

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Karan Gupta

Agenda

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

Introduction

Identifying high-value customers from amongst 1000 new Customers for Sprockets

Sprocket Central Pvt. Ltd. specializes in high-quality bikes and accessible cycling accessories to riders

To maximize its business, Sprockets needs a thorough analysis of their existing customer dataset, and then extrapolate the insights from that to identify customer trends and behavior to drive marketing for a new list of 1000 customers.

To identify and recommend high-value customers from the provided list of 1000 customers, we will first analyze the labelled data (the data of existing customers provided by Sprockets. We'll follow below methodology, discussed in depth in later parts of this presentation –

- Age-Bracket Analysis for new and old customers
- Gender-wise past bike purchase record analysis
- Wealth-segment distribution analysis for new and old customers
- State-wise car ownership analysis
- Job Industry Analysis
- RFM Modelling for Customer Segmentation

Data Exploration

Data Cleaning and Transformation

1. First step in exploration of dataset would involve identifying anomalies and remediating them for better analysis. This would involve below steps (all anomalies and steps discussed in detail in the previous email):
 - a) Cleaning inconsistent and irrelevant values –
 - i. Gender data, blank records, inconsistent ages and details of deceased customers in Customer Demographics
 - ii. Inconsistent customer id's, state identifier details in Customer Addresses
 - iii. Blank values, date formatting, and filtering out cancelled orders in Transactions dataset
 - b) Adding a new column indicating Customer Age by using their provided DOB
 - c) Filtering results for age below 100
2. After cleaning, we will:
 - a) Add a new column for Customer Age, and create brackets for categorizing age into 5 windows
 - b) Join the **Transactions**, **Demographics** and **Addresses** datasets on basis of **Customer ID** for generation of a complete dataset containing all necessary information about transacting customers
 - c) Add an external dataset to our repository from Australian Bureau of Statistics for state-wise population

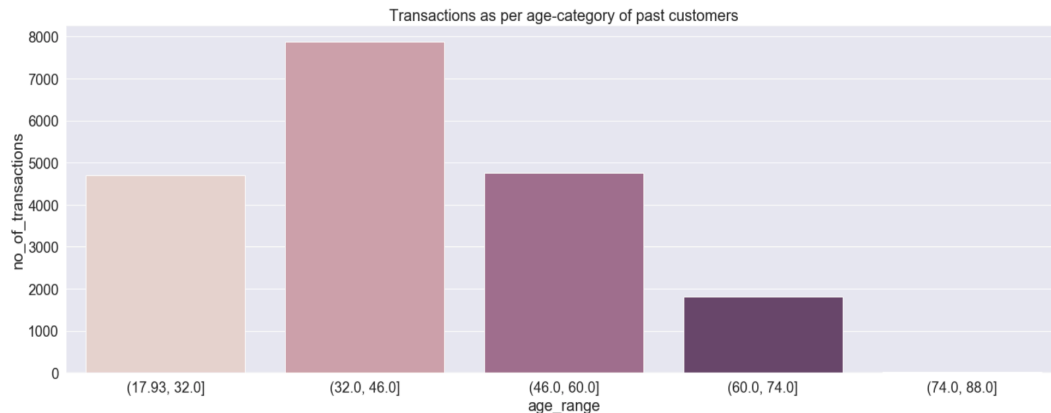
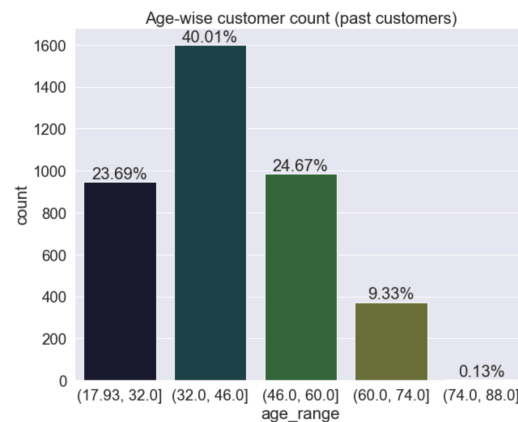
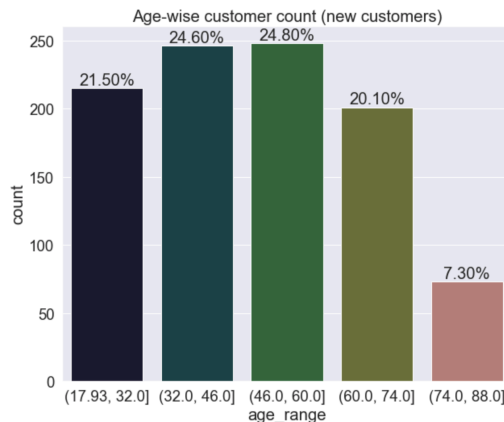
Data Exploration

Customer Age Distributions

Most Old customers are in the age group of 32 to 46

In the new customer list, the number of customers in the age group of 32 to 46 is high, but another age group has engendered, 46 to 60

As per a historic transaction analysis, the customers in the age group of 32 to 46 have been recorded to have made the most purchases with Sprockets. Thus this would be an important age group to consider for ferocious marketing



Data Exploration

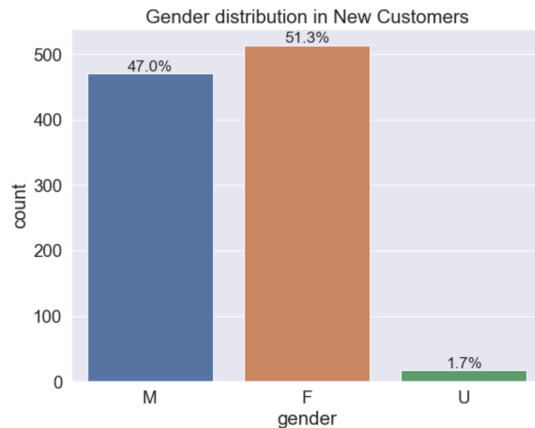
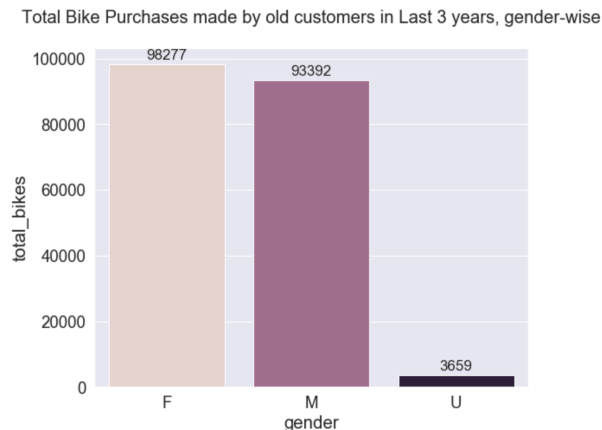
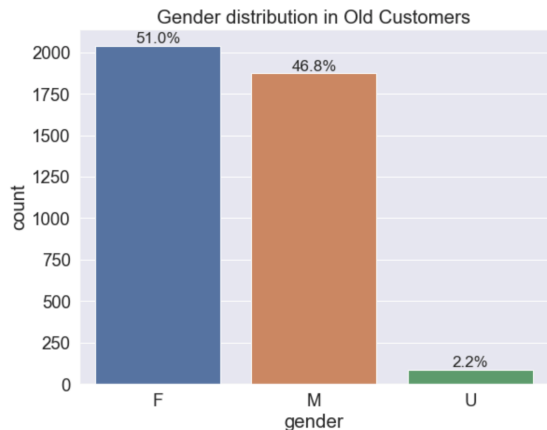
Gender-wise past bike purchase record analysis

For both the old and new customers, female population is the highest

For both new and old customers, approximately 50% of the bikes were bought by females, with the male customers accounting for approximately 48% of the purchases

The gender U accounts for as much as 2% in case of both the old and the new customers

Female customers having the highest bike purchase record, make up majority of the sales and should be effectively targeted.



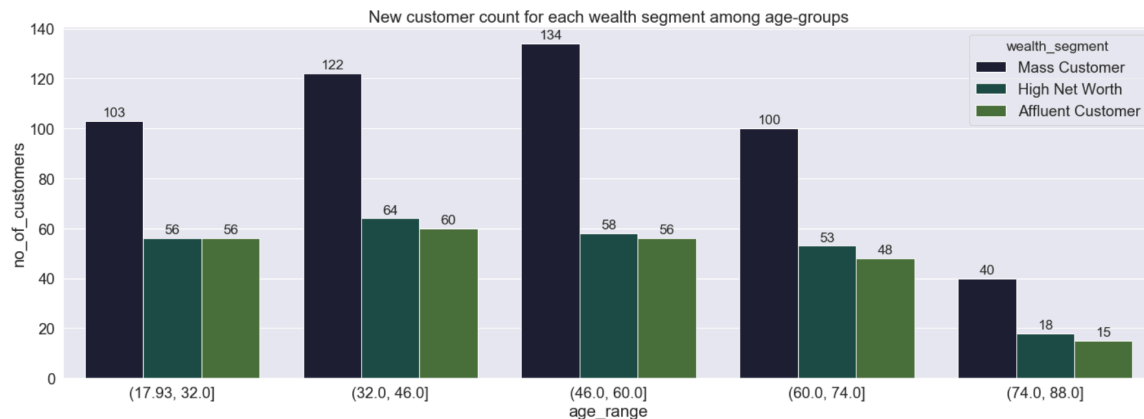
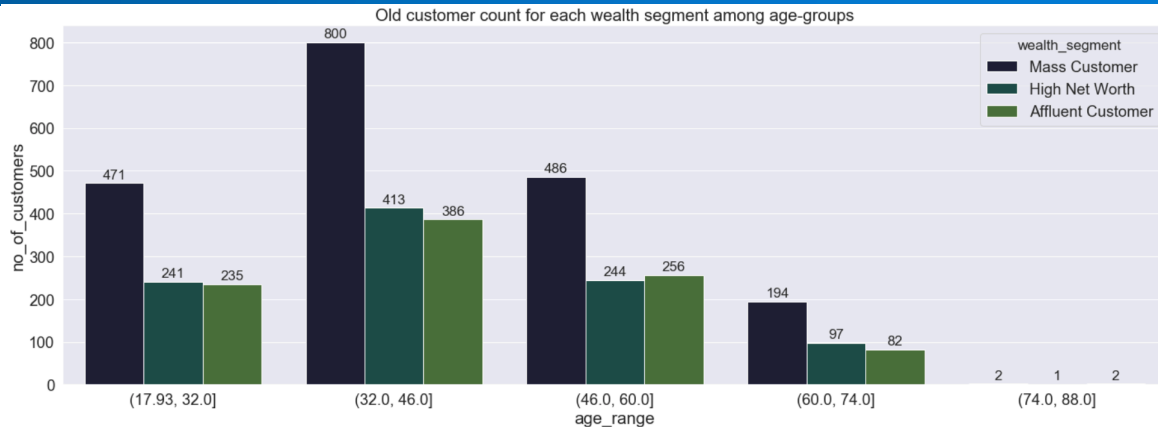
Data Exploration

Wealth Segment Analysis

In all age groups, for both old and new customers, most customers of Sprockets are ones who fall under the Mass Customer wealth segment

Also, most transactions are performed by the Mass Customer Segment, specifically by the customers of this segment who fall in the 32 to 46 age bracket

So this segment, Mass Customers in the age bracket of 32 – 46, should be considered high value customers



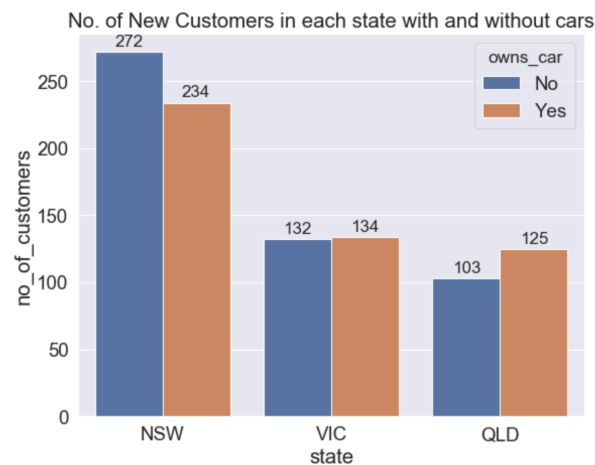
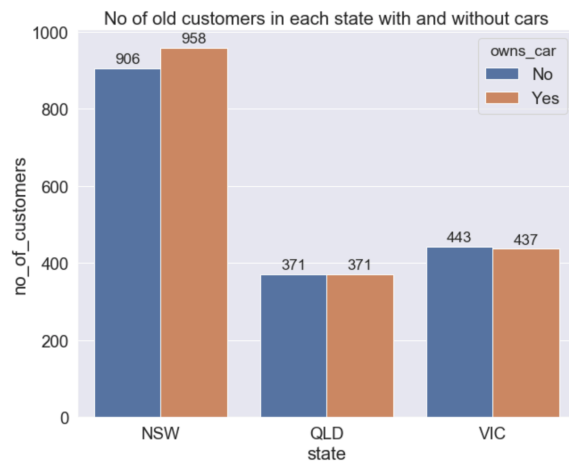
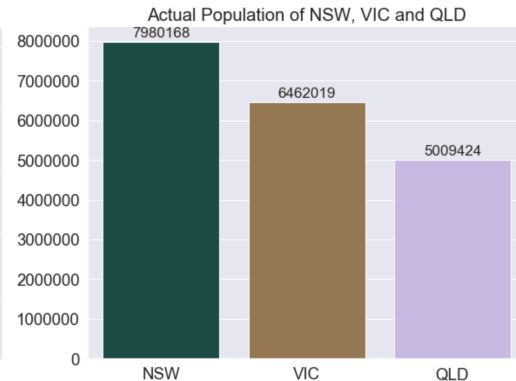
Data Exploration

State-wise car ownership

Most of the old customer population of Sprockets was from New South Wales, and a similar trend is observed amongst the new customers

Some of the old customers haven't ever transacted with Sprockets. These customers can be targeted with special schemes for encouragement

There's still a very large population in all 3 states that are not yet enrolled as Customers of Sprockets

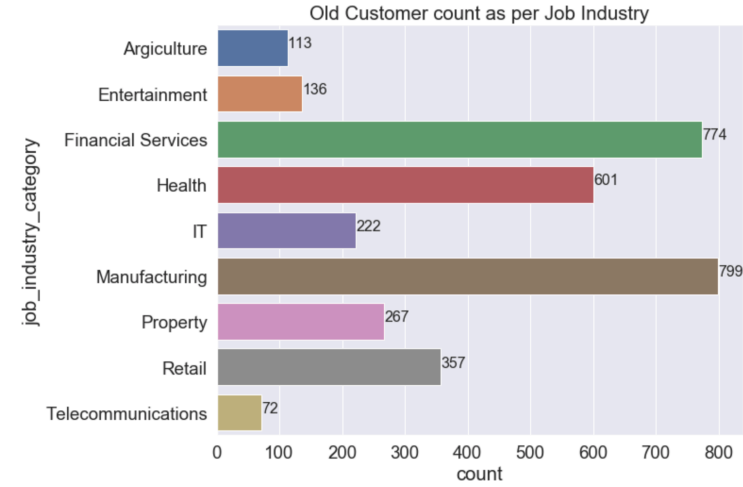
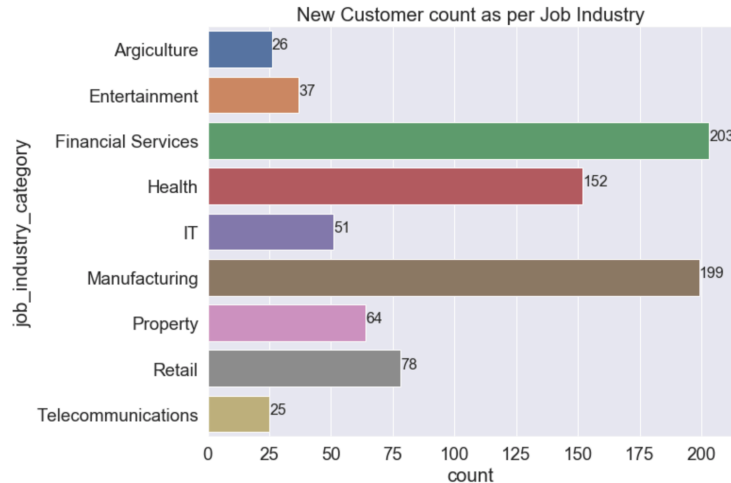


Data Exploration

Job-Industry Analysis

As per the chart, people who work in Manufacturing and Financial Services have the highest count as Sprocket's customers

Thus this granularity can be used while filtering customers as highly valuable



Model Development

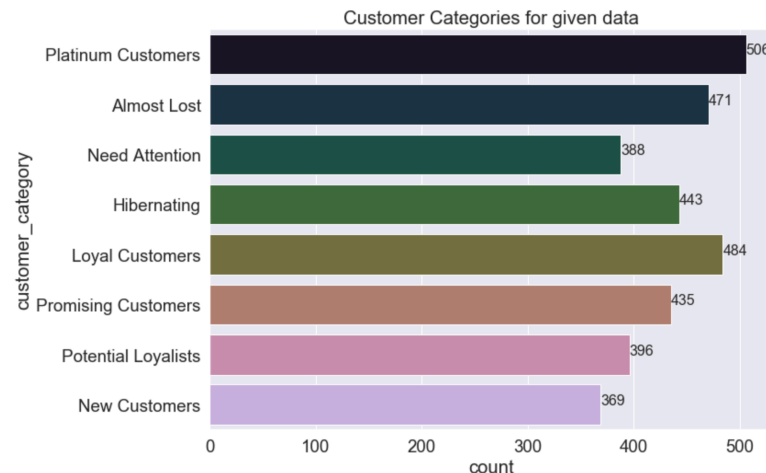
RFM Analysis and Modelling for Customer Segmentation

RFM stands for Recency, Frequency and Monetary value, each of which corresponds to a key customer trait.

There RFM metrics are important indicators of a customer's behavior because frequency and monetary value affects a customer's lifetime value and recency affects retention, a measure of engagement

This method is one of the most popular, easy to use segmentation method

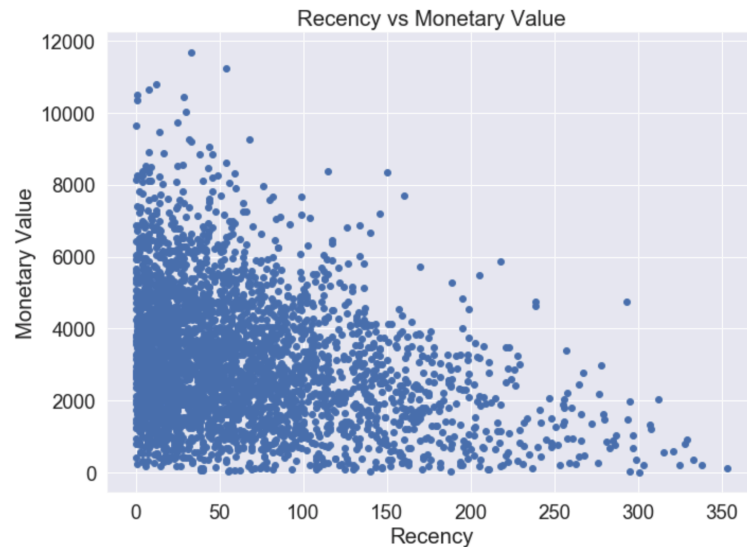
We will be using this model to cluster customers into separate classes on basis of the 3 KPI's used in this method – Recency (how recently has the customer made a purchase), Frequency (how many purchases/products has the customer completed/bought) and Monetary value (Total profit from a customer's transaction(s))



Interpretation

Recency Vs Monetary Value

From the relation between Recency and Monetary value, the Monetary Value is more for more Recent items i.e. Customers who have purchased more recently have generated more revenue/profit

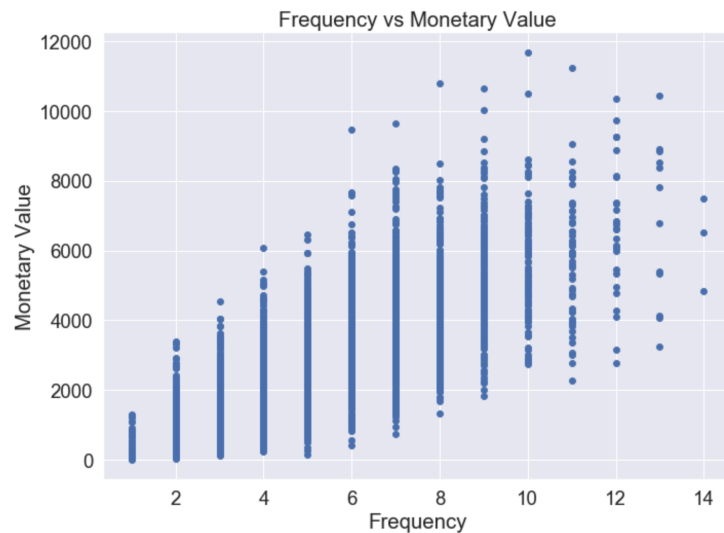


Interpretation

Frequency Vs Monetary Value

Relation between Frequency and Monetary value is quite linear, i.e. customers who have purchased more number of items, have provided more profit to Sprockets.

Thus Customers classified as Platinum and Potential Loyalists who visit frequently correlate with increased profit for the business

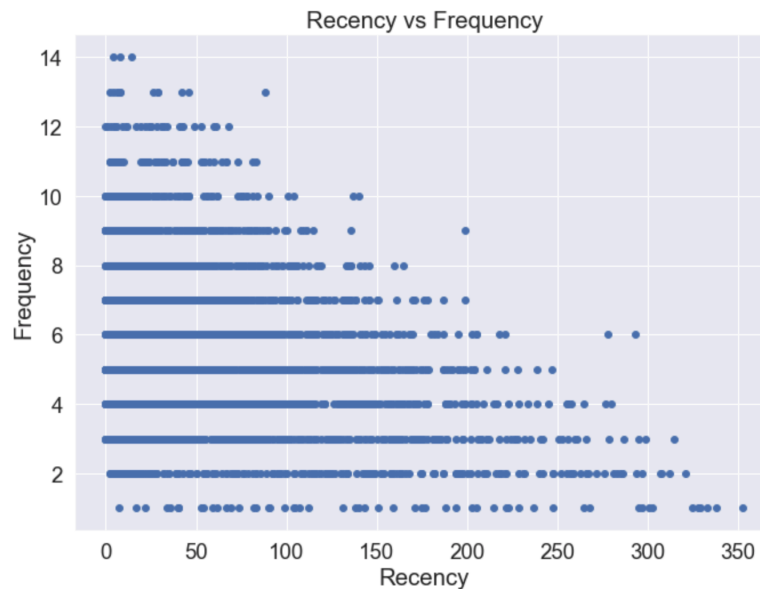


Interpretation

Recency Vs Frequency

There is sort of an inverse relationship between Recency and Frequency and rightfully so, since the customers who had visited long back have a high Recency Value, and since they haven't returned in a long time, they have a low frequency value associated.

Customers who visited more recently (0-50 days ago), have a higher frequency possibility.



Interpretation

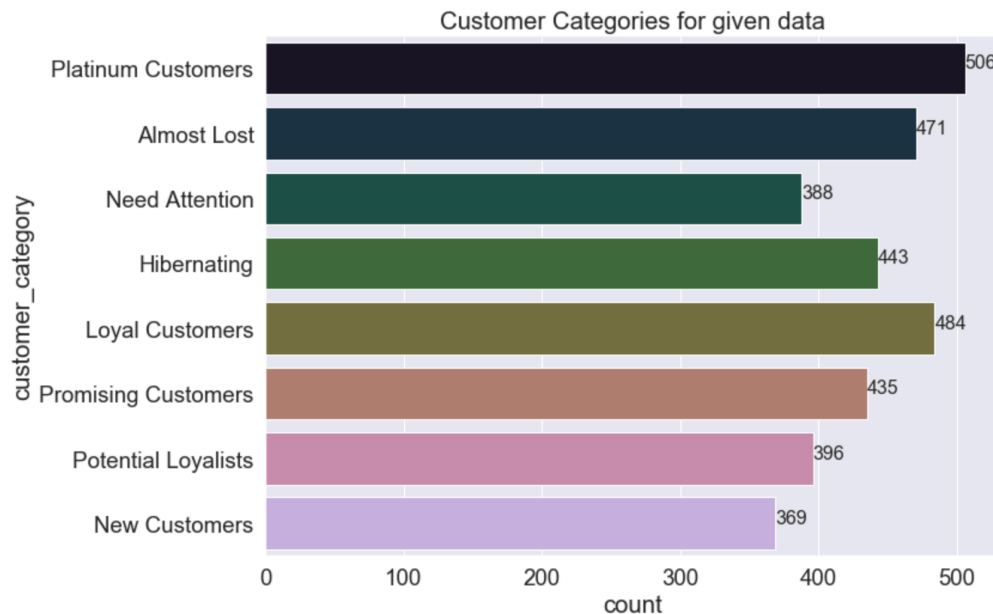
Category Division

As per our RFM Model, we will divide all customers into 8 different categories –

1. **Almost Lost** – This is the category whose RFM score is in the range of 111 to 123, indicating that these are very old customers, who have neither substantially contributed to the profit since a long time, nor have they purchased any product recently

2. **Need Attention** – RFM Score in the range of 124 – 144. These customers, though they are old, at one point they had a good frequency of buying products and had also substantially contributed to company profits. So these customers need to be encouraged to continue their business with the firm

3. **Hibernating** – These customers' RFM scores lie in the range 211 to 223. These are customers who have not transacted in a long time but had mildly contributed in raising Sprockets' revenues with their transactions in the past.



Interpretation

Category Division

4. **Promising Customers** – These are customers whose RFM lies between 223 and 244. These customers had transacted a while back and had a good frequency of ordering and had also made substantial contributions to the revenue. Hence these are called promising, since by reaching out and propagating new offers and campaigns, we can ensure their continued engagement with the business
5. **Potential Loyalists** – With their RFM score in the range of 311 tot 323, these are somewhat recent customers with average frequency and average spending amount. Through membership or loyalty programs or improved recommendations related to products, they can become Loyal or Platinum Customers
6. **Loyal Customers** – These customers have RFM in the range of 324 to 344. These customers have somewhat good recency but have had good monetary value for the firm. They are loyal, and their recency can be improved through increased digital engagement and better advertisement
7. **New Customers** – These Customers don't have a very high frequency and have not yet substantially contributed to the monetary value, but these are very recent customers who need to be introduced to offers and schemes for increasing their F and M scores. RFM score range – 411 to 423
8. **Platinum Customers** – These customers have high recency, frequency as well as they have contributed substantially to the firm's profit. Hence these customers have the highest value for Sprockets. They are classified by a strikingly high RFM score of 424 to 444