

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Data Analytics – Binod Kshetry

Agenda

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

Introduction

Identify and recommend Top 1000 Customer to target from datasets

Outline of problem

- Sprocket central is a company that specializes in high-quality bikes and cycling accessories.
- Their marketing team is looking to boost business sales by analysing provided datasets.
- Using datasets provided the aim is to analyse and identify top 1000 customers and recommend to target them to drive higher value for the company.

Contents of Data Analysis

- ❖ 'New' and 'Old' customer Age distributions.
- ❖ Job industry distributions
- ❖ Wealth segment by age category
- ❖ Bike related purchases over the past 3 years by gender
- ❖ Number of cars owned and not owned in each state
- ❖ RFM analysis and customer classification

Data Exploration

Data quality assessment and clean up

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

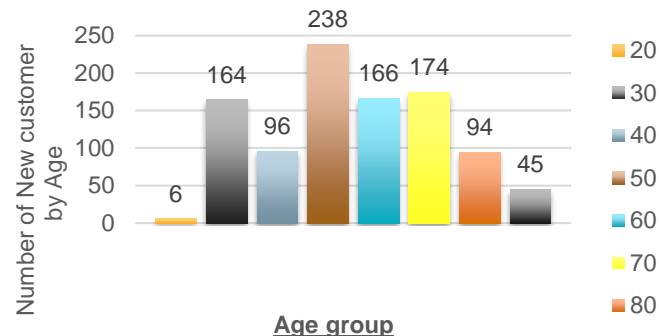
	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	DOB: Inaccurate Age: Missing	Job title: Blanks Customer ID: Incomplete	Gender : Inconsistency	Deceased Customer: Filter out	Default column: delete	
Customer Address		Customer ID: Incomplete	Status: Inconsistency			
Transactions	Profit: Missing	Online order: Blanks Brand: Blanks			Cancelled Status order: filter out	List price: format Product sold date: format

Data Exploration

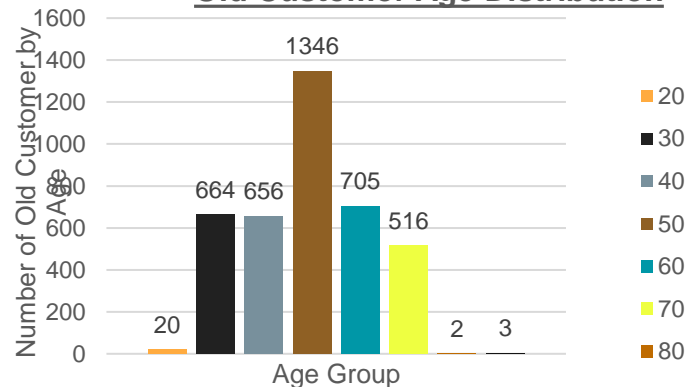
“New” and “Old” customer age distributions.

- Most customers in both New and Old are in group range of 40 – 49.
- The lowest age groups are under 20 and 80+ for both new and old customer lists
- Age group of 20 -29, 40-49 and 60-69 are the most populated in the New customer list.
- Age group of 40-49, 50-59 are most populated in Old customer list.
- The Old customer suggests age between 20-69.
- There is a steep drop in the number of customer in the age group 30-39 in the New customer list.

New Customer Age Distributions



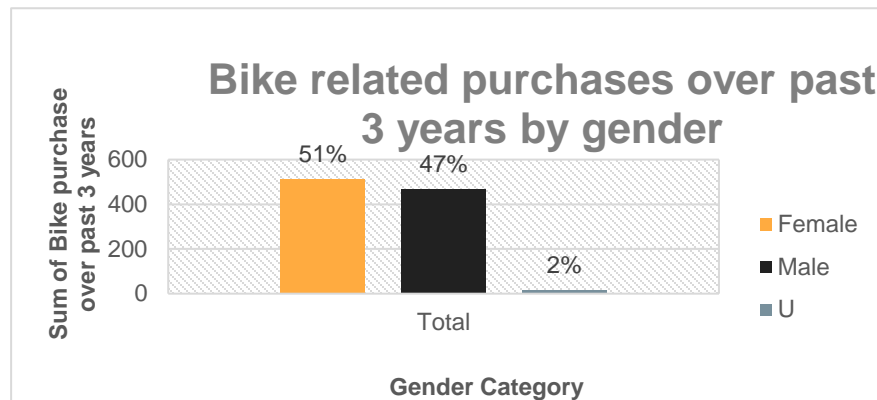
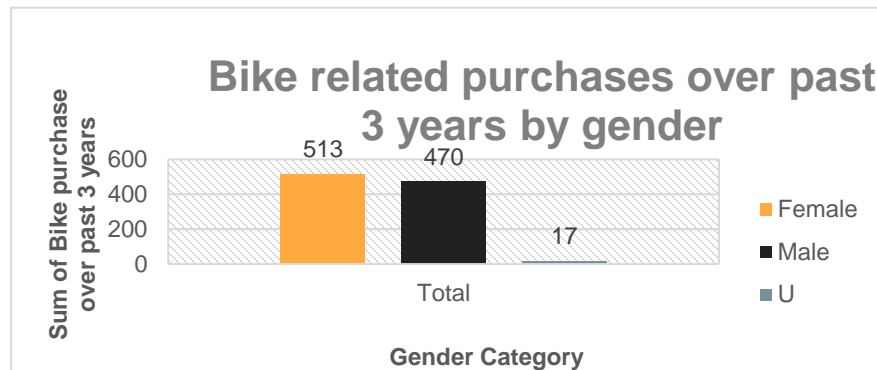
Old Customer Age Distribution



Data Exploration

Bike related purchases over the past 3 years by gender.

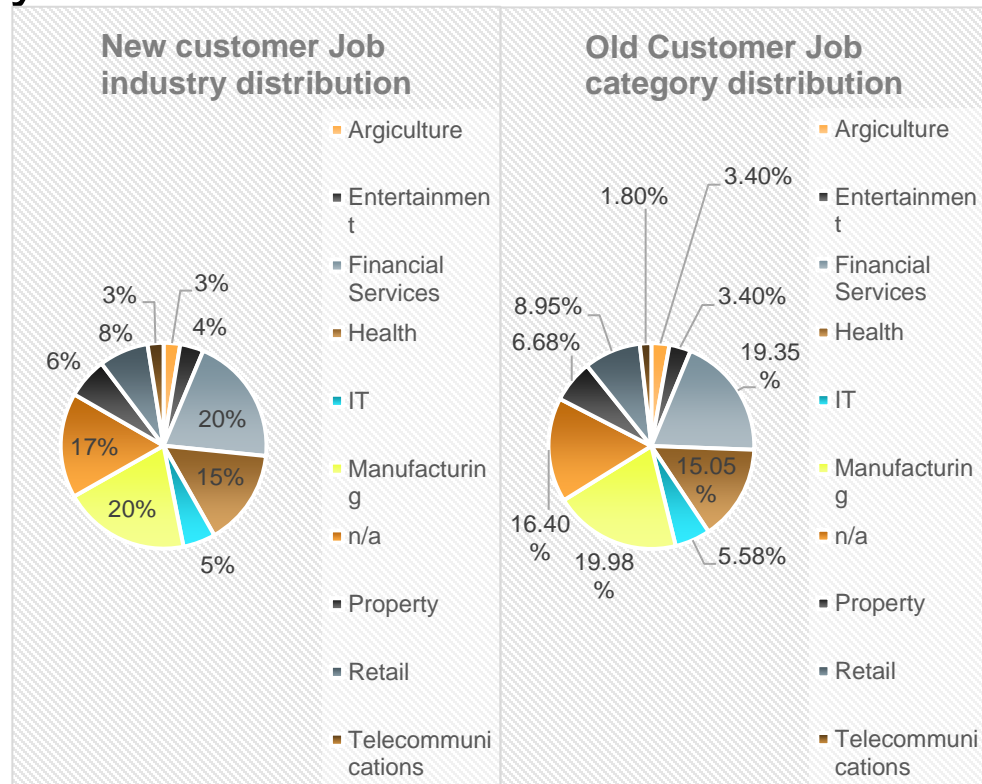
- Over the last 3 years about 51% of bike related purchases were made by female, 47% made by male and the rest 2% were made by unknown gender in New customer list.
- Numerically, female purchases are highest followed by male.
- Unknown gender bought the least percentage of bike related materials in New customer list.



Data Exploration

Job Industry Distribution

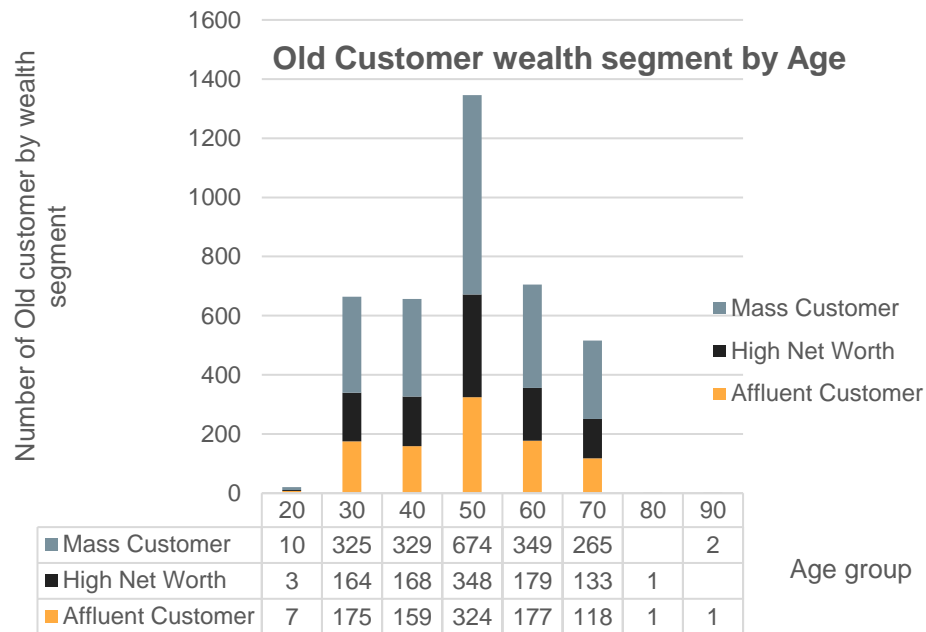
- 40% of new customers are in Manufacturing and Financial services.
- Only 3% of the new customers are from Agriculture and Telecommunications.
- Similar pattern can be seen in Old customer list, at about 19.98% and 19.35% in manufacturing and Financial services respectively.
- Around 17% of customers in both the customer lists have no record of which industry they belongs.



Data Exploration

Wealth segmentation by age category in old customer dataset

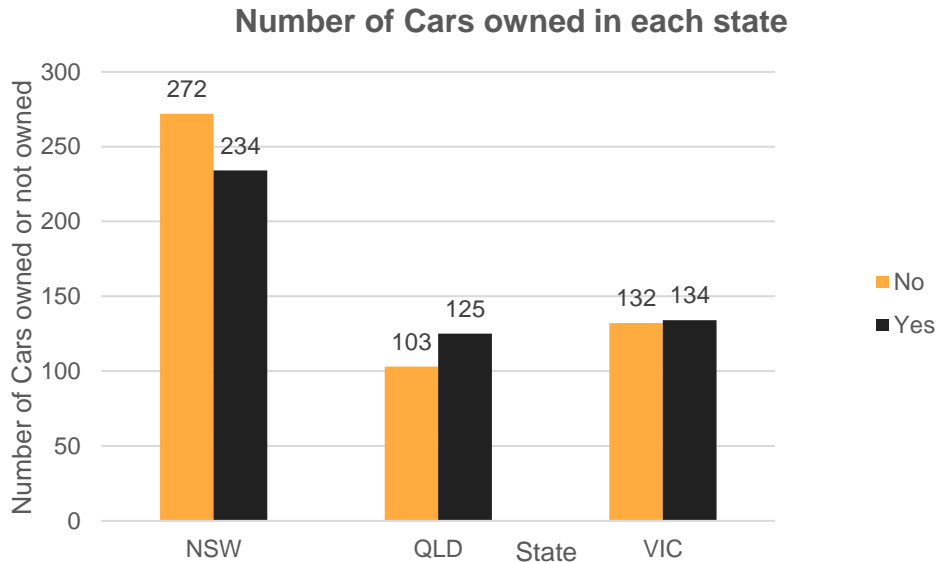
- Largest number of customers are classified as “Mass customers”.
- Age group of 40-49 has the largest number of all types of customers.
- Age groups of 70+ has the least numbers of customers in all customer types.



Data Exploration

Number of Cars owned and not owned by state

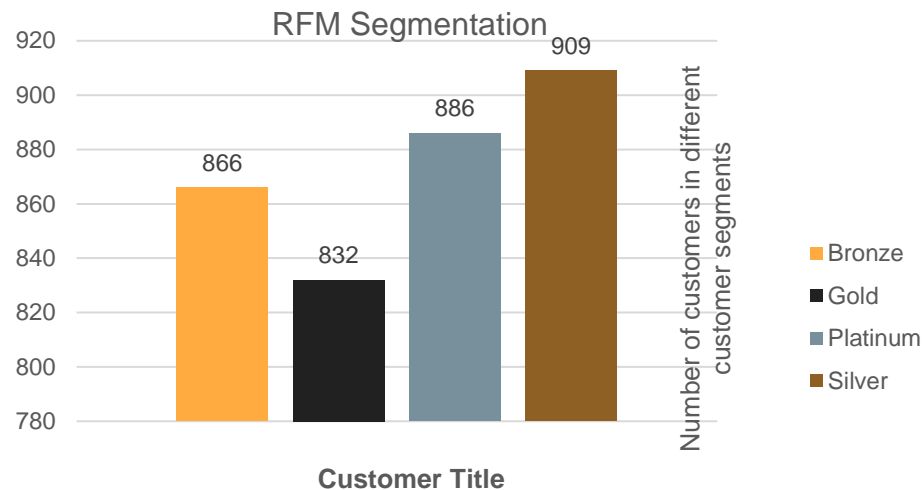
- NSW has the largest number of customer now owning cars. NSW seems to have high number of people from which data was collected.
- Victoria has almost same number of customers owning and not owning cars.
- Queensland has slightly high number of customers owning cars.



Model Development

RFM analysis and customer classification

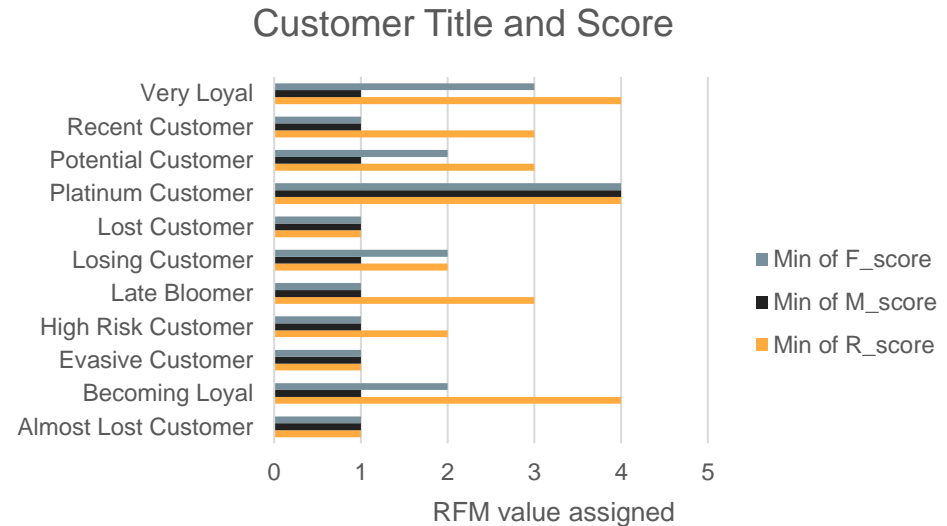
- RFM analysis is used to determine which customers a business to target to drive revenue and attain profit.
- The RFM (recency, frequency and monitory) model shows customers that have displayed high level of engagement with the business. Platinum and Gold customers are valued customers which are almost half the customer population.



Model Development

RFM analysis and customer classification

- The RFM (recency, frequency and monetary) model shows customers that have displayed high level of engagement with the business. Platinum and Gold customers are valued customers which are almost half the customer population.



Model Development

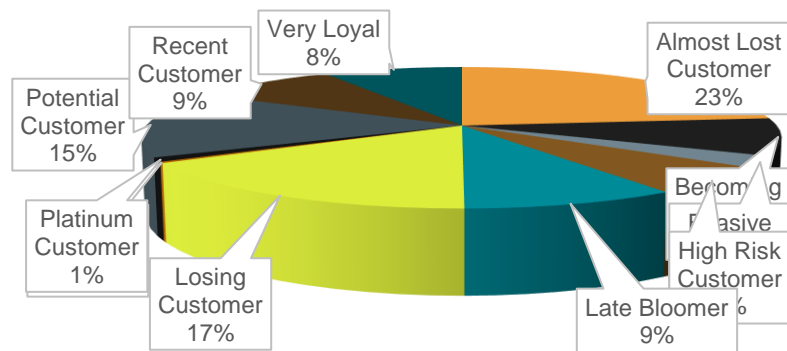
Customer Title Definition list with RFM values assigned

Rank	Customer Title	Description	RFM Value
1	Platinum	Most recent, buy often and spent most	444
2	Very loyal	Most recent, buy often and spent large sum of money	433
3	Becoming Loyal	Relatively recent, buy often and spend quite decent money	421
4	Recent customer	Bought recently, not very often, average money spent	344
5	Potential	Bought recently, never bought before, spent little money	323
6	Late Bloomer	No purchase recently but RFM value is relatively higher	311
7	Losing	Purchase was a while ago and RFM value below average	224
8	High Risk Customer	Purchase long time ago, frequency and monetary is quite high	212
9	Almost lost	Very low recency and frequency but quite high amount spent	124
10	Evasive	Low amount of recency, frequency and monetary	112
11	Lost customer	Very low RFM	111

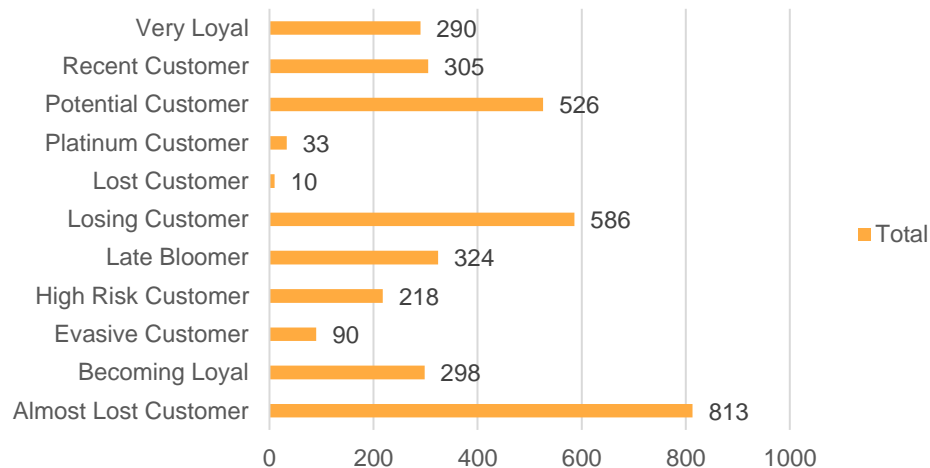
Model Development

Customer Title Distributions in dataset

Distributions of customer



Distributions of customer



Interpretation

Summary Table of the top 1000 customers to Target

Rank	Customer Title	Description	RFM Value	Number of Customers	Cumulative	Customer selection
1	Platinum	Most recent, buy often and spent most	444	33	33	<u>33</u>
2	Very loyal	Most recent, buy often and spent large sum of money	433	290	323	<u>290</u>
3	Becoming Loyal	Relatively recent, buy often and spend quite decent money	421	298	621	<u>298</u>
4	Recent customer	Bought recently, not very often, average money spent	344	305	926	<u>305</u>
5	Potential	Bought recently, never bought before, spent little money	323	526	1452	<u>74</u>
6	Late Bloomer	No purchase recently but RFM value is relatively higher	311	324	1776	0
7	Losing	Purchase was a while ago and RFM value below average	224	586	2362	0
8	High Risk Customer	Purchase long time ago, frequency and monetary is quite high	212	218	2580	0
9	Almost lost	Very low recency and frequency but quite high amount spent	124	813	3393	0
10	Evasive	Low amount of recency, frequency and monetary	112	90	3483	00
11	Lost customer	Very low RFM	111	10	3493	

Appendix

Appendix

- **The top 1000 customers discovered would have bought recently, frequently and spend the most amount of money than other customers.**