Exploratory Data Analysis (EDA) - Deep Dive

Unveiling Patterns and Insights from Data

linkedin.com/in/rusdianto1234

There is no better way to learn than to teach

-Robert Oppenheimer

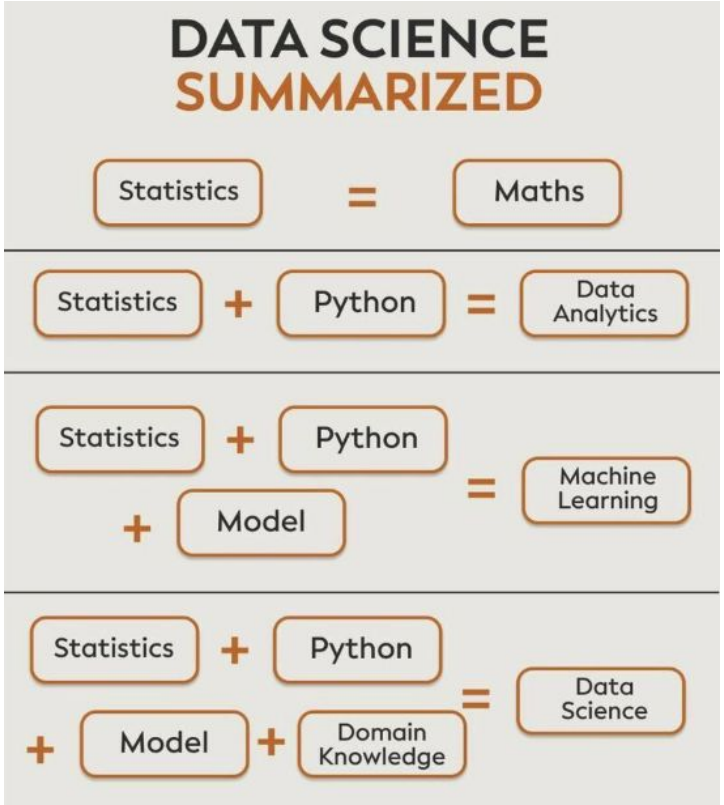
Data Analysis

Is the examination of data and relationships among variables, through both numerical and graphical methods

Why: Helps us provide clear understanding of the data to help business and to guide us making inform decision

Types:

- **Exploratory Data Analysis (EDA)**
- Confirmatory Data Analysis (CDA)



What is EDA?

A critical process of performing initial investigations on data to discover patterns, spot anomalies, and check assumptions with the help of summary statistics and graphical representations.

Role of EDA in Machine Learning:
Feature Selection | Data Preprocessing | Prepare for Model Building

What is CDA?

Evaluate the evidence using statistical tools

- Hypothesis Testing
- A/B Testing
- Regression Analysis
- Variance Analysis

EDA Process Overview

Define problem to solve or questions to answer

Data Collection:

Sources of Data: Databases, APIs, CSV files, Web Scraping, etc.

Tools: SQL, Python (Pandas, BeautifulSoup), R

Data Cleaning: Handling missing values | Removing duplicates | Addressing outliers| Correcting data types

Data Transformation: Feature scaling (Normalization, Standardization) | Encoding categorical variables| Aggregating data

Data Visualization:

Types of plots: Histograms, Box plots, Scatter plots, Bar charts, Heatmaps

Tools: Matplotlib, Seaborn, Plotly, Tableau

Statistical Analysis: CDA, Building Machine Learning Model

Deep Dive: 5G QoS EDA (Python)

Thank you