

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Division Name] - [Engagement Manager], [Senior Consultant], [Junior Consultant]

Agenda

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

Introduction

Recommend 1000 targeted new customers.

- Sprocket Central Pty Ltd is a long-standing KPMG client whom specialises in high-quality bikes and accessible cycling accessories to riders.
- Their marketing team is looking to boost business by analysing their existing customer dataset
- Using the existing 3 datasets to analyze & recommend 1000 customers should be targeted to drive the most value for the organisation.

Contents

- ❖ “New” & “Old” Customer Age distribution
- ❖ Bike related purchases over the last 3 years by gender
- ❖ Job industry distributions
- ❖ Wealth segmentation by age category
- ❖ Number of cars owned & not owned by state
- ❖ RFM analysis & customer classification

Data Exploration

Data quality assessment

Key issues for data quality assessment

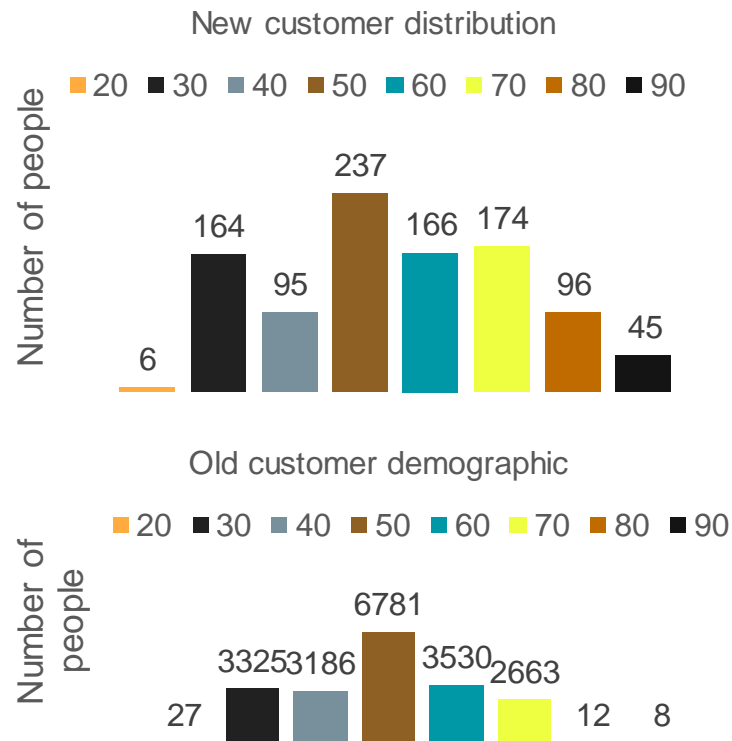
- Accuracy
- Completeness
- Consistency
- Currency
- Relevancy
- Validity
- Uniqueness

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer demographic	DOB: inaccurate Age: missing	Job title: blanks Customer id: incomplete	Gender: inconsistency	Deceased customers: filter out	Default column: delete	
Customer Address		Customer id: incomplete	States: inconsistency			
Transactions	Profit: missing	Customer id: incomplete Online order: blanks Brand: blanks			Cancelled state order: filter out	List price: format Prdocut sold day: format

Data Exploration

'New' & 'Old' Customer Age Distributions

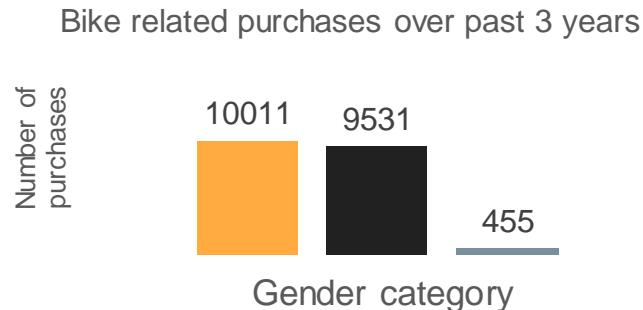
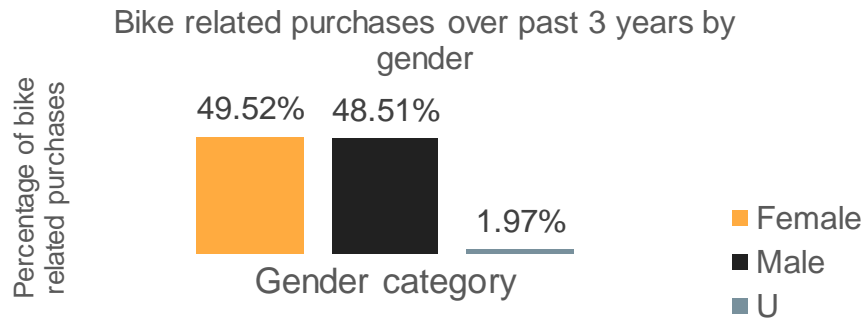
- 'New' & 'Old' Customers share the same patterns: they are majorly from 40-49.
- The lowest age groups are under 20 and over 80 for both 'New' & 'Old' customer lists.
- The 'New' customer list suggest the new target group – aged from 20 – 29 and 40- 69 are most populated.
- The ' Old' customer list suggest customers from 20 to 69.
- There is a significant fall in the 30 – 40 age group.



Data Exploration

Bike related purchases over last 3 years by gender

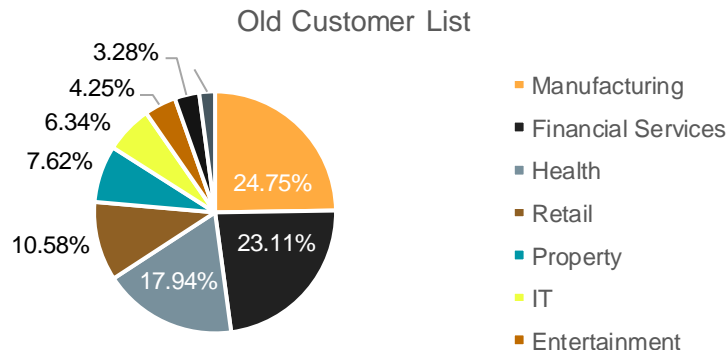
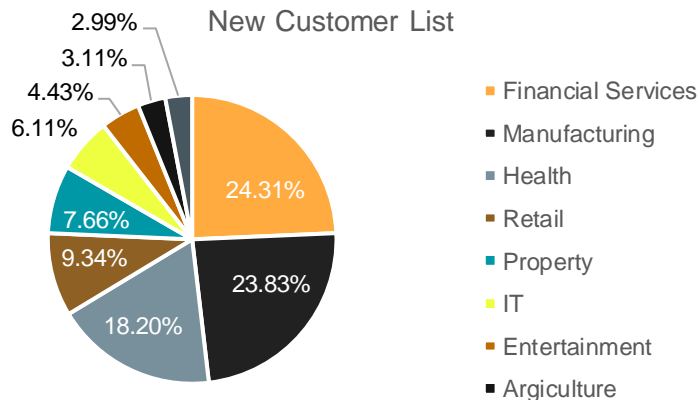
- Female make up majority of bike related sales.
- Over the past 3 years, nearly 50% of bike related purchases are made by female, compared to 48.51% purchases made by male. Approximately 2% are made by unknown gender.
- Numerically, females have 10000 purchases more than males.



Data Exploration

Job industry distribution

- 20% of 'New' customers are in Manufacturing & Financial services.
- The smallest number of customers are in Agriculture & Telecommunications at 3%.
- 'Old' customer shares the same pattern, at 19% and 21% in Manufacturing & Financial services respectively.

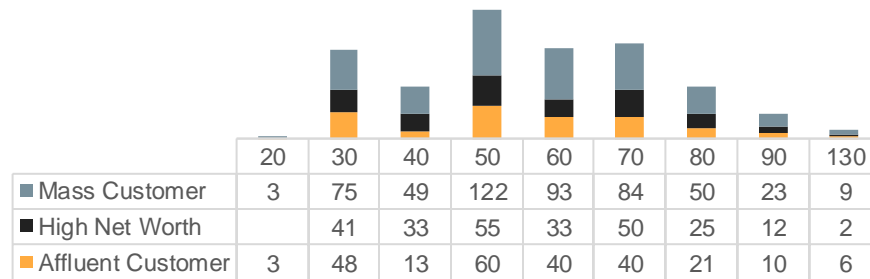


Data Exploration

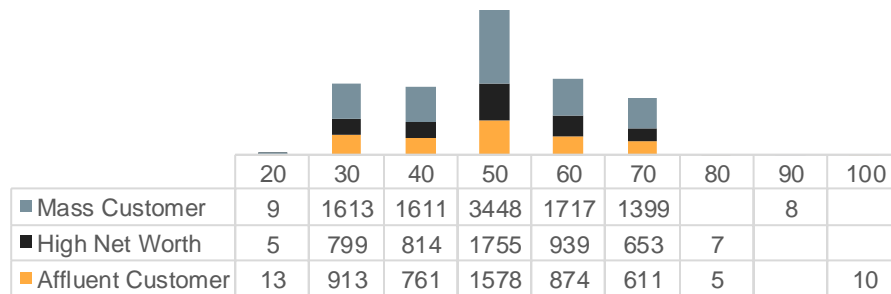
Wealth Segmentation by age category

- In all age group, the largest number of customers belongs to Mass customer category.
- The next category is the 'High Net Worth' customers.
- The 'Affluent customer' outperforms the 'High Net Worth' customer in the 40-49 age group.

New customer wealth segment by Age



New customer wealth segment by Age

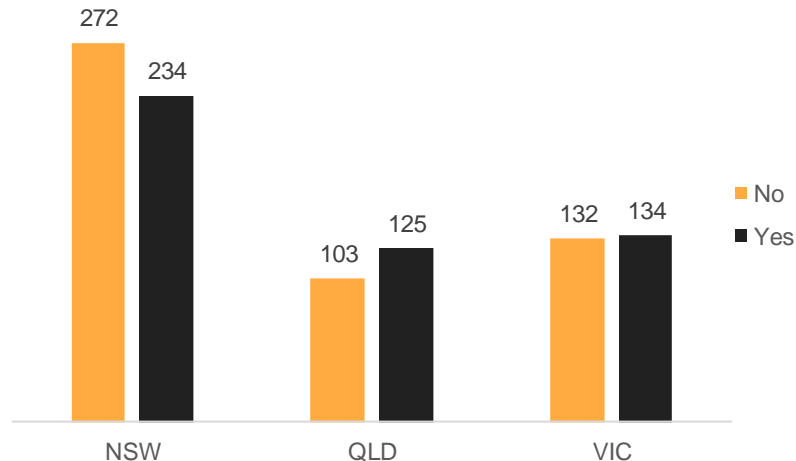


Interpretation

Number of cars owned and not owned by state

- NSW seems to have a higher number of people from which data was collected. NSW has the largest amount of people who **do not** own a car.
- Victoria is split evenly between the number of people own & do not own. However, both number are significantly lower than that of NSW.
- QLD has a considerable high number of customers that own a car.

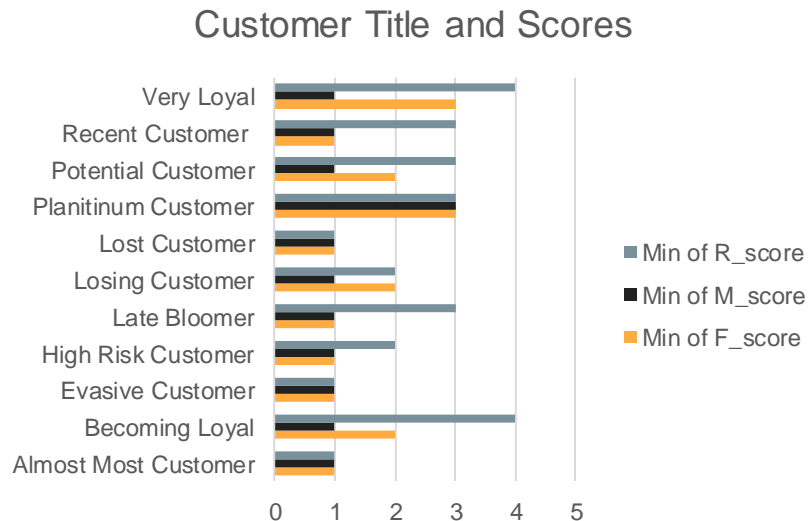
Number of cars owned in each state



Model Development

RFM Analysis & Customer Classification

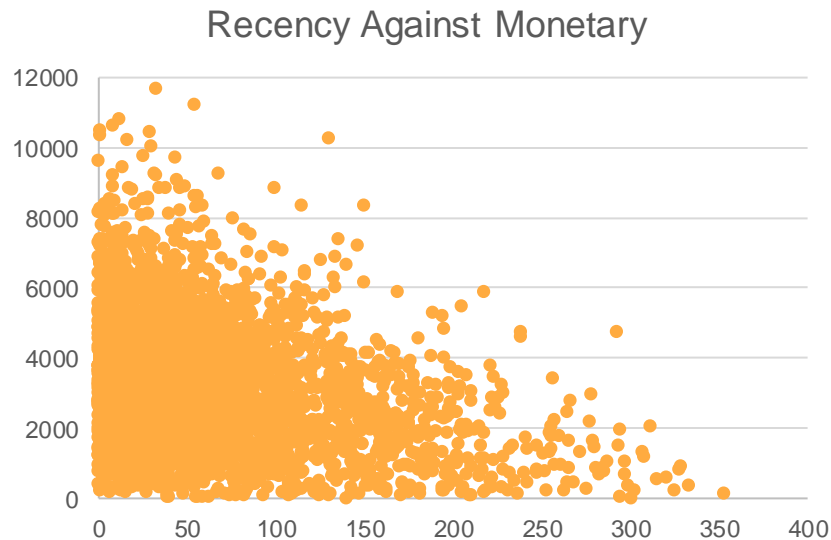
- RFM analysis is used to determine which customers a business should target to increase its revenue and value.
- The RFM (Recency, Frequency, and Monetary) model shows customers that have displayed high levels of engagement with the business in the three categories mentioned.



Model Development

Scatter-Plot based of RFM Analysis

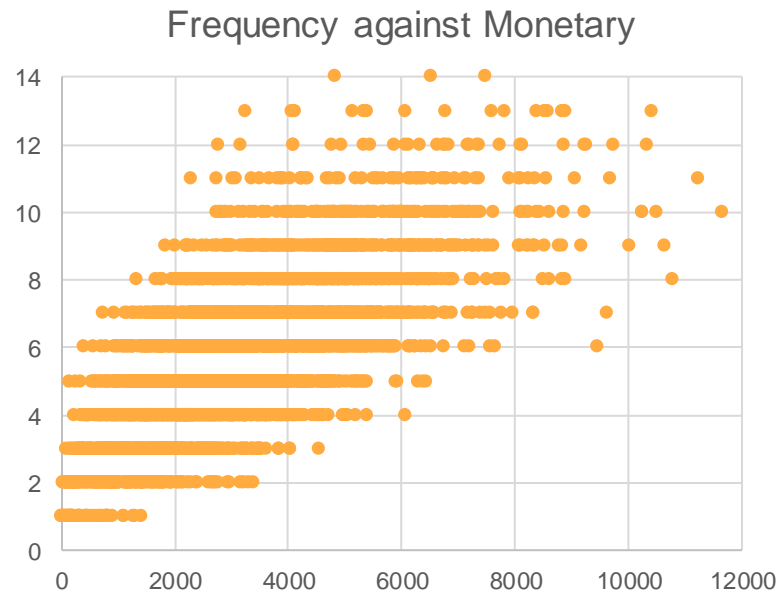
- The chart shows that customers who purchased more recently have generated more revenue, than customer who visited a while ago.
- Customers from recent past (50-100 days) also show to generate a moderate amount of revenue.
- Those who visited more than 200 days ago generate low revenue.



Model Development

Scatter-Plot based of RFM Analysis

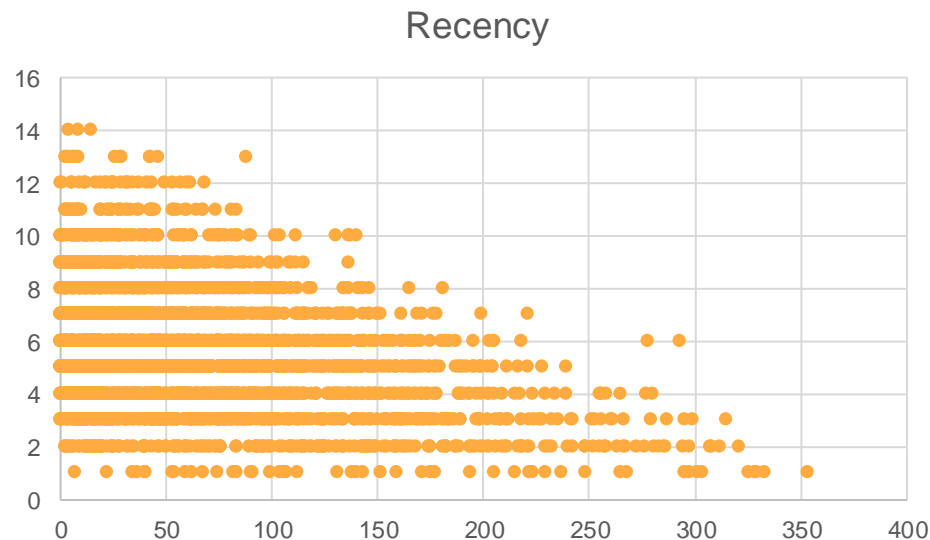
- Customers classified as “Platinum Customer”, “Very Loyal”, and “Becoming Loyal” visit frequently, which correlated with increased revenue for the business.
- Naturally, there is a positive relationship between frequency and monetary gain for the business.



Model Development

Scatter-Plot based of RFM Analysis

- Very low frequency of 0 – 2 correlated with high recency values. i.e More than 250 days ago.
- Customers that have visited more recently (0-50 days) have a higher chance of visiting more frequently (6+).
- Higher frequency has a negative relationship with recency values. Such that very recent customers are also frequent customers.



Model Development

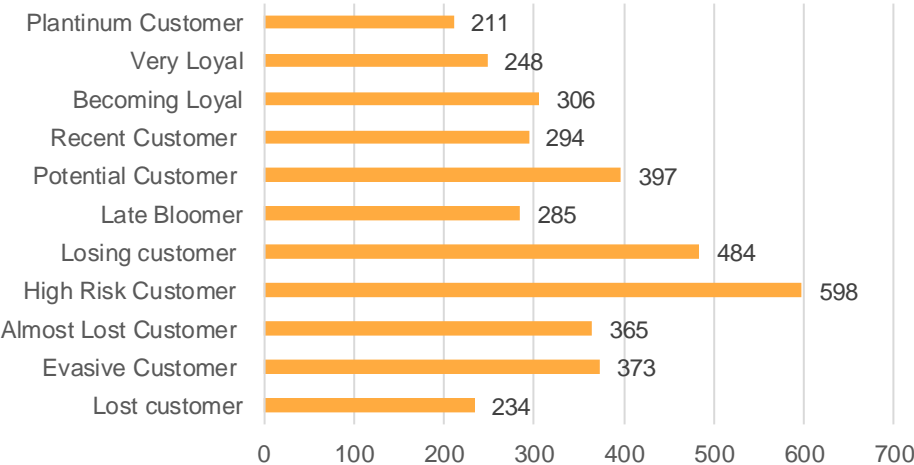
Customer Title Definition list with RFM values assigned

Rank	Customer Title	Description	RFM Value
1	Plantinum Customer	Most recent buy, buys often, most spent	444
2	Very Loyal	Most recent buy, buys often, spend large amount of money	433
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	421
4	Recent Customer	Bought recently, not very often, average money spent	344
5	Potential Customer	Bought recently, never bought before, spend 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	Purchase was long time 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

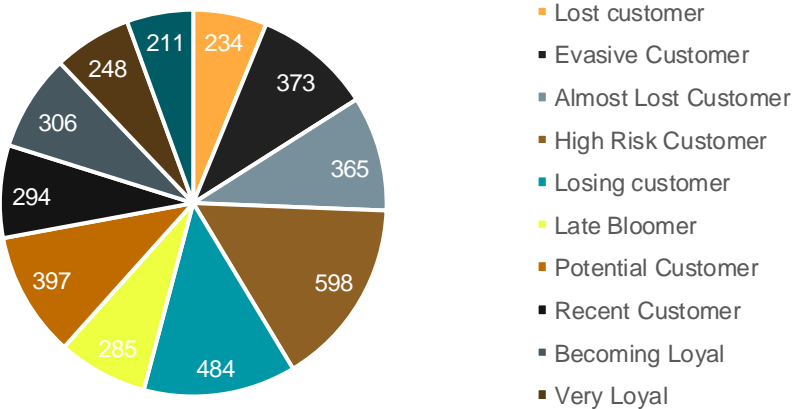
Interpretation

Customer Title Distributions in Dataset

Distribution of Customers



Distribution of Customers



Interpretation

Summary table of the Top 1000 Customers to target

Rank	Customer Title	Description	Number of customers	Culmulative	Customer Selection
1	Plantinum Customer	Most recent buy, buys often, most spent	211	211	306
2	Very Loyal	Most recent buy, buys often, spend large amount of money	248	459	248
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	306	765	211
4	Recent Customer	Bought recently, not very often, average money spent	294	1059	235
5	Potential Customer	Bought recently, never bought before, spend small amount	397	1456	0
6	Late Bloomer	No purchases recently, but RFM value is larger than average	285	1741	
7	Losing customer	Purchases was a while ago, below average RFM value	484	2225	
8	High Risk Customer	Purchase was long time ago, frequency is quite high, amount spent is high	598	2823	
9	Almost Lost Customer	Very low recency, low frequency, but high amount spent	365	3188	
10	Evasive Customer	Very low recency, Very low frequency, small amount spent	373	3561	
11	Lost customer	Very low RFM	234	3795	

Interpretation

Customer Target and Methodology

Rank	Customer Title	Description	Number of customers	Cumulati ve	Customer Selection
1	Platinum Customer	Most recent buy, buys often, most spent	211	211	306
2	Very Loyal	Most recent buy, buys often, spend large amount of money	248	459	248
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	306	765	211
4	Recent Customer	Bought recently, not very often, average money spent	294	1059	235

- Filter through the top 1000 customers by assigning the conditions discussed in the table above.
- The 1000 customers discovered would have bought recently, they have bought very frequently in the past and tend to spend more than other customers.

Appendix