Data Analyst Notes

Introduction to Data Analysis

Definition: The process of collecting, cleaning, transforming, and modeling data to discover useful insights and support decision-making.

Types of Analysis:

- Descriptive (What happened?)
- Diagnostic (Why did it happen?)
- Predictive (What will happen?)
- Prescriptive (What should be done?)

Key Tools and Technologies

Excel - Data cleaning, basic analysis, dashboards

SQL - Query databases to extract data

Python - Scripting, data analysis (pandas, numpy), visualization

R - Statistical analysis

Power BI / Tableau - Data visualization

Jupyter Notebook - Interactive Python notebooks

Git - Version control

Excel for Data Analysis

Functions: VLOOKUP, HLOOKUP, INDEX, MATCH, IF, SUMIFS, COUNTIFS

Pivot Tables

Charts & Graphs

Data Cleaning: Remove Duplicates, Data Validation

SQL Basics

SELECT, FROM, WHERE

GROUP BY, ORDER BY

JOINS: INNER, LEFT, RIGHT, FULL

AGGREGATES: COUNT, SUM, AVG, MAX, MIN

Example:

SELECT product, SUM(sales) FROM sales_data WHERE region = 'East' GROUP BY product;

Python for Data Analysis

Libraries:

- pandas - data manipulation

- numpy - numerical computing

- matplotlib, seaborn - visualization

Example:

import pandas as pd

df = pd.read_csv('data.csv')

print(df.describe())

Data Cleaning

Remove nulls or fill them (fillna(), dropna())

Remove duplicates

Convert data types

Handle outliers

Exploratory Data Analysis (EDA)

Understand the structure, trends, and patterns in data

Use: df.describe(), df.info()

Histograms, box plots, correlation matrix

Check for skewness, distribution

Data Visualization

Python: matplotlib, seaborn, plotly

Power BI / Tableau:

- Dashboards

- Slicers, filters

- KPIs

Best practices:

- Use clear titles

- Choose the right chart type

- Avoid clutter

Statistics for Data Analysts

Mean, Median, Mode

Standard Deviation, Variance

Correlation & Covariance

Probability & Distributions

Hypothesis Testing (T-test, Chi-square)

Projects/Data Sources

Kaggle

Google Dataset Search

UCI Machine Learning Repository

Soft Skills

Communication: Presenting findings clearly

Problem-solving mindset

Business acumen

Time management