TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Tran Van Loc

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Objective:

Identify 1000 Targeted Customers From Datasets For Future Activities.

Data Analysis Content:

1. Data Exploration:
   a. Customers' Age distribution.
   b. Wealth segments by Age groups.
   c. Profit by Wealth Segment.
   d. Customers' Job industry distribution.
   e. Bike-related purchases by gender in 3-year time period.
   f. Number of cars owned and not owned by state.
2. Model Development: RFM Analysis and Customer segment.
3. Interpretation: Select top 1000 targeted customers.

Data Exploration

Data Quality Assessment

- Data quality issues are discovered and listed in the table.
- More in-depth descriptions and mitigation strategies has been sent via email.

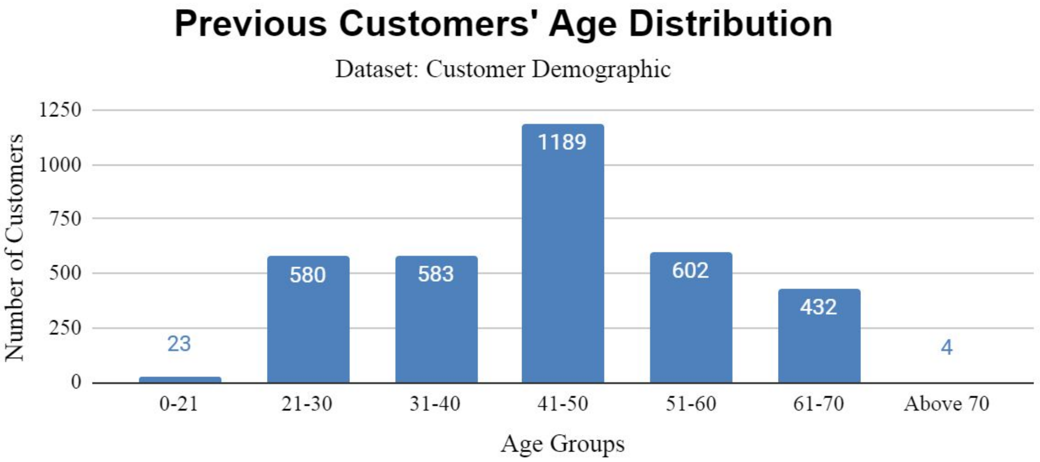
	Customer Demographic	Customer Addresses	Transactions in the past 3 months
Accuracy	- Inaccurate values in *DOB*. - Missing *age* data.		- Missing *profit* data.
Completeness	- Blanks in *job_title*. - Incomplete data in *customer_id*.	- Incomplete data in *customer_id*.	- Blanks in *online_order*. - Blanks in *brand*. - Incomplete data in *customer_id*.
Consistency	- Inconsistent values in *gender*.	- Inconsistent values in *state*.	
Currency	- Outdated values (deceased customers) in *deceased_indicator*.		
Relevancy	- Irrelevant *default* data.		- Irrelevant “Cancelled” data in *order_status*.
Validity			- Wrong data format in *list_price*. - Wrong data format in *product_first_sold_date*.
Uniqueness			

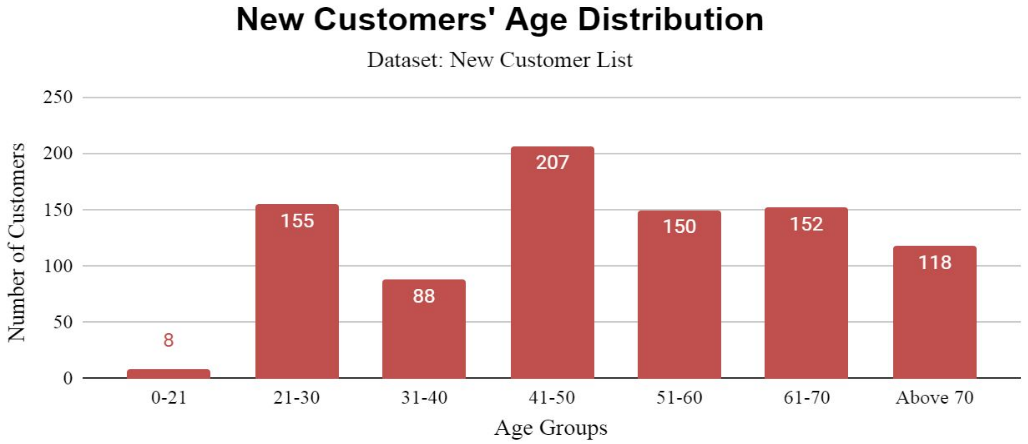
Data Exploration

Customers' Age Distribution

- Two charts show that the most populated age group is 41-50.
- One of the top lowest age groups in both charts is 0-21.
- The number of customers above 70 years old increased significantly, while the number of customers in 31-40 age group dropped.

→ 21-70 age group is suggested, especially 21-30 and 41-70.

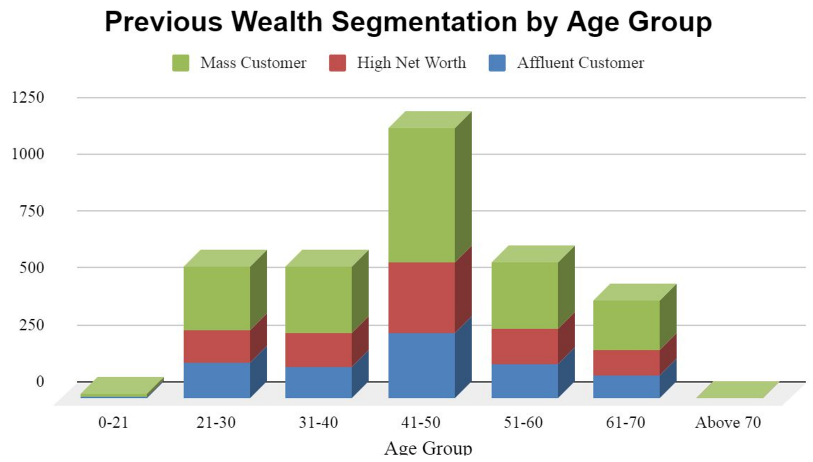


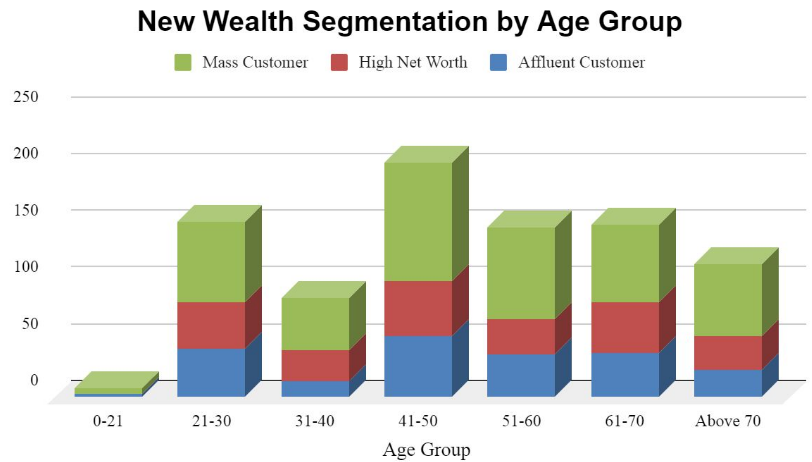


Data Exploration

Wealth Segment by Age Group

- The descending order of 3 wealth segments in number of customers are: Mass Customer, High Net Worth and Affluent Customer.
- Mass Customer wealth segment has the highest customer numbers in all age groups.
- Notice that, in 41-50 age group, Affluent Customer outnumbers High Net Worth wealth segment.

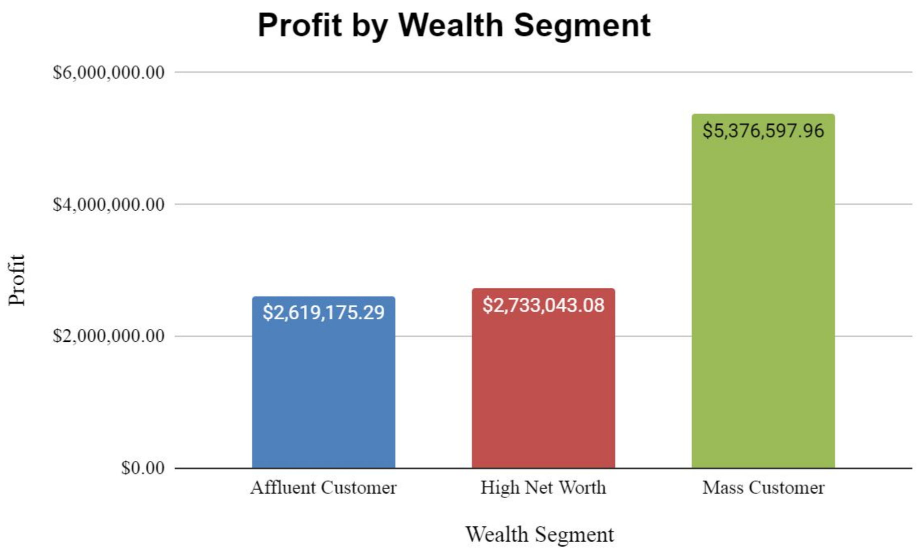




Data Exploration

Profit by Wealth Segment

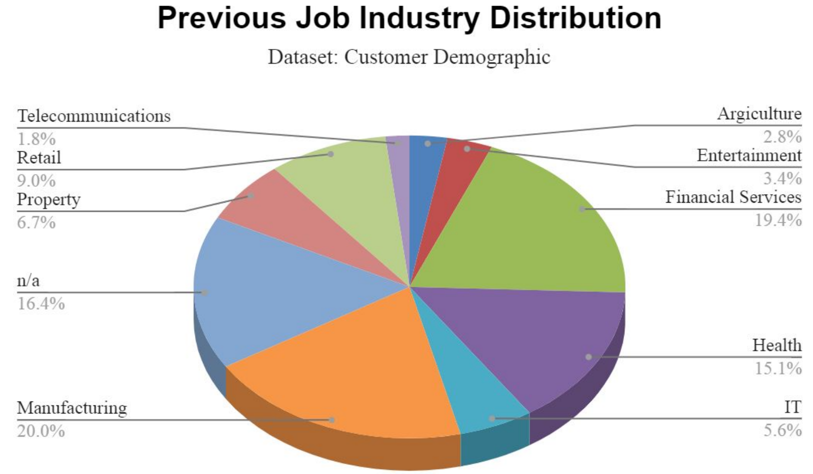
- Mass Customer wealth segment with the highest customer numbers in all age groups, also generates the highest amount of revenue.
- Profit generated by Mass Customer is around twice larger than profit by High Net Worth.
- Profit by Affluent Customer is slightly under the profit by High Net Worth.

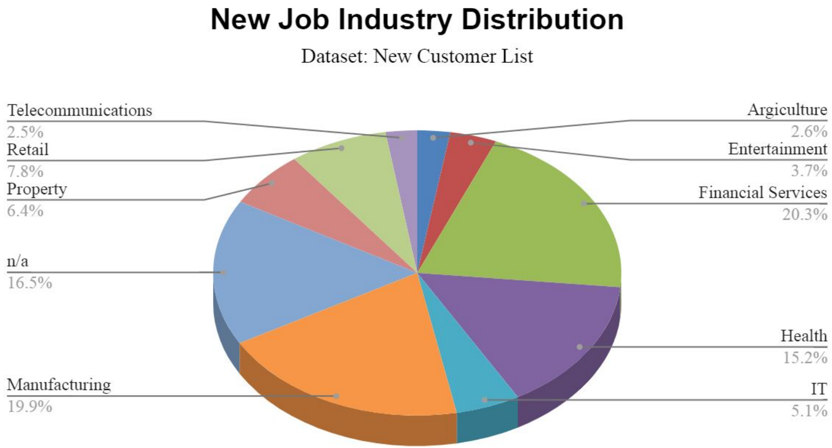


Data Exploration

Job Industry Distribution

- Manufacturing, Financial Services and Health are three industries with highest new customer numbers (approx 20%, 20% and 15% respectively).
- Two groups with smallest new customer numbers are Telecommunications (approx 2%) and Agriculture (approx 3%).
- Same pattern is applied for the previous customers' job industries, with similar percentages.

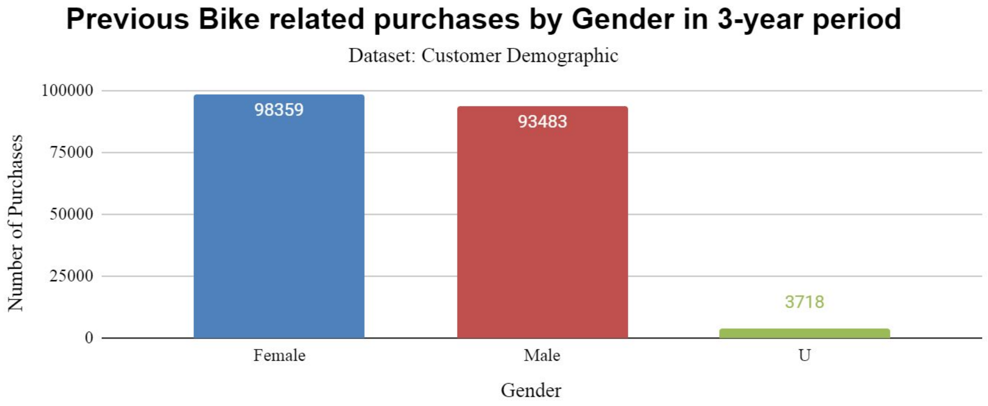


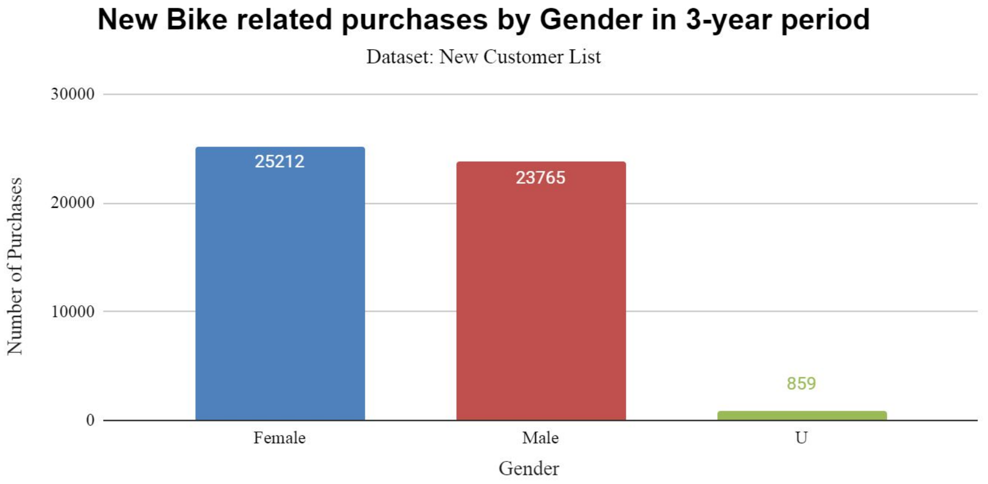


Data Exploration

Bike related purchases by Gender in 3-year period

- Numbers of purchases made by female customers are the highest in both datasets.
- Numerically, female customers purchase 3-5000 more than males.
- However, the percentages of total of both genders do not differ dramatically.

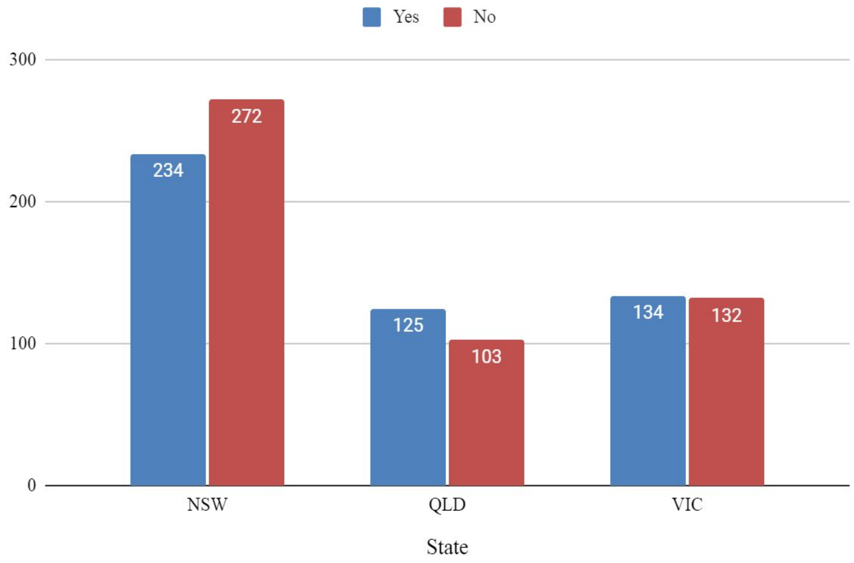




Data Exploration

Number of cars owned and not owned by State

- New South Wales is the state with the highest numbers of people do and do not have cars, which the number of cars not owned is higher by approximately 40.
- The numbers of Queensland and Victoria is significantly lower than those of New South Wales.
- The numbers of Victoria are split evenly.
- The number of customers own cars in Queensland is slightly higher than number of ones who do not.



Model Development

RFM Analysis

- The idea of RFM (recency - frequency - monetary) analysis is to segment customers by calculating and analyzing:
    + When their last purchase was.
    + How often they've purchased in the past.
    + How much they've spent overall.
- Criteria:
    1. Customers purchased recently are more likely to buy again.
    2. Customers who buy more often are more likely to buy again.
    3. Customers who spend more are more likely to buy again.
- Method:

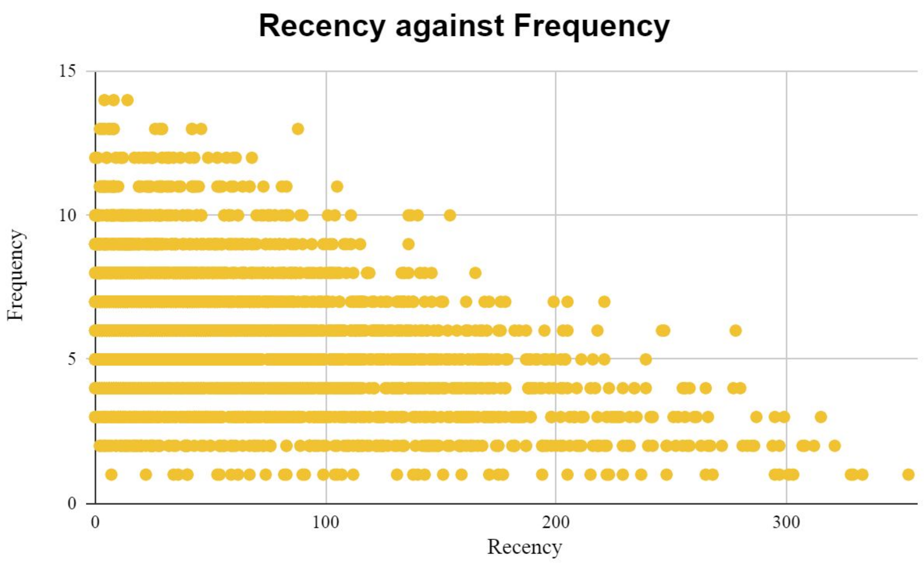
  Step 1: Base on the criteria, we rank each customer from 1 to 4 for each criteria (1 is for most recently/most frequently/spent the most).

  Step 2: Put customers into groups, each with different RFM scores and marketing approach strategies.

Model Development

RFM Analysis - Recency against Frequency

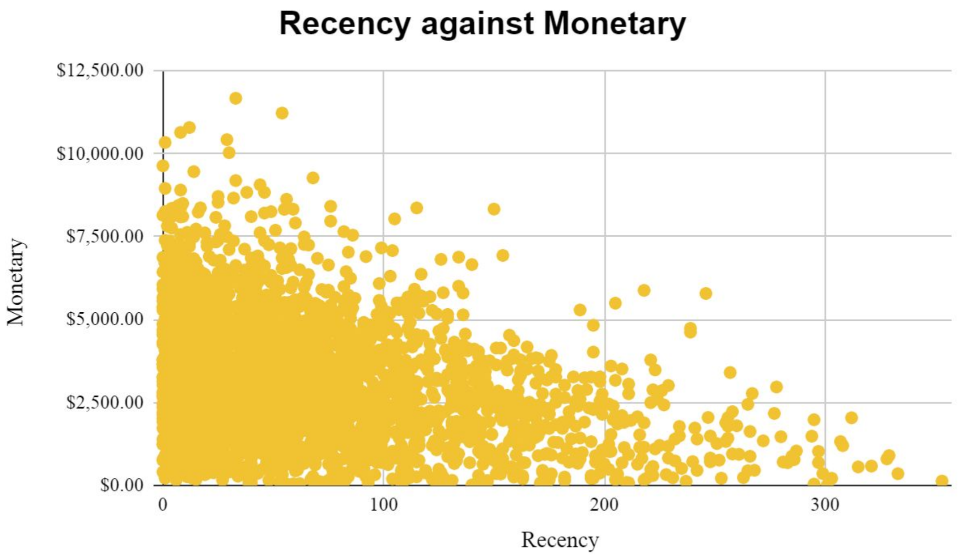
- Recency value negatively correlated with Frequency value, which mean customers who purchased recently are likely to purchase more frequently.
- It can be seen that the majority of customers who purchased in the recent 0-50 days come back more frequently (6-14 times).
- Customers who purchased more than 250 days ago have low chance to purchase again (6 times as highest).



Model Development

RFM Analysis - Recency against Monetary

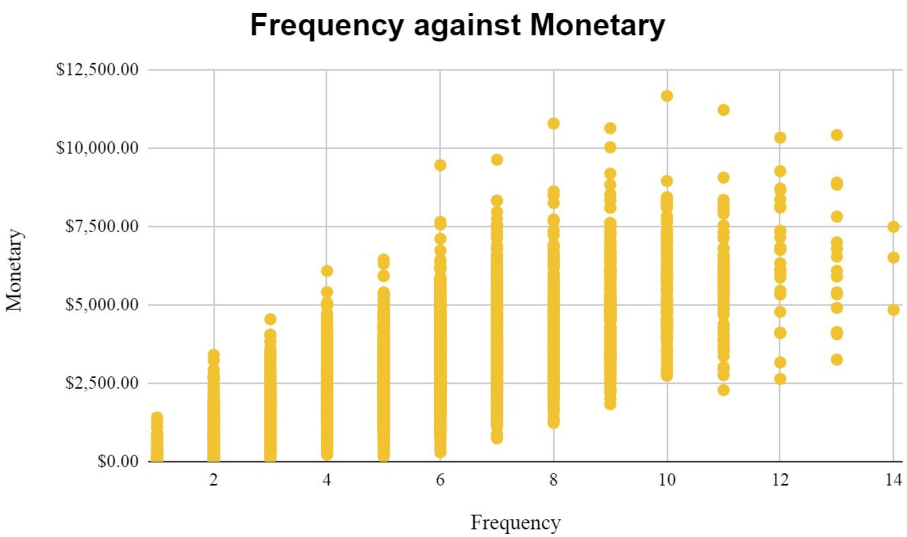
- It can be seen from the scatter chart that customers who purchased more recently would generate higher revenue.
- The amount of revenue generated drops as the recency value decreases.
- Customers who purchased more than 250 days ago generated low revenue.



Model Development

RFM Analysis - Frequency against Monetary

- Frequency value and Monetary value have a positive correlation. To one extent, customer who purchased frequently are the ones that generated high revenue.
- Majority of customers who came back frequently (more than or equal 6 times) spent up to $10,000. However, the amount of money drops when frequency value reaches 14.



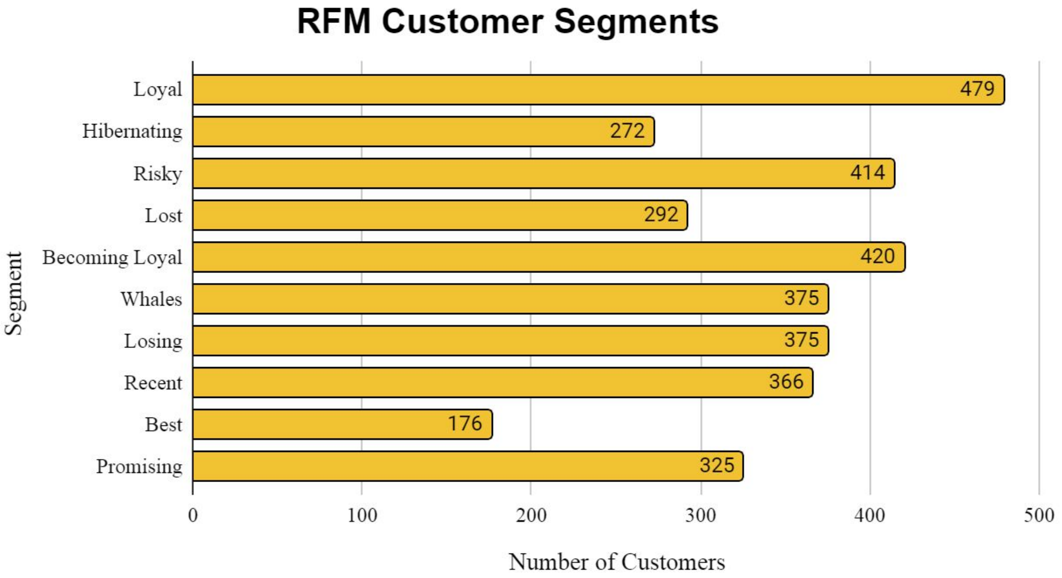
Model Development

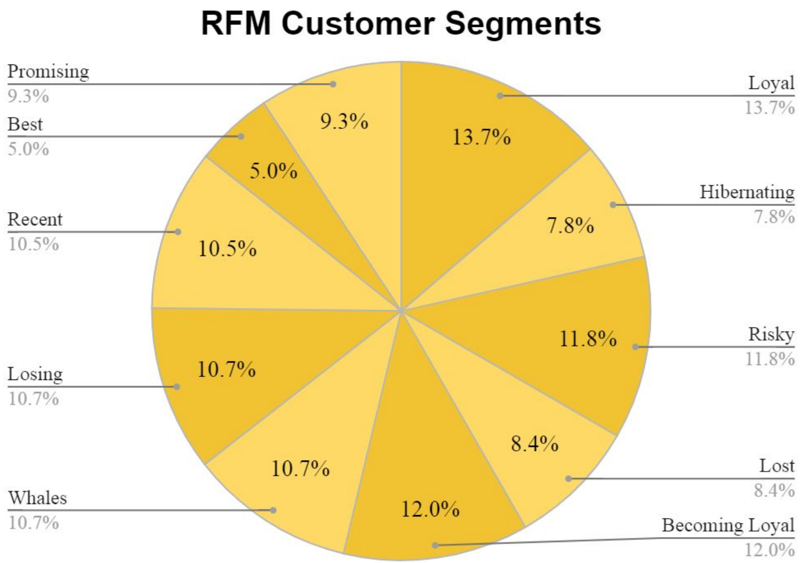
RFM Analysis - Customer Segmentation

Rank	Segment Title	Criteria
1	Best Customer	Bought most recently, most often and spent the most.
2	Loyal Customer	Bought recently, often and spent large amounts of money.
3	Becoming Loyal Customer	Bought recently, more than once and spent an adequate amount.
4	Recent Customer	Bought most recently, but not so often.
5	Promising Customer	Bought recently, never bought before, spent a small amount of money.
6	Losing Customer	Below average RFM score.
7	Risky Customer	Bought from a long time ago, but frequently and spent an adequate amount.
8	Whale Customer	Bought from a long time ago, not often but spent a large amount of money.
9	Hibernating Customer	Bought from a long time ago.
10	Lost Customer	Bought from a long time ago, only once and spent the least.

Model Development

RFM Analysis - Customer Segmentation





The bar chart shows the total number of customers in each segments

The pie chart shows the percentage of total each segments take

Interpretation

RFM Analysis - Top 1000 Customers

Rank	Segment Title	Number of Customers	Cumulative
1	Best Customer	176	176
2	Loyal Customer	479	655
3	Becoming Loyal Customer	420	1075
4	Recent Customer	366	1441
5	Promising Customer	325	1766
6	Losing Customer	375	2141
7	Risky Customer	414	2555
8	Whale Customer	375	2930
9	Hibernating Customer	272	3202
10	Lost Customer	292	3494

Interpretation

RFM Analysis - Top 1000 Customers

Rank	Segment Title	Number of Customers	Cumulative	Selection
1	Best Customer	176	176	176
2	Loyal Customer	479	655	479
3	Becoming Loyal Customer	420	1075	345

- Select 1000 customers base on the selection numbers and the criteria of 3 segments above.
- Common pattern: The targeted customers purchased recently, more than once and spent a significant amount of money.