Week 1 - Project: Data Analysis for Business Insights

Please write the answers in the 'Project Coversheet' and refer to the dataset provided for completing the tasks.

Objective:

You will apply your knowledge of data collection, cleaning, exploration, visualisation, and storytelling to analyse a real-world dataset and provide actionable insights. A data set is provided with this.

Scenario:

You are a Data Analyst working for ShopEase, an online retail company. The company wants to improve sales performance and customer satisfaction by analysing its sales data. Your task is to:

1) Collect & Clean Data - A dataset with sales transactions is provided. Your first task is to inspect the data, remove duplicates, handle missing values, and standardise formats.

2) Explore Data - Identify key trends, seasonal sales patterns, and correlations (e.g., sales vs. advertising spend).

3) Visualise Findings - Create at least three visualisations (e.g., bar charts, line graphs, heatmaps) to present key insights.

4) Analyse Customer Behaviour - Identify the most popular product categories, best-selling months, and any customer purchase trends.

5) Ethical Considerations - Check if any customer data (emails, phone numbers) should be anonymised. Ensure compliance with data privacy laws (e.g., GDPR).

6) Tell a Data Story - Summarise your findings in a brief Data Insights Report with key takeaways and recommendations for ShopEase to improve sales and customer experience.

Deliverables:

Submit your work in the project coversheet containing:

1. Cleaned Dataset (Summary in bullet points)

2. Exploratory Data Analysis Summary

3. Three Data Visualisations (Graphs/Charts)

4. Final Data Insights Report (1-2 pages)

Bonus Challenge:

Suggest two strategies to boost sales in the lowest-performing months based on data insights.

Recommended Steps or Actions to Complete the Tasks:

1) Cleaned Dataset

Steps to Complete:

1) Inspect the Dataset - Open the dataset and explore its structure. Check for the number of columns, types of data (e.g., numeric, categorical), and the number of rows.

2) Remove Duplicates - Look for duplicate rows in your dataset, which can skew analysis. Use Excel's 'Remove Duplicates' feature or Python (with pandas drop_duplicates() method) to remove them.

3) Handle Missing Values - Inspect columns for missing values. There are multiple ways to handle this:

- Drop rows with missing critical data (e.g., sales amounts).

- Fill missing values with the mean or median for numerical columns (using Excel or pandas fillna() in Python).

- Impute categorical data by replacing missing values with the most frequent category.

4) Standardise Formats - Ensure data formats are consistent (e.g., dates in YYYY-MM-DD format, numerical columns without symbols). This can be done in Excel with functions like 'Text to Columns' or via pandas (pd.to_datetime() for dates, astype() for other types).

5) Export the Cleaned Dataset - Once the dataset is cleaned, save it as a new Excel or CSV file.

2) Exploratory Data Analysis

Steps to Complete:

1) Understand the Dataset - Get familiar with the columns and their meaning. Identify numerical and categorical features. Look at the first few rows of the dataset using head() in Python or Excel's 'Preview' feature.

2) Identify Key Trends and Patterns - Start by calculating summary statistics (e.g., mean, median, standard deviation for numerical columns). Look for any obvious patterns in sales, seasonality, or product trends.

3) Find Correlations - Use correlation matrices or scatterplots to examine relationships (e.g., how sales relate to advertising spend, discount percentages, or marketing efforts).

Excel: Use the 'CORREL' function for correlation or create scatter plots.

Python: Use df.corr() to generate a correlation matrix and seaborn or matplotlib for visual exploration.

4) Write the EDA Summary - Summarise the key observations. Highlight any significant trends, such as peak sales months, correlations between sales and marketing efforts, and any unusual findings.

3) Three Data Visualisations

Steps to Complete:

1) Select Insights to Visualise: Choose 3 key findings to visualise. For example:

- Sales over time: Line graph showing sales performance across months or years.

- Sales by product category: Bar chart comparing the sales of different product categories.

- Heatmap of correlations: Correlation matrix heatmap to show relationships between various metrics (e.g., sales, advertising, discounts).

2) Create the Visualisations:

- Excel: Use built-in chart tools (e.g., insert bar, line, or scatter charts).

- Python: Use libraries like matplotlib, seaborn, or plotly to create visualisations.

3) Ensure Clarity and Insight: Label axes clearly, add titles, and use color coding if applicable to make your visualisations easy to interpret.

4) Final Data Insights Report

Steps to Complete:

1) Summarise Key Findings: Begin your report by summarising the most important insights, such as:

- What are the best-selling months or product categories?

- Are there any correlations between sales and factors like advertising spend?

- What customer behaviors or purchasing trends were identified?

2) Provide Actionable Recommendations: Based on the analysis, offer recommendations that can help the business improve sales. This could include:

- Focusing marketing efforts on peak sales periods.

- Increasing stock for best-selling products.

- Adjusting sales strategies for slow-performing months.

3) Ethical Considerations - Mention any ethical issues, such as the importance of anonymising customer data and ensuring compliance with privacy laws (e.g., GDPR).

Bonus Challenge: Strategies to Boost Sales

Steps to Complete:

1) Identify Low-Performing Months - From the data analysis, identify the months with the lowest sales.

2) Suggest Two Strategies: Based on your findings, propose two strategies that could help boost sales during those months.

- Implement targeted promotions during off-peak months.

- Offer discounts or bundle deals to encourage purchases during slow periods.