1. Introduction

In this analysis, we explored customer segmentation using the RFM (Recency, Frequency, Monetary) framework on an Online Retail dataset. The goal is to categorize customers based on their purchasing behavior to tailor marketing strategies effectively.

2. Data Exploration

We loaded the dataset, examined its structure, and performed preliminary data exploration using various libraries like Pandas, Seaborn, and Matplotlib. Initial insights into missing values and data types were obtained.

3. Feature Engineering

Amount, Frequency, Recency: Calculated and derived these key metrics to build the RFM framework.

Outlier Analysis: Detected and removed statistical outliers in Amount, Frequency, and Recency.

4. Customer Segmentation

4.1 K-Means Clustering

Applied K-Means clustering to group customers into segments based on the scaled RFM attributes.

Determined the optimal number of clusters using the elbow method and silhouette analysis.

4.2 Cluster Analysis

Analyzed the characteristics of each cluster using box plots.

Explored the relationship between Cluster Id and Amount, Frequency, Recency.

4.3 Hierarchical Clustering

Explored hierarchical clustering with different linkage methods (Single, Complete, Average).

Determined the optimal number of clusters and assigned cluster labels.

4.4 Hierarchical Cluster Analysis

Investigated the characteristics of each cluster obtained from hierarchical clustering.

Examined the relationship between Cluster Labels and Amount, Frequency, Recency.

5. Visualization

Utilized various visualizations, including box plots and dendrograms, to provide a clear understanding of the clustering results and customer behavior.

6. Conclusion

Summarized the findings from both K-Means and Hierarchical clustering.

Highlighted the implications of each cluster and potential marketing strategies.

7. Recommendations

Provided recommendations for marketing strategies based on the identified customer segments.

8. Future Work

Outlined potential areas for further analysis or improvement in the segmentation process.

9. References

Listed the libraries and tools used, along with any external sources referenced in the analysis.