DOCUMENTATION - TASK 1 CUSTOMER BEHAVIOUR ANALYSIS

The objective of this project is to analysis customer transaction data to gain insights into customer behaviour, identify patterns, and make data-driven recommendations for improving business strategies. This document explains the methodology used, assumptions made, and limitations of the analysis.

Methodology:

1.Data Preprocessing:

- **Loading Data**: Importing the dataset from a CSV file.
- **Data Cleaning**: Handling missing values, removing duplicates, and correcting data types.
- **Feature Engineering**: Creating new features, such as 'Total Purchase Amount' from 'Product Price' and 'Quantity'.
- **Data Transformation**: Converting data types, such as converting 'Purchase Date' to datetime format.

2.Exploratory Data Analysis (EDA):

- **Descriptive Statistics**: Calculating mean, median, and mode for numerical features.
- **Data Visualization**: Using histograms, box plots, and scatter plots to understand data distribution and relationships.

3.RFM Analysis:

- **Recency**: Days since the last purchase.
- **Frequency**: Number of purchases.
- **Monetary**: Total amount spent.
- **Scoring**: Assigning scores to each RFM dimension and categorizing customers based on their scores.

4.Customer Segmentation:

- **K-Means Clustering**: Grouping customers into clusters based on RFM scores.
- **Profile Analysis**: Analysing the characteristics of each cluster.

5.Insights and Recommendations:

- **Identifying Trends**: Using visualizations to understand trends in purchase behaviour.
- **Recommendations**: Providing actionable recommendations based on the analysis.

.

Assumptions:

Several assumptions were made during the analysis:

- **Complete Data**: The dataset is assumed to be complete and representative of the entire customer base.
- **Data Accuracy**: The data is assumed to be accurate and free from significant errors.
- **Stationarity**: For trend analysis, the data is assumed to be stationary or to follow a predictable trend over time.
- **Homogeneity**: Customers within a segment are assumed to have similar behavior.

Limitations:

The analysis has several limitations:

- **Data Quality**: The quality of insights is highly dependent on the quality of the data. Any inaccuracies in the dataset can lead to incorrect conclusions.
- **Limited Scope**: The analysis is limited to the features available in the dataset. Other relevant factors not included in the dataset are not considered.
- **Assumption Validity**: The assumptions made (such as stationarity and homogeneity) may not hold true in real-world scenarios, potentially affecting the analysis outcomes.
- **Simplistic Models**: The use of basic models (such as K-Means for clustering) may not capture complex customer behaviours. More advanced models could provide deeper insights.
- **Temporal Dynamics**: The analysis may not account for changes in customer behaviour over time due to external factors (e.g., economic conditions, seasonal effects).

Conclusion:

The customer behaviour analysis project aims to provide valuable insights into customer purchasing patterns and preferences. By following a structured methodology, it is possible to derive actionable recommendations for improving business strategies. However, it is important to be aware of the assumptions made and the limitations of the analysis. Future work could involve using more advanced models, incorporating additional data sources, and continuously updating the analysis to adapt to changing customer behaviours.