

Sprocket Central Pty Ltd

Data analytics approach

Uchenna Ohurogu

Agenda

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

Introduction

Identifying and Recommending which of the new 1000 Customers to be targeted in order to drive the most value for Sprocket Central

Content of Data Analysis:

- ❖ New and 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 (or not) by states.
- ❖ RFM analysis and customer classification

This will be achieved following the three (3) phases of: Data Exploration, Model Development, and Interpretation.

Data Quality Assessment and 'Clean up'

Key issues noticed in the three datasets provided

- ❖ 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 Exploration

Data Quality Assessment and 'Clean up'

SUMMARY TABLE

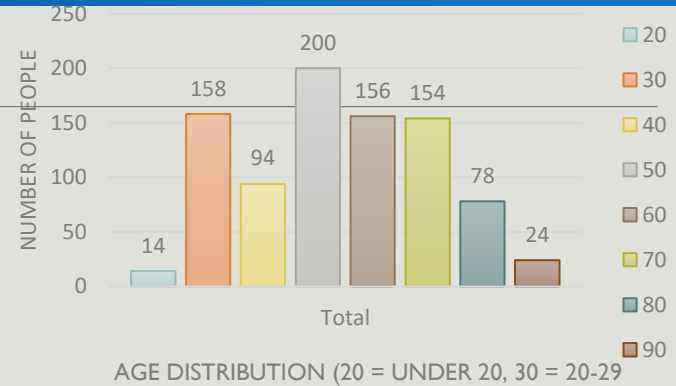
	ACCURACY	COMPLETENESS	CONSISTENCY	CURRENCY	RELEVANCY	VALIDITY
CUSTOMER DEMOGRAPHIC	DOB: Inaccurate, Age: Missing	Job title: Blank, Customer id: Incomplete	Gender: Inconsistent (An unknown 'U' value)	Deceased Customers: Filtered out	Default column: Deleted out	
CUSTOMER ADDRESS		Customer id: Incomplete	States: Inconsistent			
TRANSACTION	Profit: Missing	Customer id: Incomplete, Online order: blanks, Brand: blanks			Cancelled status order: Filtered out	List price: Format, Product sold date: Format

Data Exploration

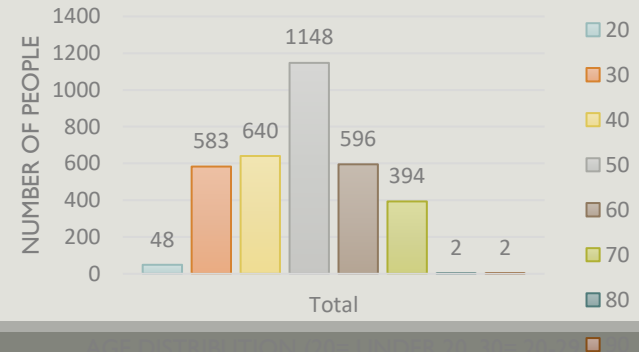
'New' and 'Old' Customer Age Distribution

- ❖ Most customers are aged between 40-49 in the adult age bracket, in 'New'. In the 'Old', majority are also in the adult age bracket.
- ❖ The lowest age group are under 20 and 70+ for both 'New' and 'Old' customer lists.
- ❖ The new customer list suggests that age groups 20-29 (30) and 40-49 (50) are the most populated.
- ❖ For the old customer list, age group 40-49 (50) is the most populated.

New Customer Age Distribution



Old Customer Age Distribution

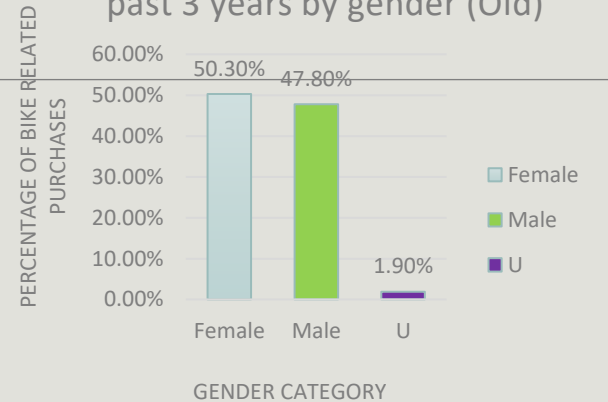


Data Exploration

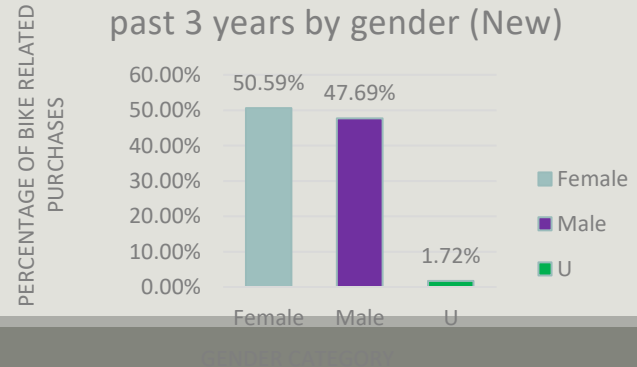
Bike related purchases over last 3 years by gender

- ❖ Over the last 3 years, about 50% of bike related purchases were made by females compared to about 48% of purchases made by males. Approximately 2% of purchases was made by unknown gender. This occurred in both the old and new customer distributions.
- ❖ Numerically, females purchases almost 10,000 more than for males.
- ❖ In summation, females make up majority of the bike related sales for the past 3 years.

Bike related purchases for the past 3 years by gender (Old)



Bike related purchases for the past 3 years by gender (New)

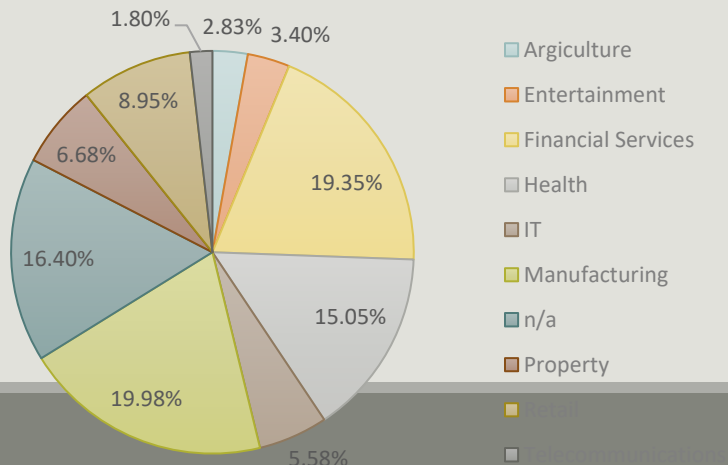


Data Exploration

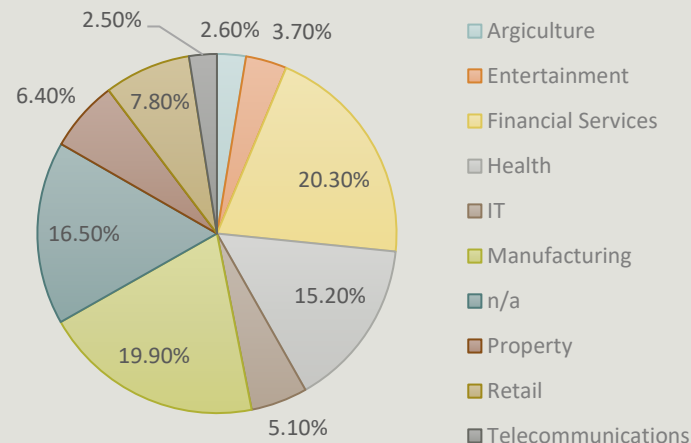
Job Industry Distribution

- ❖ There's a similar pattern of job distribution in the Old and New job industry distribution.
- ❖ The least number of customers are in Agriculture and Telecoms.

Old Customer Job Industry Distribution



New Customer Job Industry Distribution

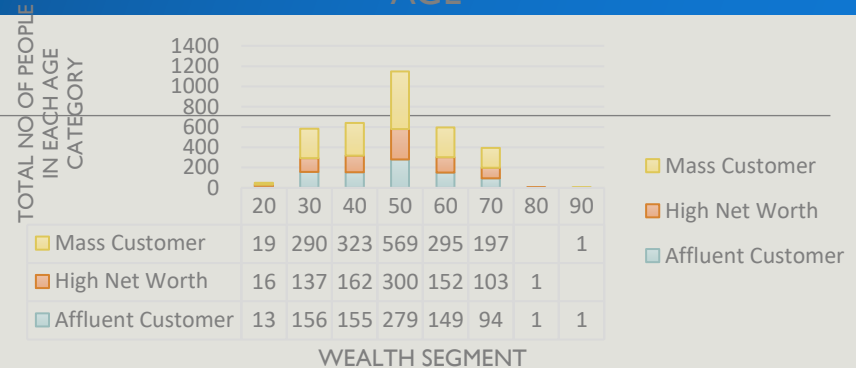


Data Exploration

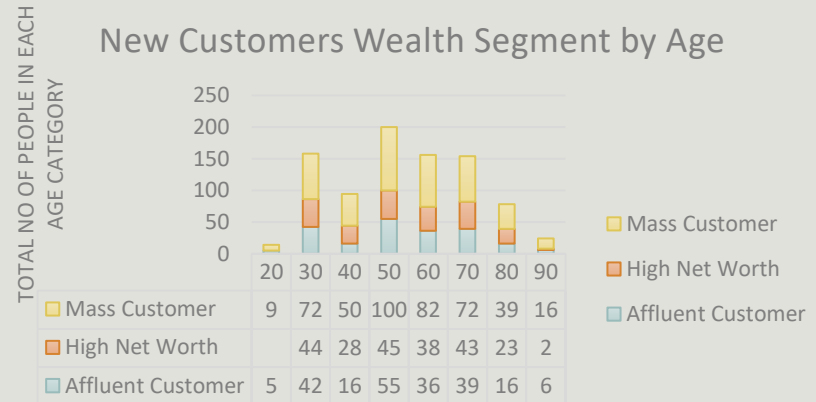
Wealth Segmentation by Age Category

- ❖ In both Old and New age distributions, the largest numbers of customers are classified as 'Mass customer'.
- ❖ The next is the 'High Net Worth' customers.

OLD CUSTOMERS WEALTH SEGMENT BY AGE



New Customers Wealth Segment by Age

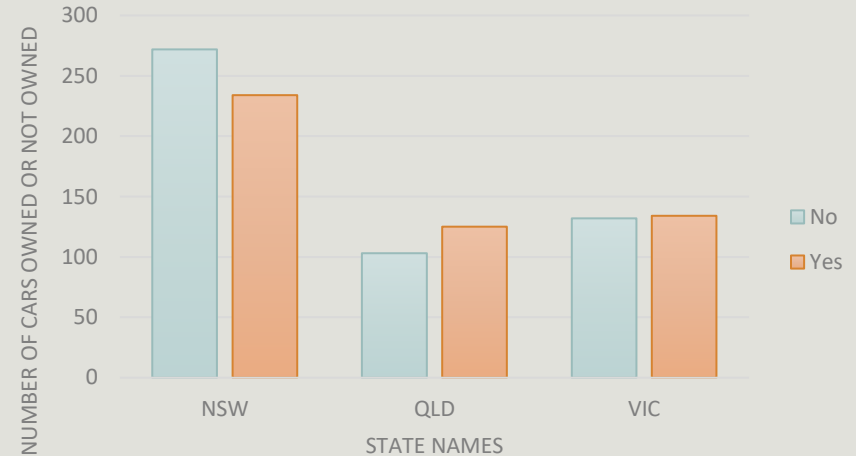


Data Exploration

Number of cars owned and not owned by state

- ❖ NSW has the largest amount of people that do not own a car. NSW seems to have a higher number of people from which data was collected in the new customer list.
- ❖ NSW also has the highest number of customers that own a car.

Number of cars owned in each state (New)



Model Development

RFM Analysis and Customer Classification

- ❖ RFM analysis was used to determine which customers Sprocket 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.

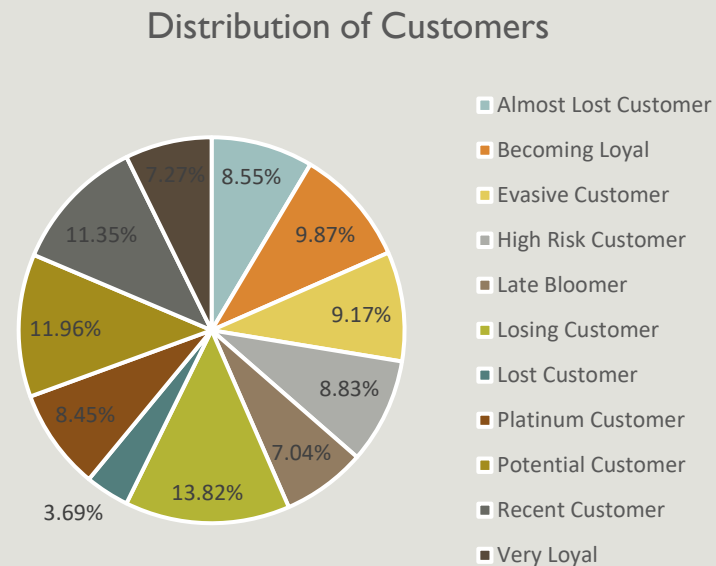
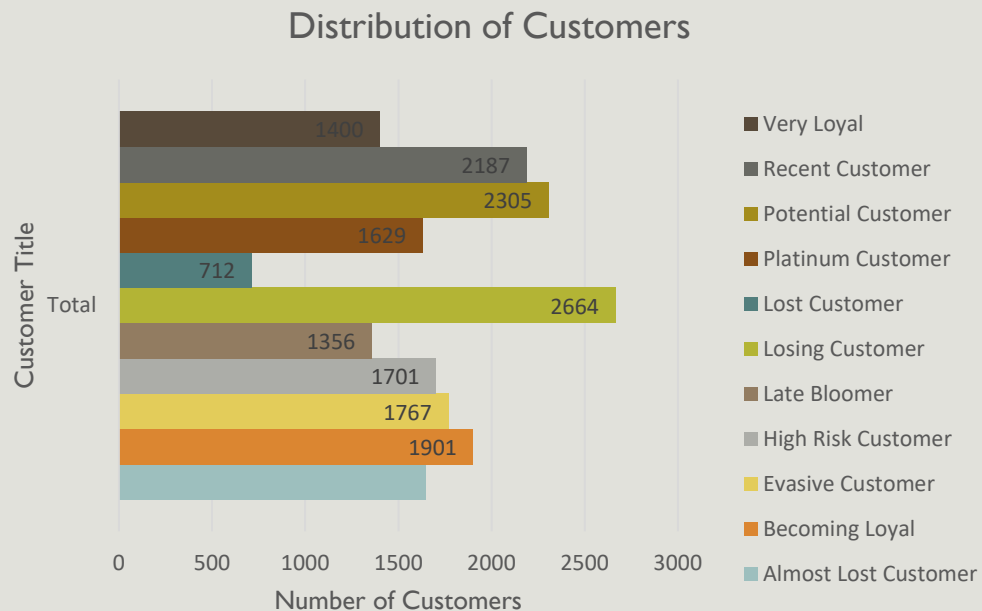
Model Development

Customer Title Definition list with RFM Values assigned

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	421
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	Purchased a while ago, below average RFM Value	224
8	High Risk Customer	Purchase was a 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

Model Development

Customer Title Distributions in Dataset



Interpretation

Summary Table of the Top 1000 Customers to Target

Rank	Customer Title	Description	Numbers of Customers	Cummulative	Customer Selection
1	Platinum Customer	Most recent buy, buys often, most spent	176	176	176
2	Very Loyal	Most recent buy, buys often, spends large amount of money	184	360	184
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	344	704	344
4	Recent Customer	Bought recently, not very often, average money spent	368	1072	296
5	Potential Customer	Bought recently, never bought before, spent small amount	355	1427	0
6	Late Bloomer	No purchases recently, but RFM Value is larger than average	333	1760	0
7	Losing Customer	Purchased a while ago, below average RFM Value	356	2116	0
8	High Risk Customer	Purchase was a long time ago, frequency is quite high, amount spent is high	360	2476	0
	Almost Lost				
9	Customer	Very low recency, low frequency, but high amount spent	326	2802	0
10	Evasive Customer	Very low recency, very low frequency, small amount spent	401	3203	0
11	Lost Customer	Very Low RFM	290	3493	0

Interpretation

Customer Target and Methodology

Rank	Customer Title	Description	Numbers of Customers	Cumulative	Customer Selection
1	Platinum Customer	Most recent buy, buys often, most spent	176	176	176
2	Very Loyal	Most recent buy, buys often, spends large amount of money	184	360	184
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	344	704	344
4	Recent Customer	Bought recently, not very often, average money spent	368	1072	296

Customer Target and Methodology

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