KPMG

Data Analytics Approach:

Sprocket Central Pty Ltd's **High Value Customers**

Discussion Point

This deck includes 4 discussion points:

- Introduction
- Data Exploration
- Model Development
- Interpretation

1.
INTRODUCTION

Let’s start with the first set of slides

Business Statement

The marketing team is **looking to boost business by analysing their existing customer dataset** to determine customer trends and behaviour.

Photo by Bill Jelen on Unsplash

Data Quality Assessment

	Transactions	CustomerDemographic	CustomerAddress
Accuracy	✔	• DOB: 1 inaccurate value • Default: all data inaccurate	✔
Completeness	• Sold Time, Standard Cost, Product Size, Order Status, Brand, Product Line, Product Class: 197 empty cells • Online Order: 360 empty cells	• Last Name: 125 cells • DOB, Tenure: 87 cells • Job Title: 506 cells • Industry Category: 656 cells	✔
Consistency	✔	Gender has values (“U”, “F”, “Femal”, “M”)	State has inconsistent data
Currency	✔	✔	✔
Relevancy	Exclude “Cancelled” data on order status table	✔	✔
Validity	Product First Sold Date has a wrong format	✔	✔

Data Manipulation

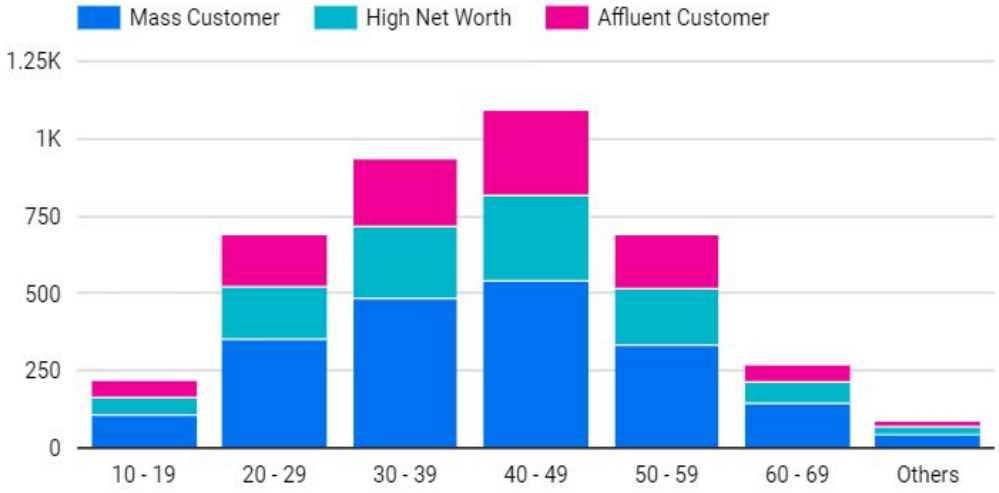
- Change product_first_sold_date type from integer to datetime, then extract time only, because the date are all the same (1970/01/01)
- Change gender's value (M) to Male and (F, Femal) to Female
- 1 DOB's inaccurate value was dropped
- Default column was dropped
- Delete all records of order_status that has "Cancelled" (orders)
- New South Wales and Victoria were changed to NSW and VIC
- Age and age_group were added into CustomerDemographic dataset
- Kpmg_mix2 dataset was created

1.

DATA EXPLORATION

Let’s start with the first set of slides

Wealth Segmentation by Age Group

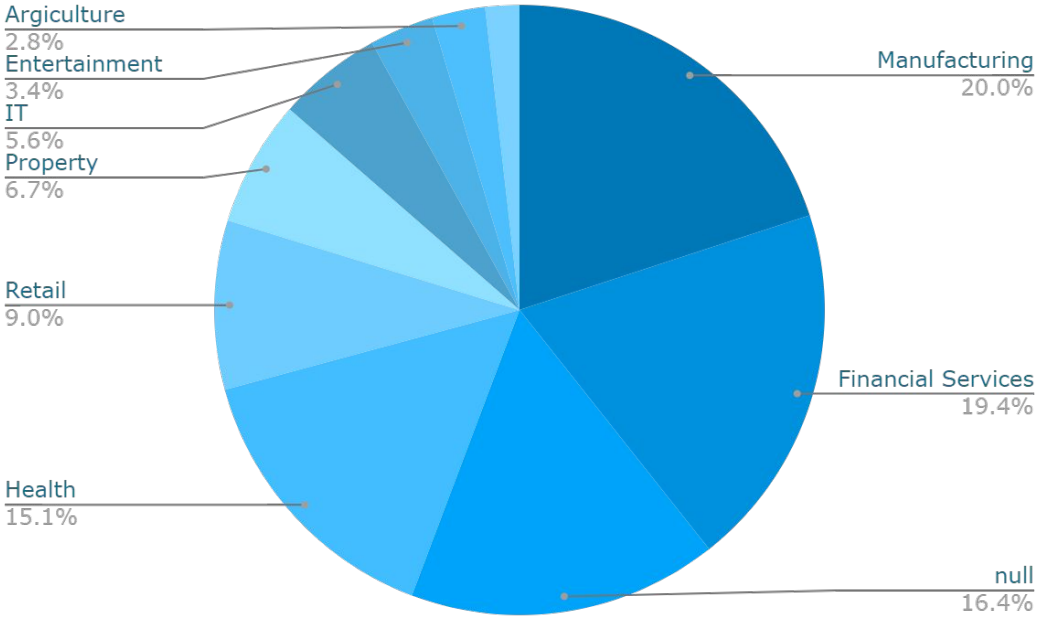


From the diagram, we can conclude that, **most customers' ages are 20 to 49 years old. Then, Mass Customer has the highest count in every age group.**

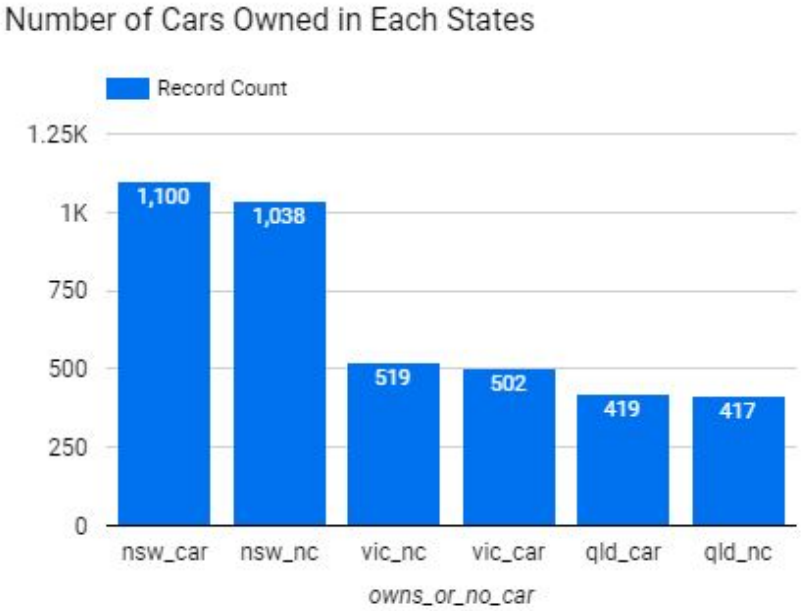
The age range from 20 to 49 was selected because if we analyze it using Pareto Column Chart, those age ranges contribute 80% of customers' sum of past 3 years bike related purchases.

Customers' Job Industry Category

Manufacturing, Financial Services, Health sectors come as the top 3 sectors of customers' job industry.



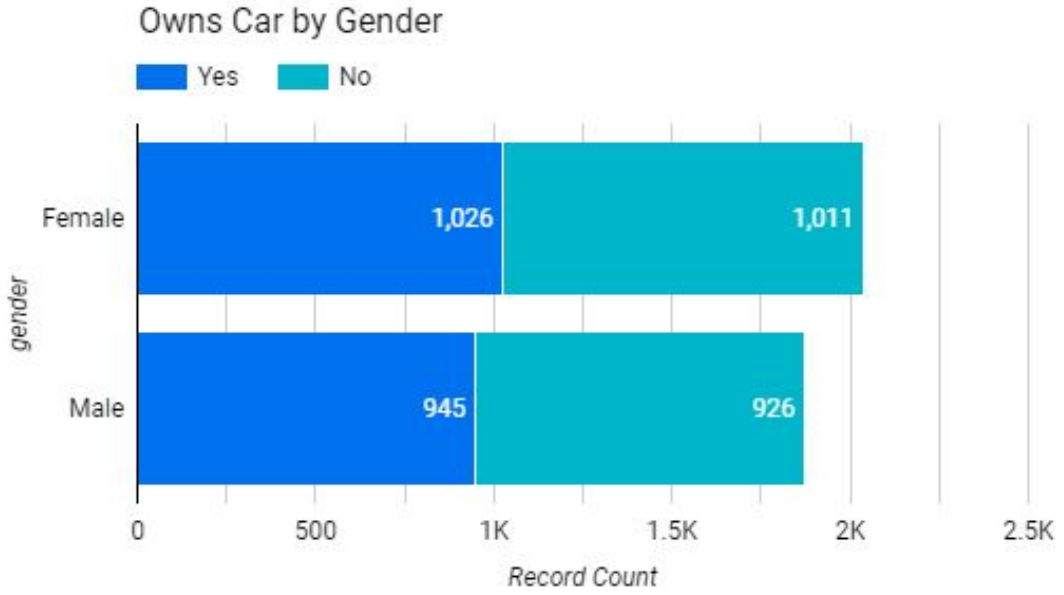
No Car = Buying a Bike?



As we can see, **New South Wales has the highest number of customers**, the second was Victoria, and so on. **Then, if we consider the bikes sold in each state, the highest number of customers also has the highest bike transactions.** The number of customers, **either owned or not owned car does have similar power buying.**

Sum and Owns Car by Gender

Sum on each gender **quite similar with female as the highest.** Also, diagram owns car by gender. Sum of female, either own car or not, slightly higher than male.



Most Customers Were...

Age

Between **20 to 49 years old** as the most powerful buyers, even for the last 3 years

Segmentation

Customers who come from **Mass Customer Segmentation**

Job Industry

Customers who work on **Manufacturing, Financial Services, and Health** sectors.

Location

Most customers **live in New South Wales (>50%)**

Gender

Either female or male, they were almost have the same power buying

2.

MODEL DEVELOPMENT & INTERPRETATION

RFM Analysis and Targeting High Value Customers

What does RFM Analysis?

RFM Analysis is an analysis using data of customer's behavior, if marketing team want to decide who will be their best customer/what customer segmentation will be chosen, RFM Analysis come as the right tool to use. RFM stands for this acronym:

(R)ecency

When customer did their last transaction

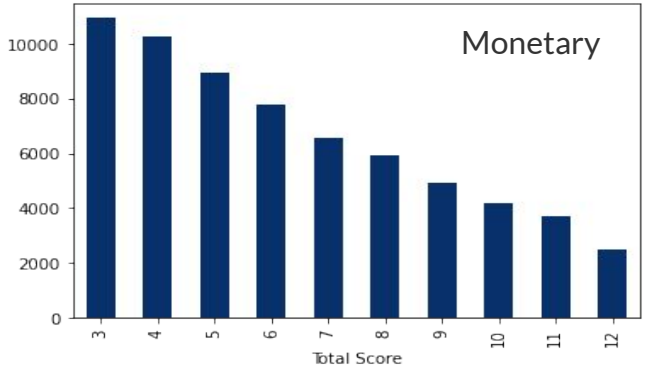
(F)requency

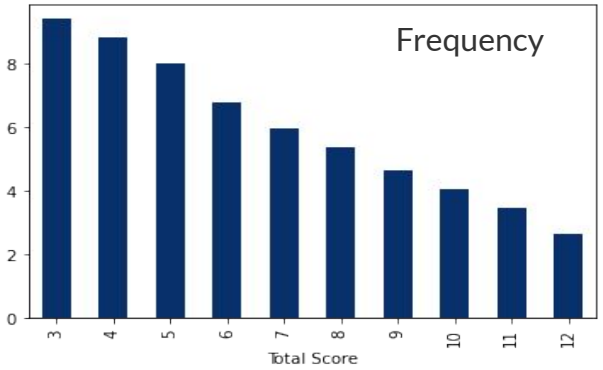
How many goods/services were used

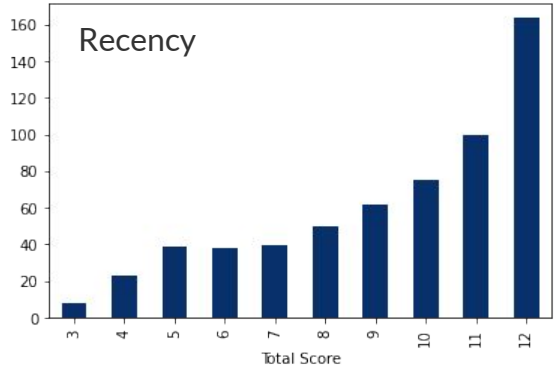
(M)onetary

How much their spending for the goods/services

RFM Analysis Result







We can see that the customer who gives the largest expenditure (monetary), the highest number of purchases (frequency), and makes the latest transactions (recency), are in **the Total Score (summation of RFMClass values) 3,4,5,6,7 as the Top 5.**

```
rfmSeg[rfmSeg['Total Score']==3]
```

customer_id	recency	frequency	monetary_value	R_Quartile	F_Quartile	M_Quartile	RFMClass	Total Score
1	7.0	11	9084.45	1	1	1	111	3
25	5.0	12	9996.33	1	1	1	111	3
37	7.0	9	11291.71	1	1	1	111	3
89	4.0	9	12741.44	1	1	1	111	3
99	2.0	9	8279.06	1	1	1	111	3
...	...	...	...	...	...	...	...	...
3402	12.0	9	10700.79	1	1	1	111	3
3420	5.0	11	13885.45	1	1	1	111	3
3427	1.0	8	12523.17	1	1	1	111	3
3433	0.0	10	10471.84	1	1	1	111	3
3470	2.0	8	14126.25	1	1	1	111	3

210 rows × 8 columns

Example of customers who have to targetted.

Thanks!

Any questions?

You can find me at:

Linkedin

Email