

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data Analytics Approach

Diana Satapathy

Agenda

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

Introduction

Stage 1: Data Quality Assessment Overview

Key Issues for Data Quality Assessment:-

- Accuracy: Correct values
- Completeness: Data fields with values
- Consistency: Values free from Contradiction
- Currency: Values up to Date
- Relevancy: Data items with value meta-data
- Validity: Data containing allowable values
- Uniqueness: Records that are duplicated

Data Quality Dimensions	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographics	<ul style="list-style-type: none">• DOB: inaccurate• Age: missing	<ul style="list-style-type: none">• Job Title: blanks• DOB: blank• Customer Id: incomplete	<ul style="list-style-type: none">• Gender: inconsistency• past_3_years_bike_related_purchases: format• tenure: format	<ul style="list-style-type: none">• Deceased customers: filter out	<ul style="list-style-type: none">• Default column: delete	
Customer Address		<ul style="list-style-type: none">• Customer id: Incomplete	<ul style="list-style-type: none">• States: inconsistency			
Transactions	<ul style="list-style-type: none">• Profit: missing	<ul style="list-style-type: none">• Customer Id: incomplete• Online orders: blank• Brand: blank	<ul style="list-style-type: none">• Customer_id: format• Transaction_id: format• Product_id: format		<ul style="list-style-type: none">• Cancelled status order: filter out• Customer_id: filter out	<ul style="list-style-type: none">• List price: format• Product sold date: format• Transaction date: format

Summary Table

An email with a detailed analysis for the above mentioned issues has been sent.

Introduction

Identify Top 1000 Potential Customers to Target

Outline of problem:-

- Sprocket Central Pty Ltd, an organization that specializes in high quality bikes and cycling accessories.
- Their marketing department is evaluating offered data in order to raise business sales.
- The goal is to analyze and view the trends in the three datasets provided and then recommend 1000 high-value customers from a list to be targeted to drive the most value for the organization.

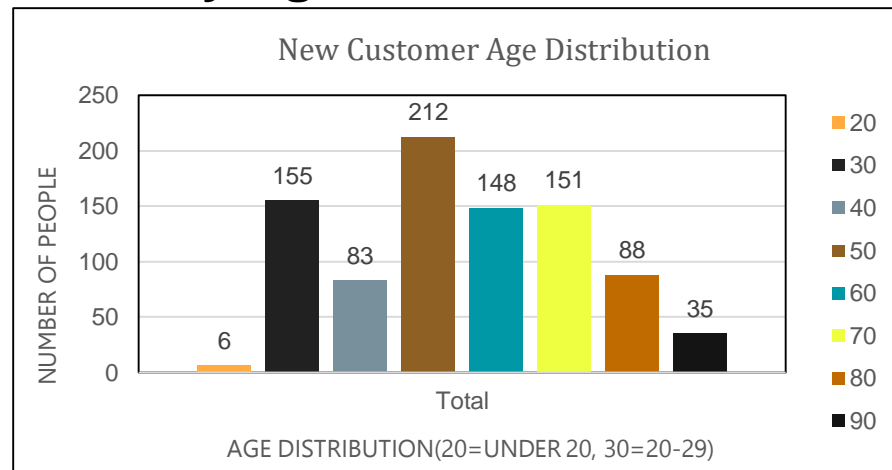
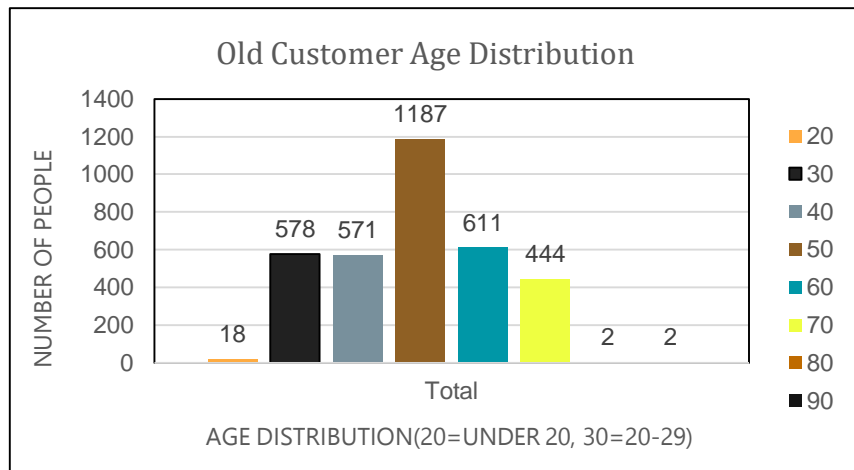
Contents of Data Analysis:-

- 'New' and 'Old' Customers Age Distribution
- Customers Gender Distribution
- Wealth Segment Distribution by age category
- Job Industry Distribution
- Distribution of Cars owned and not owned by state
- Distribution of Bike purchased in last 3 years
- RFM Analysis and Customer Segmentation

The proposed strategy is broken down into three stages: Data Exploration, Model Development and Interpretation

Data Exploration

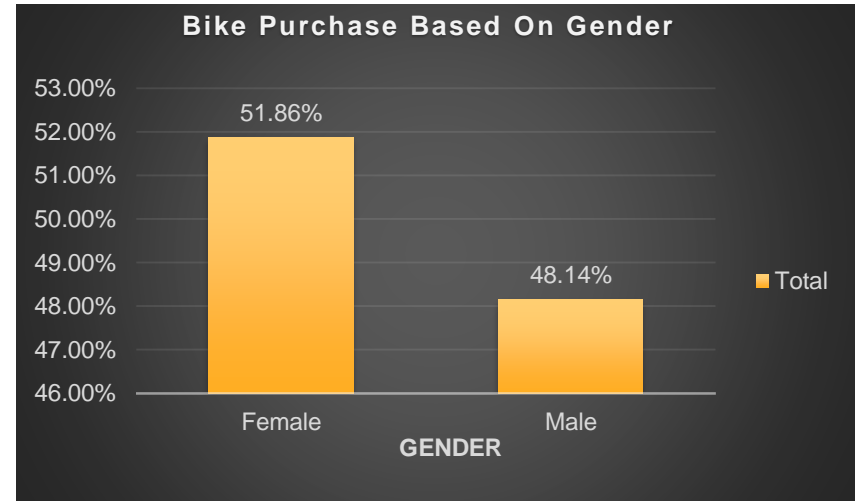
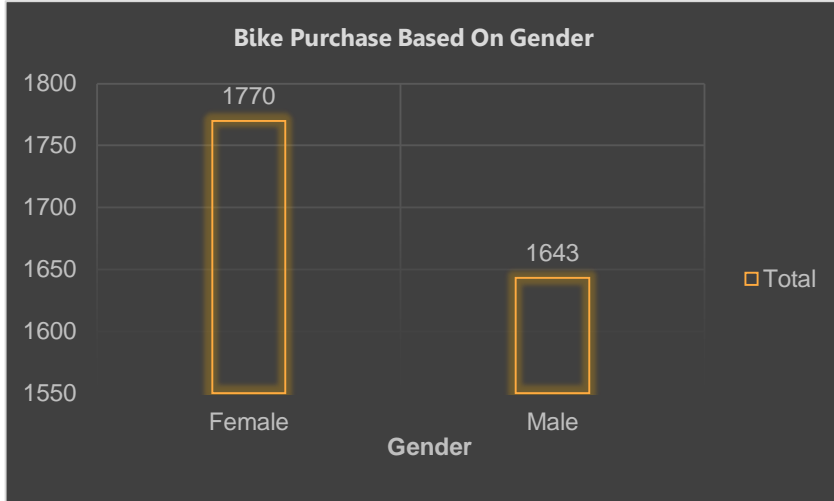
Customer Distribution by Age



- A large proportion of our customers are between the ages of 40-49 in both the charts.
- The lowest age groups in both the graphs are the ones that are under 20 and 80+ for both 'New' and 'Old' customer list.
- Customers in the age groups 20-29 and 40-49 in the 'New' customer list are most populated. The 'Old' Customer list suggests 20-59.
- Customers aged 69 and up make up a small percentage of the total in both the customer lists.
- There is a steep drop in the customers in the age group of 30-39 in 'New' customer list.

Data Exploration

Bike related purchases over last 3 years by gender

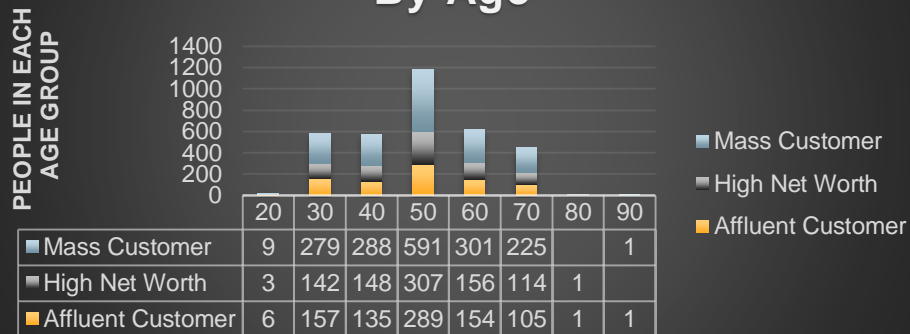


- Over the last three years about 52% of bike related purchases were made by females and 48% by males.
- Thus female members of the customers list make up of majority of bike ride sales.

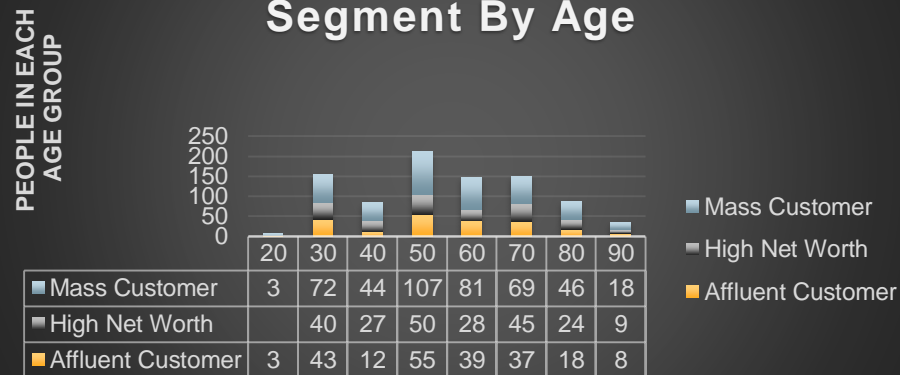
Data Exploration

Profit Distribution of Age Group by Wealth Segment

Old Customers Wealth Segment By Age



New Customers Wealth Segment By Age

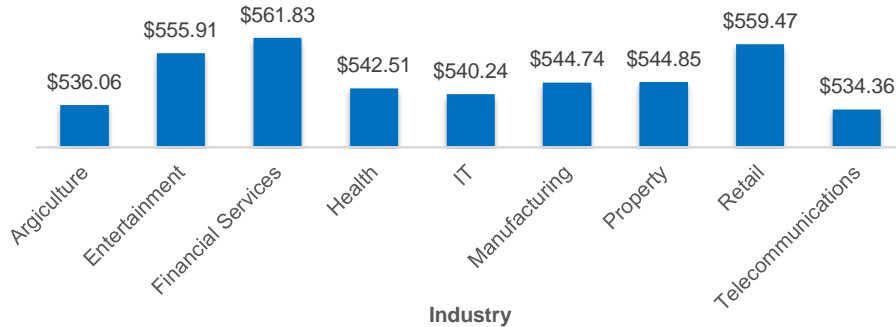


- The biggest number of customers are categorized as 'Mass Customer' in all age groups in both the customer lists.
- After Mass Customer,' the second best category is 'High Net Worth.'
- In some age groups, such as the below 20s and 20-29, in the 'Old' Customer list, 'Affluent Customer' customers have exceed 'High Net Worth'. In case of 'New' Customers list, below 29, and 40-49 age groups, Affluent Customer' customers have exceed 'High Net Worth'.

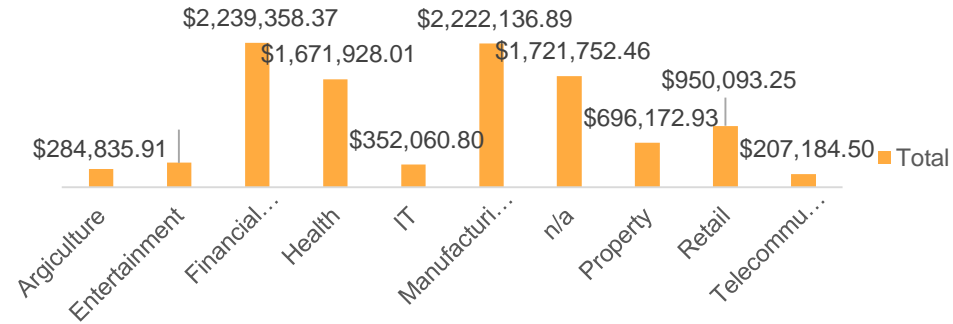
Data Exploration

Profit Distribution by Job Industry

Profit per Customer by Industry



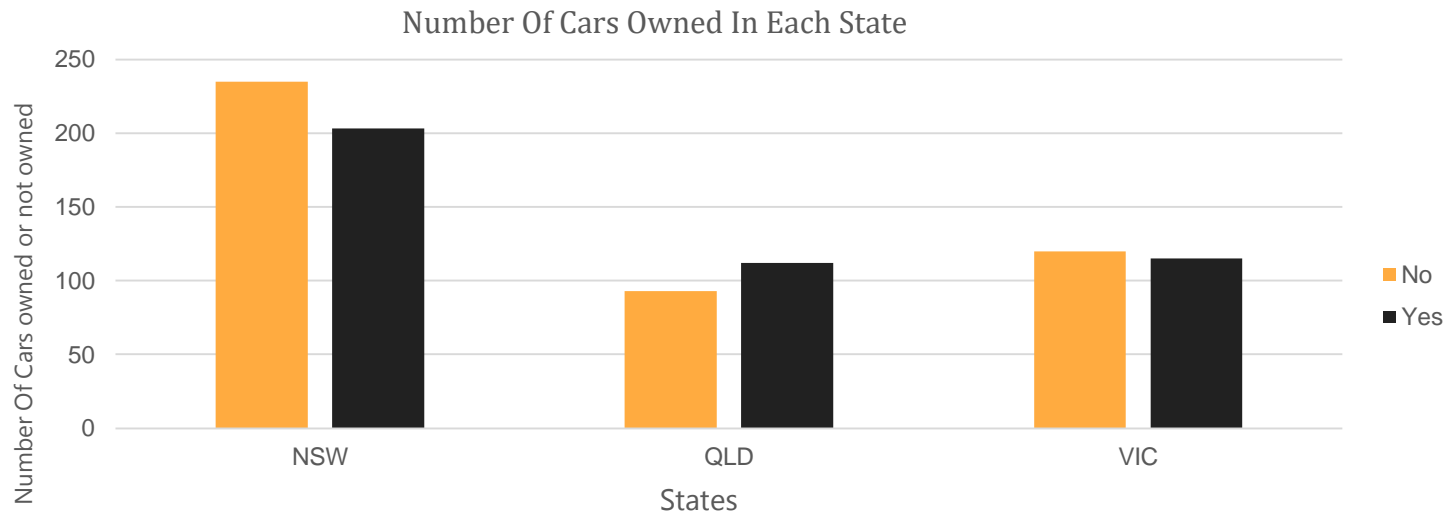
Total Profit By Industry



- The **Financial, Entertainment, and Retail industries** provide the most profit per customer for the business.
- Despite the fact that the total profit from the Retail and Entertainment industries is not the highest, we still have a high profit per customer value.
- Although the health industry is the third most profitable, its profit per customer is slightly low when compared to other profitable industries.
- The **Financial and Manufacturing are the company's most profitable industries**, with the highest profit numbers. In Addition, Health Industry is not far behind.

Data Exploration

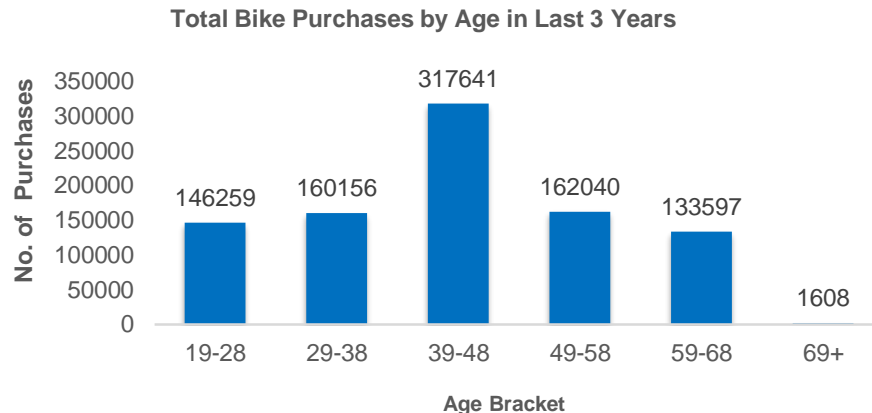
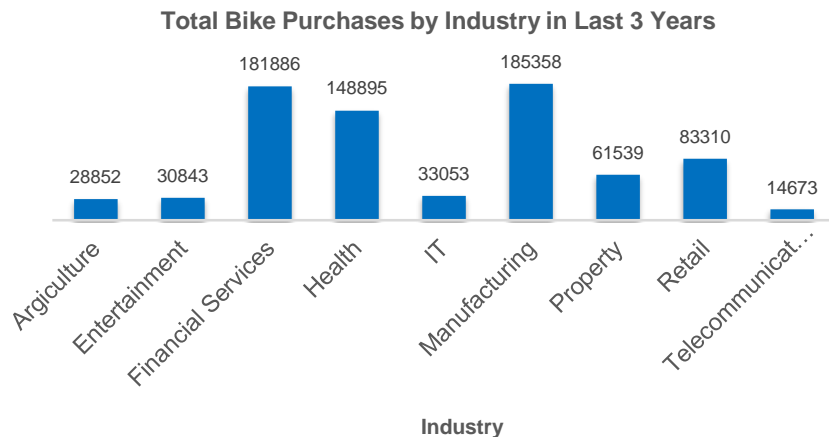
Distribution of customers owning a car by states



- The state with the most car owners is New South Wales (NSW).
- New South Wales (NSW) appears to have a larger population from which the data was obtained.
- Queensland (QLD) is almost equally split. Both of these figures are much lower than those in New South Wales.
- The number of people who do not own a car in Victoria (VIC) is slightly higher than the number of people who do.

Data Exploration

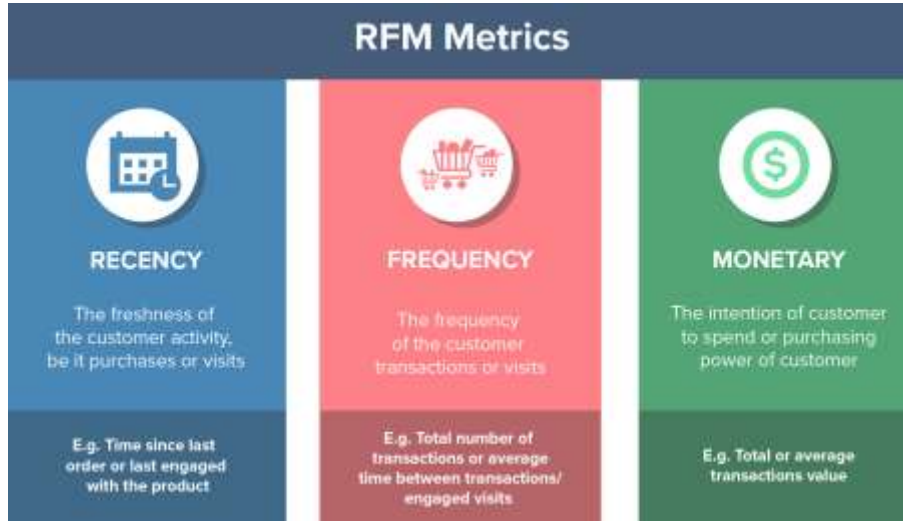
Distribution of Bike purchased in last 3 years



- The Financial, Health, and Manufacturing Industries have accounted for the majority of bike purchases in the last three years.
- The telecommunications industry makes the fewest purchases.
- When compared to other age groups, the 39-48 age group has the most bike-related purchases.
- The other age groups account for nearly half of the highest age group.

Model Development

RFM Analysis for Customer Segmentation



- RFM Analysis is used to determine which customers a business should target to increase its revenue and value
- The RFM (Recency, Frequency and Monetary) model classifies customer with different levels of engagement with business within these three categories
- The customers who are more recent, purchase frequently and spend more money are recognized high value customers

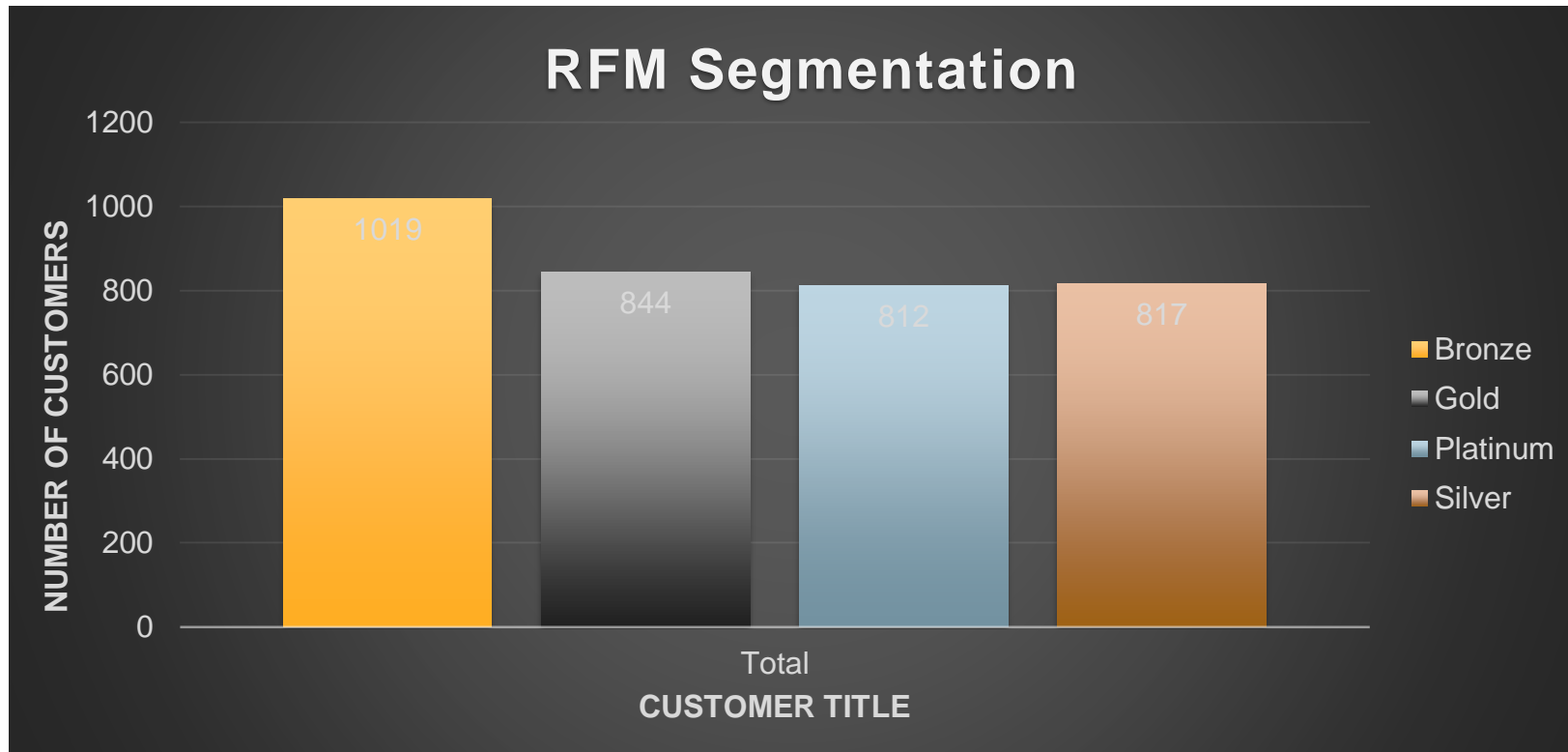
Model Development

Customer Segmentation Definition table

Customer Category	Description	RFM Value
Platinum Customer	Most recent purchase, buys often and spends large amt of money	>433
Gold Customer	Relatively recent purchase, bought more than once and spends good amount of money	<=433 and >311
Silver	Bought recently, not very often, average money spent/ Bought recently, never bought before, spent small amount	<=311 and >211
Bronze	Bought recently, never bought before, spent small amount/ Purchases was long time ago, frequency is quite high, amount spent is high/ Very low recency, low frequency, but high amount spent	<=211

Model Development

Customer Segmentation using RFM Analysis



Interpretation

Final Targeting Methodology

- The majority of customers between the ages of 29 and 50 should be targeted.
- Females have made slightly more bike-related purchases in the last three years than males.
- In all age categories, the largest number of customers are categorized as 'Mass Customer.'
- The Financial, Entertainment, and Retail industries provide the most profit per customer for the business.
- The Financial and Manufacturing are the company's most profitable industries. In addition, Health Industry is not far behind.
- Victoria (VIC) is a state where the number of people who do not own a car is slightly higher than the number of people who do
- We should start targeting customers which fall under following three categories:- Platinum, Gold and Silver Customer Titles.