Unit 6

Computer Tools in Data Processing and Application

https://github.com/sanjeevlcc/notes_2081/tree/main

6.1. Basics of Data Analysis (Spreadsheets, Power BI)

6.2. Data Visualization (Charts, Graphs, Scatter Plots)

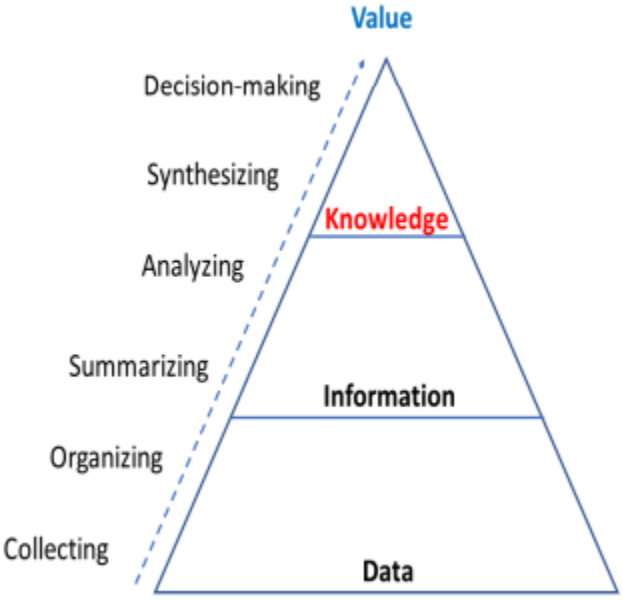
6.3. Collaboration Tools (Google Workspace, Microsoft Teams)

6.4. Using Computer Tools for Business Scenarios (Report writing, data extraction and presentation)

* *Tools for creating business reports and summaries*
* *Tools for financial data analysis and business analytic*
* *Tools for Market Trends analysis and visualization*
* *Tools for scheduling and monitoring*

6.1. Basics of Data Analysis (Spreadsheets, Power BI)

Data and Knowledge Management



- Data and knowledge management involve the processes and systems used to organize, store, and retrieve information, enabling organizations to make data-driven decisions and leverage knowledge effectively.
- Data and knowledge management (DKM) systems collect, manage, and provide controlled access to data and knowledge resources.
- These systems may also provide critical analytical and visualization capabilities to support research and decision processes.
- Data within the DKM may be at any stage of its lifecycle.
- https://github.com/sanjeevlcc/cnlabs/tree/main/Mphil-ICT/Mphil-ICT/RM/Maha%20Kumbh%20Mela%202025

    - **Data**: Raw facts and figures without context
        - (e.g., 100, "John").
    - **Information**: Processed data with meaning
        - (e.g., "John scored 100 in mathematics").
    - **Knowledge**: Insights derived from information to guide decisions.

        - **Example:**

            - **Data: "500 units sold."**

            - **Information: "500 units of product X were sold in region Y last quarter."**

            - **Knowledge: "Product X has high demand in region Y during the summer."**

What is Data Analysis?

Data analysis is the process of inspecting, cleaning, transforming, and modeling data to extract meaningful insights, support decision-making, and identify patterns or trends. It involves using statistical, mathematical, and computational techniques to interpret and derive conclusions from data.

Example of Data Analysis

https://github.com/sanjeevlcc/notes_2081/blob/main/BBA_1_Information_Technology_For_Business/Assignment/Data_Analysis_Tool.ipynb

Scenario:
A retail company wants to understand customer buying patterns to improve sales. They analyze customer purchase data from the last year.

Steps in Data Analysis:

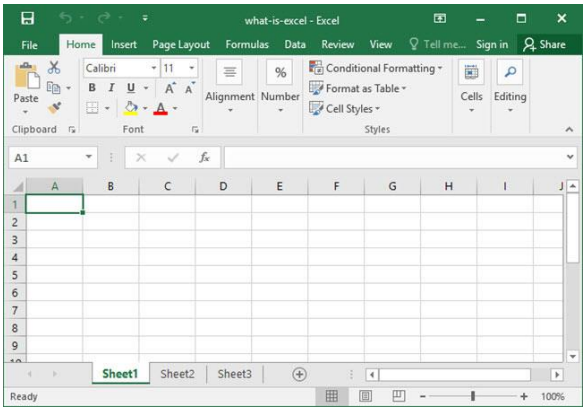
1. **Data Collection:** Gathering data from sales transactions, customer demographics, and online shopping behaviors.
2. **Data Cleaning:** Removing duplicate records, handling missing values, and correcting errors.
3. **Exploratory Data Analysis (EDA):** Identifying patterns, trends, and outliers using summary statistics and visualizations (e.g., bar charts for best-selling products).
4. **Data Transformation:** Creating new metrics, such as customer lifetime value (CLV) or average purchase frequency.
5. **Modeling & Interpretation:** Using statistical models or machine learning to predict future sales or segment customers into groups.
6. **Decision Making:** Based on insights, the company offers personalized discounts, adjusts inventory, or improves marketing strategies.

Meaningful Insight
From the analysis, the company discovers that:

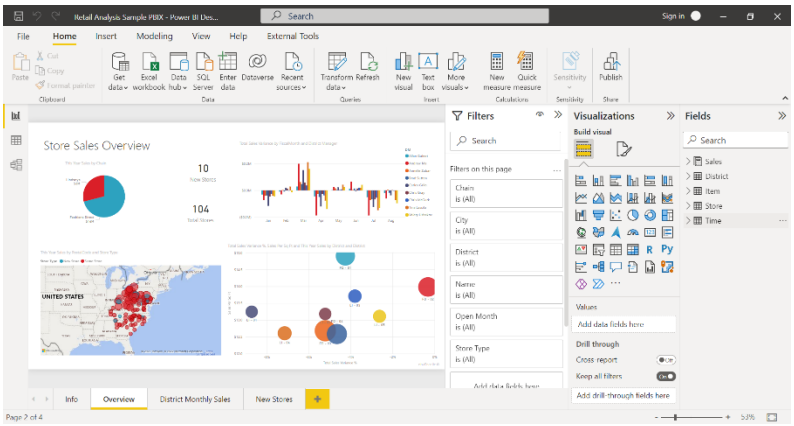
- **Peak Sales Time:** Most purchases happen between 6-9 PM.
- **Top Customers:** 20% of customers contribute to 80% of revenue (Pareto principle).
- **Product Trends:** Seasonal products (e.g., winter jackets) see a sales spike in October-December.
- **Marketing Strategy:** Personalized email campaigns increase customer retention by 15%.

By leveraging these insights, the company can optimize marketing, manage inventory better, and boost sales.

- **Spreadsheets (Excel, Google Sheets)**
  - Used for organizing, analyzing, and storing data.
  - Features: Formulas, Pivot Tables, Data Validation, Conditional Formatting.
  - Applications: Financial analysis, budgeting, sales tracking.

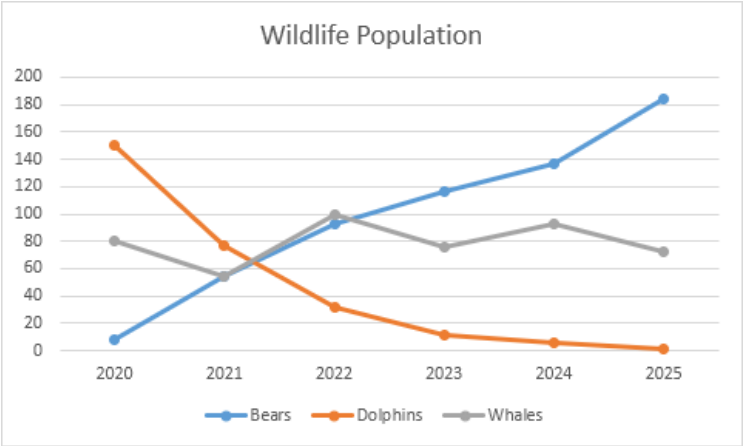


- **Power BI**
  - A business analytics tool for interactive visualizations and business intelligence.
  - Features: Dashboards, Data Modeling, AI-Powered Insights.
  - Applications: Real-time reporting, predictive analytics, KPI monitoring.



6.2. Data Visualization (Charts, Graphs, Scatter Plots)

- **Charts and Graphs**
  - Line charts: Show trends over time.
  - Bar charts: Compare categories.
  - Pie charts: Show proportions.



- **Scatter Plots**
  - Used to show relationships between two variables.
  - Helps identify trends, correlations, and outliers.

- **Tools for Visualization**
    - Excel, Google Sheets, Power BI, Tableau, Google Data Studio.

6.3. Collaboration Tools (Google Workspace, Microsoft Teams)

- **Google Workspace**
    - Tools: Google Docs, Sheets, Slides, Drive, Meet.
    - Features: Real-time editing, cloud storage, accessibility from any device.
- **Microsoft Teams**
    - Features: Chat, video conferencing, file sharing, integration with Office 365.
    - Used for business communication, remote work, and project management.

6.4. Using Computer Tools for Business Scenarios

Report writing, data extraction and presentation

- **Tools for creating business reports and summaries**: Microsoft Word, Google Docs, PowerPoint.
- **Tools for financial data analysis and business analytics**: Excel, Power BI, QuickBooks, SAP.
- **Tools for Market Trends Analysis and Visualization**: Tableau, Google Trends, Power BI.
- **Tools for Scheduling and Monitoring**: Google Calendar, Microsoft Project, Trello, Asana.

LAB

If you want to perform data analysis in Excel using the same steps as the Python program. You structure the workflow so that you can use Excel for Data Collection, Cleaning, EDA, Transformation, Modeling, and Decision Making.

Step 1: Data Collection

- Generate Random Data in Excel using formulas:
  - `Customer_ID`: `=ROW(A1)`
  - `Age`: `=RANDBETWEEN(18, 70)`
  - `Gender`: `=CHOOSE(RANDBETWEEN(1,2), "Male", "Female")`
  - `Purchase_Amount`: `=RAND()*(1000-10)+10`
  - `Category`: `=CHOOSE(RANDBETWEEN(1,4), "Electronics", "Clothing", "Groceries", "Furniture")`
  - `Transaction_Date`: `=TODAY()-RANDBETWEEN(1,365)`

Step 2: Data Cleaning

1. Remove Duplicates:
   - Select your dataset → `Data` tab → Click `Remove Duplicates`
2. Handle Missing Values:
   - Use `=IF(ISBLANK(Purchase_Amount), MEDIAN(Purchase_Amount), Purchase_Amount)`

Step 3: Exploratory Data Analysis (EDA)

- Find Category-wise Average Spending:
  - Use `Pivot Table` (`Insert` → `PivotTable`)
  - Rows: `Category`
  - Values: `Average of Purchase_Amount`
- Create Box Plot in Excel:
  - `Insert` → `Chart` → `Box & Whisker Plot`

Step 4: Data Transformation

- Age Groups:
  - Formula: `=IF(Age<=30, "Young", IF(Age<=50, "Middle-aged", "Senior"))`
- Purchase Binning:
  - Formula: `=IF(Purchase_Amount<=100, "Low", IF(Purchase_Amount<=500, "Medium", "High"))`

Step 5: Modeling & Interpretation (Regression in Excel)

- Regression in Excel (Predict Purchase Amount)
  - `Data` → `Data Analysis` → `Regression`
  - Select `Age` & `Category` as independent variables
  - `Purchase_Amount` as dependent variable
  - Click `OK` to get results like **R^2 Score, Coefficients, and P-values**

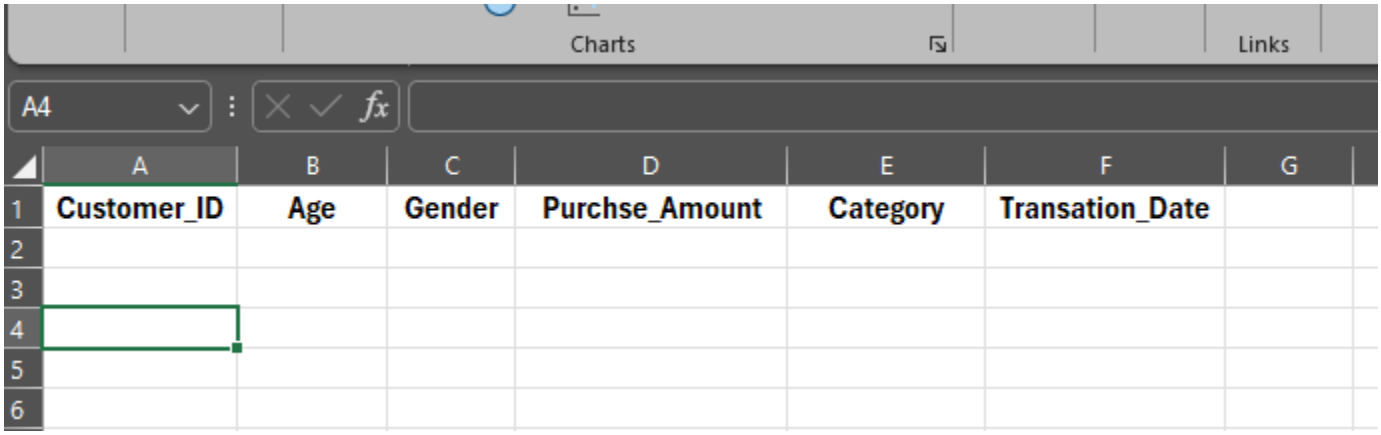
Step 6: Decision Making (Insights & Recommendations)

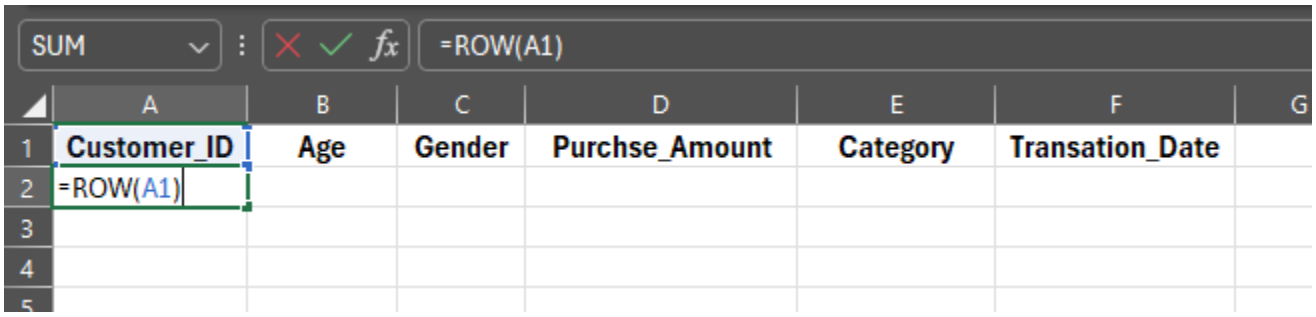
- Key Insights:
  1. Electronics has the highest spending.
  2. Senior customers prefer Furniture.
  3. Predictive model can estimate purchases.
  4. Medium spending is common.
- **Use Pivot Tables & Conditional Formatting** to identify trends for marketing and inventory strategies.

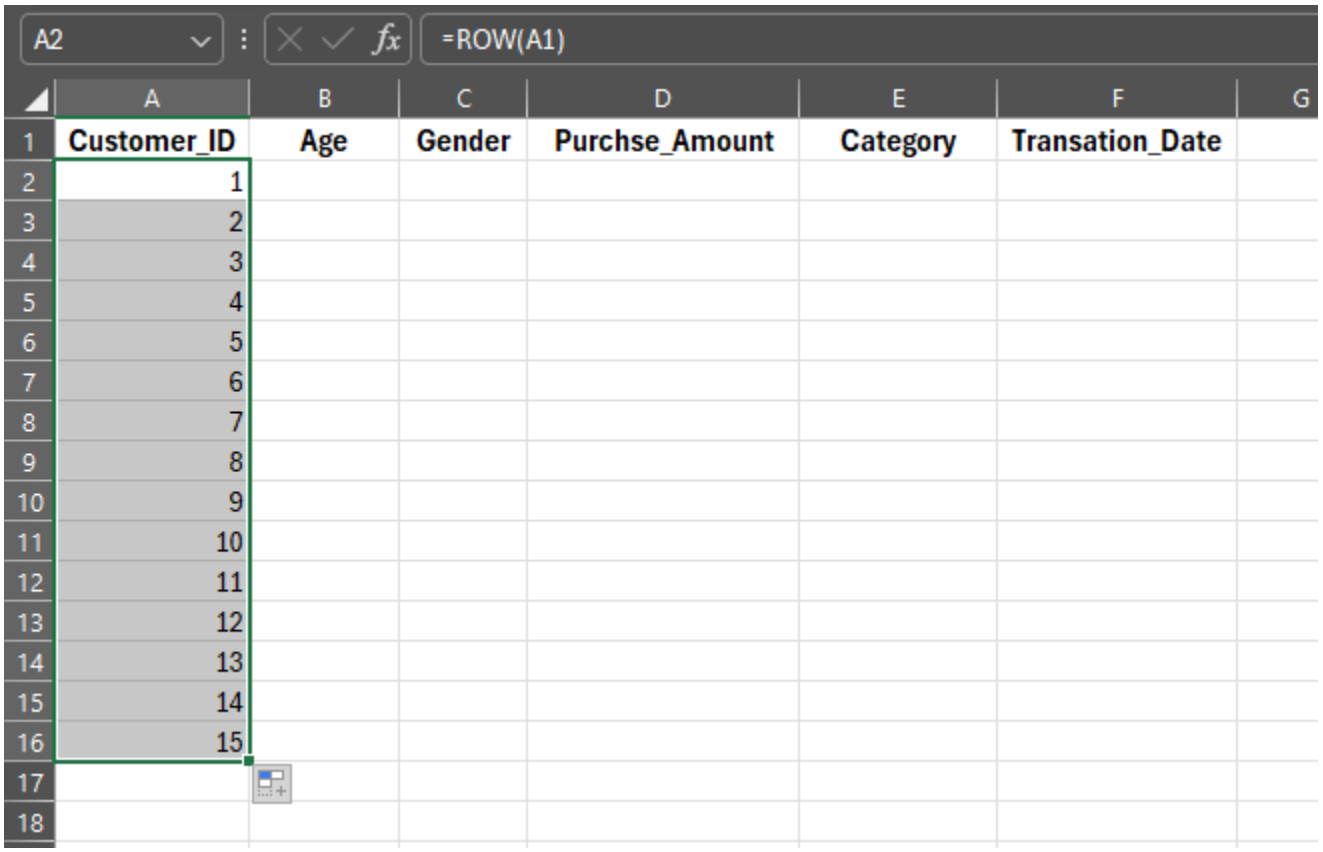
==========STEPS==========

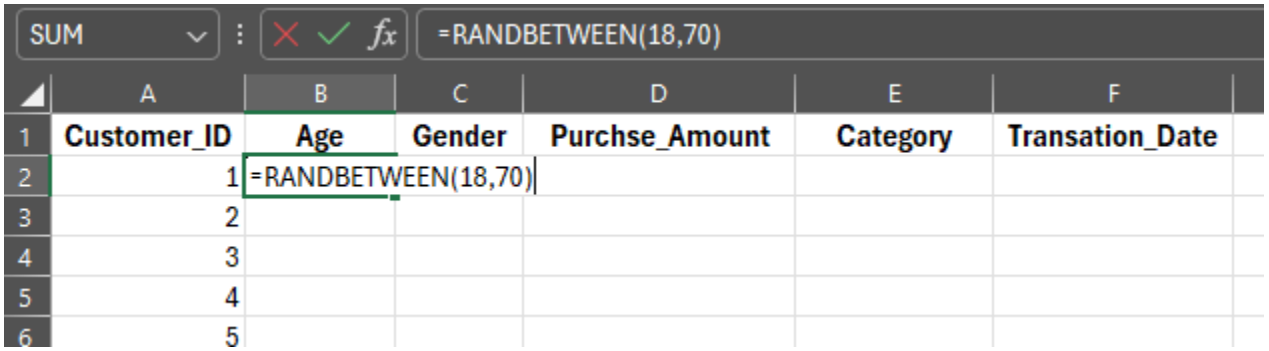
Step 1: Data Collection

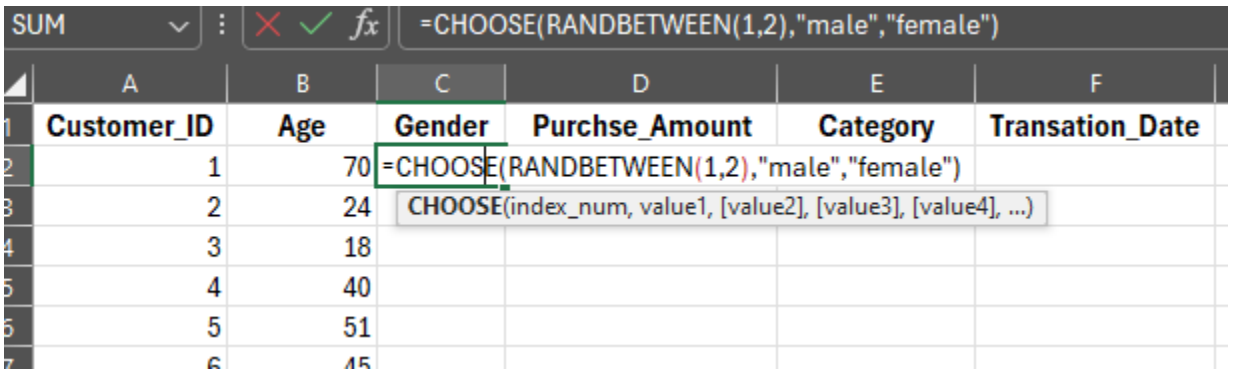
- **Generate Random Data in Excel** using formulas:
  - `Customer_ID`: `=ROW(A1)`
  - `Age`: `=RANDBETWEEN(18, 70)`
  - `Gender`: `=CHOOSE(RANDBETWEEN(1,2), "Male", "Female")`
  - `Purchase_Amount`: `=RAND()*(1000-10)+10`
  - `Category`: `=CHOOSE(RANDBETWEEN(1,4), "Electronics", "Clothing", "Groceries", "Furniture")`
  - `Transaction_Date`: `=TODAY()-RANDBETWEEN(1,365)`


Charts
Links
A4
A B C D E F G
1 Customer_ID Age Gender Purchse_Amount Category Transation_Date
2
3
4
5
6


SUM
=ROW(A1)
A B C D E F G
1 Customer_ID Age Gender Purchse_Amount Category Transation_Date
2 =ROW(A1)
3
4


A2
=ROW(A1)
A B C D E F G
1 Customer_ID Age Gender Purchse_Amount Category Transation_Date
2 1
3 2
4 3
5 4
6 5
7 6
8 7
9 8
10 9
11 10
12 11
13 12
14 13
15 14
16 15
17
18





Finally make like this ……

Customer_ID	Age	Gender	Purchase_Amount	Category	Transaction_Date
1	26	female	539.6828454	groceries	7/11/2082
2	46	FALSE	932.1317271	clothing	12/13/2081
3	47	male	37.0539436	furniture	1/4/2082
4	70	male	577.2026879	electronics	4/25/2082
5	58	female	624.0148674	groceries	12/7/2082
6	58	male	471.7724837	groceries	9/14/2082
7	51	male	131.5541493	groceries	7/29/2082
8	55	female	881.2293414	electronics	4/6/2082
9	43	female	27.80119743	clothing	11/29/2081
10	25	male	445.7808605	groceries	3/19/2082

Step 2: Data Cleaning

1. **Remove Duplicates:**
   - Select your dataset → `Data` tab → Click `Remove Duplicates`
2. **Handle Missing Values:**
   - Use `=IF(ISBLANK(Purchase_Amount), MEDIAN(Purchase_Amount), Purchase_Amount)`

Step 3: Exploratory Data Analysis (EDA)

- Find Category-wise Average Spending:
  - Use `Pivot Table` (`Insert` → `PivotTable`)
  - Rows: `Category`
  - Values: `Average of Purchase_Amount`
- Create Box Plot in Excel:
  - `Insert` → `Chart` → `Box & Whisker Plot`

Step 4: Data Transformation

- Age Groups:
  - Formula: `=IF(Age<=30, "Young", IF(Age<=50, "Middle-aged", "Senior"))`
- Purchase Binning:
  - Formula: `=IF(Purchase_Amount<=100, "Low", IF(Purchase_Amount<=500, "Medium", "High"))`

Step 5: Modeling & Interpretation (Regression in Excel)

- Regression in Excel (Predict Purchase Amount)
  - `Data` → `Data Analysis` → `Regression`
  - Select `Age` & `Category` as independent variables
  - `Purchase_Amount` as dependent variable
  - Click `OK` to get results like **R^2 Score, Coefficients, and P-values**

Step 6: Decision Making (Insights & Recommendations)

- Key Insights:
  1. Electronics has the highest spending.
  2. Senior customers prefer Furniture.
  3. Predictive model can estimate purchases.
  4. Medium spending is common.
- **Use Pivot Tables & Conditional Formatting** to identify trends for marketing and inventory strategies.

Q/A

Fill in the Blanks

Multiple Choice Questions (MCQ)

Short Questions

Comprehensive Questions

Answers

Fill in the Blanks

Multiple Choice Questions (MCQ)