

Customer **Segmentation** Report using RFM Analysis

1. Objective

The purpose of this project was to analyze customer purchasing behavior using **RFM analysis (Recency, Frequency, Monetary)** and **segment** customers into actionable groups. This segmentation allows the business to better target marketing strategies and improve customer retention.

2. Dataset

- Source: Kaggle – Online Retail dataset
 - Transactions: **Dec 2010 – Dec 2011**
 - Total records after cleaning: **166,363**
 - Unique customers: **2,997**
 - Fields used: **InvoiceNo, Quantity, InvoiceDate, UnitPrice, CustomerID, Country**
-

3. Data Preparation

Steps taken:

- Converted **InvoiceDate** from text into proper datetime format.
- Removed missing **CustomerIDs** and cancelled orders.
- Filtered out records with negative or zero Quantity/UnitPrice.
- Created **TotalPrice = Quantity × UnitPrice**.

4. RFM Metrics

For each customer:

- **Recency (R)**: Days since last purchase (lower = better).
- **Frequency (F)**: Number of unique invoices (purchases).
- **Monetary (M)**: Total money spent.

Summary:

- Recency: 1 – 698 days
- Frequency: 1 – 106 purchases
- Monetary: £6 – £168,469

5. RFM Scoring

- Scores assigned from **1 (worst) to 5 (best)**.
- Recency: most recent → higher score.
- Frequency & Monetary: higher values → higher score.
- Final RFM Score: combination of R, F, M (e.g., **555** is the best).

Example: Customers with score **555** are top **spenders, frequent, and recent buyers**.

6. Customer Segmentation

Based on the RFM scores, customers were grouped into **five segments**:

- **Champions (607 customers)**: **most valuable customers** — they purchased recently, buy often, and spend the most.
- **Loyal (693 customers)**: These customers buy frequently and consistently, but may not spend as much as Champions.
- **Potential (821 customers)**: This is the **largest segment**. They show positive signals either spending decently, buying a few times, or recently engaged but are not yet Loyal.
- **At Risk (616 customers)**: Customers who used to buy but haven't purchased recently. Their engagement is dropping, and they risk becoming Lost.
- **Lost (260 customers)**: These customers have not purchased for a very long time and show minimal activity.

7. Visualizations (Seaborn)

- **Bar Chart:** Distribution of customers across segments.
- **Heatmap:** Average Recency, Frequency, Monetary per segment.
- **Scatter Plot:** Recency vs Frequency, colored by segment.

These clearly show that:

- Champions are recent, frequent, and high-spending.
 - Potential customers form the largest group (opportunity for conversion).
 - At Risk + Lost segments together are sizable → require reactivation efforts.
-

8. Marketing Recommendations

Champions (607 customers)

- VIP treatment, exclusive offers, referral incentives.

Loyal (693 customers)

- Cross-sell and upsell, subscription bundles, loyalty programs.

Potential (821 customers)

- Convert with discounts, product recommendations, reminders.

At Risk (616 customers)

- Win-back campaigns, surveys, urgent offers.

Lost (260 customers)

- Reactivation campaigns, big discounts, or reduce effort if unresponsive.
-

9. Conclusion

The RFM analysis provided actionable segmentation of the customer base. The business should **invest in retaining Champions and Loyal customers**, while **re-engaging At Risk and Lost customers**. Converting **Potential customers** into Loyal ones represents the biggest growth opportunity. By applying these insights, marketing campaigns can be more targeted and cost-effective, improving overall customer lifetime value.