**E-Commerce Customer Segmentation Report**

**1. Executive Summary**

This report analyses customer transaction data from a UK-based online retail store to identify key customer segments for a targeted loyalty program. Using RFM (Recency, Frequency, Monetary Value) features and K-Means/hierarchical clustering, we identified three distinct segments: high-value, regular, and churned/low-value customers. The high-value segment, characterized by recent purchases, high frequency, and high monetary value, should be the primary focus of the loyalty program.

**2. Data Overview**

* **Observations:** 397,884 transactions
* **Variables:** InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country
* **Data cleaning:**
* Removed rows with missing CustomerID
* Removed rows with invalid quantities or prices
* Converted InvoiceDate to date format
* **RFM features calculated for each customer**

**3. Feature Engineering**

* **RFM Features:**
* **Recency:** Days since last purchase
* **Frequency:** Number of distinct invoices
* **Monetary Value:** Total spending
* **Log Transformation:** Frequency and Monetary Value were log-transformed to reduce skewness.

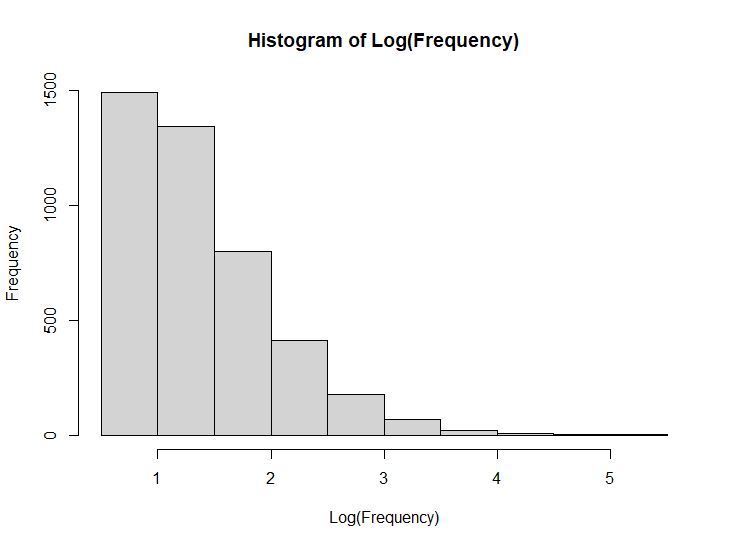
**4. Exploratory Data Analysis (EDA)**

**Distributions**

* Recency is somewhat evenly distributed.
* Log(Frequency) and Log(Monetary Value) are approximately normally distributed after transformation.

*Visuals:*

* *Histograms of Recency, Log(Frequency), Log(MonetaryValue)*

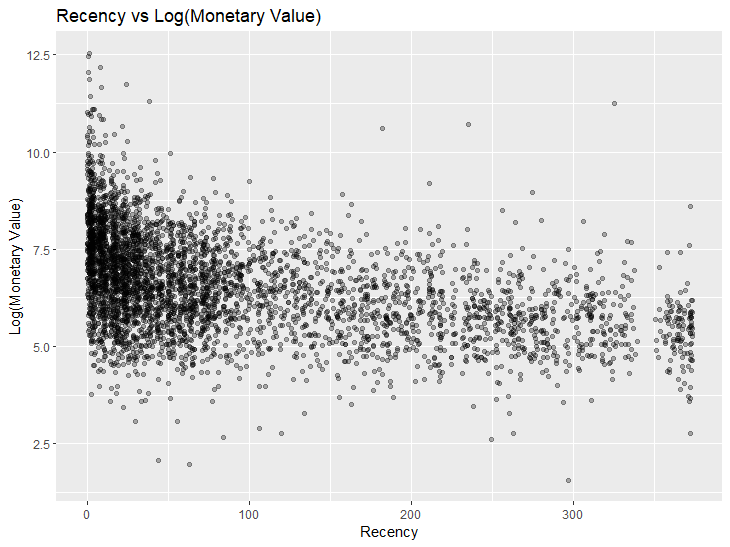
**

**Bivariate Relationships**

* Scatterplot of Recency vs Log(Monetary Value) shows the distribution of customers across these key dimensions.

*Visuals:*

* *Scatterplot of Recency vs Log(MonetaryValue)*

**

**5. Clustering Methodology**

**K-Means Clustering**

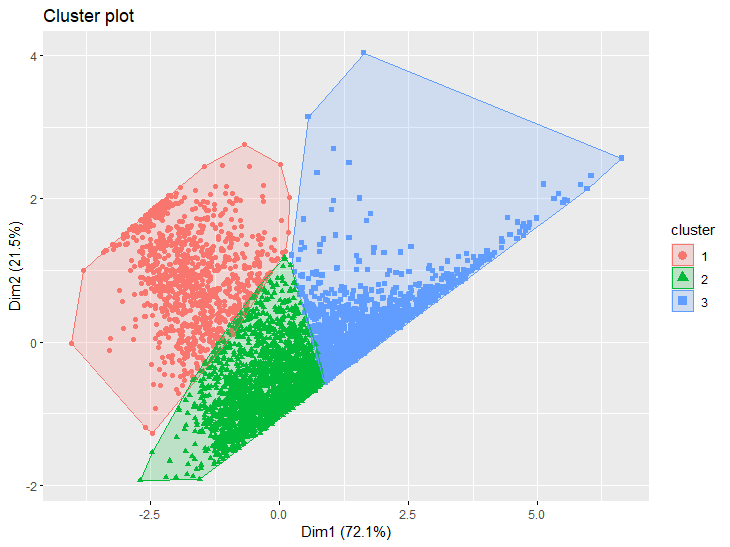
* **Optimal K:** Determined to be 3 using Elbow and Silhouette methods.
* **Cluster Visualization:** fviz\_cluster plot shows three distinct clusters.

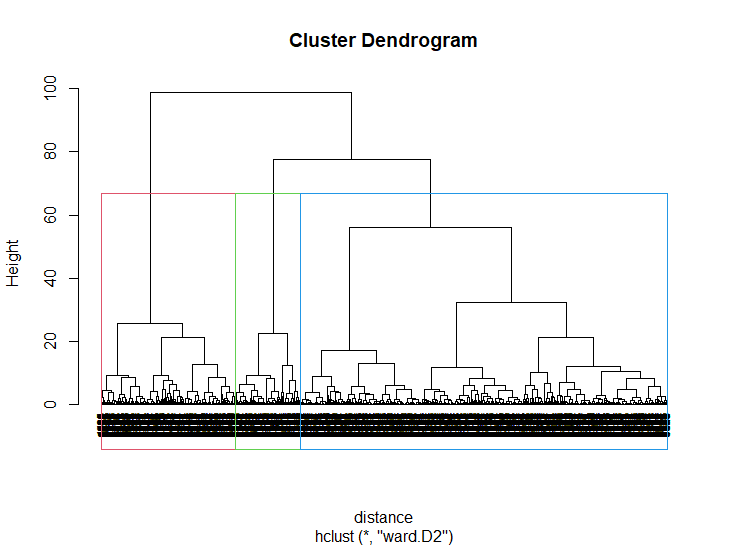
**Hierarchical Clustering**

* **Method:** Ward’s minimum variance method.
* **Dendrogram:** Visualizes the hierarchical relationships between customers.
* **Clusters:** Cut at k = 3 to create three segments.

*Visuals:*

* *K-Means cluster plot*

**

* *Hierarchical clustering dendrogram*

**6. Cluster Profiling**

**K-Means Cluster Profiles**

| **KMeans\_Cluster** | **Recency** | **log\_Frequency** | **log\_MonetaryValue** |
| --- | --- | --- | --- |
| 1 | 255 | 0.834 | 5.62 |
| 2 | 54.3 | 1.06 | 6.15 |
| 3 | 29.6 | 2.17 | 8.00 |

**Hierarchical Cluster Profiles**

| **HC\_Cluster** | **Recency** | **log\_Frequency** | **log\_MonetaryValue** |
| --- | --- | --- | --- |
| 1 | 17.5 | 2.68 | 8.74 |
| 2 | 48.9 | 1.30 | 6.58 |
| 3 | 246 | 0.81 | 5.59 |

**Cluster Descriptions**

* **High-Value Customers (Cluster 3/1):** Low recency, high frequency, high monetary value.
* **Regular Customers (Cluster 2/2):** Medium recency, medium frequency, medium monetary value.
* **Churned/Low-Value Customers (Cluster 1/3):** High recency, low frequency, low monetary value.

**7. Re-Clustering (Attempted Refinement)**

* We attempted to refine the regular customer segment (Cluster 2) by re-clustering it into two sub-segments. However, the resulting sub-segments had very similar RFM profiles. Therefore, we recommend focusing on the original three segments for the loyalty program.

**8. Recommendations**

* **Target high-value customers (Cluster 3/1) with exclusive offers, personalized recommendations, and early access to new products.**
* **Re-engage churned customers (Cluster 1/3) with targeted promotions and incentives to encourage repeat purchases.**
* **Nurture regular customers (Cluster 2/2) with loyalty rewards and cross-selling opportunities to increase their frequency and monetary value.**

**Appendix**

* R code
* Additional plots/tables

**References**

* The R Foundation for Statistical Computing
* Packages: tidyverse, lubridate, cluster, factoextra, NbClust

**End of Report**