এই মূলত ডেটা অ্যানালাইসিসের দুটি অত্যন্ত গুরুত্বপূর্ণ পদ্ধতি— **Bivariate Analysis** (দুটি ভেরিয়েবল বা কলামের মধ্যে সম্পর্ক) এবং **Multivariate Analysis** (দুইয়ের অধিক ভেরিয়েবল বা কলামের মধ্যে সম্পর্ক) নিয়ে। যখন আমরা একটি ডেটাসেটের বিভিন্ন তথ্যের মধ্যে যোগসূত্র খুঁজে বের করতে চাই, তখন এই পদ্ধতিগুলো ব্যবহার করা হয়।

নিচে আলোচনার মূল বিষয়গুলো বিস্তারিতভাবে দেওয়া হলো:

১. সংখ্যাবাচক তথ্যের মধ্যে সম্পর্ক **(Numerical to Numerical)**

যখন আপনার কাছে দুটি কলামই সংখ্যা (যেমন- বয়স এবং বেতন), তখন তাদের সম্পর্ক বোঝার জন্য নিচের টুলগুলো ব্যবহার করা হয়:

- স্ক্যাটার প্লট **(Scatter Plot):** এটি দুটি সংখ্যার মধ্যে সম্পর্ক বোঝার সবচেয়ে সহজ উপায়। 'Titanic' ডেটাসেট ব্যবহার করে দেখানো হয়েছে যে, রেস্টুরেন্টে বিল যত বেশি হয়, টিপ দেওয়ার পরিমাণও তত বাড়ে। এটি একটি পজিটিভ রিলেশনশিপ।
- মাল্টিভ্যারিয়েট বিশ্লেষণ**:** স্ক্যাটার প্লটেই আমরা রঙের (Hue) মাধ্যমে জেন্ডার, স্টাইলের মাধ্যমে স্মোকার কি না এবং সাইজের মাধ্যমে কতজন লোক খেতে এসেছে—এই সবগুলো তথ্য একটি মাত্র গ্রাফে দেখতে পারি। অর্থাৎ, আপনি এক নজরেই বুঝতে পারবেন যে একজন ধূমপায়ী পুরুষ কত বিল দিয়েছিল এবং কত টিপ দিয়েছিল।

২. সংখ্যা এবং শ্রেণিবাচক তথ্যের মধ্যে সম্পর্ক **(Numerical to Categorical)**

যখন একটি তথ্য সংখ্যা এবং অন্যটি কোনো ক্যাটেগরি (যেমন- জেন্ডার বা ক্লাসের নাম), তখন নিচের পদ্ধতিগুলো কার্যকর:

- বার প্লট **(Bar Plot):** এটি কোনো নির্দিষ্ট ক্যাটেগরির গড় মান দেখায়। যেমন- টাইটানিকের ১ম, ২য় এবং ৩য় ক্লাসের যাত্রীদের গড় ভাড়া কত ছিল। গ্রাফ থেকে দেখা যায় ১ম ক্লাসের ভাড়া স্বাভাবিকভাবেই অনেক বেশি ছিল।
- বক্স প্লট **(Box Plot):** এটি ডেটার বিস্তৃতি এবং আউটলায়ার (অস্বাভাবিক তথ্য) খুঁজে পেতে সাহায্য করে। যেমন- কোন বয়সের মানুষ কোন ক্লাসে বেশি ছিল বা কোনো বয়স্ক মানুষ ভুল করে ছোটদের ক্লাসে ছিল কি না।
- ডিস্টপ্লট **(Distplot/PDF):** এটি সবচেয়ে শক্তিশালী একটি বিশ্লেষণ। দেখানো হয়েছে কীভাবে বেঁচে যাওয়া এবং মারা যাওয়া যাত্রীদের বয়সের তুলনা করা হয়েছে। এখান থেকে একটি চমৎকার তথ্য বেরিয়ে এসেছে যে, টাইটানিকে শিশুদের বেঁচে যাওয়ার সম্ভাবনা সবচেয়ে বেশি ছিল কারণ তাদের আগে উদ্ধার করা হয়েছিল।

৩. শ্রেণিবাচক তথ্যের মধ্যে সম্পর্ক **(Categorical to Categorical)**

যখন দুটি কলামই ক্যাটেগরি (যেমন- প্যাসেঞ্জার ক্লাস এবং বেঁচে যাওয়া), তখন তাদের সম্পর্ক বোঝার উপায়:

- হিটম্যাপ **(Heatmap):** প্রথমে একটি টেবিল তৈরি করা হয় যেখানে দেখা যায় কোন ক্লাসে কতজন মারা গেছে। হিটম্যাপ সেই টেবিলটিকে রঙিন করে দেয়। রঙের গভীরতা দেখে আপনি তৎক্ষণাৎ বুঝে যাবেন কোন জায়গায় তথ্যের ঘনত্ব বেশি (যেমন- ৩য় ক্লাসে মৃত্যুর হার সবচেয়ে বেশি ছিল)।
- ক্লাস্টার ম্যাপ **(Cluster Map):** এটি আরও এক ধাপ এগিয়ে কাজ করে। এটি ক্যাটেগরিগুলোর মধ্যে মিল খুঁজে বের করে এবং সেগুলোকে কাছাকাছি নিয়ে আসে। এটি অনেকটা ফ্যামিলি ট্রির মতো দেখায় যে কোন তথ্যগুলো একে অপরের সাথে বেশি সম্পর্কিত।

৪. স্বয়ংক্রিয় এবং সময়ভিত্তিক বিশ্লেষণ

- পেয়ার প্লট **(Pair Plot):** যদি আপনার ডেটাসেটে অনেকগুলো সংখ্যার কলাম থাকে, তবে আলাদা আলাদা করে গ্রাফ না এঁকে পেয়ার প্লট ব্যবহার করলে এটি নিজেই সব কলামের সাথে সব কলামের জোড়ায় জোড়ায় গ্রাফ তৈরি করে দেয়। এটি পুরো ডেটাসেটের একটি পূর্ণাঙ্গ চিত্র একবারে দেখতে সাহায্য করে।
- লাইন প্লট **(Line Plot):** এটি মূলত সময়ের সাথে তথ্যের পরিবর্তন দেখার জন্য ব্যবহার করা হয়। যেমন- ১৯৪৯ থেকে ১৯৬০ সাল পর্যন্ত প্রতি বছর বিমানে যাত্রী সংখ্যা কীভাবে বেড়েছে, তা লাইন প্লট দিয়ে খুব সুন্দরভাবে বোঝা যায়।

উপসংহার ও পরামর্শ**:** মূল উদ্দেশ্য হলো ডেটার ভেতর থেকে লুকিয়ে থাকা গল্প বা রহস্য খুঁজে বের করা। আপনি যখন গ্রাফগুলো আঁকবেন, তখন অনেক নতুন প্রশ্ন তৈরি হবে—যেমন "কেন অমুক বন্দর থেকে আসা মানুষ বেশি বেঁচে ফিরল?"। এই প্রশ্নগুলোর উত্তর খোঁজাই হলো একজন ডেটা সায়েন্টিস্টের কাজ।

পরামর্শ দিয়েছেন যে, শুধুমাত্র কোডিং শিখলে হবে না, বিভিন্ন ডেটাসেট নিয়ে নিজে নিজে এনালাইসিস করার চেষ্টা করতে হবে। কারণ ডেটা থেকে ইনসাইট বা তথ্য বের করে আনার ক্ষমতা প্র্যাকটিসের মাধ্যমেই তৈরি হয়। পুরো প্রক্রিয়াটি অনেকটা গোয়েন্দাগিরির মতো, যেখানে আপনি গ্রাফ ব্যবহার করে তথ্যের ক্লুগুলো খুঁজে বের করছেন!