What is Data Analysis?

- Def:- Process of inspecting, cleansing, transforming, and modeling data to discover useful information is called as data analysis

Types of Data Analysis

- Descriptive Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

Descriptive Analysis

Descriptive Analysis – "What happened?"

-It looks at past data to understand what has already happened

- Example:
  You check last month’s sales and see that 500 items were sold
  🔍 You’re just *describing* the facts.

Diagnostic Analysis

Diagnostic Analysis – "Why did it happen?"

It digs deeper into the data to find out the reason behind what happened

- Example:
  Sales dropped last week. You check and find it rained heavily — fewer customers came
  🔍 You're *diagnosing* the cause

Predictive Analysis

It uses past data to make guesses or forecasts about the future

- Example:
  If sales usually increase before festivals, you can predict higher sales next Diwali
  🔍 You're *predicting* the future using patterns

Prescriptive Analysis

It uses past data to make guesses or forecasts about the future.

Example:

If sales usually increase before festivals, you can predict higher sales next Diwali.

🔍 You're *predicting* the future using patterns.

BI Tool

- Power BI
- Tableau

Popular Python Libraries for Data Analyst

- Pandas
- NumPy
- Scikit-learn
- Matplotlib
- Seaborn

Business Understanding

- Translate questions into data analysis problems is the job of Business Analyst

Case Study: Sentiment Analysis

- A clothing company receives hundreds of customer reviews online
- Some say:
- “The shirt fits perfectly!”
    - Others say:
- “Poor quality, I want a refund.”
    - The company wants to know what people really think about their products — without reading every review manually
- They used sentiment analysis tools (like AI or NLP) to automatically read and label reviews as:
- Positive

- Negative

- Neutral

Recommender Engines – What Are They?

Def:-Recommender engines are intelligent algorithms used to suggest relevant items to users based on their preferences, behaviors, or historical data

Types

1) Product Based
2) Content based

Examples of Recommender Engines

- **Netflix:** Suggests movies or shows based on what you've watched
- **Amazon**: Recommends products you might want to buy
- **Spotify**: Suggests songs or playlists based on your taste
- **YouTube**: Shows you videos similar to what you've watched before

Chatbots – What Are They?

- Def:-A chatbot is a computer program that can talk to you like a human
- It’s like a robot friend that answers your questions, helps you shop, gives support, or just

Examples of Chatbots

- Customer support on websites (“Hi! How can I help you?”)
- Ordering food from apps like Zomato or Swiggy
- Banking apps answering balance or transaction questions
- Agriculture Identifying crop diseases from images

What Are Open Source Libraries?

Definition:-Open-source libraries are collections of pre-written code made freely available for anyone to use

Importance in Modern Development:

- Accelerates development time
- Enables rapid prototyping and experimentation

Why Use Open Source?

1. Cost-Effective

- Free to use, reducing software development costs
- No licensing fees

2. Community Support

- Large, active user bases for popular libraries

3. Flexibility and Customization

- Source code can be modified to meet specific needs

4. Transparency

- Security and bugs are visible to everyone

Pandas

- Pros: Powerful data manipulation, large dataset handling
- Cons: High memory usage

NumPy

- Pros: Efficient computations, supports multi-dimensional arrays
- Cons: Limited to numerical data

Matplotlib

- Pros: Highly customizable, wide plot support
- Cons: Complex for advanced plots

Seaborn

- Pros: Simplified visualizations, integrates with Pandas
- Cons: Less customizable

Plotly

- Pros: Interactive, web-based visualizations
- Cons: Larger file sizes, complex for simple plots

Scikit-learn

- Pros: Easy to use, wide algorithm support
- Cons: Not for deep learning

TensorFlow

- Pros: Powerful for deep learning
- Cons: complex syntax

PyTorch

- Pros: Dynamic graphs, ideal for research
- Cons: Weaker deployment support

Django

- Pros: Built-in features, rapid development
- Cons: Monolithic

BeautifulSoup

- Pros: Easy to use, fast for simple tasks
- Cons: Static pages only