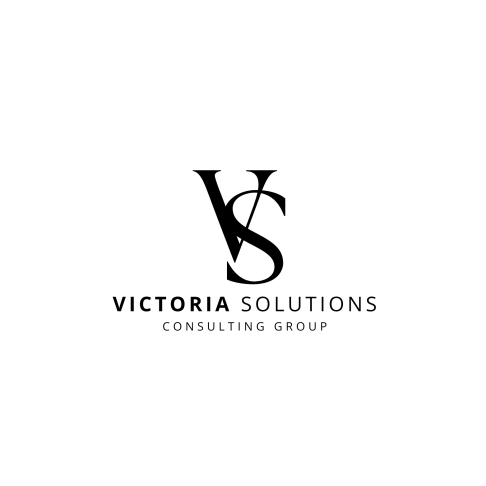
|  |  |
| --- | --- |
| Name | Mohammad Asghar |
| Contact Number | +447361131632 or +923018024440 |
| Project Title (Example – Week1, Week2, Week3) | Week 1 – ShopEase |



**Project Guidelines and Rules**

1. **Formatting and Submission**
   * **Format:** Use a readable font (e.g., Arial/Times New Roman), size 12, 1.5 line spacing.
   * **Title:** Include Week and Title (Example - Week 1: TravelEase Case Study.)
   * **File Format:** Submit as PDF or Word file to contact@victoriasolutions.co.uk
   * **Page Limit:** 4–5 pages, including the title and references.
2. **Answer Requirements**
   * **Word Count:** Each answer should be 100–150 words; total 800–1,200 words.
   * **Clarity:** Write concise, structured answers with key points.
   * **Tone:** Use formal, professional language.
3. **Content Rules**
   * Answer all questions thoroughly, referencing case study concepts.
   * Use examples where possible (e.g., risk assessment techniques).
   * Break complex answers into bullet points or lists.
4. **Plagiarism Policy**
   * Submit original work; no copy-pasting.
   * Cite external material in a consistent format (e.g., APA, MLA).
5. **Evaluation Criteria**
   * **Understanding:** Clear grasp of business analysis principles.
   * **Application:** Effective use of concepts like cost-benefit analysis and Agile/Waterfall.
   * **Clarity:** Logical, well-structured responses.
   * **Creativity:** Innovative problem-solving and examples.
   * **Completeness:** Answer all questions within the word limit.
6. **Deadlines and Late Submissions**
   * **Deadline:** Submit on time; trainees who submit fail to submit the project will miss the “Certificate of Excellence”

1. **Additional Resources**
   * Refer to lecture notes and recommended readings.
   * Contact the instructor or peers for clarifications before the deadline.

**START YOUR PROJECT FROM HERE:**

This project is based on conducting a data analysis on the sales dataset for an online retail company named ‘ShopEase’. The company wants to improve its sales performance and customer satisfaction through the analysis of the sales dataset.

The key steps of the data analysis from the beginning to the end include:

* Cleaned Dataset
* Exploratory Data Analysis (EDA)
* Data Visualisations (Three included)
* Final Data Insights Report

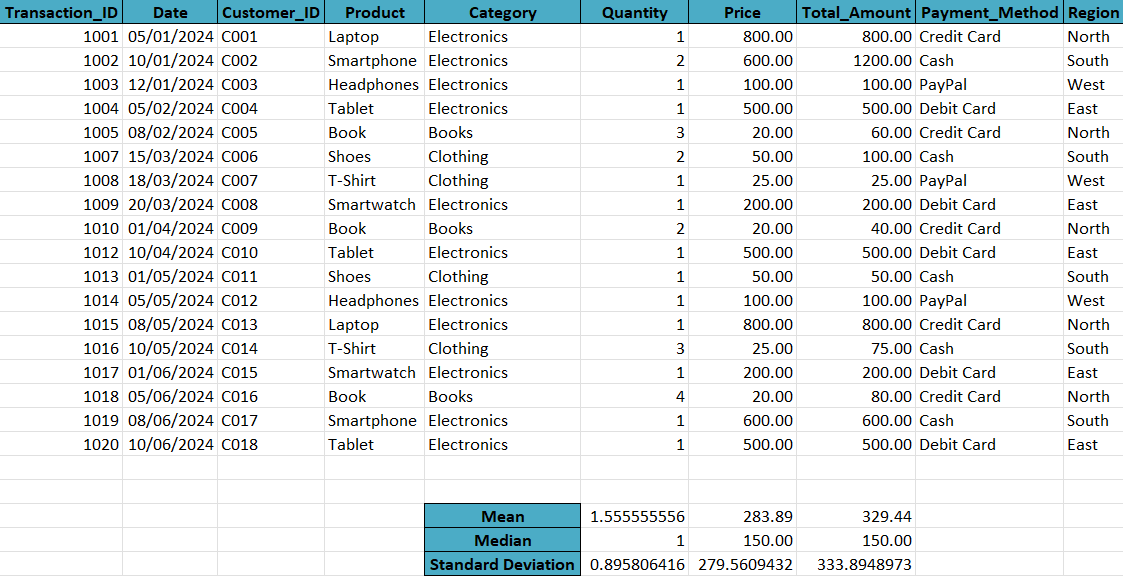
**1. Cleaned Dataset**

This step involves removing inaccuracies, filling in missing values, handling duplicates, and formatting data to make it usable for analysis.

* 1. Inspect the Dataset:
* Explored the structure of the dataset including checking number of columns and rows as well as types of data (numeric or categorical).
* Number of Columns: 10
* Number of Rows: 21 (including header row)
* Type of data: General for the whole dataset
  1. Remove Duplicates:
* The duplicate rows that were removed included the ‘C001’ & ‘C002’ rows under the ‘Customer\_ID’ column.
* Used Conditional Formatting > Highlighting Cell Rules > Duplicates to highlight the duplicates and then used the ‘Remove Duplicates’ feature to remove those duplicates.
  1. Handling Missing Values:
* Column H (Total\_Amount) had one missing value in Cell H2 which was filled up by multiplying the quantity in F2 with the Price in G2 giving us a final figure of 800.00 in H2.
  1. Standardize Formats:
* To ensure the data formats are consistent, the data type for the ‘Date’ column had to be changed from ‘General’ type to ‘Data’ type.
* Did this by first highlighting the data column (excluding the header row) and following the command ‘Data > Text to columns > Delimited > Next > Tab > Date: DMY > Finish.
* The final date format for the ‘Date’ column was in the format ‘DD/MM/YYYY’ instead of ‘YYYY-MM-DD’. This might be because I am using the format used in my home country ‘Pakistan’ which is different from the one used in the UK.
  1. Exported the Cleaned Dataset:
* After cleaning the dataset, it was saved as a new Excel file named ‘sales\_data\_cleaned’.

1. **Exploratory Data Analysis (EDA)**
   1. Understand the Dataset:

* There are 6 columns with numerical features and 4 columns with categorical features.
* The categorical featured columns include; Product, Category, Payment\_Method, & Region as they include text data types and incorporate different categories.
* For examining the first few rows of the dataset in Excel, select the rows that needs to be examined and then use the command ‘View > Freeze Panes > Freeze Panes’. This keeps the selected rows and columns visible while we can move through the rest of the worksheet.
  1. Identify Key Trends and Patterns:
* The summary statistics including the mean, median and standard deviation for numerical columns were calculated through the Excel functions.
* There is another simple way of calculating them through Power Query by using the command ‘View > Clicking Column Quality, Column Distribution & Column Profile’.
* Below is a snippet of the data file with calculated summary statistics.

****

* For all the numerical columns, the mean is greater than the median which represents a positively skewed or right skewed representation.
* The standard deviation is small compared to the mean for all numerical columns. This suggests that most of the data points are close to the average and the dataset is consistent.
  1. Find Correlations:
* In terms of examining relationships between the numerical data columns in Excel, the ‘CORREL’ function was used and we got the following Pearson correlation coefficients:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Quantity** | **Price** | **Total\_Amount** |
| **Quantity** | 1.00 | -0.45 | -0.27 |
| **Price** | -0.45 | 1.00 | 0.92 |
| **Total\_Amount** | -0.27 | 0.92 | 1.00 |

* The Total\_Amount column equates to Sales. The price against Sales represents a positive correlation with a correlation coefficient of 0.92.
* The Quantity against Price and against Sales represent a negative correlation with coefficient values of -0.45 and -0.27.
  1. Write the EDA Summary:
* In order to find the peak sale months, the sales total of each month was computed first using ‘SUMIF’ Excel function.
* Before that, the day, month and year columns were produced through the ‘Date’ column that was used for the ‘SUMIF’ function.

|  |  |
| --- | --- |
| **Month** | **Total\_Amount/Sales** |
| January | 2100 |
| February | 560 |
| March | 325 |
| April | 540 |
| May | 1025 |
| June | 1380 |

* After finding the sales total of each month from Jan till June 2024, the ‘XLOOKUP’ function was used to find the peak sale month, the second peak sale month and the third peak sale month.

|  |  |
| --- | --- |
| **Peak Sale Month** | January |
| **2nd Peak Sale Month** | June |
| **3rd Peak Sale Month** | May |

1. **Three Data Visualisations**
   1. Sales over Time:

* Total\_Amount column (Sales) over the month (time) line graph representing sales performance over first six months of the year. This was made using the table for the Month and Total\_Amount or Sales above.
* According to this graph, the month of January saw the highest sales revenue after which it had a major decline until March with sales per month of 325.
* From March 2024 onwards, the sales revenue grew drastically reaching 1380 by the month of June 2024.
  1. Sales by Product Category:
* A bar chart was made for this comparing the sales of different product categories.
* This was done using the Pivot Table feature in Excel. For the ‘Rows’, the ‘Product’ column was selected, ‘Category’ column for ‘Columns’ and for the ‘Values’ the ‘Total\_Amount or Sales’ column was selected.
* This generated the following pivot table, which was then used to create a bar chart representing different products along with their respective categories.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sum of Sales** | **Column Labels** |  |  |  |
| **Row Labels** | **Books** | **Clothing** | **Electronics** | **Grand Total** |
| Book | 180.00 |  |  | 180.00 |
| Headphones |  |  | 200.00 | 200.00 |
| Laptop |  |  | 1600.00 | 1600.00 |
| Shoes |  | 150.00 |  | 150.00 |
| Smartphone |  |  | 1800.00 | 1800.00 |
| Smartwatch |  |  | 400.00 | 400.00 |
| Tablet |  |  | 1500.00 | 1500.00 |
| T-Shirt |  | 100.00 |  | 100.00 |
| **Grand Total** | **180.00** | **250.00** | **5500.00** | **5930.00** |

* As seen through the pivot table and bar chart above, the Electronics category made the highest sales of 5500 out of which the ‘Smartphone’ generated the largest revenue amongst the other products.
* On 2nd place, we had the ‘Clothing’ category with sales revenue of 250 and the ‘Shoes’ as its product with the largest revenue.
* Lastly, the ‘Books’ category had the lowest sales of 180 as it has only one product ‘Book’.
  1. Heatmap of Correlations:
* The heatmap of correlations was made by just applying colour coding on the the correlation matrix of computed pearson correlation coefficients as given in Section 2.3. Finding Correlations above.
* The ‘Conditional Formatting’ feature within Excel was used for this. The following command was used ‘Home > Conditional Formatting > Color Scales > Green – Yellow – Red Scale.
* The colour codes used were as follows:
  + Green: +1 correlation coefficient (Positive Correlation)
  + Yellow: 0 correlation coefficient (No correlation)
  + Red: -1 correlation coefficient (Negative Correlation)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Quantity** | **Price** | **Total\_Amount** |
| **Quantity** | 1.00 | -0.45 | -0.27 |
| **Price** | -0.45 | 1.00 | 0.92 |
| **Total\_Amount** | -0.27 | 0.92 | 1.00 |

* The following heatmap was obtained:
  1. Sales within Different Regions:
* In order to visualise the sales within different regions, the pivot table feature was used again. The pivot table included the following columns from the sales data:
  + Rows: Regions
  + Columns: Payment\_Method
  + Values: Total\_Amount (Sales)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sum of Total\_Amount** | **Column Labels** |  |  |  |  |
| **Row Labels** | **Cash** | **Credit Card** | **Debit Card** | **PayPal** | **Grand Total** |
| East |  |  | 1900.00 |  | 1900.00 |
| North |  | 1780.00 |  |  | 1780.00 |
| South | 2025.00 |  |  |  | 2025.00 |
| West |  |  |  | 225.00 | 225.00 |
| **Grand Total** | **2025.00** | **1780.00** | **1900.00** | **225.00** | **5930.00** |

* The pivot table above was generated which was then used to create a bar chart representing sales within different regions for the different payment methods as illustrated below.
* The chart represents that the ‘South’ Region outperformed the other regions in terms of Sales performance with ‘Cash’ as the leading payment method in terms of raising sales.
* The ‘West’ region saw the lowest sales amongst the regions and ‘Paypal’ method was used as the payment method.

1. **Final Data Insights Report**
   1. **Summarise Key Findings:**
      1. **Best Selling Months or Product Categories:**

|  |  |
| --- | --- |
| **Month** | **Total\_Amount/Sales** |
| January | 2100 |
| May | 1025 |
| June | 1380 |

**The best-selling months were January, June, and May, with the corresponding sales values shown below:**

* The best-performing product category was ‘Electronics’, accounting for 5,500 out of 5,930 total sales.
  + 1. **Correlations:**
* Sales has a positive correlation with Price with a pearson correlation coefficient of 0.92.
* Sales has a negative correlation with Quantity with a -0.27 coefficient.
  + 1. **Customer Behaviours or Purchasing Trends:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sum of Total\_Amount** | **Column Labels** |  |  |  |  |
| **Row Labels** | **East** | **North** | **South** | **West** | **Grand Total** |
| C001 |  | 800.00 |  |  | 800.00 |
| C002 |  |  | 1200.00 |  | 1200.00 |
| C003 |  |  |  | 100.00 | 100.00 |
| C004 | 500.00 |  |  |  | 500.00 |
| C005 |  | 60.00 |  |  | 60.00 |
| C006 |  |  | 100.00 |  | 100.00 |
| C007 |  |  |  | 25.00 | 25.00 |
| C008 | 200.00 |  |  |  | 200.00 |
| C009 |  | 40.00 |  |  | 40.00 |
| C010 | 500.00 |  |  |  | 500.00 |
| C011 |  |  | 50.00 |  | 50.00 |
| C012 |  |  |  | 100.00 | 100.00 |
| C013 |  | 800.00 |  |  | 800.00 |
| C014 |  |  | 75.00 |  | 75.00 |
| C015 | 200.00 |  |  |  | 200.00 |
| C016 |  | 80.00 |  |  | 80.00 |
| C017 |  |  | 600.00 |  | 600.00 |
| C018 | 500.00 |  |  |  | 500.00 |
| **Grand Total** | **1900.00** | **1780.00** | **2025.00** | **225.00** | **5930.00** |

**Note:** During the initial stage of data analysis, I mistakenly removed two rows from the Customer\_ID column, assuming they were duplicates. However, they actually represented two separate purchases made by the same customer on different dates. I was unable to restore these rows without affecting the subsequent analysis, so I am documenting the error here for transparency.

* In terms of examining customer behaviours, a bar chart of Sales per Customer was created. A pivot table was created first through the following commands:
  + Rows: Customer\_ID
  + Columns: Region
  + Values: Total\_Amount (Sales)
* Above is the pivot table and below is the bar chart representing the data graphically:
* The chart shows that the customer with ID **‘C002’** made the largest purchase, resulting in a sale of **1200**. This customer was from the **South** region.
* In contrast, the customer with ID **‘C007’** made the smallest purchase, with a sale of just **25**, and was from the **West** region.
  1. **Provide Actionable Recommendations:**

Based on the insights from the data and visualisations above, there are some actionable recommendations that could help ShopEase boost its sales further. These include the following:

* A **monthly sales target** should be set and communicated to ensure consistent sales performance across all months.
* Analyze the **factors contributing to peak sales months** and assess whether similar strategies can be applied to improve underperforming months.
* Collect **customer feedback** to better understand their product preferences. For example, the **‘Books’** category underperformed compared to others, suggesting a need to explore the types of books customers are interested in and adjust offerings accordingly.
* For **best-selling products**, consider increasing inventory levels to meet demand and further boost sales.
* Examine **regional purchasing behaviour**. Although the **South** region had the highest sales, most transactions were made using **cash**. Ensure customers in all regions have access to a **variety of payment options** based on their preferences.
  1. **Ethical Considerations:**

It is crucial to take into account major ethical considerations while conducting a data analysis or normally working with data within an organisation. These include the following:

* Anonymise the data. Remove any identifiable information from the data so it can be shared securely such as removing names, addresses and email addresses.
* Mitigate bias: Review datasets for imbalances so that the data can be used to produce fair and factual insights leading to helpful decision making.
* Obtain consent: It is essential to inform users about how their data would be used. Even before that, it is important to obtain their consent of sharing data.
* Secure storage: Encrypt sensitive data and control access.

1. **Bonus Challenge: Strategies to Boost Sales**
   1. Identify Low-Performing Months:

* The sales values for lowest-performing months in the dataset were identified using the SMALL() function in Excel.
* After this, the ‘XLOOKUP’ function within Excel was used to find the 3 lowest performing months with the lowest sales which are shown below:

|  |  |
| --- | --- |
| **Lowest-performing Month** | March |
| **2nd Lowest-performing Month** | April |
| **3rd Lowest-performing Month** | February |

* 1. Suggest Strategies:

In order to boost sales during these low-performing months, ShopEase could consider the following strategies:

* Carefully examine consumer preferences and adjust product features to align with customer demands.
* Organize in-store events, workshops, or product demonstrations to add value and attract more shoppers.