

Data Analysis Report

1. Problem Statement The objective of this analysis is to perform customer segmentation using the Online Retail Dataset. The study focuses on identifying key customer groups based on their purchasing behavior and deriving actionable insights to enhance business strategies.

2. Methodology

- **Data Collection:** The dataset was sourced from the UCI Machine Learning Repository.
- **Data Cleaning:**
 - Removed missing CustomerID values.
 - Filtered out negative or zero quantity transactions.
 - Converted InvoiceDate to a datetime format for time-based analysis.
- **Feature Engineering:**
 - Created Recency, Frequency, and Monetary (RFM) features.
 - Segmented customers based on RFM scores.
- **Exploratory Data Analysis (EDA):**
 - Identified purchasing patterns and trends.
 - Analyzed customer segments through statistical summaries and visualizations.
- **Machine Learning Model Development:**
 - Applied clustering techniques (e.g., K-Means) for customer segmentation.
 - Evaluated different clustering methods for optimal segmentation.

3. Results & Findings

- Key insights:
 - The dataset contained **541,909 transactions** across multiple countries.
 - After cleaning, **397,924 valid transactions** were used for segmentation.
 - Customers were segmented into different groups based on RFM scores:
 - VIP Customers (High Recency, High Frequency, High Monetary)
 - Regular Customers
 - Churned Customers (Low Recency, Low Frequency, Low Monetary)
- Model performance:
 - The optimal number of clusters was determined using the Elbow Method.
 - K-Means clustering provided meaningful segmentation of customers.

- Visualizations confirmed distinct customer groups with varying purchasing behaviors.

4. Business Recommendations Based on the findings, the following actions are recommended:

- **Targeted Marketing Campaigns:**

- Offer exclusive discounts to VIP customers to enhance retention.
- Re-engage churned customers with personalized incentives.

- **Inventory Optimization:**

- Stock more products preferred by high-value customers.
- Reduce inventory for items with low engagement.

- **Customer Loyalty Programs:**

- Introduce tier-based rewards for frequent shoppers.
- Provide personalized product recommendations.

- **Model Scalability & Monitoring:**

- Implement an automated pipeline for continuous customer segmentation.
- Use real-time data updates to adjust marketing strategies dynamically.

This analysis provides valuable insights into customer behavior, helping the business improve retention and profitability through data-driven decisions.