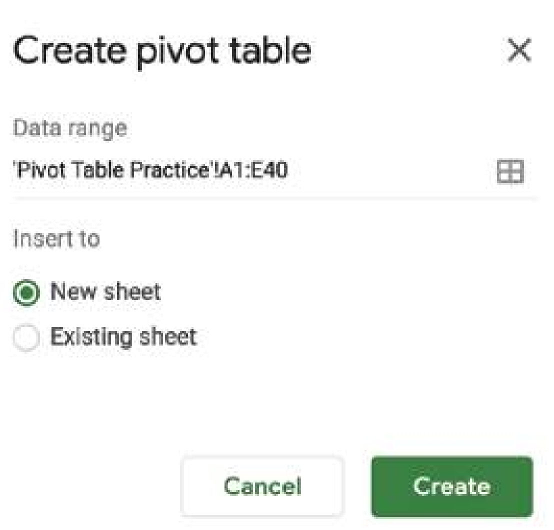
ফিল্টারিং এর কাজ শেষ হলে, Data মেন্যু বারে গিয়ে Turn off filter অপশনে ক্লিক করুন। ডেটা শিটে থাকা সব ডেটা আবার ফিরে আসবে।

Data menu > Turn off filter.

পিভট টেবিল

পিভট টেবিলের মাধ্যমে একটি ডেটা সেটের কিছু নির্দিষ্ট অংশের হিসেব দ্রুত করা যায়। পিভট টেবিল ঐ নির্দিষ্ট ডেটা গ্রুপের মধ্যে, যোগ, বিয়োগ, গড়ের মত গাণিতিক হিসাব করে।

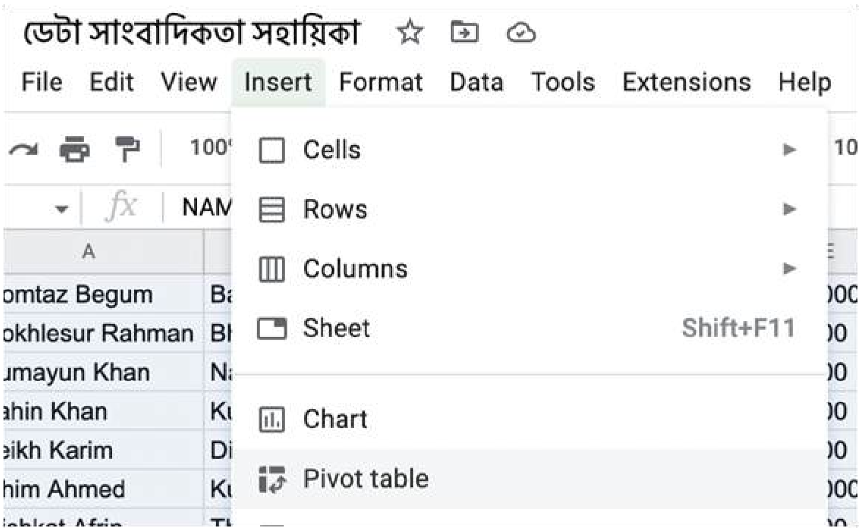
যেমন, আপনার কাছে থাকা ডেটাসেটটি আছে সেটি মূলত রাজধানি ঢাকার একটি আবাসিক এলাকার দশতলা অ্যাপার্টমেন্টে ৩৯ বসবাসকারীর। যেখানে বিভিন্ন পেশা বা বিভিন্ন জেলা বা বিভাগের মানুষ রয়েছে। পিভট টেবিলের মাধ্যমে এই ডেটা সেটটিকে পেশাওয়ারি ভাগ করতে পারবেন। জেলা বা বিভাগওয়ারিও ভাগ করা যাবে। তারপর আপনি হিসাব করতে পারবেন, একই পেশার বা একই বিভাগের কত ব্যক্তি সেখানে আছেন, তাদের মোট বেতন ও গড় বেতন কতো।



এই অনুশীলনে আগে ব্যবহৃত ডেটা সেটটিই ব্যবহার করা হবে। গুগল সিটটি আপনার গুগল ড্রাইভে কপি করুন, তারপর ব্যবহার করুন।

পিভট টেবিল তৈরি

স্প্রেডশিটের বিভিন্ন পর্যায়ের অনুশীলনে ডেটার যোগফল কিংবা গড় বের করেছেন। কিন্তু যদি আপনি জানতে চান কোন পেশার মানুষ বেশি আয় করে বা কোন পেশার মানুষের আয় কেমন। অর্থাৎ শুধু পেশাভেদে মোট বেতন বা বেতনের গড় বের করতে চান তাহলে কী করবেন?

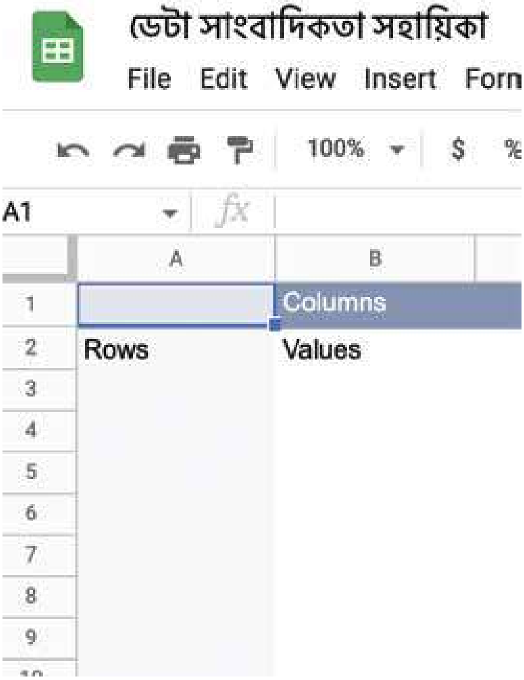


> সহজ উত্তর। আপনাকে প্রথমে ডেটা থেকে পেশাওয়ারি আলাদা গ্রুপ ভাগ করতে হবে। এবং তারপর প্রতিটি গ্রুপের ডেটা যোগ করতে হবে। পিভট টেবিলে এই কাজটি করা যায় খুব সহজে। তবে তা করতে আপনাকে প্রথমে একটি পিভট টেবিল তৈরি করতে হবে।

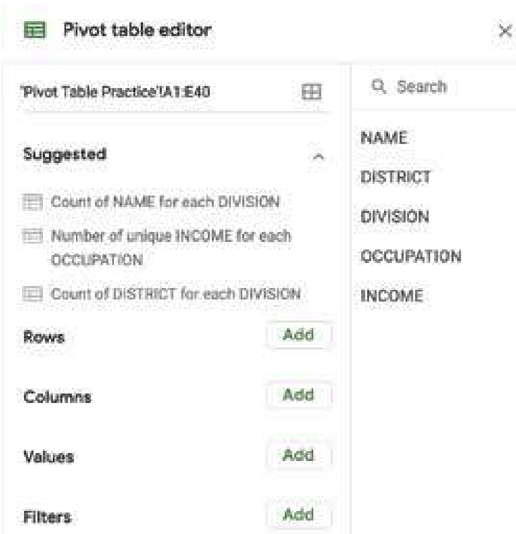
পিভট টেবিলটি তৈরি করতে, পুরো ডেটাসেটটি সিলেক্ট করুন। তারপর মেন্যু বার থেকে ডেটা মেন্যু তে যান এবং পিভট টেবিল সিলেক্ট করুন।

Data > Pivot Table

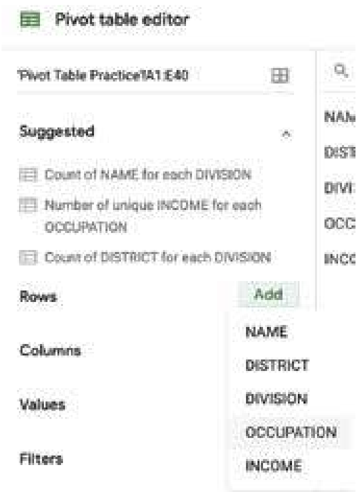
Pivot Table অপশনটি নির্বাচন করলে আপনার সামনে নতুন আরেকটি মেন্যু বক্স আসবে। যেহেতু আপনি এই প্রক্রিয়া শুরু করার আগেই পুরো ডেটাসেটটি সিলেক্ট করেছিলেন, তাই ডেটা রেঞ্জে সেলের নম্বর চলে আসবে। রেঞ্জটি ঠিক আছে কি-না যাচাই করে নিন এবং নতুন শিট তৈরি করুন।



নতুন একটি শিটের কলাম A1 B-তে আপনার পিভট টেবিলটি তৈরি হয়ে গেছে। আর স্প্রেডশিটের ডানপাশে একটি Pivot Table Editor তৈরি হয়েছে। এই মেন্যু বক্সটির মাধ্যমে পিভট টেবিলের ডেটা সাজানো এবং হিসাবের কাজটি করা হয়।



স্প্রেডশিটের ডেটা পিভট টেবিলে যুক্ত করতে Add বাটনে ক্লিক করতে হবে। প্রথম কলাম অর্থাৎ Rows-এ যে ডেটা যুক্ত করতে চান সেটি Add বাটনে ক্লিক করে নির্বাচন করুন। একইভাবে বাকি তিনটি অপশন যেমন কলাম, ভ্যালুস ও ফিল্টারে ডেটা যুক্ত করতে পারবেন। আপনি একসাথে চারটি অপশনও ব্যবহার করে বিশ্লেষণ করতে পারবেন।



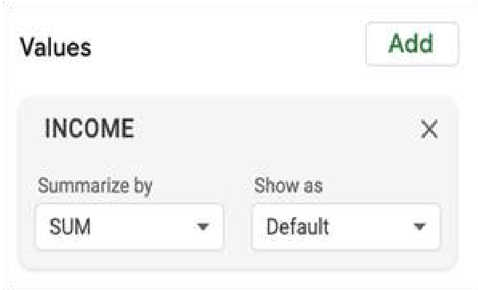
আপনার পিভট টেবিলের কাঠামো কেমন হবে তার পুরোটাই নির্ভর করে এই বিশ্লেষণ থেকে আপনি কী ধরনের তথ্য বের করতে চান। একইসাথে গুগল শিটও প্রাথমিকভাবে আপনার সামনে কিছু পরামর্শ তুলে ধরবে, কি ধরনের অপশন বেছে নিতে পারেন। তবে শেষ বিচারটি আপনিই করুন বিশ্লেষণের লক্ষের ওপর নির্ভর করে।

মূল ডেটা সেটের যে অংশের বা গ্রুপের ডেটা আপনি বিশ্লেষণ করতে চান সেই কলামটি Rows এ রাখুন এবং যে সংখ্যার ওপর ভিত্তি করে হিসাবটি তৈরি হবে সেটি Values এ রাখুন।

Rows = গ্রুপ/ বিভাগ, Values = অঙ্ক

এই পিভট টেবিল ব্যবহার করে আমরা যেহেতু পেশাওয়ারি বেতন বিশ্লেষণ করবো, সে কারণে Rows এ পেশা অর্থাৎ Occupation এর তথ্য যুক্ত করুন। একইসঙ্গে Values এ আয় বা Income যুক্ত করবো। প্রথমে Rows এ পেশা অর্থাৎ Occupation তথ্য যুক্ত করুন।

> আপনি দেখতে পাবেন পিভট টেবিলের বাম পাশে ডেটা শিটে থাকা প্রতিষ্ঠানটির সবগুলো বিভাগ একবার করে যুক্ত হয়েছে। সবগুলো বিভাগ যুক্ত হয়েছে কি-না একবার যাচাই করে নিন।



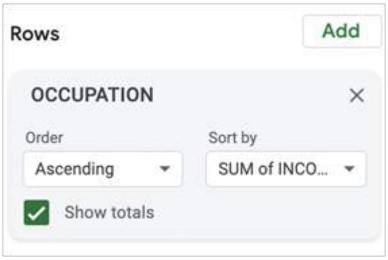
এরপর পিভট টেবিলে সংখ্যা যুক্ত করুন। এ জন্যে Values ক্লিক করুন এবং আয়ের কলামটি যুক্ত করুন। Values এ পিভট টেবিলে গাণিতিক হিসাবটা কীভাবে আপনি চান, সেটিও এই পর্যায়ে নির্ধারণ করে দিতে হবে। যোগ, গড়, মধ্যমাসহ মোট ৩০ ধরনের গাণিতিক হিসাব করার সুযোগ আছে পিভট টেবিলের এই অপশনের মাধ্যমে। তবে প্রথম অপশন হিসেবে SUM বা যোগ আসে। আপনার বিশ্লেষণের চাহিদা অনুযায়ী, এই অনুশীলনেও SUM বা যোগ অপশন নির্বাচন করতে হবে।

	A	B
1	*OCCUPATION*	SUM of INCOME
2	Business	352000
3	Doctor	287000
4	Engineer	170000
5	Government Employee	258000
6	Teaching	200000
7	**Grand Total**	**1267000**

পেশাওয়ারি মোট বেতনের পাশাপাশি পিভট টেবিল সবগুলো বিভাগ মিলে প্রতিষ্ঠানের মোট বেতনের একটি হিসাবও উল্লেখ করে।

পিভট টেবিল সাজানো

পিভট টেবিল অক্ষরের ক্রমানুসারে Rows এবং কলামের ডেটা সাজিয়ে থাকে। তবে আপনি সেটা বিশ্লেষণের বা প্রতিবেদনের সুবিধা মতো সাজাতে পারেন।



পিভট টেবিলের এই ডেটাগুলোর ক্রমানুসার Pivot Table Editor বক্স থেকে নির্ধারণ করতে পারবেন। ধরুন আপনি সবচেয়ে বেশি করা পেশার তথ্য এই পিভট টেবিল থেকে জানতে চান। সেক্ষেত্রে আপনাকে প্রতি পেশার মোট বেতন এবং বেতনের নিম্নক্রমানুসারে অপশন নির্বাচন করতে হবে। এক্ষেত্রে Sort by SUM of Salary, Ges Order Ascending।

পিভট টেবিলে স্বয়ংক্রিয়ভাবে ফলাফল চলে আসবে। এই বিশ্লেষণ পেতে আপনাকে আলাদা করে পুরো ডেটাগুলোকে আলাদা করে নির্বাচন বা সিলেক্ট করতে হবে না।

পিভট টেবিলের হিসাব পরিবর্তন

পিভট টেবিল থেকে আপনি কি ধরনের বিশ্লেষণ চাচ্ছেন, তা বুঝতে গুগল সিটের মাঝেমধ্যে ভুল হতে পারে। অথবা আপনি হয়তো ভিন্ন ধরনের কোনো বিশ্লেষণ চাচ্ছেন। এমন ক্ষেত্রে আপনি নিজেই শিটের গাণিতিক হিসাবের ধরন নির্বাচন করতে পারেন। সেটি করতে Pivot Table Editor এর Values বক্সের Summarize by অপশনে যেতে হবে।

Summarize by অপশনের মাধ্যমে বিভিন্ন সূত্র ব্যবহার করে সংখ্যাকে বিশ্লেষণ করতে পারবেন। স্প্রেডশিটে আপনি যেভাবে সূত্র লিখে দিতেন, এখানে শুধু আপনাকেই তেমন একটি সূত্র নির্বাচন বা সিলেক্ট করে দিতে হবে। অর্থাৎ পিভট টেবিল আপনার কাজ আরও অনেক সহজ করে দিচ্ছে। একাধিক বিশ্লেষণ একবারে করতে পারছেন।

সাংবাদিকতার তথ্য বিশ্লেষণে যেসব হিসাব প্রায়শই ব্যবহৃত হয়

- SUM/**যোগ :** গ্রুপের মোট সংখ্যা বা মানের যোগফল
- Count A/**কাউন্ট এ :** গ্রুপের মোট ডেটার সংখ্যা (শুধুমাত্র টেক্সট কলামের জন্য)
- Count/**কাউন্ট :** গ্রুপের মোট ডেটার সংখ্যা (শুধুমাত্র টেক্সট সংখ্যার জন্য)
- Average/**গড় :** গ্রুপের গড় মান
- Max/**সর্বোচ্চ :** গ্রুপের সর্বোচ্চ সংখ্যা/মান
- Min/**সর্বনিম্ন :** গ্রুপের সর্বনিম্ন সংখ্যা/মান
- Median **:** গ্রুপের মধ্যম মান/সংখ্যা

ধরুন এবার আপনি পেশাগুলোর গড় বেতন বের করবেন। তাহলে Summarize by অপশনে SUM এর বদলে AVERAGE লিখুন। পিভট টেবিলে এবার আপনি একই দলের ভিন্ন গাণিতিক বিশ্লেষণ দেখতে পাবেন। যোগফলের পরিবর্তে পেশাওয়ারি গড় বেতন।

	A	B
1	*OCCUPATION*	AVERAGE of INCOME
2	Business	88000
3	Doctor	57400
4	Engineer	85000
5	Government Employee	43000
6	Teaching	40000

পিভট টেবিলে বিশ্লেষণের পূর্ণরূপ

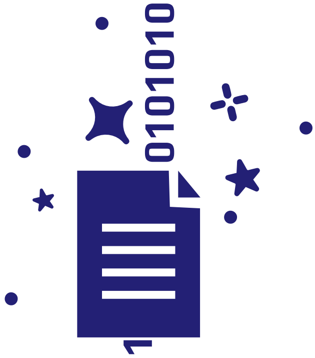
যদি আপনি প্রতিটি পেশার আয়ের যোগফল, গড় এবং কর্মরত ব্যক্তির সংখ্যা একসঙ্গে দেখতে চান, তাহলে পিভট টেবিলে একাধিক অপশন যুক্ত করতে পারবেন। এই অনুশীলনের ক্ষেত্রে Value বক্সে প্রথমে যোগফল যুক্ত করতে হবে। এরপর একইপদ্ধতিতে গড় এবং তারপর মোট সংখ্যা বা Count। এভাবে তিনটি কলাম আলাদা আলাদাভাবে যুক্ত করুন

	A	B	C	D
1	*OCCUPATION*	SUM of INCOME	AVERAGE of INCOME	COUNTA of OCCUPATION
2	Business	352000	88000	8
3	Doctor	287000	57400	8
4	Engineer	170000	85000	7
5	Government Employee	258000	43000	8
6	Teaching	200000	40000	8
7	**Grand Total**	**1267000**	**57590.90909**	**39**

এখন আপনি একবারে পেশাওয়ারি মোট বেতন, গড় বেতন এবং প্রতি পেশায় থাকা মোট ব্যক্তির সংখ্যা একবারে পিভট টেবিলে উঠে এসেছে।

ডেটা বিশ্লেষণে স্প্রেডশিটের পরের ধাপে চাইলে আপনি পাইথনও শিখতে পারেন। এই কিউআর কোডটি স্ক্যান করলে জিআইজেএনের যে টিপশিটটি পাবেন, সেখানে অনলাইনে পাইথন শিক্ষার বিস্তারিত আছে।

0101010101010101010101010101010101010101010101010101010101010



অধ্যায় ০৮

ডেটা থেকে গল্প

> **ডেটা স্টোরি শুধু ডেটা নিয়ে স্টোরি নয়। কিংবা এসব প্রতিবেদন শুধু ডেটার বিশ্লেষণেই শেষ হয় না। ডেটাও মানুষের গল্প বলে। সংখ্যার পেছনে জীবন থাকে। আর সেটি লেখার মাধ্যমে তুলে ধরতে হয়। একজন প্রতিবেদক যত সহজ গল্পে তার ডেটা বিশ্লেষণের ফলাফল পাঠক বা দর্শকের সামনে তুলে ধরবে একটি ডেটা প্রতিবেদন ততো সফল ও শক্তিশালী হবে। কেননা, মানুষ সংখ্যা শুধু শোনেন বা জানেন, কিন্তু গল্প অনুভব করেন।**

মনে রাখবেন, ডেটা সাংবাদিকতা মানে শুধু পরিসংখ্যান দিয়ে প্রতিবেদনটিকে ভরিয়ে ফেলা নয়। গল্পটা শেষ পর্যন্ত মানুষের। সেখানে যতটা না সংখ্যা থাকবে, তার চেয়ে বেশি থাকবে বিশ্লেষণ, যেন ডেটার সঙ্গে মানুষের জীবনের সম্পর্কটা বোঝা যায়। সংখ্যাগুলোকে মূল লেখা থেকে সরিয়ে প্রযোজ্য জায়গায় চার্টে বা ভিজ্যুয়ালাইজেশনের মাধ্যমে সহজবোধ্য করে তুলে ধরুন। চিত্র হবে গল্পের অংশ। ডেটা হবে চিত্রের উপজীব্য। কিন্তু শুধু ডেটাই স্টোরি নয়। ডেটা, মানুষ, সাক্ষাৎকার, যাচাই, বিভিন্ন সূত্রের ব্যবহার এবং চিত্রায়ণ সবকিছু মিলেই একটি ডেটা প্রতিবেদন। অনলাইনের আশীর্বাদকে কাজে লাগান। ইন্টারঅ্যাকটিভ চিত্রায়ণ ব্যবহার করে ডেটা সাংবাদিকতাকে আকর্ষণীয় ও বোধগম্য করে তুলুন।

> ডেটাকে গল্পে রূপান্তরের বেশকিছু কৌশল বৈশ্বিক সাংবাদিকতায় ব্যবহৃত হয়। যা অনুসরণ করে একজন প্রতিবেদক একটি বড় ডেটা শিটের বিশ্লেষণকে একটি মানবিক গল্পে পরিণত করতে পারে। এই অধ্যায়ে সে সব কৌশল নিয়ে আলোচনা করা হবে।

ডেটাকে প্রশ্ন করে শুরু

প্রথমেই ডেটা থেকে সংবাদ খুঁজে বের করতে হবে। এ জন্য ডেটাকে প্রশ্ন করার পরামর্শ দিয়েছেন ইন্টারনিউজের ডেটা উপদেষ্টা ইভা কন্সট্যানটারাস। কেননা, ডেটা থেকে প্রাপ্ত উত্তরই আপনার প্রতিবেদনের ভবিষ্যৎ নির্ধারণ করে দেবে। শিরোনাম থেকে শুরু করে এই প্রতিবেদনে কী কী তথ্য আপনি যুক্ত করবেন, কার সাক্ষাৎকার নিবেন এই সবই প্রশ্নোত্তরের মাধ্যমে জানা যাবে।

আর তা জানতে ডেটাকে চার ধরনের প্রশ্ন করা হয়

- সমস্যা-সংক্রান্ত প্রশ্ন
- প্রভাব-সংক্রান্ত প্রশ্ন
- কারণ-সংক্রান্ত প্রশ্ন
- সমাধান-সংক্রান্ত প্রশ্ন

সমস্যা-সংক্রান্ত প্রশ্ন ডেটা শিটের বিশ্লেষণ থেকে আপনার প্রথম যে প্রশ্নের উত্তর খুঁজে বের করতে হবে তা হলো, সমস্যাটা কত বড়? এর বিস্তৃতি কত দূর? সমস্যাটি কী সমাধানের দিকে এগোচ্ছে না-কি আরও প্রকট হচ্ছে?

এই তিন প্রশ্ন আপনাকে আপনার প্রতিবেদনের গুরুত্ব এবং প্রয়োজনীয়তা বুঝতে সাহায্য করবে। একইসাথে প্রতিবেদন তৈরির প্রক্রিয়া শুরুর আগে আপনি যে পূর্বানুমান করেছিলেন সেটি সঠিক না ভুল, তা যাচাই করা যাবে। তবে পূর্বানুমান ভুল হলেও, আপনার বিশ্লেষণ বৃথা যাবে না বা আপনার প্রতিবেদন শেষ হয়ে যাবে না। আপনার ডেটা বিশ্লেষণে যে ফলাফল এসেছে তার ওপর ভিত্তি করেই তৈরি হবে প্রতিবেদন। ধরুন, আপনি বায়ু দূষণ নিয়ে প্রতিবেদন তৈরি করবেন। এজন্য আপনাকে প্রথমেই জানতে হবে, দেশে বায়ু দূষণ বাড়ছে না কমছে? আগের বছরগুলোতে কেমন ছিল? দেশের বায়ু দূষণের অবস্থা বিশ্বের অনন্যা দেশের তুলনায় কি পর্যায়ে আছে? আর এসব ডেটার মাধ্যমে সমস্যা কতটা প্রকট সেটা বোঝা যাবে।

প্রভাব-সংক্রান্ত প্রশ্ন

ডেটার সঙ্গে প্রশ্নোত্তরের এই পর্যায়ে আপনি প্রতিবেদনের মুখ বা চরিত্রের সন্ধান পাবেন। তবে সে জন্য আপনাকে এ সংক্রান্ত প্রশ্ন করতে হবে।

- সমস্যাটির কারণে কারা ক্ষতিগ্রস্ত হয়েছে?
- কীভাবে ক্ষতিগ্রস্ত হয়েছে?
- আক্রান্তদের সবাই কী সমান ভাবে ক্ষতিগ্রস্ত হচ্ছে?
- না-কি কিছু অংশ বেশি, কিছু অংশ কম ক্ষতিগ্রস্ত হয়েছে?
- ক্ষতিগ্রস্তদের বয়স কত, পেশা কী, কোথায় থাকেন?

এখানেও উদাহরণ স্বরূপ বায়ু দূষণের প্রতিবেদনটি ব্যবহার করা যেতে পারে। ধরুন, বায়ু দূষণের কারণে শিশুসহ সবস্তরের মানুষ ক্ষতিগ্রস্ত হচ্ছে। কিন্তু কোন বয়স, পেশা বা এলাকার মানুষ কতটা আক্রান্ত হচ্ছে, তা এই প্রশ্নের মাধ্যমে চিহ্নিত করতে হবে। হাসপাতাল বা চিকিৎসকের ডেটা সংগ্রহ করে আপনি এই বিশ্লেষণ করতে পারবেন।

কারণ-সংক্রান্ত প্রশ্ন

প্রতিবেদনের ভিলেন বা খলনায়ক সন্ধান পেতে ডেটাকে কারণ সংক্রান্ত প্রশ্নগুলো করতে হবে। এক্ষেত্রেও তিনটি ভিন্ন প্রশ্ন রাখতে হবে ডেটা সেটের সামনে।

প্রশ্নোত্তরের এই অংশে প্রথমেই ডেটা থেকে সমস্যার কারণ জানার চেষ্টা করুন, জানতে চান

- সমস্যার কারণ কী?
- এই সমস্যার জন্যে দায়ী কে বা কারা?
- কী কারণে সমস্যাটা প্রকট হয়েছে?

এখন প্রশ্ন হচ্ছে, ডেটাকে এই প্রশ্ন করবেন। ধরুন বায়ু দূষণের জন্য দায়ী কে, সেটি খুঁজতে আপনাকে প্রথমেই কীভাবে বায়ু দূষণ হচ্ছে, তা চিহ্নিত করতে হবে। কোন উৎস থেকে দূষণের মাত্রা কত, সেটি আবহাওয়া বিভাগের ডেটাকে প্রশ্ন করে বের করতে পারবেন। আরেকটি ডেটা দিয়েও এক্ষেত্রে সমস্যার কারণগুলোকে প্রমাণ করতে পারবেন। যেহেতু বায়ু দূষণের ব্যাপকতা বুঝতে আপনি প্রথমেই গত কয়েক বছরের দূষণের ডেটা বের করেছেন, বছরওয়ারি সেই ডেটার সঙ্গে বায়ু দূষণের উৎসগুলোর আনুপাতিক হার বের করলে সেটি করতে পারবেন। যেমন, দশ বছরে শহরে যে গাড়ির সংখ্যা বা ইটের ভাটা যে হারে বেড়েছে, বায়ু দূষণের মাত্রা বৃদ্ধির সঙ্গে তার সামঞ্জস্য খুঁজে পাবেন।

সমাধান-সংক্রান্ত প্রশ্ন

একটি প্রতিবেদন লেখার শেষ ভাগে সাধারণত প্রতিবেদন সংশ্লিষ্ট সমস্যাটির সমাধান উল্লেখ করা হয়। ঠিক একইভাবে ডেটাকে প্রশ্ন করার শেষ পর্যায়ে সমাধান-সংক্রান্ত প্রশ্নগুলো সামনে আনা হয়। ডেটাকে প্রশ্ন করা হয়।

ডেটাকে প্রশ্ন করা হয়

- এই সমস্যার সমাধান কী?
- সমাধান উদ্যোগের কার্যকারিতা কীভাবে পরিমাপ করা হবে?
- এর আগে এই ধরনের সমস্যার সমাধান কোথায় হয়েছে?
- এবং তারা কীভাবে তার সমাধান করছে?

প্রতিবেদনে উল্লেখ করা সমস্যার সমাধান আপনি ডেটার মাধ্যমে দিতে পারবেন। বায়ু দূষণের প্রতিবেদনটি ধরা যাক। বায়ু দূষণ কমানোর জন্যে বিশেষজ্ঞরা স্বাভাবিকভাবেই এর উৎস বন্ধের পরামর্শ দিবেন। কিন্তু সেটি কতটা কার্যকর তা আপনি ডেটার মাধ্যমে দেখাতে পারেন। ধরুন, আপনি প্রতিবেদনে ইট ভাটা বন্ধ এবং হাইব্রিড গাড়ি ব্যবহার বাড়ানোকে সমাধান হিসেবে উল্লেখ করেছেন। সেক্ষেত্রে যে শহরে হাইব্রিড গাড়ির ব্যবহার দিনদিন বাড়ছে, সেই শহরের বায়ু দূষণের ডেটা এবং গাড়ির নিবন্ধনের ডেটা বিশ্লেষণ করে আপনি দেখাতে পারেন কীভাবে হাইব্রিড গাড়ির ব্যবহার দূষণ কমায়। একইভাবে ইটের ভাটা বন্ধ হয়েছে, এমন শহরে দূষণ কমার ধারাবাহিক মাত্রা আর ভাটা বন্ধের ডেটার সামঞ্জস্যতা তুলে ধরতে পারবেন।

> তবে সব ডেটা যে পাবেন তা নয়, অনেক ডেটাই আপনি হয়তো পাবেন না। আবার ডেটা পেলেও আপনার প্রশ্নের উত্তরের প্রয়োজনীয় ডেটা হয়তো আপনি পাবেন না। কিন্তু প্রশ্নটি করতে হবে। কেননা প্রশ্ন না করলে, আপনি ডেটাই খুঁজবেন না।

ডেটা থেকে প্রশ্নের উত্তর যখন আপনি পেয়ে যাবেন, তখন ডেটা ও তার বিশ্লেষণের বিষয়ে আপনার একটু পূর্ণ ধারণা আসবে। এরপর প্রতিবেদন লেখার কাজ শুরু করতে হবে। তবে তার আগে আপনার দর্শক বা পাঠক সম্পর্কে আপনার ধারণা থাকা প্রয়োজন।

তাই পাঠক বা দর্শক সম্পর্কে কিছু প্রশ্নের উত্তর আগে জেনে নিন। কেননা আপনার প্রতিবেদন লেখার ধরন নির্ধারিত হবে পাঠক বা দর্শকের শ্রেণী, পেশা, বয়স ও শিক্ষাগত যোগ্যতার ওপর নির্ভর করে।

পাঠক বা দর্শক

- দর্শক বা পাঠক কারা?
- তারা কী ধরনের এবং কী পরিমাণ তথ্য গ্রহণ বা ধারণ করতে পারবেন?
- কোন মাধ্যম ব্যবহার করে তারা প্রতিবেদনটি দেখবেন বা পড়বেন?
- কী ধরনের চিত্রায়ণ প্রতিবেদনে যুক্ত করা হচ্ছে?
- দর্শক বা পাঠকের কাছ থেকে কী ধরনের প্রতিক্রিয়া আশা করা হচ্ছে?

ডেটার সহজ ব্যাখ্যা

ডেটা প্রতিবেদন পাঠকের কাছে সহজ করে উপস্থাপনের সবচেয়ে কার্যকর পদ্ধতি হচ্ছে সংখ্যাকে মানুষের জীবনের সাথে মিলিয়ে বিশ্লেষণ করা। সংখ্যাকে এমনভাবে তুলে ধরতে হবে, যেন পাঠকের কাছে তা দুর্বোধ্য মনে না হয়। আর লেখার ভাষা হতে হবে, পরিষ্কার, সহজ, সরাসরি ও মানবিক। ডেটাকে সহজ করে উপস্থাপনের বেশ কিছু কৌশল বিশ্বের বড় বড় সব সংবাদকক্ষে অনুসরণ করা হয়। সেগুলো নিচে উল্লেখ করা হলো–

পূর্ণসংখ্যা

সংখ্যাতাত্ত্বিক বিশ্লেষণের অনেক ক্ষেত্রেই ফলাফলে ভগ্নাংশ আসে। কিন্তু প্রতিবেদন লেখার সময় ভগ্নাংশ উল্লেখের কোনো প্রয়োজন নেই। কেননা নির্দিষ্ট ঐ সংখ্যা পাঠকের কাছে আলাদা করে কোনো অর্থ তৈরি করে না। পাঠক তা মনে রাখতে পারে না। তাই এমন সংখ্যা লিখুন যা পাঠক মনে রাখতে পারবে। এ জন্যে ভগ্নাংশকে পূর্ণ সংখ্যায় লিখে প্রায় যুক্ত করুন। যেমন, চলতি অর্থবছরের বাজেটে সরকার উন্নয়ন ব্যয়ে বরাদ্দ করেছে **২,৫৯,৬১৭** কোটি টাকা। এমন একটি বড় সংখ্যা প্রতিবেদনে উল্লেখ করলে তা পড়তে পাঠককে বেগ পেতে হবে বা মনে রাখতে পারবে না। কিন্তু আপনি যদি বলেন প্রায় ২ লাখ ৬০ হাজার কোটি টাকা প্রায় সব ধরনের পাঠক বা দর্শকই তা মুহূর্তে বুঝতে পারবে এবং মনে রাখতে পারবে।

হার

শতকরা হিসাবকে হারে প্রকাশ করা হয়। কখনো তাকে শতাংশ বলা হয়, কখনো বা ভাগ। দুটো একই হিসাব নির্দেশ করে। কোনো বড় সংখ্যা থেকে কিছু অংশ আলাদা করে বোঝানোর ক্ষেত্রে এই শতকরা হিসাব ব্যবহার করা হয়। কেননা, সরাসরি সংখ্যায় বললে পাঠকের কাছে তা আরো কঠিন মনে হয়। যেমন আপনি যদি বলেন, চলতি বছর ৬ লাখ ৭৮ হাজার কোটি টাকার বাজেটের মধ্যে ২ লাখ ৬০ হাজার কোটি টাকা উন্নয়ন ব্যয়ে বরাদ্দ করেছে সরকার। এই তথ্য পাঠকের কাছে খুব একটা বোধগম্য হয় না। কিন্তু আপনি যদি শতকরা হিসাব বের করেন, তবে আপনি দেখতে পারবেন উন্নয়ন ব্যয়, মোট বাজেটের ৩৮ ভাগ। তবে আপনার এমন দর্শক বা পাঠকও থাকতে পারেন, যার এই শতাংশের হিসেবে ধারণা নেই। তাই আপনি যদি বলেন, চলতি বছর সরকার যত টাকা ব্যয় করবে, তার প্রতি ১০০ টাকার ৩৮ টাকা ব্যয় হবে উন্নয়ন খাতে। তাহলে, তা বুঝবে না এমন কেউ থাকবে না। শতকরা হারকে সহজে উপস্থাপনের আরেকটি উদাহরণ হতে পারে এমন, ধরুন আপনি প্রতিবেদনে লিখেছেন দেশের উপকূলীয় ৪৮ শতাংশ নারী জরায়ুসংক্রান্ত অসুস্থতায় ভুগছেন। এর বদলে আপনি যদি লিখেন, উপকূলের প্রতি দুইজন নারীর একজন জরায়ুসংক্রান্ত অসুস্থতায় ভুগছেন, তাহলে তা শুধু দর্শককে বুঝতেই সহায়তা করবে না, রোগের ভয়াবহতার মাত্রাকেও নির্দেশ করবে।

অনুপাত

শতকরা হারের মতো অনুপাতও একই হিসাব প্রকাশ করে। এই হিসাবও শতাংশ অথবা ভাগে প্রকাশ করা হয়। ধরুন দেশের পর্যটন শিল্প গেল বছর ৭৬ হাজার ৬৯০ কোটি টাকার ব্যবসা করেছে। অর্থাৎ দেশের অর্থনীতিতে এ খাতের অবদান ছিলো ৭৬ হাজার ৬৯০ কোটি টাকা। কিন্তু সেটি দেশের অর্থনীতিতে বা উৎপাদনে কতটুকু অবদান রাখছে, তা বোঝাতে আপনাকে মোট দেশজ উৎপাদন বা জিডিপির সঙ্গে পর্যটনের অনুপাত হিসাব উপস্থাপন করতে হবে।

বিষয়বস্তু

বাক্যের সাবজেক্ট বা বিষয়বস্তু হিসেবে মানুষ বা কোনো ব্যক্তিকে উপস্থাপন করা, পরিসংখ্যান কে নয়। যখন মানুষকে বাক্যের সাবজেক্ট হিসেবে সামনে আনবেন, তখন ক্রিয়াপদ ব্যবহার করা সহজ হবে। যেমন, স্বাস্থ্য অধিদপ্তর থেকে তথ্য এলো, মাতৃমৃত্যু ২০১২ সালের ৫০ দশমিক ৯ শতাংশ থেকে কমে ২০২২ সালে ২৫ শতাংশ হয়েছে। এই তথ্যটিতে সাবজেক্ট মৃত্যু। কিন্তু আপনি যদি বলেন গত এক দশকে প্রসবকালীন সময়ে নারীর মৃত্যু দ্বিগুণ কম। এখানে সাবজেক্ট নারী, একজন মানুষ। এটি পাঠকের কাছে বেশি সহজে পৌঁছাবে।

তুলনা

সব ধরনের প্রতিবেদনকে সহজবোধ্য করতে তুলনা প্রকাশ করা হয়। বিশেষ করে যেখানে সংখ্যা থাকে। ডেটা সাংবাদিকতার ক্ষেত্রে এমন তুলনার প্রচলন বেশি। যেমন, আপনার অনুসন্ধানে দেখা গেল বছরে দেশ থেকে ৯০ হাজার কোটি টাকা পাচার হচ্ছে অথবা ঘুষ বা অবৈধ লেনদেন হচ্ছে। এই টাকাটা কত বড় পরিমাণ তা বোঝাতে আপনি তুলনা ব্যবহার করে বলতে পারেন, এই টাকা দিয়ে তিনটি পদ্মা সেতু নির্মাণ করা সম্ভব। এর ফলে পাঠকের সামনে সংখ্যার বিশালত্বটা দৃশ্যমান হবে।

সহজ ও বোধগম্য ভাষা

সংবাদ প্রতিবেদন লেখার যে মৌলিক নিয়ম, ডেটা সাংবাদিকতার বেশিরভাগ ক্ষেত্রে সেগুলো অনুসরণ করলেই একটি শক্তিশালী ও সহজ পাঠযোগ্য প্রতিবেদন তৈরি করা সম্ভব। তাই শব্দ ব্যবহার ও বাক্য গঠনে, সহজ ও বোধগম্যতা যাচাই করে নেয়া অবশ্যই প্রয়োজন।

ডেটা প্রতিবেদন লেখার কাঠামো

ইন্টারনিউজের জন্য ডেটা সাংবাদিকতা শেখার একটি ম্যানুয়াল[৪৫] তৈরি করেছেন ইভা কন্সট্যানটারাস। সেখানে ডেটা প্রতিবেদন লেখার ৪টি প্রচলিত কাঠামোর কথা উল্লেখ করা হয়েছে। দেশে দেশে ডেটা সাংবাদিকতার ক্ষেত্রে এই কাঠামোগুলোর যে কোনো একটি অনুসরণ করে প্রতিবেদন লেখা বা সাজানো হয়। তবে কোন ক্ষেত্রে কোন ধরনের কাঠামো অনুসরণ করা হবে, তা পুরোটাই নির্ভর করে ডেটা ও সংবাদের ওপর।

১. তুলনা বা বৈপরীত্য

ডেটা বিশ্লেষণে যখন বিপরীতমুখী দুইটি পৃথক ফলাফল থাকে, তখন এই কাঠামো ব্যবহার করা হয়। বিশ্লেষণের ওপর ভিত্তি করে যখন প্রতিবেদনে কোনো গোষ্ঠীর অসংগতি তুলে ধরা হবে, তখন গোষ্ঠীর বিষয়টিও এই কাঠামোতে উল্লেখ করা হয়। যাতে দর্শক বা পাঠক দুইটি বিপরীতধর্মী পরিস্থিতি পাশাপাশি দেখে ঘটনার ব্যপকতা সম্পর্কে ধারণা পায়। সেই সাথে কি কারণে এমন ঘটনা ঘটে, তা বুঝতে পারে। কেননা দুইটি ভিন্ন অবস্থানে সংশ্লিষ্ট পক্ষগুলোর সিদ্ধান্ত বা কর্মকাণ্ড আলাদা।

বিপরীতমুখী বৈশিষ্ট্য

দুই পক্ষের প্রধান এবং চমকপ্রদ বৈশিষ্ট্যগুলো দিয়ে গল্পের শুরু করতে হবে। কিন্তু এক পক্ষের বৈশিষ্ট্য অন্য পক্ষের থেকে একদম ভিন্ন হতে হবে। যেমন: একই অবস্থান থেকে একসঙ্গে শুরু করা দুটি উদ্যোগের গল্প, যার একটি সফল হয়েছে, অন্যটি ব্যর্থ।

অভিন্ন ইতিহাস

দুইটি উদ্যোগ বা বিষয়ের ইতিহাস ও উৎপত্তি এক হতে হবে।

ভিন্ন সিদ্ধান্ত

দুইটি পৃথক উদ্যোগে নেয়া সিদ্ধান্ত বা কর্মকাণ্ড বিপরীতমুখী হতে হবে।

ধারাবাহিকতা

ঐ বিপরীতমুখী সিদ্ধান্তের কারণে কীভাবে একটি উদ্যোগ সফল হয়েছে আর অন্যটি ব্যর্থ হয়েছে, তার ধারাবাহিকতা বর্ণনা করতে হবে।

ভিন্নতার কারণ

একই অবস্থান থেকে শুরু করে কীভাবে দুইটি পক্ষের সিদ্ধান্ত আলাদা হলো, তা উল্লেখ করতে হবে।

সমাধান

দুই পক্ষ বা উদ্যোগকে আবার একই জায়গায় নিয়ে আসার সম্ভাব্যতা যাচাই, সম্ভাব্য ফলাফল ও সংশ্লিষ্ট উদাহরণ তুলে ধরতে হবে।

---

[৪৫] *ARIJ hosts Data Journalism Training Manual (in Arabic) - Arab Reporters for Investigative Journalism (ARIJ).* (2020b, November 30). Arij | أريج. Retrieved October 12, 2022, from https://en.arij.net/materials/arij-hosts-data-journalism-training-manual-in-arabic/

২. ধারাক্রম

কোনো জটিল বিষয়ে প্রতিবেদন তৈরির ক্ষেত্রে এই কাঠামোটি অনুসরণ করা হয়। কেননা, এসব প্রতিবেদনে প্রতিটি ডেটা এক একটি ভিন্ন ভিন্ন সমস্যার চিত্র তুলে ধরে এবং একটি ডেটার সঙ্গে অন্য ডেটাটি ধারাবাহিকভাবে সম্পর্কযুক্ত থাকে। ফলে একটি সমস্যার সঙ্গে অন্য আরেকটি সমস্যা সম্পর্কিত থাকে। তাই সেগুলো ধারাবাহিকভাবে বিশ্লেষণ বা চিত্রায়ণ করে এগোলেই কেবল দর্শক বা পাঠক প্রতিবেদনটি সহজে বুঝতে পারে।

সংবাদ সূচনা

বিশ্লেষণে বের হয়ে আসা একাধিক অসংগতি একবারে সংক্ষিপ্ত আকারে প্রথমেই তুলে ধরতে হবে।

৩. প্রেক্ষাপট

প্রতিবেদনের তথ্যসূত্র, সংশ্লিষ্ট পুরোনো তথ্য ও ইতিহাস থাকবে দ্বিতীয় অংশে

ধাপে ধাপে ব্যাখ্যা

ঘটনা বা বিষয়বস্তুর ধাপভিত্তিক বিশ্লেষণ করতে হবে প্রতিবেদনের এই অংশে। এই ধারাবাহিক বর্ণনা পাঠককে গল্পের শেষ পর্যন্ত নিয়ে যাবে। প্রতিটি ধাপে উদাহরণ এবং সংশ্লিষ্ট সাক্ষাৎকারও যুক্ত করতে হবে।

সমাধান

গল্পের শেষ হবে সম্ভাব্য সমাধানের উপায় ও বিকল্পগুলো নিয়ে। এক্ষেত্রে সমস্যা যে ধাপে এগিয়েছে, সমাধানটাও সেভাবে পর্যায়ক্রমে দেখানো হবে। এ সংশ্লিষ্ট একটি উদাহরণও তুলে ধরতে হবে।

৪. ব্যাখ্যামূলক

গল্প বলার এই পদ্ধতিতে পাঠক বা দর্শককে সংবাদ শুধু জানানো নয়, বিস্তারিত বোঝানো হয়। কীভাবে, কোন কোন ধাপ পেরিয়ে সংবাদটি সংবাদে পরিনত হয়েছে সেই ধারণা এই ধরনের প্রতিবেদনে বলা হয়।

১। পাঠকদের খবর বোঝাতে সাহায্য করুন।

২। প্রতিবেদনের আলোচনাকে প্রাসঙ্গিক রাখুন ।

৩। গ্রাফিক বা অ্যানিমেটেড ভিডিওর মাধ্যমে ব্যাখ্যা তৈরি করুন ।

৪। প্রতিবেদনে নিম্নলিখিত বিষয়গুলো থাকলে পার্থক্য বুঝতে সুবিধা হবে–

- ঘটনাটা কি ?
- এটি কীভাবে ঘটেছে?
- বিষয়টি কতটা খারাপ বা ভাল?
- ঘটনার সঙ্গে পাঠক কীভাবে সম্পর্কিত??
- এর থেকে আমরা কী শিখতে পারি?

ডেটা প্রতিবেদনের ইন্ট্রো বা সংবাদ সূচনা

ডেটা প্রতিবেদনের প্রথম বাক্যটি প্রতিবেদনের বাকি আর সব বাক্যের থেকে বেশি গুরুত্বপূর্ণ। একটি ভালো ইন্ট্রো বা সংবাদ সূচনার ওপর নির্ভর করে পাঠক আপনার প্রতিবেদনটি পড়তে আগ্রহী হবে কি হবে না। আর এ কারণেই ইন্ট্রো বা সংবাদ সূচনার ভাষা ও বাক্য গঠন, সহজ, সাবলীল ও প্রাসঙ্গিক হতে হয়। আপনার প্রতিবেদনের সবচেয়ে গুরুত্বপূর্ণ অনুসন্ধান বা তথ্যটি অবশ্যই সংবাদের সূচনায় থাকতে হবে। ইন্ট্রো লেখার সময়ে আরও কয়েকটি বিষয় গুরুত্ব দিতে হয়–

১) ডেটা প্রতিবেদনে সংখ্যা আর সংখ্যার বিশ্লেষণ থাকলে, সংবাদ সূচনা লেখার সময় সংবাদ বা ঘটনার বিষয়কে গুরুত্ব দিতে হবে, সংখ্যাকে নয়।

২) মানুষের ওপর সংখ্যার প্রভাব কী, সেটি সংবাদ সূচনায় থাকতে হবে।

৩) মানবিক গল্প বলতে হবে, পরিসংখ্যান নয়।

৪) পাঠক বা দর্শককে চমক দিতে হবে।

৫) বাক্য এমন হবে, যেন প্রতিবেদনের শেষ পর্যন্ত পাঠকের আগ্রহ থাকে।

ডেটা সাংবাদিকতায় সাধারণত চার ধরনের সংবাদ সূচনা বা ইন্ট্রো লেখা হয়

সংখ্যানির্ভর

এই ধরনের সংবাদ সূচনায় সংখ্যা দিয়ে প্রতিবেদনের প্রথম বাক্য শুরু হয়। তবে একইসাথে সংখ্যার মাধ্যমে পাঠক যেন ঘটনার প্রভাবটা বুঝতে পারে, তা নিশ্চিত করতে হয়। সংখ্যা প্রত্যক্ষ অথবা পরোক্ষ দুইভাবেই উল্লেখ করা যেতে পারে। বাক্যের গঠনের ওপর নির্ভর করবে সংখ্যা কীভাবে উপস্থাপন করা হবে। একইসঙ্গে সংখ্যার উপস্থাপনে একটি ধারাবাহিকতা রাখতে হবে। যাতে পাঠক সংখ্যার সঙ্গে ঘটনার যোগসূত্র খুঁজে পায় এবং পরবর্তী সংখ্যাতাত্ত্বিক বিশ্লেষণ পড়তে আগ্রহী হন।

গল্পনির্ভর

একটি মানবিক গল্প দিয়ে এই সংবাদ সূচনার শুরু হয়। যেখানে প্রতিবেদনের বিষয়টি কীভাবে সাধারণ মানুষের জীবনে প্রভাব ফেলছে, তা তুলে ধরা হয়। বিশেষ করে, ঐ সব ভুক্তভোগীর জীবনের একটি মুহূর্ত সহজ ও প্রাঞ্জল ভাষায় তুলে ধরতে হবে। যাতে পাঠক সংবাদ সূচনাটি পড়ে বিষয়ের প্রভাব ও ব্যাপকতা কল্পনা করতে পারে।

> গল্প বাছাইয়ের ক্ষেত্রে এমন মানুষের গল্প নিতে হবে যিনি আপনার ডেটাসেটের অংশ। অর্থাৎ ডেটাসেটকে পুরোপুরিভাবে প্রতিনিধিত্ব করে এমন ব্যাক্তির গল্প নির্বাচন করতে হবে। এক্ষেত্রে সবচেয়ে ভালো হয়, এমন কারো জীবনের গল্প বলা, যিনি প্রতিবেদন-সংশ্লিষ্ট বিষয়টির প্রথম থেকে শেষ পর্যন্ত জড়িত থাকে এবং তার গল্পের মাধ্যমে পুরো প্রতিবেদনটিকে তুলে ধরা।

দৃশ্যনির্ভর

এই ধরনের সংবাদ সূচনায় প্রতিবেদককে তার লেখার মাধ্যমে ঘটনার বাস্তব চিত্র ফুটিয়ে তুলতে হয়। ঘটনাস্থল, সময় ও সংশ্লিষ্ট চরিত্রগুলোকে এমনভাবে উপস্থাপন করতে হয়, যেন পাঠক গল্পের শুরুতেই ঘটনার সঙ্গে নিজেকে সম্পৃক্ত মনে করতে পারে এবং সেখানকার পরিস্থিতি কেমন ছিলো তা কল্পনা করতে পারে। যে ধরনের ডেটা প্রতিবেদনে পরিবেশ ও ঘটনাস্থল খুব গুরুত্বপূর্ণ সেখানে এই ধরনের সংবাদ সূচনা কার্যকর ভূমিকা রাখে।

সারাংশনির্ভর

সারাংশনির্ভর সংবাদ সূচনা প্রতিবেদনের মূল ধারণা বা আইডিয়া সম্পর্কে পাঠককে অবহিত করে এবং এর সংবাদ মূল্য বুঝতে সহায়তা করে। এই ধরনের সংবাদ সূচনা মূলত সোজাসাপ্টা সংবাদের ক্ষেত্রে ব্যবহৃত হয়। এক্ষেত্রে প্রতিবেদনে উল্টো পিরামিড কাঠামো অনুসরণ করা হয়। যেখানে ডেটা বিশ্লেষণের সবচেয়ে গুরুত্বপূর্ণ তথ্য সংবাদ সূচনায় উল্লেখ করতে হয়।

ডেটা প্রতিবেদনে সহজ আর সাবলীল গল্প বলার দক্ষতা বৃদ্ধিতে তিরিসা অ্যান্ডেরুডের কার্যকর সাতটি পরামর্শ জানতে, কিউআর কোডটি স্ক্যান করুন।

ডেটাভিত্তিক প্রতিবেদনে সাক্ষাৎকার

নিয়মিত সাংবাদিকতার সাক্ষাৎকার থেকে ডেটা সাংবাদিকতায় করা সাক্ষাৎকার আলাদা হতে পারে। নিয়মিত প্রতিবেদনের ক্ষেত্রে অনেক সময় সংবাদ বা ঘটনা কভার করতে গিয়ে তাৎক্ষণিকভাবে সাক্ষাৎকার নিতে হয়। প্রতিবেদককে সোর্সের ওপর বেশি নির্ভর করতে হয়, যেহেতু তিনি বিষয়টি সম্পর্কে সামান্য ধারণা রাখেন। তবে ডেটা সাংবাদিকতায় প্রতিবেদক যে বিষয়ে সাক্ষাৎকার নেবেন, তিনি ঐ বিষয়ে একরকম বিশেষজ্ঞে পরিণত হন, সাক্ষাৎকার নেবার আগেই। কেননা বিশাল ডেটাসেট দীর্ঘ সময় ধরে বিশ্লেষণের কারণে, সংশ্লিষ্ট বিষয়ে সম্পূর্ণ ধারণা থাকে তার। সাক্ষাৎকারের মাধ্যমে শুধু ঘটনার ব্যাপকতাকে বিশ্লেষণ করা হয়। সংশ্লিষ্টদের দৃষ্টি আকর্ষণ করা হয়। আর তাই ডেটা সাংবাদিকতায় সাক্ষাৎকার সবসময় ডেটা বিশ্লেষণের পর করা হয়।

ডেটা প্রতিবেদনের সাক্ষাৎকারের উদ্দেশ্য	
১	ডেটা বিশ্লেষণে চিহ্নিত সমস্যা মানুষের জীবনে যে প্রভাব ফেলছে তা তুলে আনা।
২	ডেটা বিশ্লেষণে চিহ্নিত সমস্যার কারণ পর্যালোচনা।
৩	সমস্যার জন্য দায়ীদের জবাবদিহি আওতায় আনা।
৪	একই ঘটনায় জড়িত দুইপক্ষকে সামনে নিয়ে আসা ।
৫	ডেটানির্ভর সমাধান খুঁজে বের করা।
৬	আরও ডেটা খোঁজা।

ডেটা সাংবাদিকতায় চার ধরনের সাক্ষাৎকার রয়েছে

প্রভাব-সংক্রান্ত

প্রতিবেদন সংশ্লিষ্ট ঘটনায় যারা আক্রান্ত বা ক্ষতিগস্ত, তাদের সাক্ষাৎকার প্রভাব-সংক্রান্ত সাক্ষাৎকার হিসেবে বিবেচিত হয়। এসব সাক্ষাৎকার প্রতিবেদন তৈরির প্রথম পর্যায়ে নেয়া হয় এবং প্রতিবেদন লেখার সময়েও এসব সাক্ষাৎকার প্রথমদিকে উপস্থাপন করা হয়। যেখানে সাক্ষাৎকার প্রদানকারীরা, ডেটাসংশ্লিষ্ট বিষয়টি তাদের জীবনে কীভাবে প্রভাব ফেলছে, তা তুলে ধরেন বা জানান। একটি ডেটা কীভাবে অন্যটির সঙ্গে সম্পর্কযুক্ত এবং ঘটনার পেছনের কারণগুলো কীভাবে পর্যায়ক্রমে মানুষের জীবনকে প্রভাবিত করে তা এই ধরনের সাক্ষাৎকারে জানানো হয়। একইসঙ্গে এই সাক্ষাৎকারের মাধ্যমে, ঘটনার জন্যে কারা দায়ী, তাও উন্মোচিত হয়। তাই প্রভাব সংক্রান্ত প্রতিবেদনগুলো ডেটা সাংবাদিকতার জন্যে খুবই গুরুত্বপূর্ণ। যার সাহায্যে প্রতিবেদনে পাঠক বা দর্শকের দৃষ্টি আটকে রাখা যায়।

ব্যাখ্যামূলক

প্রতিবেদনে যখন বিষয়সংশ্লিষ্ট বিশেষজ্ঞ বা বিশ্লেষকের মতামত ও পর্যবেক্ষণ নেয়া হয়, তখন সেটি ব্যাখ্যামূলক সাক্ষাৎকার। এই সাক্ষাৎকারে বিশেষজ্ঞরা আপনার ডেটা বিশ্লেষণে পাওয়া ফলাফলের কারণ ব্যাখ্যা করেন। কেন ঘটনাটি ঘটছে বা এর ভবিষ্যৎ কী? তা এই সাক্ষাৎকারের মাধ্যমে জানা যায়। ব্যাখ্যামূলক সাক্ষাৎকারের ক্ষেত্রে সাধারণত সরকার, বেসরকারি প্রতিষ্ঠানের প্রতিনিধি, স্বাধীন বিশ্লেষক, গবেষক এবং শিক্ষকদের বিবেচনা করা হয়। এক্ষেত্রে সাক্ষাৎকারের পাশাপাশি, তাদের ব্যাখ্যার পেছনে সংশ্লিষ্ট ডেটাও সংগ্রহ করতে হবে।

জবাবদিহিমূলক

প্রতিবেদন-সংশ্লিষ্ট ঘটনার জন্য যারা দায়ী কিংবা যাদের দায়িত্ব বিষয়গুলো দেখভাল করা, তাদের থেকে জবাবদিহিমূলক সাক্ষাৎকার নেয়া হয়। এই সাক্ষাৎকারটি প্রতিবেদনের জন্য অত্যন্ত প্রয়োজনীয়। এই জবাবদিহিমূলক সাক্ষাৎকার ছাড়া প্রতিবেদনটি অসম্পূর্ণ থাকে। কর্তৃপক্ষ কিংবা দোষী পক্ষের বক্তব্য নেওয়ার সময় তাদের সামনে তথ্য-প্রমাণ ও ডেটা বিশ্লেষণ তুলে ধরুন এবং তাদের আত্মপক্ষ সমর্থনের পর্যাপ্ত সুযোগ দিন। এই ধরনের সাক্ষাৎকারে সংশ্লিষ্ট ব্যাক্তি যখন তার যুক্তি তুলে ধরবেন, তখন সেই যুক্তি প্রমাণের জন্য প্রয়োজনীয় ডেটা, তথ্য-প্রমাণ উপস্থাপন বা সরবরাহ করতে বলতে হবে। একইসঙ্গে এইসব সাক্ষাৎকারে যে সব সংখ্যা উল্লেখ করবে, তা যাচাই করে প্রচার বা প্রকাশ করতে হবে। কেননা তিনি ইচ্ছাকৃতভাবে ভুল তথ্য দিয়ে প্রতিবেদককে বিভ্রান্ত করতে পারেন। এমন সাক্ষাৎকারে open-ended প্রশ্ন করা যাবে না। ডেটা উল্লেখ করে সরাসরি উত্তর দিতে হয়, এমন প্রশ্ন করুন।

সমাধানমূলক

যেসব ব্যক্তি বা প্রতিষ্ঠান সংশ্লিষ্ট বিষয়ের সমস্যা বা ত্রুটি সমাধানে কাজ করে তারা সমাধানমূলক সাক্ষাৎকার দিয়ে থাকেন। এই সাক্ষাৎকারের জন্য এমন ব্যাক্তি খুঁজে বের করতে হবে যিনি আপনার প্রতিবেদনে চিহ্নিত সমস্যার একটি দীর্ঘমেয়াদি সমাধানের রূপরেখা দিতে পারবেন। এ ক্ষেত্রে এমন ব্যক্তি বা প্রতিষ্ঠানের সঙ্গে কথা বলা উচিত যারা সরাসরি মাঠ পর্যায়ে এমন সমস্যা সমাধানের কাজের সাথে সম্পৃক্ত। এ ক্ষেত্রে এমন ব্যক্তিদের সাথে কথা বলা যায়, যারা আগে একই ধরনের সমস্যায় আক্রান্ত হয়েছিল এবং এখন তা কাটিয়ে উঠেছে। এমন ব্যক্তিদের সাক্ষাৎকারের মাধ্যমে দুইটি ভিন্ন সময়ে সমস্যাটির প্রকটতা সম্পর্কে ধারণা পাওয়া যায়। একইসঙ্গে তারা কীভাবে সেই সমস্যা কাটিয়ে উঠেছেন সে অভিজ্ঞতাও জানা যায়।

শক্তিশালী সাক্ষাৎকারের জন্য	
১	এমন প্রশ্ন করুন যেন সংখ্যা দিয়ে উত্তর দেয়া যায়।
২	প্রতিবেদনের পূর্বানুমান প্রমাণ কিংবা তা ভুল প্রমাণ করার উদ্দেশ্য নিয়ে প্রশ্ন তৈরি করুন।
৩	যে প্রশ্ন করবেন, নিশ্চিত করুন তার সঠিক উত্তর আপনার আগেই জানা আছে।
৪	প্রশ্নকে কয়েকটি ধাপ বা ক্যাটাগরিতে ভাগ করুন। প্রত্যেক ক্যাটাগরিতে কিছু প্রশ্ন রাখুন।
৫	প্রশ্ন বা প্রতিবেদন-সংক্রান্ত প্রয়োজনীয় তথ্য বা নথি সংগ্রহ করুন এবং একটি সাক্ষাৎকার নেয়ার সময় সাথে রাখুন।

ডেটাকে প্রশ্ন করে মানবিক প্রতিবেদন তৈরির আরো কৌশল জানতে ইভা কন্সট্যানটারাসের সাক্ষাৎকারটি পড়তে ও শুনতে কিউআর কোডটি স্ক্যান করুন।

কেস স্টাডি

বাংলাদেশে ৯৭ ভাগ নারী নির্যাতনের মামলায় কোনো সাজা হয়নি। নারী ও শিশু নির্যাতন মামলা নিয়ে প্রথম আলোয় প্রকাশিত আলোচিত এই সিরিজ[৪৬] বিপুল পরিমাণ ডেটা সংগ্রহ ও বিশ্লেষণ করা হয়। কিন্তু এই সিরিজটির অন্যতম বৈশিষ্ট্য ছিল লিখনশৈলী। ডেটানির্ভর স্টোরিকে কীভাবে প্রথাগত রিপোর্টিংয়ের সঙ্গে মিলিয়ে একটি মানবিক প্রতিবেদনে পরিণত করতে হয়, একাধিক অ্যাঙ্গেল বের করতে হয়–এ নিয়ে লিখেছেন কুর্‌রাতুল আইন তাহমিনা।

অনুসন্ধানের জন্মকথা

প্রথম আলোর প্রতিবেদক আসাদুজ্জামান। কাজের সূত্রে পুরান ঢাকার জজকোর্ট এলাকায় তাঁর দিন কাটে। সেখানেই নারী ও শিশু নির্যাতন দমন আইনে আসা মামলাগুলোর ট্রাইব্যুনাল বা বিশেষ আদালত বসে। সেসব অপরাধের ভুক্তভোগী বা তাদের আত্মীয়-পরিজনকে প্রতিদিন বিচারের আশায় দিনের পর দিন ঘুরতে দেখেন আসাদ। তাদের নিরুপায় আক্ষেপ আর হয়রানির কথা শোনেন। অন্যদিকে পুলিশ বা আইনজীবীদের মুখে শোনেন এসব মামলার বড় অংশই ভুয়া।

> তখন মহানগরসহ ঢাকা জেলার জন্য ট্রাইবুন্যাল ছিল পাঁচটি। এই পাঁচটি বিশেষ আদালতকক্ষে রাখা বছরওয়ারি বিচারিক নিবন্ধন খাতার পাতা উল্টে আসাদ দেখেন, ব্যাপকসংখ্যক মামলা চলছে তো চলছেই। নিষ্পন্ন মামলাগুলোর রায়ে আসামির পর আসামি অব্যাহতি অথবা খালাস পাচ্ছে।

[৪৬] https://www.prothomalo.com/bangladesh/%E0%A6%A8%E0%A6%BE%E0%A6%B0%E0%A7%80-%E0%A6%93-%E0%A6%B6%E0%A6%BF%E0%A6%B6%E0%A7%81%E0%A6%B0%E0%A6%BE-%E0%A6%AC%E0%A6%BF%E0%A6%9A%E0%A6%BE%E0%A6%B0-%E0%A6%AA%E0%A6%BE%E0%A7%9F-%E0%A6%A8%E0%A6%BE

ভেতরের কথাটা কী? মামলায় সাজার হার কত? কেন এতো অব্যাহতি আর খালাস? ধর্ষণ বা হত্যার মতো গুরুতর অপরাধে করা মামলাগুলো কতখানি মিথ্যা হতে পারে? মিথ্যা মামলার অভিযোগের সত্যতা কী? তা ছাড়া, এতো মামলা কেন ঝুলে থাকছে?

এসব প্রশ্নের উত্তর, গোনাগাঁথা হিসাব বা সুনির্দিষ্ট ব্যাখ্যা-প্রমাণের হদিস মেলে না। এদিকে নিত্যদিন দেখা ভুক্তভোগীদের আকুতি আসাদকে পীড়িত করে। তিনি নিবন্ধন খাতা থেকে ছয়টি গুরুতর অপরাধের মামলার তথ্য টোকা শুরু করেন। মামলার নথিপত্র সংগ্রহ করতে থাকেন। সরকারি উকিলদের সাক্ষাৎকার নিতে থাকেন। হাতে টালি করে হিসাবপত্র করতে চেষ্টা করেন।

কিন্তু কাজটা কঠিন হয়। গল্পটা বলবেনই বা কীভাবে? অতঃপর আসাদ তাঁর মোটসোটা ডায়েরিটা নিয়ে আমার কাছে হাজির হন। যে প্রতিবেদনের ধারণা জন্ম নিয়েছিল আসাদের মনের একান্ত তাগিদে, সেটা ক্রমে রূপ নেয় এক গন্ধমাদন পর্বতে। উপাত্তভিত্তিক এই অনুসন্ধানী কাজ চলে টানা প্রায় দুই বছর। তারপরও কিন্তু কাজ শেষ হয় না।

গুছিয়ে তথ্য-উপাত্ত সংগ্রহ

আসাদ আমার কাছে এসেছিলেন ২০১৬ সালের আগস্টের মাঝামাঝি। তখন তাঁর মৌলিক তথ্য সংগ্রহের ধারাটি গুছিয়ে দিলাম। ছয়টি গুরুতর অপরাধের মামলা দেখার সিদ্ধান্ত চূড়ান্ত হলো– ধর্ষণ, ধর্ষণজনিত মৃত্যু, দলবদ্ধভাবে ধর্ষণ ও তার ফলে ভুক্তভোগী নারী বা শিশুর মৃত্যু অথবা আহত হওয়া, সম্ভ্রমহানি করে আত্মহত্যায় প্ররোচনা, যৌনপীড়ন এবং যৌতুকের জন্য হত্যা বা হত্যাচেষ্টা।

প্রতিটি মামলার কী কী তথ্য আসাদ নিবন্ধন খাতা থেকে নেবেন, তার ছক চূড়ান্ত করলাম। তথ্যগুলো হচ্ছে– ট্রাইব্যুনালের নম্বর; ট্রাইব্যুনালে মামলা আসার বছর; ট্রাইব্যুনালে মামলার নম্বর; থানার নাম ও মামলা দায়েরের বছরসহ থানায় মামলার নম্বর এবং সরাসরি ট্রাইব্যুনালে আসা মামলাসহ মামলার উৎস; আইনে অপরাধের ধারা; আসামির সংখ্যা; মামলার তৎকালীন অবস্থা; নিষ্পন্ন হয়ে থাকলে সেটার সাল-তারিখ; সাজা হয়ে থাকলে সেটার বিবরণ।

এভাবে ২০১৬ সালের আগস্টের শেষ ভাগে আসাদ ভাগগুলো সুনির্দিষ্ট করে ট্রাইব্যুনাল ধরে বছরওয়ারি নিবন্ধন খাতাগুলো থেকে শুমারির ধাঁচে তথ্য-উপাত্ত টোকা শুরু করলেন। ভিত্তিবছর ধরা হয়েছিল ২০০২ সালকে। কেন?

১৯৮৩ সালে ফৌজদারি দণ্ডবিধির বাইরে নারী নির্যাতন (নিবর্তক শাস্তি) অধ্যাদেশ নামের একটি বিশেষ আইন হয়েছিল। ১৯৯৫ সালে এসেছিল নারী ও শিশু নির্যাতন (বিশেষ বিধান) আইন। অবশেষে ২০০০ সালের ১৪ ফেব্রুয়ারি ভালোবাসা দিবসে জন্ম নেয় নারী ও শিশু নির্যাতন দমন আইন।

১৯৯৫ সালের আইনে বিশেষ আদালত গঠিত হয়েছিল। সর্বশেষ আইনে প্রতি জেলায় অন্তত একটি করে "নারী ও শিশু নির্যাতন দমন ট্রাইব্যুনাল" স্থাপনের বিধান হলো। আগের আইনে করা চলমান মামলাসমেত বিশেষ আদালতগুলো এই ট্রাইব্যুনালে রূপ নিলো।

> ঢাকা জেলার ১ নম্বর ট্রাইব্যুনালটি আদিতে গঠিত হয়েছিল ১৯৯৬ সালে। ২০০২ সালকে ভিত্তিবছর হিসেবে বেছে নেওয়ার কারণ হচ্ছে, নতুন আইনের পর সে বছরই আরও দুটি ট্রাইব্যুনাল যুক্ত হয়। পরের বছর আরও দুটি গঠিত হলে ঢাকার মোট ট্রাইব্যুনাল দাঁড়ায় পাঁচটি।

আদালতের বিচারিক নিবন্ধন খাতা উন্মুক্ত তথ্য। সবাই সেটা দেখতে পারেন। আদালতগুলোয় বিশেষ সূত্রের সহায়তায় আসাদ দিনের কাজের বাইরের সময়টা জুড়ে তথ্য-উপাত্ত তোলার সুযোগ পেয়েছিলেন। ২০১৬ সালের ডিসেম্বরে সে বছরের অক্টোবর পর্যন্ত আসা মামলার তথ্য টোকা শেষ হলো। প্রায় ১৫ বছরের মোটামুটি শুমারি সাঙ্গ হলো। আমরা ক্ষান্ত দিলাম।

চার মাসে আসাদের ছোট ছোট কিন্তু স্পষ্ট হস্তাক্ষরে তিনটি মোটা ডায়েরি ভরে উঠেছিল। পাশাপাশি প্রথম আলোর কম্পোজ বিভাগের দক্ষ সহকর্মীরা তথ্য-উপাত্ত এক্সেলে তুলে প্রাথমিক উপাত্তভাণ্ডার গড়ে তুলেছিলেন।

যাচাই এবং যাচাই

পর্বতটি গন্ধমাদন। নাছোড়বান্দা অধ্যাবসায়ী আসাদুজ্জামান বীর হনুমান। তিনি একা এটা মাথায় করে তুলে এনেছিলেন বটে, কিন্তু আরও কাজ ছিল বেশুমার।

> প্রথম থেকেই উপাত্তভাণ্ডার গোছানোর কাজে যুক্ত হয়েছিলেন প্রথম আলোর প্রশিক্ষণ বিভাগে আমার সহকর্মী প্রণব ভৌমিক। ক্রমে যোগ দিলেন বিভাগের আরেক সহকর্মী আরিফা আলম এবং আমাদের প্রশিক্ষণার্থী ইনটার্ন নাসরিন আক্তার ও মুশফেকা ইসলাম।

আগস্ট ২০১৭ সাল পর্যন্ত নিবিড় তত্ত্বাবধানে এবং আসাদের সহযোগিতায় তাঁরা উপাত্ত যাচাই এবং ঝাড়াই-বাছাইয়ের কাজ করলেন। এ কাজটি হলো দুই ধাপে। প্রথম ধাপে এক্সেলের প্রাথমিক উপাত্তভাণ্ডারটি আসাদের নোটখাতার সঙ্গে মিলিয়ে দেখে ফাঁকগুলো চিহ্নিত করা হলো।

> তারপর শুরু হলো যাচাইয়ের দ্বিতীয় ধাপ। নিবন্ধন খাতা থেকে উপাত্ত টুকেছেন আসাদ একা হাতে। সেটারও তো যাচাই প্রয়োজন। যাচাইবাহিনী এবার আদালতে গিয়ে খাতার সঙ্গে তথ্য মেলালেন। প্রতিটি খটকা, ভুল ও বদলের নোট দিলেন।

তথ্য-সংগ্রহ ও যাচাইয়ের পর্যায়ে কতগুলো ঘাটতি বা দুর্বলতা স্পষ্ট হয়েছিল। যেমন, ট্রাইব্যুনালগুলোতে তথ্য সংরক্ষণের কাজ সব মূলত হাতে লিখে করা হয়।

ঢাউস ব্যবহারজীর্ণ খাতাগুলো নামিয়ে দপ্তর সহকারী বা অন্য কেউ মামলার বিচারিক তথ্য তোলেন। তোলায় ভুল বা ফাঁক নজরে এসেছে। দুটি ট্রাইব্যুনালের একাধিক বছরের খাতার বেশ কিছু পাতা এতটাই ছেড়া যে তথ্য তোলা যায়নি।

অতঃপর আমরা উপাত্তভাণ্ডার সংশোধন করলাম। মামলার সংখ্যা এবং তথ্যে কিছু যোগ-বিয়োগ হলো। শেষতক প্রায় ১৫ বছরে আসা নমুনা মামলার সংখ্যা দাঁড়াল ৭ হাজার ৮৬৪টি।

বিশ্লেষণ তথা অরণ্যে পথ খোঁজা

এক্ষেলে প্রাথমিক উপাত্তভাণ্ডারটি ছিল বাংলায়। সেখানে জড়ো হওয়া তথ্য-উপাত্তের যথাযথ বিশ্লেষণের জন্য সামাজিক বিজ্ঞানের গবেষণায় ব্যবহৃত এসপিএসএস সফটওয়্যার ব্যবহার করা দরকার হয়ে পড়ল। উপাত্তভাণ্ডারটি ইংরেজিতে রূপান্তরিত করা হলো। তথ্য-উপাত্তর সূচকগুলোতে কোড বসাতে হলো। তারপর এসপিএসএস-১৭ সফটওয়্যারে নিয়ে উপাত্তভাণ্ডারের বিন্যাস চূড়ান্ত করা হলো।

> টুকটাক ফাঁক বেরোনো এবং যাচাই ও সংশোধনের কাজ শেষ হচ্ছিলো না। এই পর্যায়ে নোটসহ সংশোধনী চূড়ান্ত করা হতে লাগল আমাদের বিশ্লেষণের আকর ইংরেজি উপাত্তভাণ্ডারে। সম্পাদকের পরামর্শ ও অনুমোদনক্রমে নিযুক্ত দক্ষ দুজন আমাদের চাহিদামতো এ কাজগুলো সম্পন্ন করলেন। তাঁরা চাহিদামতো বিশ্লেষণের ফলাফল বা টেবিলগুলো করে দিলেন।

আমরা প্রত্যেক ধরন বা ভাগ তথা ক্ষেত্রের তথ্য-উপাত্তের প্রাথমিক হিসাব বা ফ্রিকোয়েন্সির ছকগুলো দেখে বিশ্লেষণের রূপরেখা স্থির করলাম। কখনো সর্বোচ্চ তিনটি ক্ষেত্র পর্যন্ত তথ্য-উপাত্ত পরস্পরের সঙ্গে মিলিয়ে বা ক্রস করে দেখলাম। উপাত্তনির্ভর সার্বিক ও বিশেষ চিত্র পাওয়া সম্ভবপর হলো।

উপাত্তে প্রাণসঞ্চার

কিন্তু শুকনো উপাত্ত তো পরিস্থিতি বোঝাবে না। উপাত্তে প্রাণ দেয় এবং উপাত্তের অর্থ-তাৎপর্য-ব্যাখ্যা ও সত্যতা তুলে ধরে মানুষ, মানুষের অভিজ্ঞতা ও গল্প। কাজটির সূচনা পর্ব থেকেই আমরা সংশ্লিষ্ট পক্ষগুলোর ওয়াকিবহাল কয়েকজনের সঙ্গে, কয়েকজন ভুক্তভোগীর সঙ্গে বিশদ কথা বলেছিলাম। কিছু মামলার নথিপত্র যা পাওয়া যায়, আসাদ জোগাড় করছিলেন।

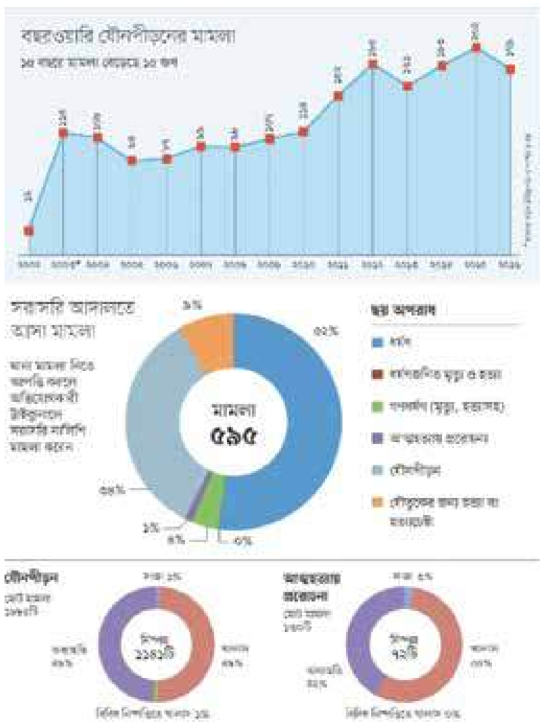
> গবেষণাপর্বের এসব সাক্ষাৎকার এবং তথ্য-উপাত্তের প্রাথমিক বিশ্লেষণ থেকে কিছু প্রবণতাকে গুরুত্বপূর্ণ বলে চিহ্নিত করতে পারলাম। সেগুলোর দিকনির্দেশনা অনুসারে আমরা মোট ৬৫টি মামলার নথিপত্র জোগাড় করলাম, বেছে নিলাম।

ততদিনে প্রতিবেদক গোলাম মর্তুজাকে আমাদের দলে পেয়েছি। ২০১৭ সালের অক্টোবর নাগাদ আসাদ ও মর্তুজা বাছাই করা এই মামলাগুলোর নিবিড় খোঁজখবর করতে মাঠে নামলেন। প্রত্যন্ত অঞ্চলে ছড়িয়ে থাকা ভুক্তভোগী, তদন্তকারী কর্মকর্তা বা আইনজীবীর সাক্ষাৎকার নিলাম মুঠোফোনে।

> এভাবে আমরা কখনো মামলায় আপসের ভেতরের কথার খোঁজ পেলাম, ভুক্তভোগীদের মানবিক পরিস্থিতি জানতে-বুঝতে পারলাম। এমনকি পুলিশের খুঁজে না পাওয়া অথবা তদন্ত-বিচারের ফোকর গলে বেরিয়ে যাওয়া আসামির সঙ্গে কথা বলতে পারলাম। ধর্ষণ মামলায় বিচার না পাওয়া ভুক্তভোগী একজন অল্পবয়সী মেয়ের সাক্ষাৎকার নিয়েছিলাম নিবিড় গোপনীয়তায়, তার কথা ভুলতে পারার নয়।

নিবিড় অনুসন্ধানের এই ৬৫ মামলার নথি এবং সাক্ষাৎকারলব্ধ তথ্য দিয়ে আমরা নিজেরাই এক্সেলে আলাদা করে বিশদ উপাত্তভাণ্ডার পাতলাম এবং বিশ্লেষণ করলাম। পাশাপাশি রাষ্ট্রপক্ষের আইনজীবীসহ এ ক্ষেত্রে দীর্ঘকাল কাজ করা আইনজীবী, আন্দোলনকর্মী, আইনজ্ঞ, পুলিশ কর্মকর্তা, গবেষকসহ ওয়াকিবহাল এবং বিশেষজ্ঞ মানুষজনের সাক্ষাৎকার নিয়ে চললাম।

নিবিড় অনুসন্ধানে পাওয়া তথ্য ও মানুষজনের কথা আমাদের প্রতিবেদনগুলোর প্রাণ হয়ে উঠল



ইনফোগ্রাফিক: প্রথম আলো

ফসল তোলা

এটা অনেক আগেই স্পষ্ট হয়ে গিয়েছিল যে, আমাদের ধারাবাহিকভাবে কয়েকটি প্রতিবেদন করতে হবে। পর্ব এবং সেগুলোর মূল প্রতিবেদন ও পার্শ্বপ্রতিবেদনের সম্ভাব্য থিম বা মূল বিষয়বস্তু নিয়ে বিস্তারিত চিন্তাভাবনা ও পরিকল্পনা চলমান ছিলো তথ্য-উপাত্ত সংগ্রহের শুরু থেকেই। ২০১৭ সালের নভেম্বর নাগাদ ছয়টি পর্বে প্রতিবেদনের মূল ধারণা ও অবয়বের পরিকল্পনা চূড়ান্ত হলো। লেখা ও সম্পাদনার কাজ শুরু হলো। পাশাপাশি অধিকতর যাচাই, নতুন তথ্য সংযোজন ও পুনর্লিখনের কাজ চলতে লাগল।

> ২০১৮ সালের ৮ মার্চ নারীদিবসে সার্বিক অনুসন্ধান ও ফলাফল নিয়ে একটি মূল প্রতিবেদন ও একটি পার্শ্বপ্রতিবেদনসহ ধারাবাহিকের প্রথম পর্ব প্রকাশিত হলো। তথ্য-উপাত্ত দিয়ে বলা গেলো যে নির্বাচিত ছয়টি অপরাধে নিষ্পন্ন মামলার মাত্র তিন শতাংশে সাজা হয়েছে। মূল প্রতিবেদনটি সেদিনের পত্রিকায় শীর্ষ প্রতিবেদন ছিলো। এর সঙ্গে সুলতানা কামালের একটি অভিমত গেলো।

যাচাই, নতুন তথ্য সংযোজন, পুনর্লিখন ও সম্পাদনা শেষে একেকটি করে পর্ব বেরোতে লাগলো। সে বছরের ২৩ মে শেষ পর্বটি বেরোলো। প্রতিটি পর্বে পার্শ্ব প্রতিবেদন ছিল। চতুর্থ পর্বের দুটি প্রতিবেদন করেন মূল দলের বাইরে থেকে মানসুরা হোসাইন এবং শেখ সাবিহা আলম। ষষ্ঠ ও শেষ পর্বটি ছিলো করণীয় সম্পর্কে। সেটার সঙ্গে একটি বিশেষজ্ঞ অভিমতও প্রকাশিত হয়। পরে ১০ জুন একটি সম্পাদকীয় হয়। আরও পরে রায়, তদন্ত প্রতিবেদনসহ নথিপত্রে, আইনজীবী বা পুলিশ কর্মকর্তাদের বক্তব্যে বৈরি ভাষা ও ভাব নিয়ে একটি উপসম্পাদকীয় প্রকাশিত হয়। এই ধারাবাহিকে অধিকাংশ ক্ষেত্রেই ভুক্তভোগীর নাম-পরিচয়ে নিচ্ছিদ্র গোপনীয়তা রক্ষা করতে হয়েছে। আলোকচিত্র প্রকাশের সুযোগ প্রায় ছিলো না। আঁকা ছবি এবং ইনফোগ্রাফ ছিলো মূল ভরসা। পেছনে তাকিয়ে বলা যায়, এ ক্ষেত্রে আরও সৃজনশীলতার সুযোগ ছিলো। তা নিয়ে আরও ভাবনা-পরিকল্পনা করতে হতো।

লেগে থাকার অর্থ সময়, ধৈর্য এবং খাটুনি

সবগুলো কাজ মিলিয়ে অনুসন্ধানী দলে নিবিড়ভাবে যুক্ত ছিলেন ছয়জন, আমি নিজেসহ। যাচাইয়ে বাড়তি পেয়েছি তিনজনের সহায়তা; প্রাথমিক উপাত্তভাণ্ডার তৈরিতে কম্পোজ বিভাগের সহায়তা। সম্পাদনা ও উপাত্ত বিশ্লেষণের কাজে মূল দলের দুজন ছাড়া আরও দুজন সহায়তা করেছেন। প্রাথমিক ও চূড়ান্ত সম্পাদনায় মূল দলের দুজনসহ অন্তত আরও তিনজন যুক্ত থেকেছেন।

> এতজনকে পাওয়া গিয়েছে সম্পাদকের এবং প্রতিষ্ঠানের সর্বাত্মক সমর্থন থাকায়। বাইরের কর্মীদের পারিশ্রমিক এবং মূল দলের কর্মীদের প্রণোদনাসহ ফটোকপি ইত্যাদিসহ কিছু বাড়তি খরচ লেগেছে। তবে সেটা বিরাট কিছু নয়। প্রতিষ্ঠানের বড় বিনিয়োগ ছিলো নিজস্ব কর্মীদের প্রায় দুই বছর কাজটির জন্য পর্যাপ্ত সময় ছেড়ে রাখা। আসাদসহ রিপোর্টাররা অবশ্য এ সময় নিয়মিত কাজও চালিয়ে গেছেন। তাঁরা ফাঁকে ফাঁকে কাজটি করেছেন, আবার টানা সময় ছাড়ও পেয়েছেন।

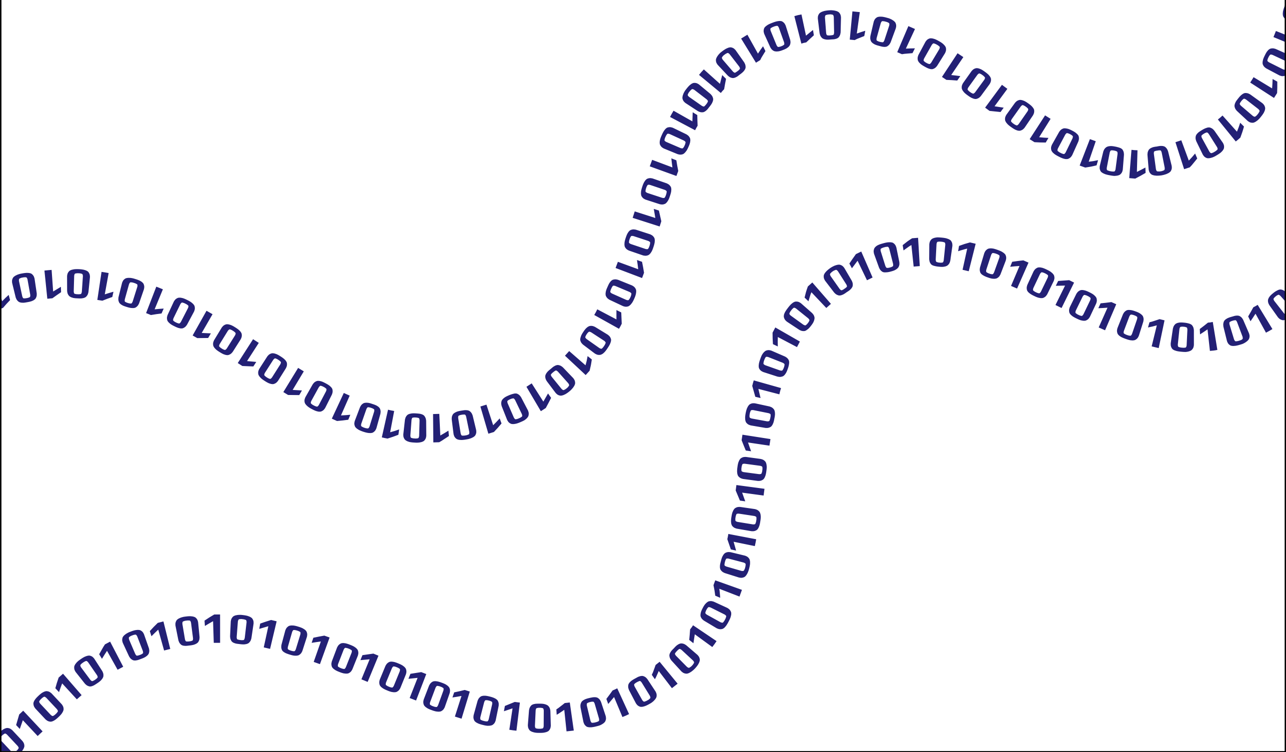
এতো সহযোগিতা এবং সহায়তা সত্ত্বেও মূল দলের কর্মীরা সব সামলাতে হিমসিম খেয়েছি। কাজটা নামানোর চাবিকাঠি যদি বলতে হয়, তো সেটা ছিল লেগে থাকা-নিষ্ঠা ও ধৈর্য্য। প্রত্যেকে ভালোবেসে কাজটা করেছেন। কাজের মূল তাগিদের সঙ্গে সম্পৃক্ত হয়েছেন। এভাবেই বিক্ষিপ্ত কর্মীরা মিলে একটি দলে কাজে মনেপ্রাণে এক হয়ে গেছেন।

আমাদের সবচেয়ে বড় প্রাপ্তি ও পুরস্কার হয়েছে, একটি ধূসর ক্ষেত্রে প্রাথমিক সুনিশ্চিত তথ্য-উপাত্ত ও দিকনির্দেশ দিতে পারার তৃপ্তি। আমরা তথ্য-উপাত্ত দিয়ে প্রমাণ করেছি, প্রায় ১৬ বছরে গুরুতর অপরাধে নিষ্পন্ন মামলার ৯৭ শতাংশে কোনো সাজা হয়নি। আমাদের প্রতিবেদনগুলো প্রতিষ্ঠিত করতে পেরেছে যে, অন্তত এই ছয়টি গুরুতর অপরাধে মিথ্যা মামলা হওয়ার অভিযোগ ধোপে টিকবে না। আমরা পুলিশি তদন্তে চূড়ান্ত প্রতিবেদনের সুবাদে অব্যাহতি এবং বিচারে খালাসের কারণগুলো গুনেগেঁথে তার সুনির্দিষ্ট ব্যাখা কিছু দিতে পেরেছি। আমরা থানার তদন্ত, মামলার সাক্ষ্যপ্রমাণ হাজির করা, শুনানিকে কার্যকর করার দায়িত্বে রাষ্ট্রপক্ষের উকিল ও পুলিশের গাফিলতি চিহ্নিত করতে পেরেছি। থানা ও ট্রাইব্যুনাল ধরে পরিস্থিতির তথ্য দিয়েছি; অপরাধের ধরন ধরে নিবিড় বিশ্লেষণ করেছি। আপসের বাস্তবতা উদাহরণ ধরে দেখাতে পেরেছি।

আরেক বড় তৃপ্তি, ভুক্তভোগীদের পাশে দাঁড়ানোর অনুভূতি। এই অর্জনগুলোর কারণেই প্রতিষ্ঠান এই বড় কাজটি করার সুযোগ করে দিয়েছেন। হয়তো ব্যক্তি কোনো ভুক্তভোগীর জীবনে অথবা পরিস্থিতি বড় কোনো ইতিবাচক বদল আমরা আনতে পারিনি। তবে প্রতিবেদনগুলো সংশ্লিষ্ট কর্তাব্যক্তিসহ মানুষজনকে নাড়া দিয়েছে, ভাবিয়েছে। একাধিক কর্তৃপক্ষ আমাদের সঙ্গে যোগাযোগ করেছেন। কাজটি নিয়ে প্রকাশ্যে আলাপ-আলোচনা করেছেন। দু'একটি সিদ্ধান্তের আলোচনাও কানে এসেছে। সমস্যাটি এত মৌলিক ও ব্যাপক যে, রাতারাতি বদল সম্ভব নয়। কিন্তু আমাদের কাজটি সদ্ব্যবহার করার সুযোগও কিন্তু তাৎক্ষণিক নয়। সম্পাদকের সিদ্ধান্তে প্রথমা প্রকাশন থেকে ২০১৮ সালে সাজা মাত্র তিন শতাংশ নামের একটি পুস্তিকায় পুরো কাজটি সংরক্ষিত করা হয়েছে।

আগেই যেমনটা বলেছি, এ কাজে যাচাই ও সংশোধনের তাগিদ কখনো শেষ হয়নি। পুস্তিকাটি গোছানোর সময় ১৯৯৫ ও ২০০০ সালের আইনের ধারায় কিছু বিভ্রান্তির সুযোগ ছিল। বিভ্রান্তি আমাদের হয়েছেও। পুস্তিকাটিতে আমরা সেটা সংশোধন করেছি। ভুল ও সংশোধনের আলাদা নোটও দিয়েছি। এখন এই লেখার সুবাদে দু-একটি ভুল ও ব্যাখ্যার অভাব চোখে পড়েছে। সেই সংশোধনী দাগিয়ে রেখেছি। অর্থাৎ কাজটি এখনো আমাদের তাড়িয়ে বেড়াচ্ছে। সেটাও কিন্তু সাংবাদিকের জন্য একটা বড় অর্জন। সামনে এই বিষয়ে আরও কাজ করবো, করতে হবে। তখন এই কাজটি আমাদের ভিত্তি দিকনির্দেশ হয়ে থাকবে। আর লেগে থেকে কাজ করা ছাড়া মানুষের জন্য কাঙ্ক্ষিত বদল আনার প্রক্রিয়ায় অংশী হওয়ার দ্বিতীয় কোনো উপায় সাংবাদিকের আছে কি?

প্রথম আলোয় প্রকাশিত নারী ও শিশু নির্যাতন মামলার সিরিজের প্রথম প্রতিবেদনটি পড়তে কিউআর কোডটি স্ক্যান করুন



অধ্যায় ০৯

চিত্রায়ণ

চিত্রায়ণ

> অল্প কিছু চর্চা ও অনুশীলনের মাধ্যমে সাংবাদিকেরা ডেটা ও ভিজ্যুয়ালাইজেশন নিয়ে কাজের কলাকৌশল শিখে নিতে পারেন, যেগুলো তাদের প্রতিবেদনকে আরও সমৃদ্ধ করবে

তথ্যে ভরপুর এই বিশ্বে, সাংবাদিকেরা যে কোনো বিষয়কে ডেটা ভিজ্যুয়ালাইজেশনের মাধ্যমে দ্রুত ও স্পষ্টভাবে ফুটিয়ে তুলতে পারেন।

> আপনি যদি সময়ের সাপেক্ষে একটি বিষয়ের প্রেক্ষাপট বর্ণনা করতে চান (পরিস্থিতি কী ভালো হয়েছে, খারাপ হয়েছে, না-কি আগের মতোই আছে), বা (ম্যাপের মাধ্যমে) দেখাতে চান ইভেন্টগুলো কোথায় কোথায় ঘটছে, বা সংক্ষেপে বলতে চান এখন কী হচ্ছে (ভিন্ন ভিন্ন দোকানে খাদ্যসামগ্রীর দাম) তাহলে একটি ডেটা ভিজ্যুয়ালাইজেশন পাঠককে বিষয়টি সহজে বুঝতে সাহায্য করতে পারে।

আর এভাবে ডেটা নিয়ে কাজের জন্য আপনার কোডিং জানা বা গণিতে উচ্চতর ডিগ্রী থাকারও প্রয়োজন নেই। অল্প কিছু চর্চা ও অনুশীলনের মাধ্যমে সাংবাদিকেরা ডেটা ও ভিজ্যুয়ালাইজেশন নিয়ে কাজের কলাকৌশল শিখে নিতে পারেন, যেগুলো তাদের প্রতিবেদনকে আরও সমৃদ্ধ করবে।

ডেটাসেট খোঁজা

ডেটা ভিজ্যুয়ালাইজেশন নিয়ে কাজ শুরু করতে গেলে আপনার শুরুতেই প্রয়োজন হবে একটি ডেটাসেট। ডেটাসেটটি পেয়ে যাওয়ার পর, আপনার এখন নিজেকেই জিজ্ঞাসা করতে হবে যে, আপনি ভিজ্যুয়ালাইজেশনের মাধ্যমে কী দেখাতে চান। কীভাবে এটি আপনার প্রতিবেদনকে সমৃদ্ধ করবে, বা এই চার্টটি কি আলাদাভাবেই একটি গ্রাফিক্স হিসেবে পোস্ট করা যাবে সোশ্যাল মিডিয়ায় বা নিউজলেটারে?

> উত্তর পেয়ে যাওয়ার পর, আপনি এবার ভিজ্যুয়ালাইজেশনের জন্য ডেটা প্রস্তুতের কাজ শুরু করতে পারেন। উপরে যেসব সাইটের কথা বলা হয়েছে, সেখান থেকে আপনি ডেটা ডাউনলোড করতে পারবেন .csv (কমা সেপারেটেড ভ্যালুজ) বা মাইক্রোসফট এক্সেল ফাইল (সাধারণত .xls ev .xlsx ফাইল) আকারে। আপনার যদি মাইক্রেসফট এক্সেল অ্যাকাউন্ট না থাকে তাহলে গুগল শিটও ব্যবহার করতে পারেন। যেটি বিনামূল্যে ব্যবহার করা যায় এবং প্রয়োজনীয় বেশিরভাগ বিশ্লেষণের কাজ করা যায়।

কিন্তু এই ডেটা আপনি ভিজ্যুয়ালাইজ করতে পারবেন না, যদি সেটি বিশৃঙ্খল অবস্থায় থাকে। বিশৃঙ্খল ডেটাসেটের অবস্থা ঠিক তেমনই থাকে যেমনটি শুনতে মনে হচ্ছে। যেমন: উদাহরণ হিসেবে আপনার ডেটাসেটের একটি কলামে যদি একাধিক টাকার অঙ্ক বা সংখ্যা থাকে যেগুলো একটি নির্দিষ্ট ফরম্যাটে নেই (কিছু জায়গায় হাজারের অঙ্কের পর কমা আছে, কিছু জায়গায় নেই, কিছু জায়গায় সংখ্যাগুলো কথায় লেখা আছে) তাহলে সেটি বিশৃঙ্খল ডেটা।

ওপেন রিফাইনের মাধ্যমে ডেটা ক্লিনিংয়ের কৌশল সম্পর্কে আরও জানতে চাইলে কিউআর কোডটি স্ক্যান করে মিডিয়া হ্যাকের টিপশিটটি পড়ে নিতে পারবেন।

ডেটাসেটটি যদি ছোট হয় (১০০ বা তার কম লাইন), তাহলে হয়তো আপনি গুগল শিটে নিজে নিজেই সংখ্যাগুলো ঠিকঠাক করে নিতে পারবেন। কিন্তু আপনার ডেটাসেটটি যদি অনেক বড় হয়, তাহলে ওপেন রিফাইনের মতো টুল ব্যবহার বেশি কাজের হবে। ওপেন রিফাইন খুবই শক্তিশালী একটি টুল। এটি শিখে নিতে হয়তো কিছুটা সময় লাগবে। কিন্তু এটা খুবই কাজের হবে, যদি আপনাকে নিয়মিত এ ধরনের বড় ডেটাসেট নিয়ে কাজ করতে হয়।

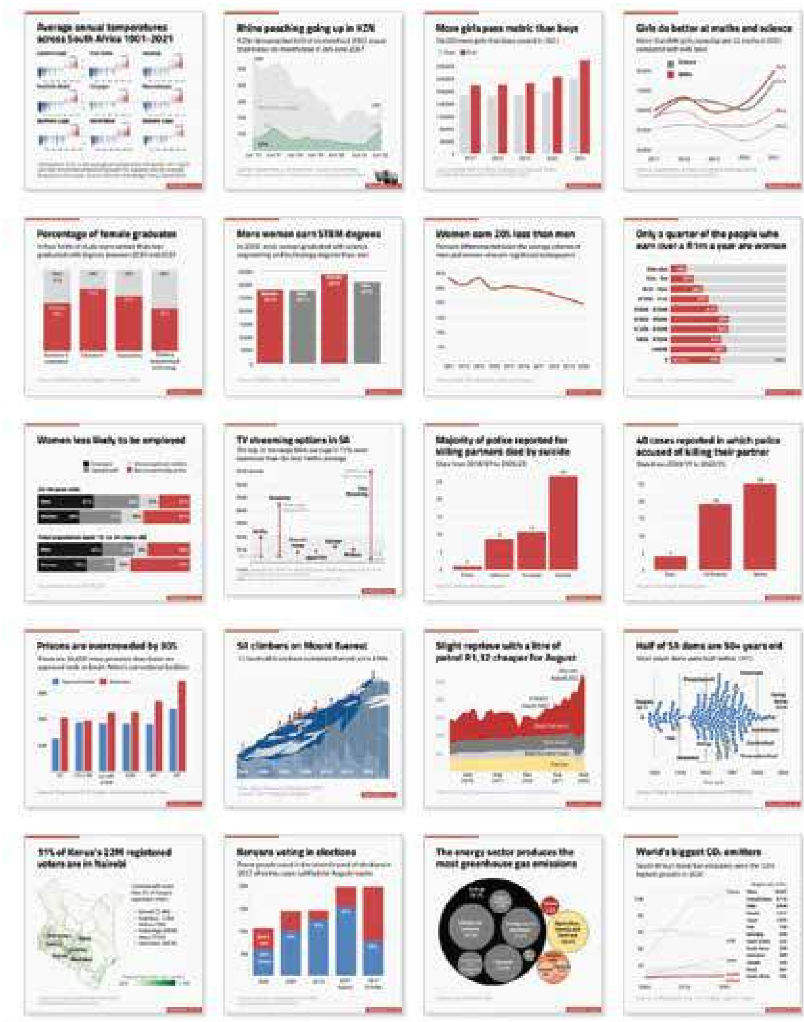
> আপনি যদি ডেটাসেটটি .csv বা এক্সেল ফাইল আকারে না পেয়ে সেটি কোনো বাৎসরিক রিপোর্টের অংশ হিসেবে বা প্রেস রিলিজ হিসেবে পিডিএফ ফাইল আকারে পেয়ে থাকেন; তাহলে ডেটাগুলো সেখান থেকে .csv বা এক্সেল ফাইলে বদলে নিতে পারেন ট্যাবুলা (ফ্রি অ্যাপ) বা এডোবি অ্যাক্রোব্যাট (পেইড অ্যাপ) ব্যবহার করে।

ভিজ্যুয়াল ভাবনা

একটি চার্ট বানানোর সময় আপনাকে অবশ্যই মাথায় রাখতে হবে যে, আপনি পাঠকের মনোযোগ আকর্ষণের জন্য প্রতিযোগিতা করছেন। ফলে আপনাকে সেটি রাখতে হবে সাধারণ ও সহজবোধ্য অবস্থায়।

একটি গ্রাফে সাধারণত থাকে একটি শিরোনাম, দুইটি অক্ষ, বার বা লাইন চার্ট দিয়ে উপস্থাপন করা ডেটা পয়েন্ট, লিজেন্ডের মাধ্যমে চলকরাশিগুলোর ব্যাখ্যা এবং অবশ্যই তথ্যসূত্র ও গ্রাফিক্সের প্রকাশকের নাম।

এখানে কিছু উদাহরণ থাকছে



ইনফোগ্রাফিক: আউটলায়ার

গ্রাফগুলোতে হয়তো কিছু উপাদানের সেট থাকতেই হবে, কিন্তু সেগুলো কখনোই একঘেঁয়ে হওয়া যাবে না। ভিজ্যুয়ালাইজিং শুরুর সময় এই বিষয়গুলো আপনার মাথায় রাখা উচিত–

> আপনার চার্ট বা ম্যাপটি কী ইন্টারঅ্যাকটিভ হবে নাকি স্ট্যাটিক? কোভিড-১৯[৪৭] সংখ্যার মতো প্রতিদিন হালনাগাদ করা হয় এমন বিষয়ের জন্য ইন্টারঅ্যাকটিভ চার্টই ভালো। তবে স্ট্যাটিক চার্টগুলো হোয়াটসঅ্যাপ, টেলিগ্রাম, সিগন্যাল, টুইটার ও ফেসবুকের মতো প্ল্যাটফর্মগুলোতে বেশি সহজে শেয়ার করা যায়।

> আপনার ভিজ্যুয়ালাইজেশনের বক্তব্য কি স্পষ্ট হয়েছে? ভিজ্যুয়ালাইজেশনে খুব বেশি কিছু বলতে যাওয়ার চেষ্টা করবেন না। বরং শুরু করুন একটি সহজ, স্পষ্ট বক্তব্য দিয়ে। তারপর প্রয়োজনে সেখানে আরও বিস্তারিত তথ্য যোগ করুন। চার্টটি বানিয়ে ফেলার পর সেটি সমালোচনামূলক দৃষ্টি দিয়ে পর্যালোচনা করুন এটি কি আসলেই আপনার বক্তব্য বা উন্মোচনটি ঠিকঠাক ফুটিয়ে তুলতে পেরেছে? কোনো বন্ধু বা সহকর্মীকে চার্টটি দেখিয়ে নেওয়া একটি ভালো অভ্যাস, যারা আপনাকে বলবে, সেখান থেকে তারা কী বুঝেছেন। ফলে আপনি নিশ্চিত হতে পারবেন যে, আপনার বক্তব্যটি সবার কাছে পৌঁছাচ্ছে।

> ব্যবহৃত ফন্টটি কি যথেষ্ট বড় এবং রঙগুলো কি যথেষ্ট শক্তিশালী? অক্ষের সঙ্গে থাকা লেখাগুলো পড়ার জন্য আপনার পাঠককে যেন ভিজ্যুয়ালাইজেশনটি জুম করে দেখতে না হয় বা খুবই হালকা রঙের লেবেল পড়তে না হয়। মনে রাখুন আপনি হয়তো গ্রাফটি ল্যাপটপ বা কম্পিউটারে তৈরি করছেন। কিন্তু আপনার বেশিরভাগ পাঠকই হয়তো ভিজ্যুয়ালাইজেশনটি দেখবেন তাদের মোবাইল ফোনে।

ডেটা-চিত্রায়ণ সম্পর্কে আরও জানতে চাইলে মার্কিন অর্থনীতিবিদ জনাথন শোয়েবিশের "বেটার ডেটা ভিজ্যুয়ালইজেশন" গ্রন্থটি পড়তে পারেন। কিউআর কোডটি স্ক্যান করে বইটি কিনতে পারবেন।

৪৭ Otter, A. (2022b, July 5). Covid-19 in South Africa. The Outlier. Retrieved October 12, 2022, from https://theoutlier.co.za/covid-19/81382/covid-19-sa-vaccination-tracker

রং বাছাই

চার্ট বানানোর বেশিরভাগ টুলেই আক্ষরিক অর্থে লাখ লাখ রঙ আছে। প্রায়ই আমরা প্রলুব্ধ হই যত বেশি সম্ভব রঙের ব্যবহারের জন্য। কিন্তু এটি অনেক ক্ষেত্রেই চার্টটিকে পাঠকদের জন্য কম সহজবোধ্য করে তুলতে পারে। কার্যকরীভাবে রঙের ব্যবহার শেখার জন্য অভিজ্ঞতার প্রয়োজন হয়। কিন্তু সাধারণ কিছু মৌলিক দিক-নির্দেশনা আছে, যেগুলো মাথায় রাখা উপকারী হতে পারে–

- শুরু করুন যত কম সম্ভব রঙ নিয়ে। পরবর্তীতে প্রয়োজন হলে আপনি সব সময়ই সেখানে আরও রঙের ব্যবহার করতে পারবেন।

- 'চার্টে ব্যবহৃত রঙগুলো কী উপস্থাপন করছে' তা চিন্তা করুন। যদি সেগুলো কয়েকজন ভিন্ন ভিন্ন রাজনীতিবিদের মতো আলাদা বিষয়কে উপস্থাপন করে তাহলে তাদের প্রত্যেকের জন্য আলাদা রঙ ব্যবহার করুন। আবার যদি ম্যাপে একটি অঞ্চলে নির্দিষ্ট রাজনৈতিক দলের ভোটপ্রাপ্তির তারতম্য উপস্থাপন করা হয়, তাহলে আপনি হয়তো একই রঙের হালকা থেকে গাঢ় শেডের তালিকা বেছে নিতে পারেন।

- মূল একটি পর্যবেক্ষণকে আলাদাভাবে ফুটিয়ে তোলার জন্য রঙের ব্যবহার করুন। যেমন, আপনার চার্টের বেশিরভাগ জায়গায় ধূসর রঙ ব্যবহার করে প্রধান ডেটা পয়েন্টগুলো হাইলাইট করতে পারেন লাল রঙ দিয়ে।

- রঙগুলির প্রতিটি আলাদাভাবে চোখে পড়ে কী না খেয়াল করুন। অর্থাৎ একটি আরেকটির চেয়ে হালকা বা গাঢ় হয়ে যাচ্ছে কি না সেটি নজরে রাখতে হবে। আলাদা করার মতো যথেষ্ট উপাদান না থাকলে নীল ও সবুজের মতো রঙগুলি প্রায়ই অনেক মানুষের কাছে একইরকম মনে হয়। নিশ্চিত করুন যে রঙগুলি একে অপরের চেয়ে যথেষ্ট আলাদা, যেন সবাই তাদের পার্থক্য বুঝতে পারে।

- রঙের ব্যবহার করুন ধারাবাহিকতা বজায় রেখে। যেমন, রঙের বদল থেকে মানুষ কী বুঝতে পারে, তা বিবেচনা ছাড়া চার্টে অ্যানোটেশনের জন্য ভিন্ন ভিন্ন রঙের ব্যবহার করবেন না। যেখানে সম্ভব, সব ধরনের লেবেলের জন্য একটি রঙ ব্যবহার করুন, যেমন সহজে পাঠযোগ্য কালো রঙ।

- রঙের ব্যবহার করুন উদ্দেশ্য/লক্ষ্য নিয়ে। যদি একাধিক রঙের ব্যবহার করতেই হয় তাহলে নিশ্চিত করুন যে, সেগুলো পাঠককে কিছু বাড়তি উপকারী অন্তর্দৃষ্টি দিচ্ছে। যেমন, অ্যানোটেশন ও লেবেলের ক্ষেত্রে একাধিক রঙের ব্যবহার তখনই করবেন যখন সেগুলো পাঠককে চার্টের বিভিন্ন উপাদান আরও ভালোভাবে বুঝতে সাহায্য করবে।

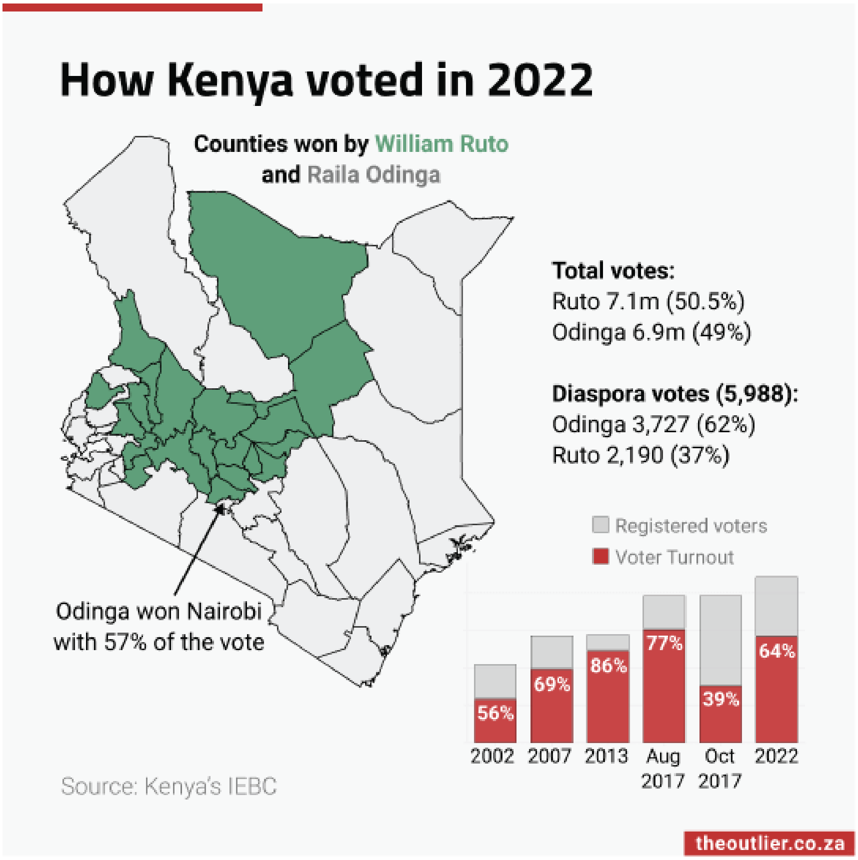
- আপনি যদি রঙের ব্যবহারে অভিজ্ঞও হয়ে থাকেন, তারপরও সুসংজ্ঞায়িত রঙের একটি সংগ্রহ দিয়ে শুরু করা কার্যকরী হতে পারে। শুরু করার জন্য সবচেয়ে ভালো একটি জায়গা: সিন্থিয়া ব্রিউয়ারের কালারব্রিউয়ার সেট[৪৮] :https://colorbrewer2.org/

[৪৮] ColorBrewer: Color Advice for Maps. (n.d.-b). Retrieved October 12, 2022, from https://colorbrewer2.org/

- বর্ণান্ধতার কথাও বিবেচনা করুন। সবাই একইভাবে রঙ দেখে না। রঙসংক্রান্ত বেশিরভাগ টুলেই একটি অপশন পাবেন। যার মাধ্যমে আপনি পরীক্ষা করে নিতে পারবেন যে, আপনার নির্বাচন করা রঙগুলো একজন বর্ণান্ধ ব্যক্তির কাছে কেমন লাগবে। যেমন, কালারব্রিউয়ারে আছে একটি “কালারব্লাইন্ড সেফ” অপশন।

উদাহরণ

সাধারণত, সমগোত্রীয় মানের মধ্যে তারতম্য বোঝানোর জন্য সবচেয়ে ভালো উপায় একই ধারার রঙের একটি তালিকা ব্যবহার করা। এই ধরনের ম্যাপের ক্ষেত্রে এটি প্রায়ই ব্যবহৃত হয় যেখানে উঁচুমাত্রার মান বোঝাতে গিয়ে একইরঙের ক্রমশ গাঢ় শেড ব্যবহার করা হয়।



ইনফোগ্রাফিক: আউটলায়ার

ভিন্ন ভিন্ন শ্রেণীবিভাগ বোঝাতে পুরোপুরি আলাদা ধরনের রঙ ব্যবহার করাই ভালো। যেমন, এই ম্যাপে ২০২২ সালের কেনিয়ান নির্বাচনে একজন প্রার্থী কোন কোন জায়গায় জিতেছেন

তা হাইলাইট করার জন্য সবুজ রঙ ব্যবহার করা হয়েছে। একইসঙ্গে, ভোটার উপস্থিতির বিষয়টি স্পষ্ট করার জন্য ব্যবহার করা হয়েছে গাঢ় লাল। যেখানে সেটি ধূসর রঙের মাধ্যমে উপস্থাপিত মোট নিবন্ধিত ভোটারের মধ্যে স্পষ্টভাবে ফুটে উঠেছে। সাধারণত লেবেলের ক্ষেত্রে একটি রঙই ধারাবাহিকভাবে ব্যবহার করা উচিৎ। যেমন এই উদাহরণের ক্ষেত্রে, প্রার্থীদের নামের রঙটি লেখা হয়েছে ম্যাপে উপস্থাপিত রঙের দিকে ইঙ্গিত করে।

কেনিয়ার নির্বাচন নিয়ে তৈরি উপরের ডেটাচিত্রটি কীভাবে তৈরি হয়েছে, তা জানতে চাইলে কিউআর কোডটি স্ক্যান করুন

ডেটাচিত্র

লাইন ও বার চার্ট

সবচেয়ে বেশি ব্যবহৃত দুইটি মূল চার্ট হলো লাইন ও বার চার্ট। দুটিতেই থাকে একটি এক্স-অক্ষ (যেখানে তুলনা করা তথ্যগুলো দেখানো হয়) এবং একটি ওয়াই-অক্ষ (আপনার তুলনার বিষয়টির পরিমাপজনিত তথ্য দেখানো হয়)।

> সময়ের সাথে সাথে কীভাবে ডেটার বদল হয়েছে–এই ধরনের চলমান তথ্য উপস্থাপনের জন্য লাইন চার্ট বেশ ভালো। যেমন, উদাহরণ হিসেবে বলা যায় জ্বালানির মূল্যবৃদ্ধি, মুদ্রাস্ফীতি বা শেয়ার বাজারের ওঠানামার মতো বিষয়ের কথা।

আমরা বার চার্ট ব্যবহার করি বিভাগভিত্তিক তথ্য উপস্থাপনের জন্য। যেমন, ভিন্ন ভিন্ন দেশে জ্বালানির মূল্য,[৪৯] প্রতিবছরে তাপমাত্রার পরিবর্তন, স্কুলের ফলাফল, রাজনৈতিক প্রার্থীদের কে কত ভোট পেয়েছেন বা মাসিক অপরাধ পরিসংখ্যানে থাকা খুনের ঘটনা। লাইন ও বার চার্ট তৈরি করার দারুন দুইটি টুল ফ্লোরিশ[৫০] ও ডেটার‍্যাপার[৫১]। দুইটিই ব্যবহার করা যায় বিনামূল্যে। এবং এখানে আপনি নিজের ডেটাসেট আপলোড করে সঙ্গে সঙ্গেই ভিজ্যুয়ালাইজেশন শুরু করতে পারেন।

৪৯ The Outlier on. (2022c, March 9). Twitter. Retrieved October 12, 2022, from https://twitter.com/outlierafrica/status/1501527234649247745

৫০ https://flourish.studio/

৫১ https://www.datawrapper.de/

ভিডিও টিউটোরিয়াল দেখে ডেটা বিশ্লেষণ ও চিত্রায়ণ শিখতে চাইলে কিউআর কোডটি স্ক্যান করে আফ্রিকা ডেটা হাবের টিউটোরিয়ালগুলো দেখে নিতে পারেন।

অন্যান্য চার্ট

ডেটা ভিজ্যুয়ালাইজ করার আরও অনেক পথপদ্ধতি আছে। এখানে থাকছে এমন অল্প কিছু অপশন

> একটি পূর্ণাঙ্গ বিষয়ের অংশগুলো দেখানোর জন্য কার্যকরী হতে পারে পাই চার্ট। যেমন, একটি নির্দিষ্ট শহর, দেশ বা অঞ্চলের জনসংখ্যা লিঙ্গভিত্তিকভাবে ভেঙে দেখানোর জন্য পাই চার্ট ব্যবহার করা যেতে পারে। তবে সতর্ক থাকুন, পাইয়ের "টুকরা" এতো বেশি করবেন না, যেন তা সহজে বোঝা কঠিন হয়। একইসঙ্গে, মনে রাখুন যে, পাই চার্ট "একটি পূর্ণাঙ্গ বিষয়ের অংশগুলো" উপস্থাপন করে। ফলে এখানকার সবগুলো টুকরা মিলে যেন তা সবসময় ১০০ ভাগই হয়। ঠিক একটি পাইয়ের মতো।

কয়েকটি ডেটাসেট একে অপরের সঙ্গে কীভাবে সম্পর্কযুক্ত–তা বোঝানোর জন্য উপকারী হতে পারে স্ক্যাটার প্লট। যেমন, স্কুল পরীক্ষার ফলাফল বনাম তারা কোন শহরের স্কুল থেকে এসেছে। স্ক্যাটার প্লট উপকারী কারণ এটি তৃতীয় আরেকটি মাত্রা যোগ করে। এক্স ও ওয়াই অক্ষে আপনি পয়েন্ট যোগ করছেন। পাশাপাশি সেই পয়েন্টটির আকার ও আয়তন এখানে তৃতীয় মাত্রা যোগ করছে।

আরও অনেক ধরনের চার্ট আছে, যেগুলোর অনেকগুলোই তৈরি করা হয় কয়েকটি মূল চার্টের সমন্বয়ে। যেমন, এরিয়া চার্ট অনেকটা লাইন চার্টের মতোই, কিন্তু তারা আয়তন-সংক্রান্ত ধারণা আরও ভালোভাবে ফুটিয়ে তোলে। সময়ের সঙ্গে সঙ্গে মূল একটি বিষয়ের অন্তর্গত কোনো অংশ কীভাবে পরিবর্তিত হয়েছে–তা দেখানোর ভালো উপায় হতে পারে স্ট্যাকড বার বা এরিয়া চার্ট।

মানচিত্র তৈরি

ডেটা ভিজ্যুয়ালাইজের সবচেয়ে জনপ্রিয় টুল ম্যাপ বা মানচিত্র। নির্বাচনী ফলাফলের মতো ডেটা উপস্থাপনের জন্য এগুলো প্রায়ই ব্যবহৃত হয়। কিন্তু এটি আরও বিস্তৃত পরিসরের ডেটা ভিজ্যুয়ালাইজেশনের জন্য ব্যবহার করা যায়। যেমন স্বাস্থ্য সংকট, যুদ্ধ, নাগরিক সংঘর্ষ ও প্রাকৃতিক দুর্যোগের মতো বিষয়ের ক্ষেত্রে।

ইনফোগ্রাফিক: আউটলায়ার

ডেটা ভিজ্যুয়ালাইজেশনের জন্য সবচেয়ে জনপ্রিয় দুই ধরনের ম্যাপ হলো কোরোপ্লেথ ম্যাপ ও বাবল ম্যাপ। কোরোপ্লেথ ম্যাপে আপনি বিভিন্ন রঙের শেড দিয়ে ভিন্ন ভিন্ন অঞ্চল, জেলা বা মিউনিসিপ্যালিটিকে উপস্থাপন করেন। বাবল ম্যাপে আপনি কোনো পয়েন্টারের আকার দিয়ে জনসংখ্যার ঘনত্বের মতো ডেটা; বা রাজধানী, কারখানা ও ঘটনার অবস্থান ভিজ্যুয়ালাইজ করেন।

ডেটার‍্যাপার ও ফ্লোরিশ শুধু লাইন ও বার চার্ট ভিজ্যুয়ালাইজ করার জন্যই জনপ্রিয় নয়, এগুলো ম্যাপিং করার জন্যও দারুন টুল। এসব টুলে ম্যাপ নিয়ে কাজের আলাদা বিষয়টি হলো এই ডেটায় অক্ষাংশ ও দ্রাঘিমাংশসহ অবস্থান-সংক্রান্ত তথ্য থাকতে হয়। অ্যাপগুলোতে যদি আপনার পছন্দসই ম্যাপের টেমপ্লেট না থাকে, তাহলে আপনি শূন্য থেকেও নিজের ম্যাপ তৈরি করে নিতে পারেন। কোড জানা না থাকলেও চিন্তা করবেন না। mapshaper.org[৫২] ও ফ্লোরিশের "ব্ল্যাংক ম্যাপস" অপশন দিয়ে শুরু করলে আপনি সহজেই এসব বাধা উতরে যেতে পারবেন।

[৫২] mapshaper

কোডিং দক্ষতা ছাড়াই ডেটা চিত্রায়ণের আরো কৌশল শিখতে চাইলে কিউআর কোডটি স্ক্যান করে মিডিয়া হ্যাকের টিপশিটটি দেখে নিতে পারেন।

ডেটাচিত্র প্রকাশ

চার্ট বা ম্যাপ বানিয়ে ফেলার পর ডেটার‍্যাপার বা ফ্লোরিশ থেকে আপনি .jpeg বা .png ফরম্যাটের স্থির ছবি আকারে এটি ডাউনলোড করতে পারেন। একইসঙ্গে আপনি সেটি এমবেড কোড বা আইফ্রেম ব্যবহার করে ওয়েবসাইটের কোনো লেখার মধ্যেও প্রকাশ করতে পারেন।

ফ্লোরিশের আরেকটি বাড়তি অপশন আছে। এটি আপনাকে গ্রাফিক্সটি .svg (স্কেলেবল ভেকটর গ্রাফিক্স) ফাইল আকারেও ডাউনলোড করতে দেয়, যেটির ওপরে আপনি ফিগমার মতো অন্য ডিজাইন টুল দিয়ে আরও কাজ করতে পারেন।

সহায়ক টুল

ফ্লোরিশ, ডেটার‍্যাপার ও ম্যাপশেপার ডট ওআরজি থেকে শুরু করে টাবুলা, অ্যাক্রোব্যাট প্রো ও ওপেন রিফাইন পর্যন্ত; বেশ কিছু টুল নিয়ে আমরা কথা বলেছি। আপনার যদি এরইমধ্যে এসব টুল নিয়ে কাজের অভিজ্ঞতা থাকে, তাহলে এখানে থাকছে আরও কিছু টুলের হদিস। ডেটা সাংবাদিকতাসংক্রান্ত টুল ও রিসোর্সের একটি বড় তালিকা।[৫৩]

[৫৩] Otter, A (2022e, August 21). The big list of data journalism tools and resources | by Alastair Otter | May, 2022 | Media Hack | Media Hack. Medium. Retrieved October 12, 2022, from https://inside.mediahack.co.za/data-journalism-tools-and-resources-8df1c3d96c13

010101010101010101010101010101010101010101010101010101010

অধ্যায় ১০

নৈতিকতা

নৈতিকতা

> ডেটা সাংবাদিকতায় নৈতিক বিবেচনা সাংবাদিকতার অন্য বাকি ক্ষেত্রগুলোর থেকে আলাদা নয়। এখানেও সংবাদে ডেটার যথার্থতাকে নৈতিকতার বড় মানদণ্ড হিসেবে বিবেচনা করা হয়। তবে একইসাথে নতুন ধারার এই সাংবাদিকতায় জনস্বার্থ এবং ব্যক্তিগত গোপনীয়তার অধিকারের ভারসাম্য রাখাটা খুব জরুরি।

কেননা, আপনি যখন একটি ডেটাসেটে কয়েক লাখ মানুষের ডেটা বিশ্লেষণ করে একটি প্যাটার্ন বা ট্রেন্ড খুঁজে বের করছেন, তখন আপনি সেই ডেটাসেটে কয়েক লাখ মানুষের ব্যক্তিগত ডেটা নিয়ে নাড়াচাড়া করছেন। সেই তথ্যের গোপনীয়তা রক্ষা করাও আপনার দায়িত্ব।

> ডেটা নিয়ে কাজ করা সাংবাদিকদের সবচেয়ে বড় উদ্বেগের বিষয় ডেটার নির্ভুলতা। স্প্রেডশিটের বিশ্লেষণ আপনাকে চার্ট এবং মানচিত্র ঠিক বানিয়ে দেবে, কিন্তু আপনার সংখ্যা যদি ভুল থাকে, তবে তা পুরো অন্যপথে প্রবাহিত হবে। তাই প্রতিবেদককে সংখ্যার উৎসের বিশ্বাসযোগ্যতা যেমন যাচাই করতে হবে, একইসাথে যে প্রক্রিয়ায় গ্রাফিকাল উপস্থাপনটি হচ্ছে সেখানে কোনো ম্যানিপুলেশনের ঝুঁকি আছে কি-না কিংবা কোনো ভুল হচ্ছে কি-না সেটা সমানভাবে যাচাই করতে হবে। কেননা মনে রাখতে হবে ডেটা সাংবাদিকতায় ঐ ডেটাচিত্রগুলো সবচেয়ে গুরুত্বপূর্ণ। পাঠক বা দর্শক সেগুলোকেই প্রথমে দেখে এবং আপনার প্রতিবেদন সম্পর্কে আগ্রহী হয়।

সংখ্যার নৈতিকতা

ডেটার নৈতিকতা নিশ্চিতে আপনাকে ডেটাকে সেই একই প্রশ্ন জিজ্ঞাসা করতে হবে, যেভাবে আপনি আপনার বাকি আর দশটি প্রতিবেদনের জন্যে প্রাপ্ত তথ্যের উৎসকে প্রশ্ন করেন। যে ব্যক্তি আমাকে ডেটা দিচ্ছে তার স্বার্থ কি? কীভাবে এই তথ্য সংগ্রহ করা হয়েছে, এবং কে বা কারা (বা কখন বা কোথায়) এই ডেটা সংগ্রহ করা হয়েছে? ডেটা যদি কোনো জরিপ থেকে এসে থাকে, তবে জানতে হবে কি ধরনের প্রশ্ন করে এমন ডেটা পাওয়া গেছে।

প্রশ্নগুলি কীভাবে সাজানো হয়েছিলো? কত মানুষের ওপর জরিপ চালিয়ে এই ফলাফল পাওয়া গেছে। কোন সময়কালে এই জরিপ পরিচালিত হয়েছে। যে সূত্রের মাধ্যমে আমি এই ডেটা পাচ্ছি তার কী এমন জরিপ পরিচালনা করার সামর্থ্য আছে কিংবা এই তথ্য সরবরাহ করার সামর্থ্য কি আছে। একইসঙ্গে আপনি অন্য কোনো স্বাধীন উৎস থেকে একই ধরনের ডেটা পেতে পারেন কি-না সেই চেষ্টা করতে হবে, ডেটার বিশ্বাসযোগ্যতা, নির্ভুলতা এবং নৈতিকতার প্রশ্নে। এক্ষেত্রে তথ্য অধিকার আইন একটি কার্যকর সমাধান হতে পারে। কারণ আপনার হাতে এরই মধ্যে প্রয়োজনীয় ডেটা চলে এসেছে। আপনি জানেন কারা এই ডেটা সংরক্ষণ করেন বা কাদের কাছে এই ডেটাসেট আছে। আপনার হাতে থাকা ডেটা যাচাই করতে, একই তথ্যের জন্যে আপনি তথ্য অধিকার আইনে আবেদন করতে পারেন। এই প্রক্রিয়ার মাধ্যমে আপনি আরও নতুন ডেটা আপনার বিশ্লেষণের জন্য সংগ্রহ করতে পারবেন।

প্রেক্ষাপট

ডেটা উপস্থাপনের ক্ষেত্রে, প্রসঙ্গ বা প্রেক্ষাপট অত্যন্ত গুরুত্বপূর্ণ। আপনি যদি শুধু বিশ্লেষণে প্রাপ্ত তথ্য পাঠকের সামনে তুলে ধরেন, তবে তা পাঠকের কাছে কোনো মানে তৈরি করে না। কেননা সংখ্যা নিজে থেকে কিছু বলে না। পৃথকভাবে উপস্থাপিত একটি সংখ্যা পাঠককে বলবে না, সেটি বেশি বা কম, কিংবা খারাপ বা ভালকে নির্দেশ করছে। আপনাকে একটি সূচক বা মানদণ্ডের আওতায় সংখ্যাকে ফেলতে হবে। একটি ঐতিহাসিক প্রেক্ষাপটের মধ্যে, ব্যক্তি বা দিনে সংখ্যাগুলিকে তাদের উপস্থাপন করলে, তা আরও অর্থবহ করতে সহায়তা করে। ধরুন আপনি যে কোনো একটি নিত্যপন্যের বাজার দামের বৈশ্বিক তুলনা করলেন যে, ঐ নির্দিষ্ট পণ্যটির কোন দেশে কত দাম। এ ক্ষেত্রে শুধু পণ্যটির দাম দেখে পাঠক সঠিক চিত্রটি পাবে না, যে কোনো দেশের জনগণের পণ্যটি কিনতে বেশি কষ্ট হচ্ছে কোথায় কম। এ জন্যে আপনাকে একইসাথে সংশ্লিষ্ট দেশগুলোর গড় মজুরি বা মাথাপিছু আয় উল্লেখ করতে হবে। বিনিময় মূল্যও বিবেচনায় নিতে হবে।

সাক্ষাৎকার

ডেটার প্রাসঙ্গিকতা, পেক্ষাপট এবং নির্ভুলতা নিশ্চিতের জন্য সংখ্যার পিছনে থাকা মানুষের সাথে সাক্ষাতকার প্রয়োজন হয়। এক্ষেত্রে দুই পক্ষের সাক্ষাৎকার নিতে হবে, যারা ডেটা সংগ্রহ করেছে এবং ডেটা যাদের নিয়ে কথা বলছে। এর ফলে আপনার ডেটার নির্ভরযোগ্যতা অনেকখানি নিশ্চিত হবে। সংশ্লিষ্টদের সাথে এই সাক্ষাৎকার একটি ডেটা প্রতিবেদনকে যেমন শক্তিশালী করতে পারে, তেমনি এই গোষ্ঠী প্রতিবেদককে বিভ্রান্তও করতে পারে, ভুল পথে পরিচালিত করতে পারে। এমন ক্ষেত্রে সাংবাদিককে নিজস্ব বিচার বুদ্ধি দিয়ে সিদ্ধান্ত নিতে হবে, কী করলে তার প্রতিবেদনটি নৈতিকতার সাথে আপস করবে না।

ভিজ্যুয়াল উপস্থাপনা

ডেটাচিত্র তৈরির ক্ষেত্রে বেশকিছু বিষয় বিবেচনায় রাখতে হয়। কেননা এটি প্রতিবেদনের শেষ এবং সবচেয়ে গুরুত্বপূর্ণ ধাপগুলোর একটি। কেননা, প্রতিবেদন পড়ার আগে, পাঠক আপনার ডেটাচিত্রটিতেই আগে চোখ বুলিয়ে নেবে।

তাই শুধু নির্ভুল চিত্রায়ণেই প্রতিবেদকের কাজ শেষ হবে না। পাঠক যেন গ্রাফিক্স দেখে বিভ্রান্ত না হয়, সেদিকে খেয়াল রাখতে হবে। আপনার ভিজ্যুয়ালাইজেশন কি ব্যক্তিগত তথ্য বা ছবি প্রকাশ করে কি-না সেটি যাচাই করতে হবে। যদি করে, তবে তা কারো ক্ষতির কারণ হচ্ছে কি-না সেটা বিবেচনায় নিতে হবে। আবার যদি আপনি কোনো জনগোষ্ঠীকে চিত্রায়ণ থেকে বাদ দেন, কেন বাদ দিচ্ছেন, এর সংবাদ মূল্য কী সেগুলোও যাচাই করে নেয়া গুরুত্বপূর্ণ।

ডেটায় পরিবর্তন

তথ্য পরিবর্তনশীল। আজকে যে তথ্য বা ডেটা নিয়ে আপনি বিশ্লেষণ তৈরি করলেন এবং একটি সিদ্ধান্তে উপনীত হলেন, তা সপ্তাহখানেক পরেই বদলে যেতে পারে, ডেটা বদলে যাওয়ার কারণে। তবে এতে আপনার প্রতিবেদন ভুল হয়ে যাবে না। তবে ডেটা যে পরিবর্তন হতে পারে, সেটি প্রতিবেদনে উল্লেখ করে পাঠককে জানানো একজন সাংবাদিকের নৈতিক দায়িত্ব। তবে বর্তমানে উৎসে ডেটার পরিবর্তন স্বয়ংক্রিয়ভাবে প্রতিবেদনের ডেটাকে বদলে দেয়ার প্রযুক্তিও ব্যবহৃত হচ্ছে সাংবাদিকতায়। তবে আপনি যদি সেই প্রযুক্তি ব্যবহার করতে নাও পারেন, পাঠককে ডেটা বদলের খবর জানাতে, ডেটার মূল উৎস প্রতিবেদনের সাথে জুড়ে দিতে পারেন। আগ্রহী পাঠক সেখান থেকে পরবর্তী অগ্রগতি জেনে নিতে পারবে।

অনুসন্ধানের স্বচ্ছতা

প্রতিবেদনের বিশ্বাসযোগ্যতা এবং নৈতিকতা নিশ্চিত করতে পাঠকের কাছে স্বচ্ছ হতে হবে। পাঠককে অনুসন্ধানের সঙ্গী করুন। অনুসন্ধানের প্রতিটি ধাপ প্রতিবেদনে এমনভাবে ব্যাখ্যা করুন, যেন পাঠক তা পড়ে বুঝতে পারে কি প্রক্রিয়ায় তথ্য সংগ্রহ করা হয়েছে, কীভাবে বিশ্লেষণ করে এমন তথ্য তুলে আনা হয়েছে। উন্মুক্ত ডেটার ক্ষেত্রে তথ্যের সূত্রও প্রতিবেদনে উল্লেখ করা যেতে পারে। একইসঙ্গে ডেটা বিশ্লেষণ-পদ্ধতিও বিস্তারিত জানান। এর বাইরে যেকোনো অনুসন্ধানী ডেটা সাংবাদিকতা নিয়ে পাঠকের মনে সাধারণত যে সব প্রশ্ন আসে, বা অতীতে প্রতিবেদন-সংক্রান্ত যে ধরনের প্রশ্ন পেয়েছেন, তার আলোকে কিছু উত্তর আগে থেকেই প্রতিবেদনে উল্লেখ করুন। যাতে আগ্রহী পাঠক আপনার এবং আপনার প্রতিবেদনের নৈতিকতা নিয়ে প্রশ্ন তুলতে না পারে। স্বচ্ছতা না থাকলে নিউ ইয়র্ক টাইমস বা ওয়াশিংটন পোস্টের মত গণমাধ্যমও সমালোচনার মুখে পড়ে।

ম্যাসাচুসেটস অ্যামহার্স্ট বিশ্ববিদ্যালয়ের সহকারী অধ্যাপক রদ্রিগো জামিথ, তার এক গবেষণায়[৫৪] নিউইয়র্ক টাইমস বা ওয়াশিংটন পোস্টের ২০১৭ সালের প্রথমার্ধের ১৫৯টি প্রতিবেদন বিশ্লেষণ করে দেখেছেন যে, এই দুই প্রভাবশালী গণমাধ্যমের সাংবাদিকেরা তাদের তথ্য সংগ্রহের কৌশল এবং বিশ্লেষণের পদ্ধতি পাঠকের সামনে ব্যাখ্যা করতে ব্যর্থ হয়েছেন। ঐ গবেষণায় দেখা গেছে যে, প্রতি ১০টির মধ্যে প্রায় ৯টি প্রতিবেদনেই ব্যবহৃত ডেটাসেটের সাথে ডেটার মূল উৎসের লিঙ্ক যুক্ত করা হয়নি। আবার যেগুলোতে করা হয়েছে, তার বেশ কিছুতে পাঠকের প্রবেশের অধিকার রাখা হয়নি। অন্যদিকে মাত্র ২০ ভাগ

[৫৪] Zamith, R. (n.d.-b). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. Retrieved October 12, 2022, from https://www.tandfonline.com/doi/abs/10.1080/21670811.2018.1554409?journalCode=rdij20

প্রতিবেদনে ডেটা বিশ্লেষণের পদ্ধতিগত ব্যাখ্যা দেয়া হয়েছে। বাকিগুলোতে পাঠক জানতে পারেনি কিসের ভিত্তিতে যাচাই করে সংশ্লিষ্ট প্রতিবেদক একটি সিদ্ধান্তে উপনীত হয়েছেন। জ্যামিথের ভাষায় ডেটা সাংবাদিকদের এমন প্রবণতা নতুন ধারার এই সাংবাদিকতাকে ঝুঁকিতে ফেলতে পারে। তাই ডেটা সাংবাদিকতায় স্বচ্ছ থাকার বিকল্প নেই।

ডেটা ও উৎসের নিরাপত্তা

ডেটা সাংবাদিকতায় অনেক তথ্য বা ডেটা নিয়ে প্রতিবেদককে কাজ করতে হয়। বিশাল সেই ফাইলের নিরাপত্তা নিশ্চিত করার দায়িত্ব প্রতিবেদকেই নিতে হবে। কেননা সেখানে অনেকে স্পর্শকাতর এবং ব্যক্তিগত তথ্য থাকতে পারে। এ জন্য ডেটা বিশ্লেষণ এবং প্রতিবেদন প্রকাশ পরবর্তী সময়েও ডেটাকে নিরাপদে রাখতে হবে, যদি না উন্মুক্ত ডেটা হয়। ফাঁস হওয়া তথ্য যেভাবে সুরক্ষা দেয়া হয়, ডেটার সোর্সকেও সেভাবে সুরক্ষিত রাখার ব্যাপারে সাংবাদিকদের সতর্ক থাকতে হবে।

ডেটা সাংবাদিকতায় নৈতিকতা সম্পর্কে আরও জানতে চাইলে কিউআর কোডটি স্ক্যান করে এথিক্যাল জার্নালিজম নেটওয়ার্কের এই অনলাইন কোর্সটি অনুসরণ করতে পারেন।

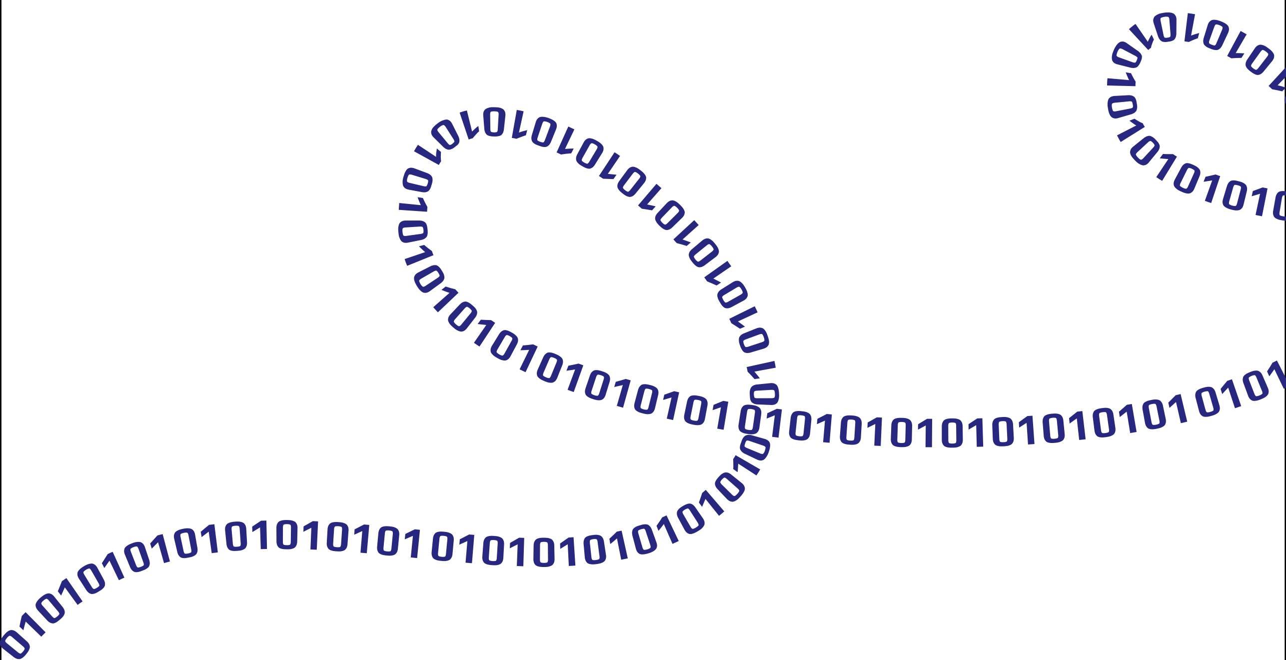
ডেটা সাংবাদিকতায় নৈতিকতার বাইরেও বেশ কিছু ঝুঁকি রয়েছে। প্রতিবেদন তৈরির বিভিন্ন ধাপে যার মুখোমুখি আপনাকে হতে হবে। এ জন্য সেসব ঝুঁকির বিষয়ে আগে থেকে সতর্ক থাকা প্রয়োজন। ঝুঁকি চিহ্নিত করার কৌশল জানার পাশাপাশি তা কীভাবে মোকাবেলা করতে হবে, তাও জানতে হবে। ডেটা জার্নালিজম হ্যান্ডবুকে প্রচলিত সাংবাদিকতার সব ঝুঁকিকেই ডেটা সাংবাদিকতার ঝুঁকি হিসেবে চিহ্নিত করেছে। এর মধ্যে উল্লেখযোগ্য কয়েকটি ঝুঁকি কীভাবে মোকাবেলা করবেন, তা জেনে নেয়া যাক।

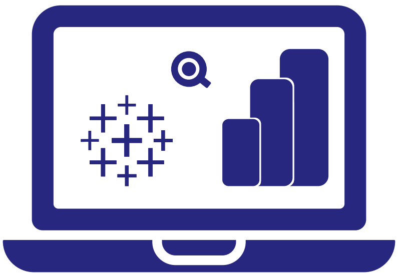
ডেটা সাংবাদিকতার ঝুঁকি

বিষয়	ঝুঁকি	প্রতিকার
বিষয়	ভুল উৎস হতে ডেটা সংগ্রহ	ডেটা সংগ্রহের আগে সম্ভাব্য উৎসের বিশ্বাসযোগ্যতা যাচাই করা
তথ্য	স্প্রেডশিটে ভুল ডেটা সংযোজন	স্প্রেডশিটে সয়ংক্রিয়ভাবে ডেটা যুক্ত করার পর, ম্যানুয়ালি পুনরায় যাচাই করা
আইনগত	আইনগত নিষেধাজ্ঞা আছে এমন ডেটা ব্যবহার	ডেটার উৎস এবং আইনগত বিধিনিষেধ নিয়ে আইনজীবীর সঙ্গে আলোচনা করা
প্রেক্ষিত	প্রতিবেদনে ডেটা ব্যবহারে ভুল প্রেক্ষিত বিবেচনায় নেয়া	প্রতিবেদনের প্রেক্ষিত ও ডেটার সামঞ্জস্যতা নিয়ে বিশেষজ্ঞ কিংবা সম্পাদকের সঙ্গে আলোচনা করা
স্প্রেডশিট	স্প্রেডশিটে কোনো সেল খালি থাকা	স্প্রেডশিটে স্বয়ংক্রিয়ভাবে ডেটা যুক্ত করার পর, হাতে পুনরায় ক্লিন করা
বিশ্লেষণ	ভুল সূত্র ব্যবহার	ডেটাকে করা প্রশ্ন অনুযায়ী সূত্র ব্যবহারে সতর্কতা অবলম্বন করা। প্রয়োজনে বিশেষজ্ঞের পরামর্শ নেয়া।
প্রাইভেসি	ব্যক্তিগত ডেটা প্রকাশ	ব্যক্তিগত ডেটা প্রকাশ করতে গিয়ে আইন লঙ্ঘন করছেন কি-না যাচাই করে নিন।
ডিজিটাল নিরাপত্তা	অনলাইন সার্ভার বা গুগল ড্রাইভ থেকে ডেটা হ্যাকিং	নজরদারি এড়াতে ভিপিএন ব্যবহার করা। শক্তিশালী পাসওয়ার্ড ব্যবহার করা।

মনে রাখা ভালো, শুধু ডেটা সাংবাদিকতা নয়, যে কোনো সাংবাদিকতাতেই নৈতিকতাই আপনার সবচেয়ে বড় রক্ষাকবচ। ডেটা সাংবাদিকতার সুবিধা হলো, বেশিরভাগ ক্ষেত্রেই আপনি পাবলিক বা উন্মুক্ত ডেটাসেট নিয়ে কাজ করছেন। এ ধরনের ডেটায় আইনি ঝুঁকি কম। কিন্তু লিক বা ফাঁস হওয়া ডেটার ক্ষেত্রে আপনাকে তথ্য ও উৎস যাচাইয়ের কঠোর মানদণ্ড মেনে চলতে হবে, যেমনটা সাংবাদিকতার আদর্শ চর্চায় করা হয়ে থাকে । স্পর্শকাতর ডেটা বা নথির বেলায় অবশ্যই সম্পাদক, প্রয়োজনে আইনজীবীর সঙ্গে পরামর্শ করে নিন। আর ডিজিটাল ডেটার ক্ষেত্রে কারো প্রাইভেসি বা ব্যক্তিগোপনীয়তা ক্ষুণ্ন হচ্ছে কি না খেয়াল করুন। আইনিভাবে কোন কোন ডেটা বা তথ্য ব্যক্তিগত হিসেবে বিবেচিত, তা মাথায় রাখুন। প্রাইভেসি-সচেতন হোন।

ডেটা সাংবাদিকতায় ডিজিটাল নিরাপত্তাবিষয়ক ঝুঁকি ও তা মোকাবেলার আরও কৌশল জানতে কিউআর কোডটি স্ক্যান করে ডেটা জার্নালিজম হ্যান্ডবুকের টিপশিটটি দেখুন।





অধ্যায় ১১

কীভাবে ডেটা সাংবাদিক হবেন?

কীভাবে ডেটা সাংবাদিক হবেন

> নিউইয়র্ক টাইমসের অ্যারনপিল হোফারের মতে এই প্রশ্ন প্রচলিত সাংবাদিকতা থেকে ভিন্ন কিছু নয়। শুধু প্রতিবেদনের বিস্তৃতিটা একটু বড় হতে হবে আর বড় কাজের ইচ্ছেটা থাকতে হবে।

আপনি ডেটা সাংবাদিকতা জানেন না। ভাবছেন শুরু করবেন কীভাবে? বিখ্যাত ডেটা সাংবাদিক পল ব্র্যাডশ[৫৫] বলেছেন, "কোথা থেকে একজন ডেটা সাংবাদিক শুরু করবেন? অবশ্যই প্রথম উত্তর হলো 'ডেটা থেকে'; কিন্তু আরেকটি উত্তর আছে, 'একটি প্রশ্ন থেকে'।" অর্থাৎ ডেটাকে প্রশ্ন করে। নিউইয়র্ক টাইমসের অ্যারনপিল হোফারের মতে এই প্রশ্ন প্রচলিত সাংবাদিকতা থেকে ভিন্ন কিছু নয়। শুধু প্রতিবেদনের বিস্তৃতিটা একটু বড় হতে হবে, আর বড় কাজের ইচ্ছেটা থাকতে হবে। ডেটা সাংবাদিক হতে আপনাকে কোডিং জানতে হবে না। অর্থনীতি কিংবা বাণিজ্যের শিক্ষার্থী হতে হবে না। আপনাকে আগে থেকেই গণিতে পারদর্শীও হতে হবে না। হ্যাঁ, এই জ্ঞানগুলো থাকলে, ডেটা সাংবাদিকতায় নিঃসন্দেহে আপনি অনেকটা এগিয়ে যাবেন। তবে প্রাথমিক এই সব জ্ঞান ছাড়াও আপনি ডেটা সাংবাদিকতা রপ্ত করতে পারবেন।

কী ধরনের দক্ষতা লাগবে

ডেটা সাংবাদিকতা শুরুর জন্য আপনাকে কে–ডেটা সায়েন্স নিয়ে পড়তে হবে, এমন নয়। তথ্য সংগ্রহ, বিশ্লেষণ ও লেখার চিরাচরিত পদ্ধতিগুলোই এখানে প্রযোজ্য। বাড়তি বলতে কিছু প্রযুক্তিগত টুল ব্যবহারের দক্ষতা। এতক্ষণ আগের অধ্যায়গুলো পড়ে নিশ্চয়ই বুঝতে পেরেছেন।

- শুরু করার জন্য আপনাকে এক্সেল বা গুগল শিটসে হাত পাকিয়ে নিতে হবে
- অনলাইন ডেটা সংগ্রহের জন্য স্ক্র্যাপিং এবং কনভার্সন টুলগুলো রপ্ত করতে হবে
- স্প্রেডশিটের পাশাপাশি অনলাইনে চার্ট ও গ্রাফ তৈরির টুলগুলো ব্যবহার জানতে হবে এবং
- প্রচুর চর্চা করতে হবে।

[৫৫] Bradshaw, P. (2017b, May 31). How to be a data journalist. The Guardian. Retrieved October 12, 2022, from https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide

এখানে ডেটা সাংবাদিকতা শুরু করার জন্য কয়েকটি পরামর্শ দেয়া হলো যা হয়তো আপনার কাজে লাগবে

বড় পরিসরে ভাবুন

সাংবাদিক হিসেবে আপনার ভাবনার পরিধি বড় করতে হবে। একটি সাধারণ খবরকে বৃহৎ পরিসরে দেখতে পারা বা দেখতে চাওয়ার ক্ষমতা একজন ডেটা সাংবাদিককে অনেকখানি এগিয়ে দেয়। ধরুন, ট্রেডিং কর্পোরেশন অব বাংলাদেশ (টিসিবি) এর সর্বশেষ বাজার দরের তালিকাটি সংগ্রহ করে দেখলেন বছরের ব্যবধানে সব পণ্যের দাম গড়ে ৩০ ভাগের মত বেড়ে গেছে। আপনি চাইলেই এই তথ্য দিয়েই সংবাদ লিখতে পারেন। কিন্তু যদি আপনি টিসিবির গেল ২০ বছরের ডেটা বিশ্লেষণ করে বের করেন যে, এবাবের মতো দাম বৃদ্ধি আগে কখনও হয়নি। তবে সেটি একটি সাড়া জাগানো প্রতিবেদন হবে। এই যে আপনি একদিনের বাজার দরের বদলে ২০ বছরের বাজার দর দেখতে চাইলেন, ডেটা সাংবাদিক হবার জন্যে এই চাওয়াটাই খুব বেশি প্রয়োজন।

আগে প্রশ্ন ঠিক করুন

ডেটাসেট পাবার পর, সেখান থেকে আপনি কী পেতে চান, সেটি জানাটা গুরুত্বপূর্ণ। কেননা, আপনি কী জানতে চান তার ওপর নির্ভর করবে ডেটাকে আপনি কী প্রশ্ন করবেন। যেমন, আগের উদাহরনে বাজার দরের টিসিবির ২০ বছরের ডেটা নিয়ে আলোচনা হয়েছে। ঐ ডেটাশিট সংগ্রহ করা হয়েছিলো এটা জানতে যে, এ বছরই কি সবচেয়ে বেশি হারে দাম বাড়লো না-কি, আগেও এমন হয়েছে। সেই তথ্য জানতে আপনাকে বছরওয়ারি কি হারে মূল্যবৃদ্ধি হয়েছে, ডেটাকে সেই প্রশ্ন করতে হবে। এভাবে প্রতিটি ডেটাসেট বিশ্লেষণের প্রশ্ন ঠিক করে নিতে হবে।

ছোট ডেটা সেট দিয়ে শুরু করুন

ছোট ডেটাসেট দিয়ে ডেটা সাংবাদিকতার অনুশীলন শুরু করা বুদ্ধিমানের কাজ। কারণ ছোট ডেটাসেট পাওয়া সহজ এবং এমন ডেটা বিশ্লেষণ করা অপেক্ষাকৃত সহজ। প্রাথমিক কাজের এই সাফল্য আপনাকে ডেটা সাংবাদিকতায় আরো উৎসাহী করবে। আপনি অল্প সময়ের মধ্যে আপনার বিশ্লেষণ শেষ করতে পারবেন। কেননা প্রাথমিক পর্যায়ে একটি ডেটা প্রতিবেদনের জন্য সংবাদ কক্ষ থেকে আপনি হয়তো বেশি সময় পাবেন না।

স্প্রেডশিটে দক্ষতা বাড়ান

ডেটা সাংবাদিকতার মূল অস্ত্র বিশ্লেষণ দক্ষতা। আর এই দক্ষতা নির্ভর করে দুইটি বিষয়ের ওপর। প্রথমত, আপনার প্রশ্ন। দ্বিতীয়ত, সেই প্রশ্নের উত্তর জানতে এক্সেল বা স্প্রেডশিটের মাধ্যমে ডেটা বিশ্লেষণের কৌশল জানা। যার পুরোটাই নির্ভর করে বিভিন্ন সূত্রের ওপর। এটা ঠিক এক্সেল বা স্প্রেডশিট ডেটা সাংবাদিকতার প্রাথমিক টুলস। কিন্তু এটিই ডেটা সাংবাদিকতা

শুরুর দিকের সবচেয়ে কার্যকর টুলস। স্প্রেডশিট ব্যবহারে আপনি যত পারদর্শী হবেন, পরবর্তী ধাপের কাজগুলো আপনার জন্য ততো সহজ হবে। পরবর্তী পর্যায়ে পাইথন, আর-এর মতো বেশকিছু টুলস রয়েছে, যা ব্যবহার করে আপনি বৃহৎ ডেটাসেট নিয়ে কাজ করতে পারবেন। এসব কৌশল শেখার অনলাইন কোর্সও আছে কয়েকটি। যেখান থেকে আপনি বিনামূল্যে পাইথন ও আর শিখতে পারবেন।

ডেটাকে প্রশ্ন করতে জানুন

ডেটাকে প্রশ্ন করা এবং সেই প্রশ্নের উত্তর বের করতে, স্প্রেডশিটে কী সূত্রও প্রয়োগ করতে হবে, তা জানা ডেটা বিশ্লেষণের জন্য অত্যন্ত গুরুত্বপূর্ণ। এ বিষয়ে দক্ষতা অর্জন করতে ডেটাশিট নিয়ে অনুশীলন করুন। তবে এই প্রশ্ন করার দক্ষতা বৃদ্ধিতে সবচেয়ে কার্যকর অনুশীলন হচ্ছে ডেটা প্রতিবেদন পড়া। অন্যান্য ডেটা প্রতিবেদন বিশেষ করে পশ্চিমা বিশ্বের ডেটা প্রতিবেদনগুলো পড়লে জানা যাবে, কীভাবে সেখানকার সাংবাদিকরা ডেটাকে প্রশ্ন করে নতুন বিশ্লেষণ বের করেন।

দেখুন, পড়ুন

বড় সংবাদপত্রগুলো কী ধরনের ডেটা সাংবাদিকতা করছে, তারা কী ধরনের ডেটা নিয়ে কাজ করছে, সে বিষয়ে নিয়মিত খোঁজ খবর রাখুন। যেহেতু বাংলাদেশে সেভাবে ডেটা সাংবাদিকতার চর্চা হয় না, পশ্চিমা গণমাধ্যমগুলো অনুসরণ করুন। ভারতীয় ডেটা সাংবাদিকতা অনুসরণ করতে পারেন। সেখান থেকে আইডিয়া গ্রহণ করুন। যেহেতু ভারত এবং বাংলাদেশের সংবাদের প্রেক্ষাপট ও প্রাসঙ্গিকতা অনেকটা একইরকম, সে কারণে সেখানকার ডেটা সাংবাদিকরা যে ধরনের প্রতিবেদন তৈরি করছে, সে সম্পর্কে ধারণা নেয়া যেতে পারে। আপনি চাইলে ছোট আকারে বাংলাদেশে একই ধরনের ডেটা দিয়ে একটি বিশ্লেষণ দাঁড় করাতে পারেন।

ভিজ্যুয়ালাইজেশন তৈরি বুঝে নিন

এক্সেল বা স্প্রেডশিট ব্যবহার করেই আপনি চার্ট তৈরি বা ডেটা চিত্রায়ণের প্রাথমিক কাজগুলো করতে পারবেন। শুরুতে এগুলোতে দক্ষতা অর্জন করুন। এক্ষেত্রে ফ্লারিশ, ডেটার‌্যাপ, ইনফোগ্রামসহ আরও বেশকিছু অনলাইন টুলস এর বিষয়ে এই সহয়িকার চিত্রায়ণ অধ্যায়ে উল্লেখ করা হয়েছে। সেগুলো রপ্ত করুন। তবে প্রথমেই নয়, ধাপে ধাপে। প্রথমে স্প্রেডশিটের মাধ্যমে ডেটা চিত্রায়ণে দক্ষতা অর্জন করুন।

বিশেষজ্ঞের পরামর্শ নিন

ডেটা সাংবাদিকতায় সফলতার একটি বড় অংশই হচ্ছে অনুশীলন। তবে অনুশীলনের বিভিন্ন পর্যায়ে আপনি আটকে যেতে পারেন। সেসব সমস্যার সমাধান আপনি অনলাইনে খুঁজতে পারেন। ইউটিউবে সমস্যা ধরে ধরে সমাধানের বেশকিছু কার্যকর

ভিডিও রয়েছে। এর বাইরে সরাসরি বিশেষজ্ঞের পরামর্শ নিন। কেননা অনেক ক্ষেত্রে আপনি বিশ্লেষণের যে সমস্যাটিতে আটকে যাচ্ছেন সেটির সমাধান পেতে হয়তো কয়েক সেকেন্ড প্রয়োজন, কিন্তু না জানার কারণে সেখানে কয়েকদিন আটকে থাকতে হতে পারে। এমন সমস্যা এড়াতে যে কোনো প্রয়োজনে বিশেষজ্ঞের পরামর্শ নিন।

ডেটার সঙ্গে জীবনের সম্পর্ক খুঁজুন

এই সহয়িকার বিভিন্ন পর্যায়ে ডেটার সঙ্গে মানুষের জীবনের সম্পর্কের গুরুত্বকে তুলে ধরা হয়েছে। ডেটা সাংবাদিক হতে চাইলে এই দক্ষতা আপনাকে অর্জন করতে হবে। ডেটার সঙ্গে যেসব গল্প জড়িয়ে আছে, সেগুলো চিহ্নিত করে প্রতিবেদনে উল্লেখ করতে হবে। এই কাজটি আপনি যতো ভাল পারবেন, আপনার প্রতিবেদন ততো বেশি সহজবোধ্য ও পাঠকপ্রিয় হবে। প্রতিবেদন লেখার এই দক্ষতা অর্জনের জন্য শীর্ষ সংবাদমাধ্যমগুলোতে প্রকাশিত ডেটা প্রতিবেদনগুলো পড়ুন।

রিপোর্ট করুন

ডেটা সাংবাদিকতা শুধু শ্রেণীকক্ষের পাঠ নয়। এর অনেক কিছুই আপনি চর্চা করতে করতে শিখবেন। তাই উপরে আলোচনা করা ডেটা সাংবাদিকতার প্রাথমিক জ্ঞান ও কৌশলগুলো রপ্ত করে, তার প্রয়োগ করুন। দুই একটি ছোট প্রতিবেদন তৈরি করুন। এতে ডেটা সাংবাদিকতায় আপনার আত্মবিশ্বাস বাড়বে। একইসাথে আপনার প্রতি সংবাদ কক্ষ ও সম্পাদকের আস্থাও বাড়বে। পরবর্তী কাজগুলোতে আরও বেশি সহায়তা পাবেন।

তারপর ধীরে ধীরে আরো জটিল প্রযুক্তিগত টুলগুলো শিখতে থাকুন। তবে শুরুটা বরাবরই একটু কঠিন বলে মনে হতে পারে। কিন্তু এই সহায়িকার প্রথম অধ্যায়ে যা বলা হয়েছে তা নিশ্চয় মনে আছে? বেশিরভাগ সাংবাদিকই ডেটা সাংবাদিকতা শিখেছেন নিজে নিজে। অনলাইনে বিভিন্ন রিসোর্স ঘাঁটুন এবং ভিডিও দেখে চর্চা করুন। মনে রাখবেন, চর্চা করতে করতে করতেই টুলগুলো আপনার আয়ত্বে আসবে। মনে অসংখ্য প্রশ্ন তৈরি হবে। ডেটা না থাকলেও আপনি তখন নিত্য নতুন উপায়ে আমাদের অভিজ্ঞতাকে ডেটায় রূপ দিতে শিখবেন। আর এভাবে হয়তো আপনার হাত ধরেই ডেটা সাংবাদিকতা প্রতিষ্ঠিত হবে এ দেশে।

ডেটা সাংবাদিকতায় শুরু করার ১২টি সহজ কৌশল জানতে কিউআর কোডটি স্ক্যান করে জিআইজেএন-এর টিপশিটটি দেখুন।

গোটা বিশ্বেই স্বচ্ছতা ও জবাবদিহি প্রতিষ্ঠায় বিশেষ করে দুর্নীতি, অর্থপাচার বিদেশে অবৈধ সম্পদ অর্জন ও রাজস্ব ফাঁকির ঘটনা উদ্‌ঘাটনে ডেটা সাংবাদিকতা বড় অবদান রাখছে। কেননা ডেটাভিত্তিক প্রতিবেদন অনেক বেশি সুনির্দিষ্ট ও প্রায় অকাট্য। একইসাথে ডেটা সাংবাদিকতা নতুন নতুন অনুসন্ধানের ক্ষেত্র তৈরি করে জবাবদিহি ও বস্তুনিষ্ঠতার সুযোগকে আরও সমৃদ্ধ করে। যার বড় উদাহরণ ২০১৫ সালে ইন্টারন্যাশনাল কনসোর্টিয়াম অব ইনভেস্টিগেটিভ জার্নালিস্টস (আইসিআইজে) প্রকাশিত পানামা পেপারস, পেগাসাস পেপারস, প্যান্ডোরা পেপারস ইত্যাদি। যার মাধ্যমে কোটি কোটি তথ্যকে বিশ্লেষণ করে দেশে দেশে গভীরতর অনুসন্ধানের মাধ্যমে ক্ষমতার অপব্যবহার ও দুর্নীতি, বিশেষ করে অর্থপাচারের গোপন এক অধ্যায়কে উদ্‌ঘাটন করা সম্ভব হয়েছে। যদি সাংবাদিকতার লক্ষ্য হয় তথ্যের অবাধ প্রকাশ, স্বচ্ছতা ও জবাবদিহি নিশ্চিত করা এবং জনগণকে বস্তুনিষ্ঠ ও অকাট্য তথ্য প্রদান করা, তাহলে ডেটা সাংবাদিকতা সেই প্রচেষ্টায় অন্যতম সময়োপযোগী হাতিয়ার।

978-984-35-3808-6