**Introduction to Data Analytics**

**What is Data Analytics?**

Data Analytics involves examining raw data to uncover patterns, trends, and insights. It helps businesses and organisations make better decisions by transforming data into actionable information.

**Importance in Decision-Making**

* Provides insights to improve business processes.
* Helps predict trends and customer behavior.
* Supports evidence-based decision-making.

**Difference Between Data Analytics, Data Science, and Business Intelligence**

* **Data Analytics**: Focuses on analysing data for insights and trends.
* **Data Science**: Broader field including analytics, machine learning, and data modeling.
* **Business Intelligence (BI)**: Uses tools and dashboards to visualise and monitor historical data for business performance.

**Types of Data Analytics**

1. **Descriptive Analytics**:
   * **What it does**: Summarises past data.
   * **Example**: Sales reports showing last month’s revenue.
2. **Diagnostic Analytics**:
   * **What it does**: Identifies reasons for past events.
   * **Example**: Analysing why sales dropped in a region.
3. **Predictive Analytics**:
   * **What it does**: Forecasts future trends based on past data.
   * **Example**: Predicting next quarter’s sales.
4. **Prescriptive Analytics**:
   * **What it does**: Recommends actions to achieve desired outcomes.
   * **Example**: Suggesting the best marketing strategy to boost sales.

**Understanding Data**

1. **Data Types**
   * **Qualitative Data**: Descriptive, non-numeric (e.g., names, colors, opinions).
   * **Quantitative Data**: Numeric, measurable (e.g., age, salary, height).
2. **Structured, Semi-structured, and Unstructured Data**
   * **Structured**: Organised in rows/columns (e.g., spreadsheets, databases).
   * **Semi-structured**: Partially organised with some structure (e.g., JSON, XML).
   * **Unstructured**: No fixed format (e.g., images, videos, free-text).
3. **Data Sources**
   * **Internal**: Comes from within an organisation (e.g.,databases).
   * **External**: Comes from outside (e.g., APIs, web scraping, surveys).
4. **Data Formats**
   * **Examples**:
     + **CSV**: Plain text with data in rows/columns.
     + **Excel**: Spreadsheet files with more advanced features.
     + **JSON**: Text format for data exchange, readable by machines and humans.
     + **SQL**: Query language for structured databases.

**Data Collection and Cleaning** involves gathering and preparing data for analysis.

1. **Data Collection Methods**:
   * **Surveys**: Collecting responses directly from people.
   * **APIs**: Accessing data from web services.
   * **Scraping Tools**: Extracting data from websites.
2. **Data Cleaning**:
   * Fixes issues like missing data, outliers, duplicates, and inconsistencies.
   * **Techniques**:
     + **Imputation**: Filling in missing data.
     + **Scaling**: Adjusting data range (e.g., normalising).
     + **Encoding**: Converting categorical data into numeric format.

**Data Visualisation**

The process of turning data into visual formats like charts, graphs, and maps to help understand trends, patterns, and insights.

**Principles of Visualisation**

1. **Clarity**: Ensure the message is clear and easy to understand.
2. **Accuracy**: Represent data truthfully without distortion.
3. **Simplicity**: Avoid clutter; keep visuals clean and straightforward.
4. **Relevance**: Use visuals that fit the audience and purpose.

**Choosing the Right Chart Type**

* **Bar Charts**: Compare categories or groups.
* **Line Charts**: Show trends over time.
* **Pie Charts**: Display proportions (use sparingly).
* **Scatter Plots**: Show relationships between two variables.
* **Heatmaps**: Represent data intensity or density.

**Avoiding Misleading Graphs**

1. Use consistent scales.
2. Start axes at zero unless justified.
3. Avoid exaggerated visual effects.
4. Clearly label data and sources.

**Visualisation Tools**

1. **Python Libraries**:
   * **Matplotlib**: Basic charts with high customisation.
   * **Seaborn**: Simplifies complex visualisations, works well with Matplotlib.
   * **Plotly**: Interactive and web-friendly visualisations.
2. **BI Tools**:
   * **Tableau**: Drag-and-drop visuals, powerful analytics.
   * **Power BI**: Integrates well with Microsoft tools, easy reporting.
   * **Google Data Studio**: Free, connects with Google services for dashboards.