

Customer Segmentation using RFM Analysis

Understanding customer behavior to drive data-driven
decisions

Python | SQL | Power BI

Prathmesh Ambulge

GitHub : : <https://github.com/Prathamesh-14-a>

➤ **Business Problem:**

- Businesses often treat all customers the same
- This leads to inefficient marketing and poor retention
- Customer behavior varies significantly across purchase patterns

➤ **Project Objective :**

- Segment customers based on **Recency, Frequency, and Monetary value**
- Identify **high-value**, **at-risk**, and **inactive** customers
- Enable targeted marketing and retention strategies

➤ **Dataset Overview :**

○ **Content:**

- Source: Online Retail Transaction Data
- Records: 3,92,692 transactions
- Customers: 4,338 unique customers
- Time period: Dec 2010 – Dec 2011

○ **Key Fields Used:**

- InvoiceNo
- CustomerID
- InvoiceDate
- Quantity
- UnitPrice

Raw dataset contained 541909 records. After removing cancellations, invalid quantities/prices, and null customers, the final dataset used for analysis contains 392692 records.

➤ Data Cleaning (Python):

- Removed records with missing CustomerID

```
# Removing Rows having Null customerID
df = df[df['CustomerID'].notna()].copy()
✓ 0.2s
```

- Removed cancelled transactions (InvoiceNo starting with 'C')

```
#Taking non canceled InvoiceNo into consideration
df_filtered = df[~df['InvoiceNo'].str.startswith("C", na= False)].copy()
✓ 1.6s
```

- Removed zero or invalid quantity/price entries

```
#Taking Rows which have UnitPrice and Quantity greater than 0
df_filtered = df_filtered[df_filtered['UnitPrice'] > 0].copy()
df_filtered = df_filtered[df_filtered['Quantity'] > 0].copy()
✓ 0.1s
```

➤ Data Cleaning (Python):

- Created TotalAmount = Quantity × UnitPrice

```
df_cleaned['TotalAmount'] = df_cleaned['Quantity'] * df_cleaned['UnitPrice']  
✓ 0.0s
```

- Capped extreme transaction values at the **99.5th percentile** to prevent bulk orders from distorting RFM metrics

```
upper = df_cleaned['TotalAmount'].quantile(0.995)  
df_cleaned['TotalAmount'] = df_cleaned['TotalAmount'].clip(upper=upper)
```

```
✓ 0.0s
```

➤ **RFM Metrics Definition**

- **Recency:** Days since last purchase
- **Frequency:** Number of unique purchases
- **Monetary:** Total spend per customer
- RFM metrics aggregated using **SQL**
- Snapshot date chosen as last transaction date

➤ **RFM Scoring & Segmentation**

- Customers scored from 1–5 for each RFM metric
- Quantile-based scoring approach
- Combined scores used to form customer segments
- 5 buckets (NTILE)

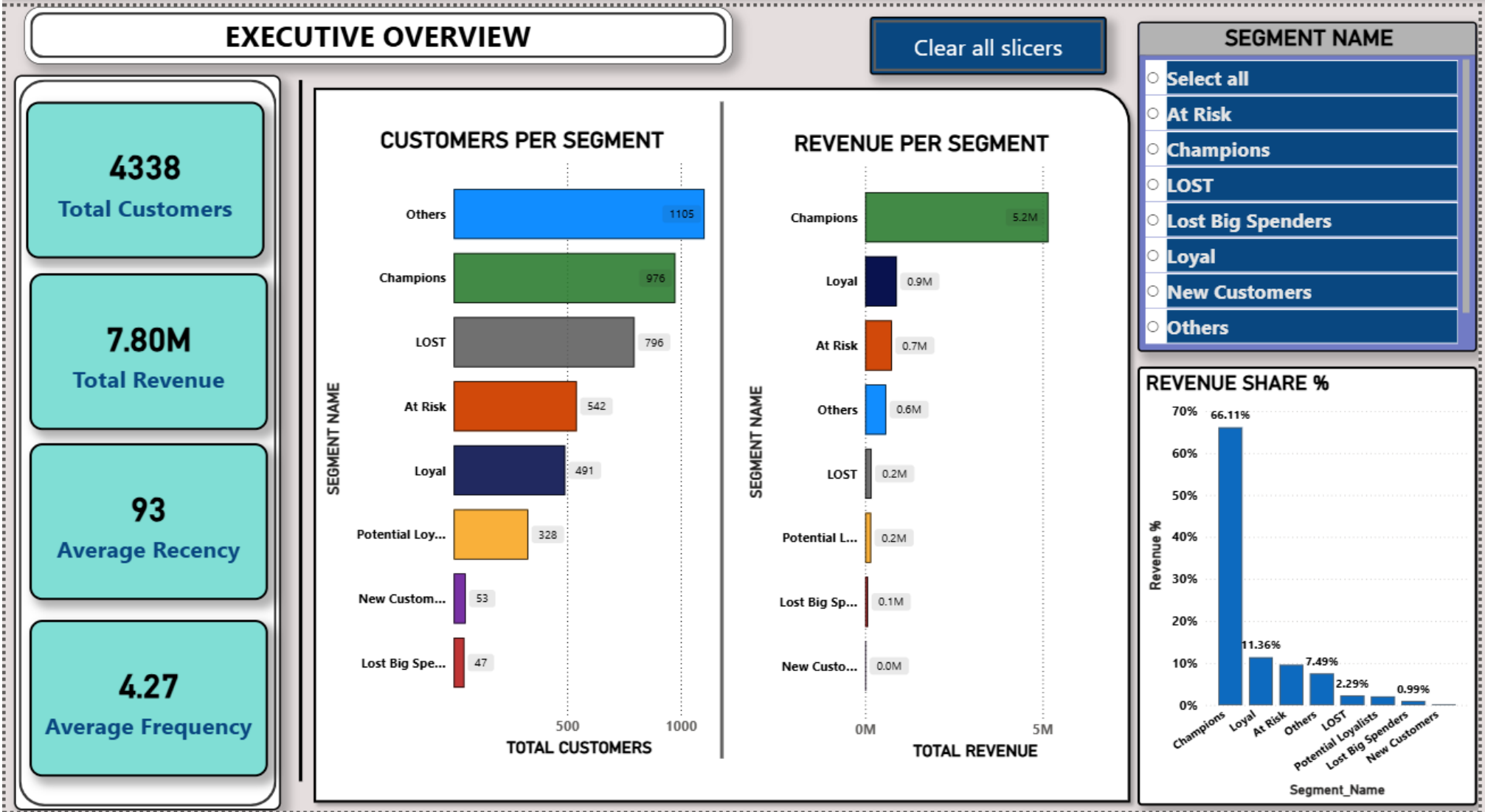
Segment	R_Score	F_Score	M_Score	Meaning
Champions	4 - 5	4 - 5	4 - 5	Best Customers
Loyal	3 - 5	4-5	3-5	Repeat Buyers
Potential Loyalists	4 - 5	2 - 3	2 - 3	Could Become Loyal
New Customers	5	1	1 - 2	Recently Acquired
At Risk	1 - 2	3 - 5	3 - 5	Slipping Away
Lost Big Spenders	1	1 - 2	4 - 5	High Value , Inactive
Lost	1 - 2	1 - 2	1 - 2	Churned

➤ **Dashboard Overview (Power BI)**

- Interactive Power BI dashboard for segment-level analysis
- KPIs:
 - Customer count
 - revenue
 - RFM behavior
- Filters enable dynamic exploration by segment

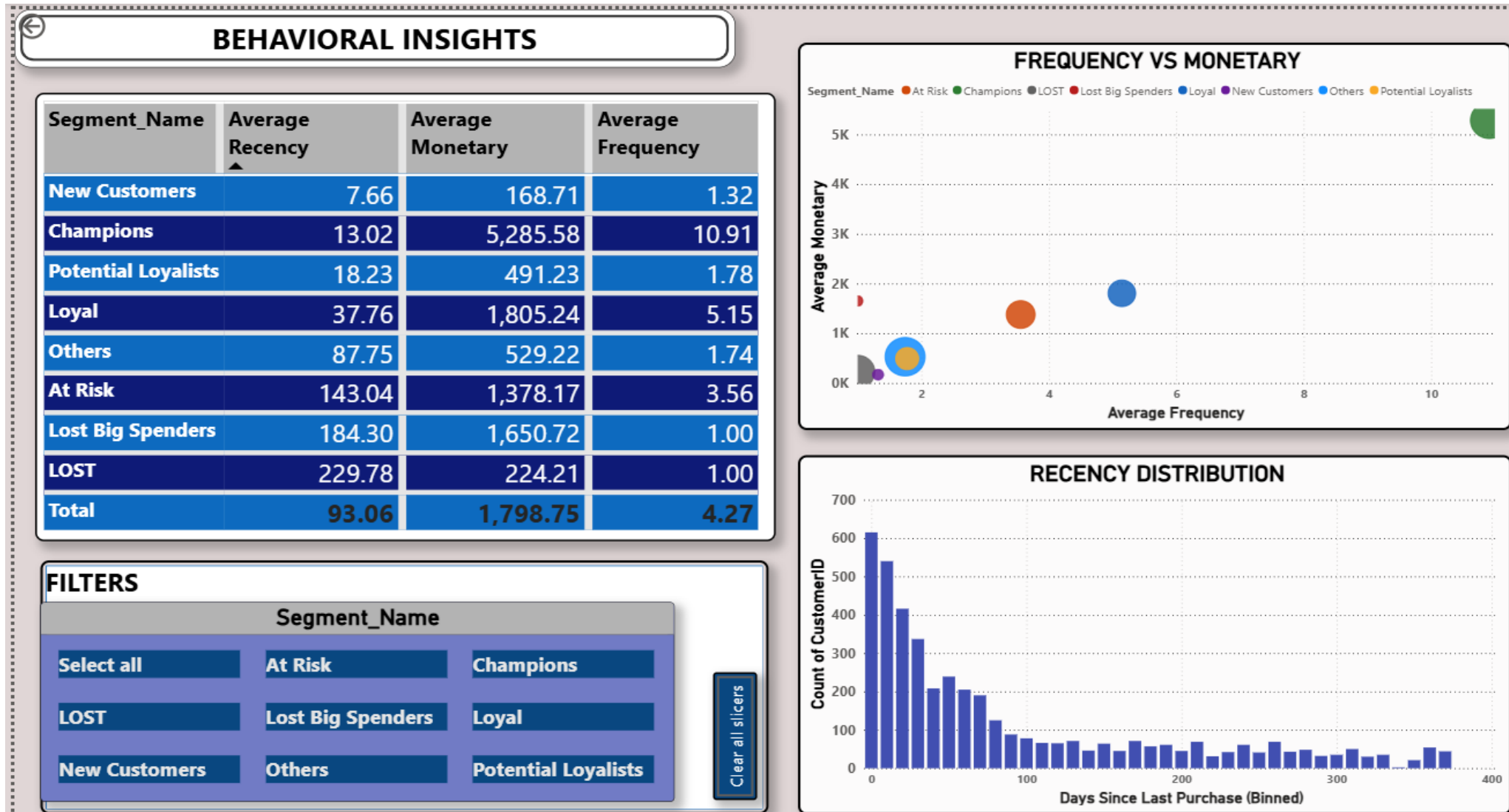
➤ Dashboard: Executive Overview

- Champions contribute 66.11% of total revenue
- Majority of customers are concentrated into Champions and Loyal
- High revenue concentration among top segments



➤ Dashboard: Behavioral Deep Dive

- Champions show lowest recency and highest monetary value
- At-Risk customers have high past value but declining activity
- Lost segment shows low frequency and long inactivity



➤ **Key Insights**

- A small percentage of customers contribute a large share of revenue
- Champions and Loyal segments show low recency and high frequency
- Dormant customers form a significant portion but contribute minimal revenue
- Clear opportunity for re-engagement campaigns

➤ **Recommendations**

- Run targeted reactivation campaigns for **At Risk** segment
- Focus retention efforts on **Champions & Loyal Customers**
- Avoid spending marketing budget on long-inactive customers
- Use RFM segmentation as a recurring monitoring framework

➤ **Conclusion**

- RFM analysis provides a simple yet powerful framework
- Enables data-driven customer segmentation
- Supports strategic marketing and retention decisions