

TheAnalyticsTeam

# Sprocket Central Pty Ltd

Data analytics approach  
Sri Hartina

# Agenda

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

# Introduction

## Identify and Recommend High Value Customers

### Outline of Problem

- Sprocket Central is a company that specializes in bikes & cycling accessories.
- Their marketing team is looking to boost business by analysing their provided customer dataset
- Using the existing 3 datasets provided to recommend which of these 1000 new customers should be targeted to drive the higher value for the company.

### Contents of Data Analysis

- New and Old Customers by age distributions
- Bike related purchases over the last 3 years by gender
- New and Old Customers Job industry distributions
- Wealth segmentation by age category
- Number of cars owned and not owned by state
- RFM analysis and customer classification

# Data Exploration

## Data Quality Assessment

Assessment of data quality and completeness in preparation for analysis

### Data Quality Dimension

- Accuracy:** Inaccurate data
- Completeness:** Data field with blank values
- Consistency:** Spelling text value differently with same meaning
- Currency:** Values up to date
- Relevancy:** Irrelevant field data
- Validity:** Format value

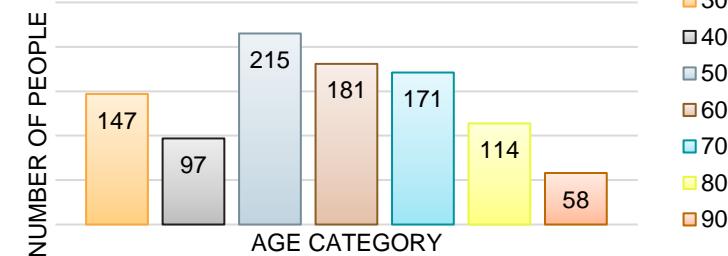
Worksheet Name	Data Quality Issue					
	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Transaction		online order & brand: blanks				list price & product first sold date: format
Customer Demographic	DOB: Inaccurate	DOB & job title: blanks	gender: inconsistency	deceased customer: filter out	default: column delete	DOB: format
Customer Address			state: inconsistency			

# Data Exploration

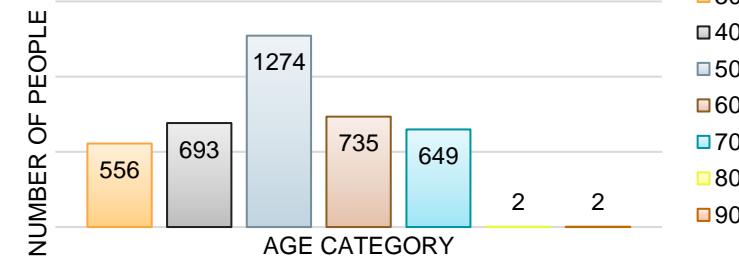
## New and Old Customer by Age Distribution

- Most customers in New and Old Customers are aged between 41-50 years.
- The lowest customers in New Customer is aged between 81-90 years.
- The lowest customers in Old Customer is aged between 71-90 years.
- The 3 group of age in New Customer most populated are aged between 41-70 years.
- The 3 group of age in Old Customer most populated are aged between 31-60 years.

New Customer Age Distribution



Old Customer Age Distribution



# Data Exploration

## Bike Related Purchases over Last 3 Years by Gender

- Bike related purchases over last three years mostly made by females: 483.009 about 51%.
- Purchases made by males about 473.314 about 49% almost 10.000 less than female.

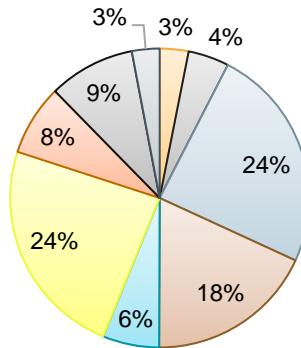
Gender	Number of Purchases	Number of Purchases (%)
Female	483009	51%
Male	473314	49%
<b>Grand Total</b>	<b>956323</b>	<b>100%</b>

# Data Exploration

## Job Industry Distribution

- The most job industry in Old and New Customer are in Manufacturing and Financial Services which is more than 20%
- The smallest job industry in Old and New Customer are in Telecommunications and Agriculture which is only 2% and 3%

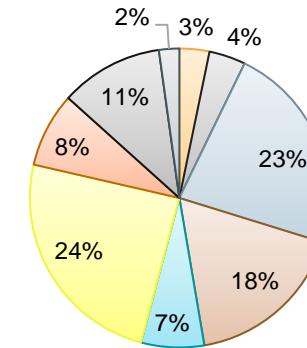
Job Industry in New Customer



Legend:

- Argiculture
- Entertainment
- Financial Services
- Health
- IT
- Manufacturing
- Property
- Retail
- Telecommunications

Job Industry in Old Customer

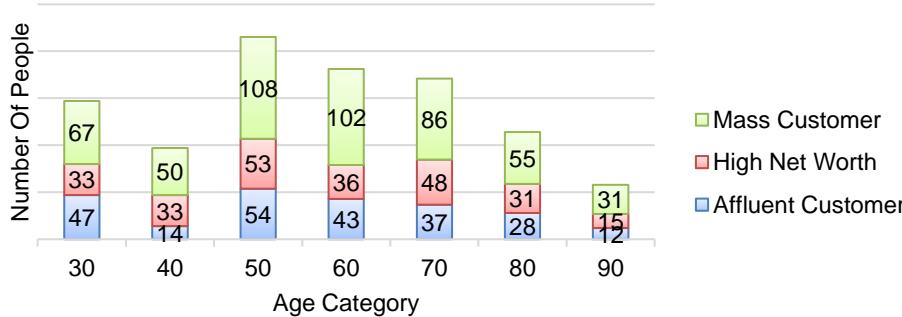


# Data Exploration

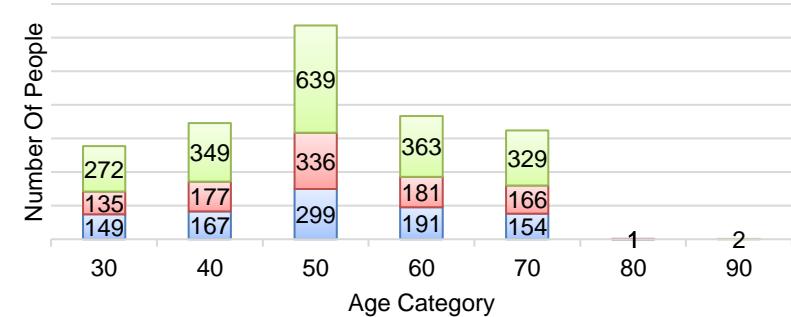
## Wealth Segmentation by Age Category

In all age category both in New and Old customers, the largest number of customers are classified as "Mass Customer"

New Customer Wealth Segmentation By Age



Old Customer Wealth Segmentation By Age

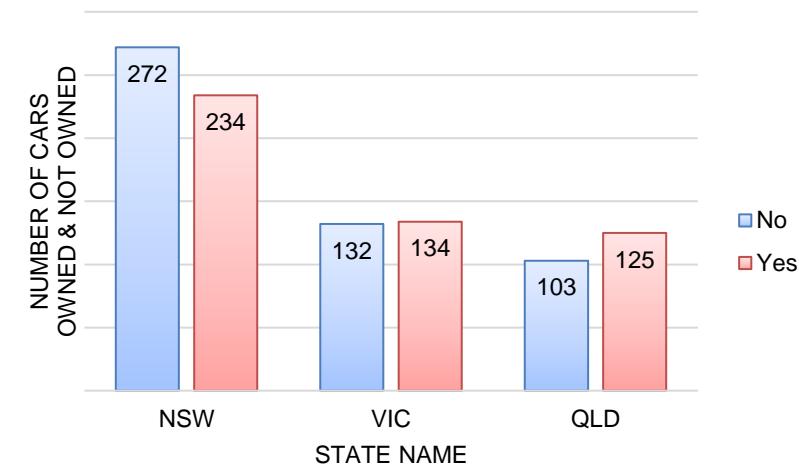


# Data Exploration

## Number of Cars Owned and Not Owned in Each State

- Based on the data, New South Wales (NSW) is a country that has the most population. Besides NSW has the largest amount of people that do not own a car.
- Victoria (VIC) has more people who own cars than those who do not. but the difference in numbers is not significant
- Queensland (QLD) has the lowest number of population. QLD has more people who own cars than those who do not.

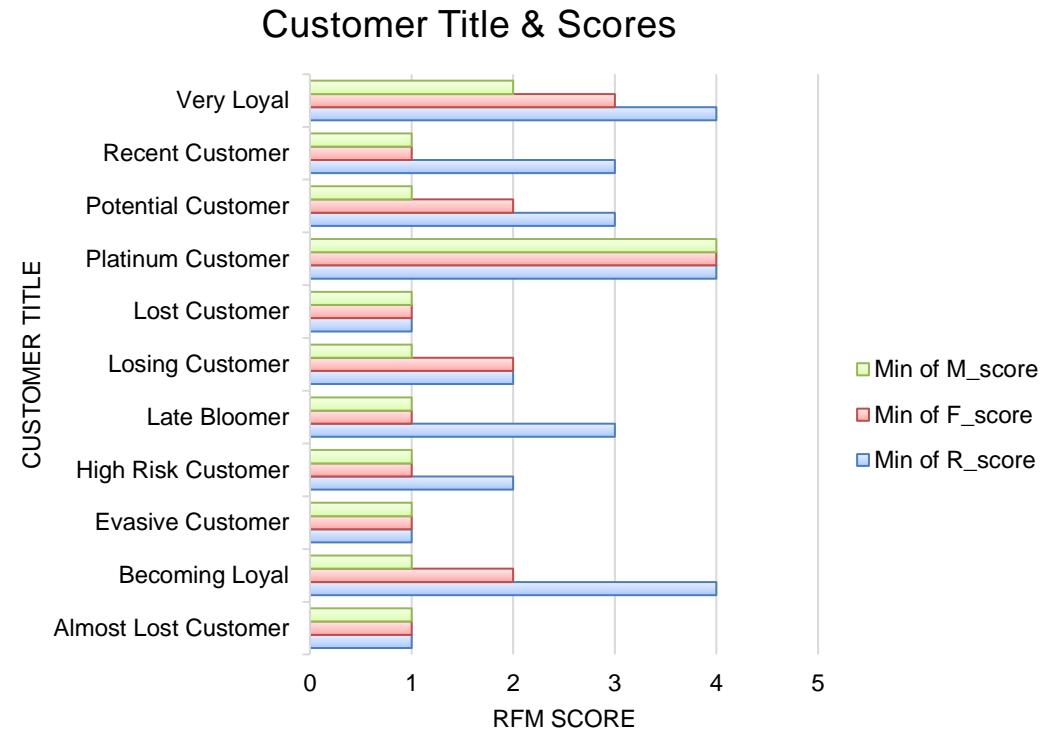
Number of Cars Owned in Each State



# Model Development

## RFM Analysis and Customer Classification

- RFM analysis is a marketing method that is used to identify the best customers to increase its revenue and value.
- RFM analysis is objectively evaluate and group customers based on the recency, frequency, and monetary amount of their most recent transactions.

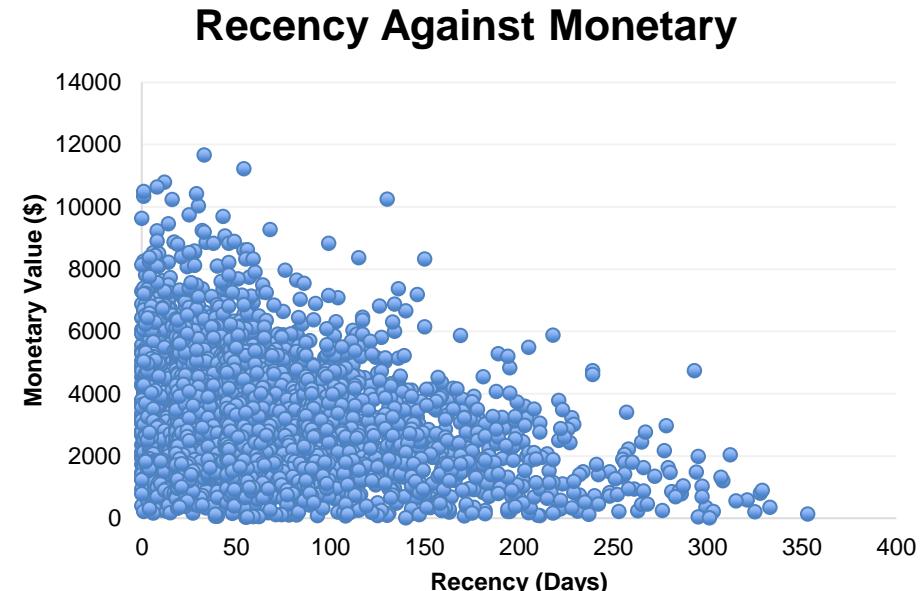


# Model Development

## Scatter Plot based of RFM Analysis

### Recency Against Monetary

- The chart shows that customers who purchased recently have generated more revenue, than customers who made a purchase a while ago.
- Customers from recent past (50-100 days) also show to generate a moderate amount of revenue.
- Customers who purchased more than 200 days ago generate low revenue.
- There is a negative relationship between monetary and recency value.

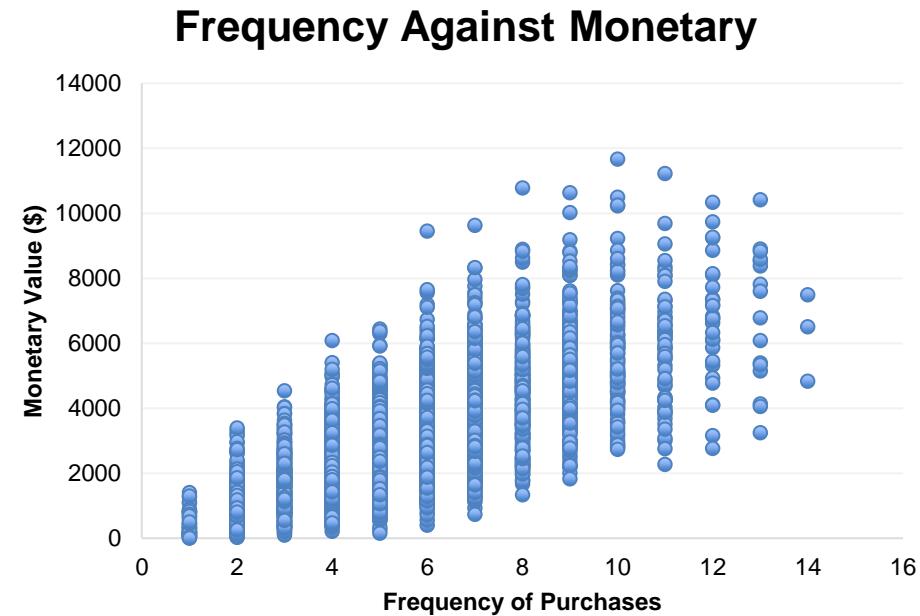


# Model Development

## Scatter Plot based of RFM Analysis

### Frequency Against Monetary

- Customers who make a one-time purchase generate lower revenue.
- The higher the frequency of purchases made by customers the more revenue generated.
- Naturally, there is a positive relationship between frequency and monetary gain for the business.

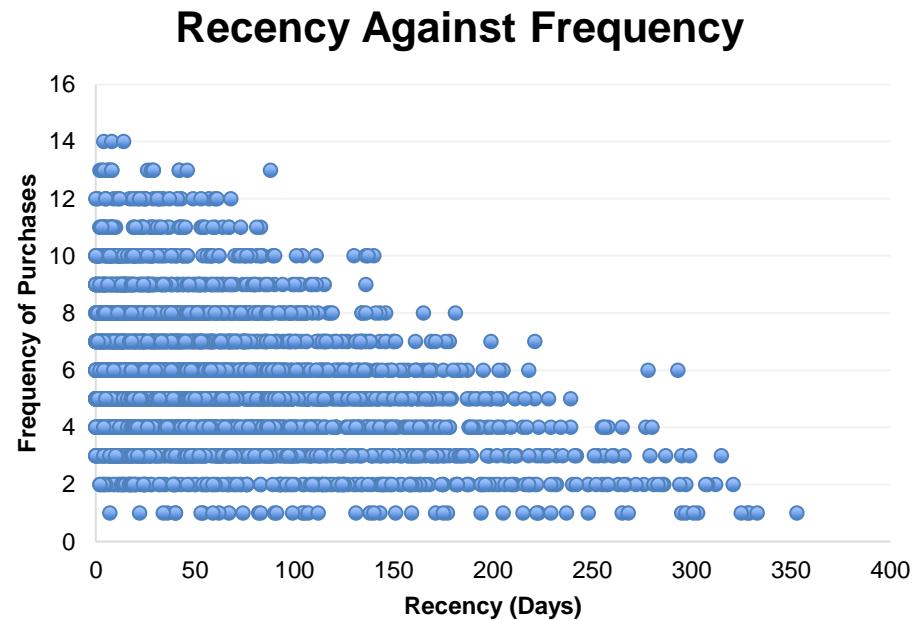


# Model Development

## Scatter Plot based of RFM Analysis

### Recency Against Frequency

- Customers who make purchases (> 250 days) tend to only make up to 2x purchases.
- Customers who make purchases on time (0-50 days) make purchases more often.
- The higher the purchase frequency, the more recently (in time of day) the customer makes a purchase. So that customers on recent days make purchases more often.
- There is a negative relationship between frequency and recency value.



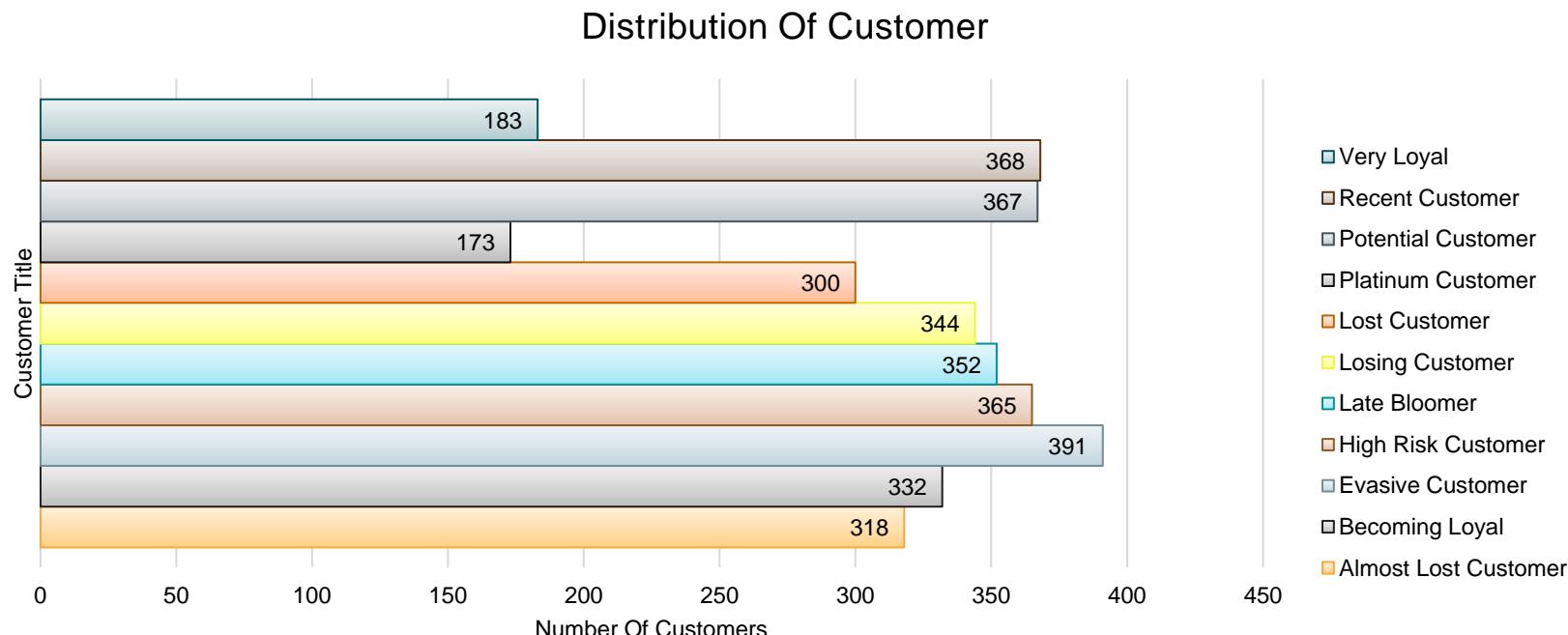
# Model Development

## Customer Title Definition with RFM Value

Rank	Customer Title	Description	RFM Value
1	Platinum Customer	Most recent buy, buys often, most spent	444
2	Very Loyal	Most recent buy, buys often, spends large amount of money	433
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	413
4	Recent Customer	Bought recently, not very often, average money spent	344
5	Potential Customer	Bought recently, never bought before, spent small amount	323
6	Late Bloomer	No purchases recently, but RFM value is larger than average	311
7	Losing Customer	Purchases was a while ago, below average RFM value	224
8	High Risk Customer	Purchases was long ago, frequency is quite high, amount spent is high	212
9	Almost Lost Customer	Very low recency, low frequency, but high amount spent	124
10	Evasive Customer	Very low recency, Very low frequency, small amount spent	112
11	Lost Customer	Very Low RFM	111

# Model Development

## Customer Title Distribution



# Interpretation

## Summary Table of The Top 1000 Targeted Customers

Rank	Customer Title	Description	Number of Customers	Cumulative	Customer Selection
1	Platinum Customer	Most recent buy, buys often, most spent	173	173	173
2	Very Loyal	Most recent buy, buys often, spends large amount of money	183	356	183
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	332	688	332
4	Recent Customer	Bought recently, not very often, average money spent	368	1056	312
5	Potential Customer	Bought recently, never bought before, spent small amount	367	1423	0
6	Late Bloomer	No purchases recently, but RFM value is larger than average	352	1775	0
7	Losing Customer	Purchases was a while ago, below average RFM value	344	2119	0
8	High Risk Customer	Purchases was long ago, frequency is quite high, amount spent is high	365	2484	0
9	Almost Lost Customer	Very low recency, low frequency, but high amount spent	318	2802	0
10	Evasive Customer	Very low recency, Very low frequency, small amount spent	391	3193	0
11	Lost Customer	Very Low RFM	300	3493	0

# Thank You