TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Michael Ojo

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 long-standing KPMG client whom specializes in high-quality bikes and accessible cycling accessories to riders.

- Their marketing team is looking to boost business sales by analyzing provided datasets to determine customer trends and behavior.

- Using the 3 datasets provided the aim is to analyze and study the behavior of the old customers and recommend to the client which customers from the new 1000 customers should be focused on.

This will be done with the three phases of Data Exploration, Model Development and Interpretation.

Contents of Data Analysis/Model Development

- Data Quality Assessment
- RFM analysis and customer valuation
- ANOVA Analysis to determine the significant influence factors on Profit
- Regression Analysis on Profit vs Tenure
- Regression Analysis on Tenure vs age
- Visualizations

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

Summary Table

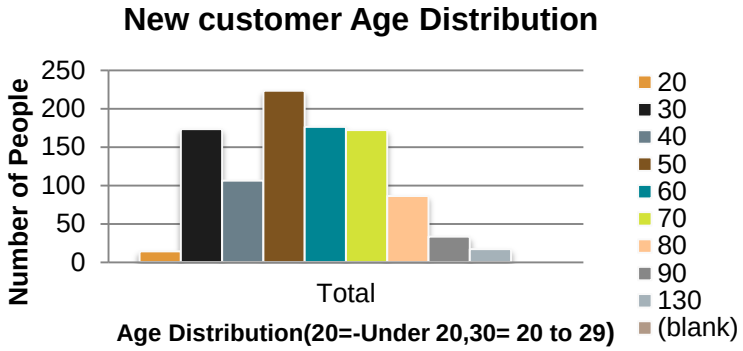
	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Customer Demographic	DOB: Inaccurate Age: missing	Job title: blanks Customer id: Incomplete, duplicates	Gender: Inconsistent	Deceased Customers: filter out	Default column: delete	
Customer Address		Customer id: incomplete	States: inconsistency			
Transactions	Profit: missing				Cancelled status order: filter out	List price: format Product sold date: format

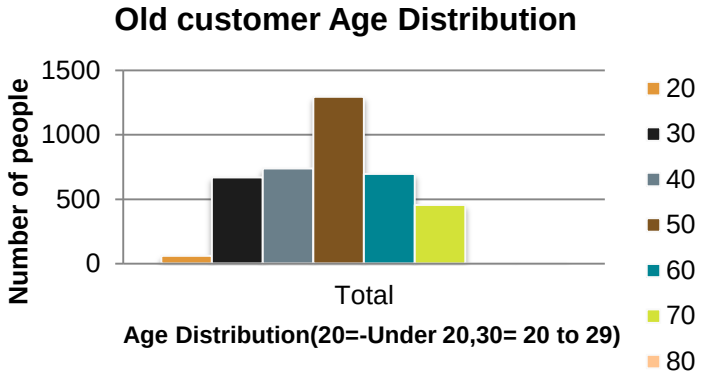
A complete analysis has been sent via mail

Data Exploration

‘New’ and ‘Old’ Customer Age Distributions

- Most customers are aged between 40-49 in ‘New’. In ‘Old’, majority are also aged between 40-49.
- The lowest age groups are under 20 and 80+ for both ‘New’ and ‘Old’ customer lists.
- The ‘New’ customer list suggests that age groups 20-29 and 40-69 are most populated.
- There is a steep drop of customers in the 30-39 age group in ‘New Customer list’ and an increase in the 20-29 age group.

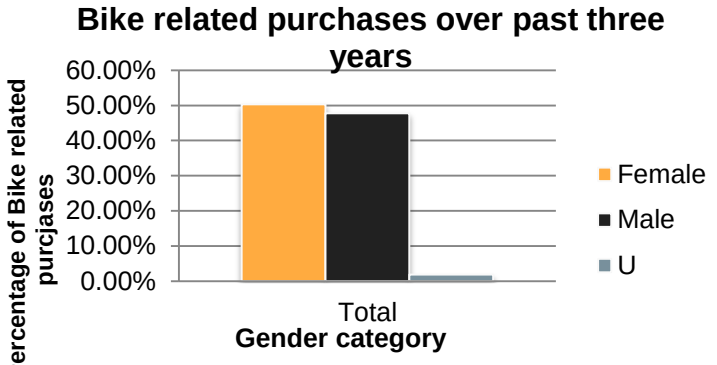


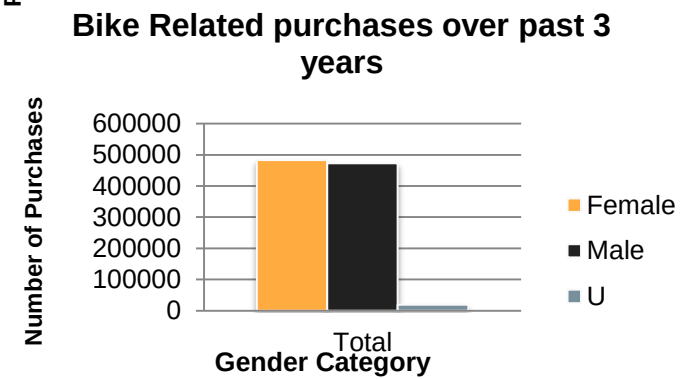


Data Exploration

Bike related purchases over the last 3 years by gender

- Over the last three years about 50% of bike related purchases were made by females to 48% of purchase made by males. Approximately 2% were made by unknown gender
- Numerically, females purchase almost 10,000 more than males
- Females make up majority of bike related sales

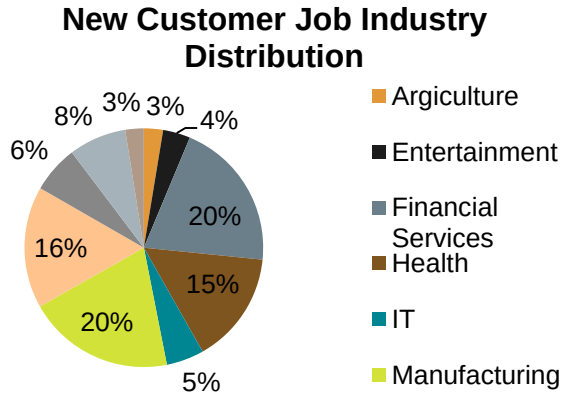


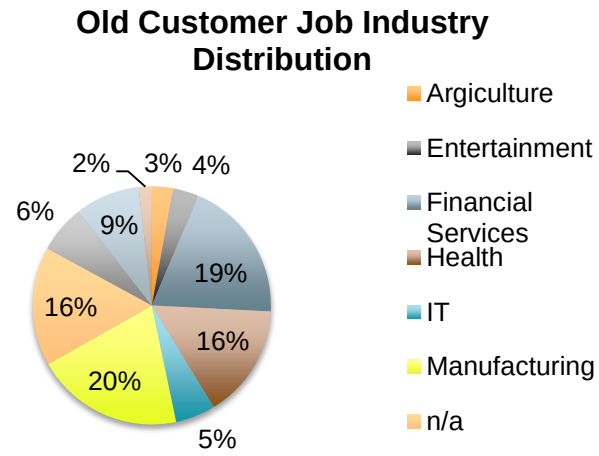


Data Exploration

Job Industry Distribution

- 20% of ‘New’ customers are in Manufacturing and Financial Services
- The smallest number of customers are in Agriculture and Telecommunication at 3%
- Similar pattern in ‘Old’ customer list, at 20% and 19% in Manufacturing and Financial respectively

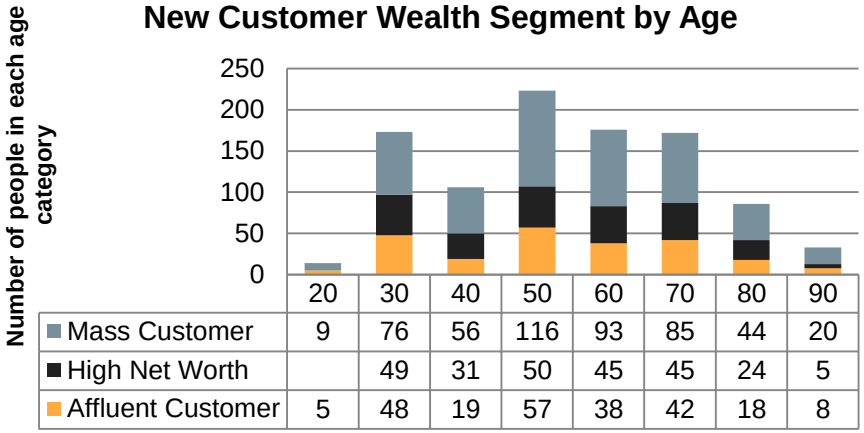




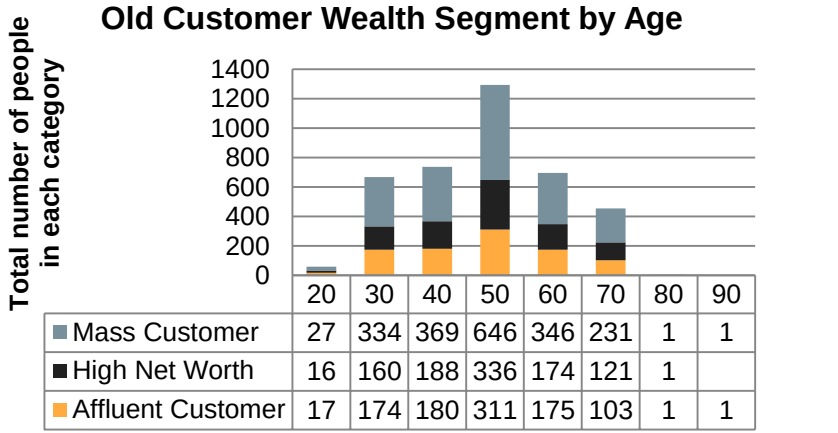
Data Exploration

Wealth Segmentation by Age category

- In all age categories, the largest number of customers are classified as ‘Mass Customer’.
- The next category is the ‘High Net Worth’ Customers.
- The ‘Affluent Customer’ outperforms the ‘High Net Worth’ Customer in the 40-49 age group.



	20	30	40	50	60	70	80	90
Mass Customer	9	76	56	116	93	85	44	20
High Net Worth		49	31	50	45	45	24	5
Affluent Customer	5	48	19	57	38	42	18	8

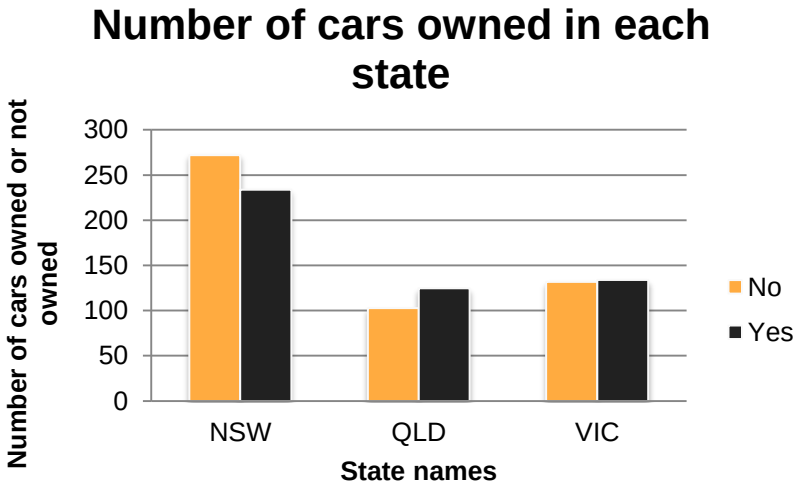


	20	30	40	50	60	70	80	90
Mass Customer	27	334	369	646	346	231	1	1
High Net Worth	16	160	188	336	174	121	1	
Affluent Customer	17	174	180	311	175	103	1	1

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
- Victoria is also split quite evenly. But both numbers are significantly lower than those of NSW
- QLS has a relatively high number of customers that own a car



Model Development

RFM model Analysis and Customer Classification

- RFM analysis is used to determine which customers a business should target to increase its revenue and value.
- It uses sales data to segment a pool of customers based on their purchasing behavior.
- The resulting customer segments are neatly ordered from most valuable to least valuable. This makes it straightforward to identify best customers.
- 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

ANOVA Analysis on Profit and wealth Segment

WEALTH

One-way ANOVA: Sum of Profit versus Wealth Segment

Method

Null hypothesis	All means are equal
Alternative hypothesis	Not all means are equal
Significance level	α = 0.05
Rows unused	1

Equal variances were not assumed for the analysis.

Factor Information

Factor	Levels	Values
Wealth Segment	4	0, Affluent Customer, High Net Worth, Mass Customer

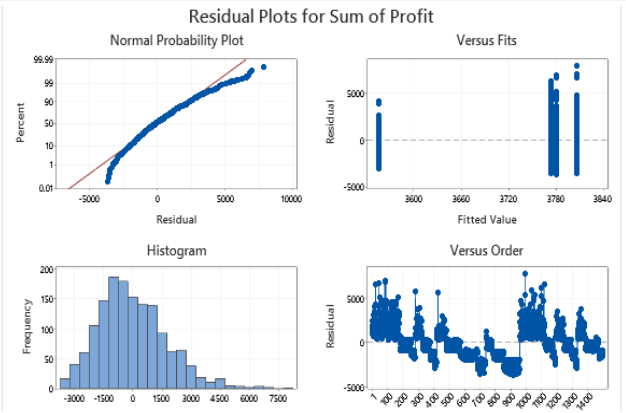
Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Wealth Segment	3	252.222	0.34	0.798

Model Summary

R-sq	R-sq(adj)	R-sq(pred)
0.07%	0.00%	0.00%

One-way ANOVA: Sum of Profit versus Wealth Segment



We can see from the analysis that Wealth segment **do not** really have a significant effect on profit as the means from each segment are statistically equal with a **p-value of 0.798** and a very low **R2 of 0.07%.** Also the residual analysis ofcourse did not meet the Normality assumption and has a lot of irregularities.

Model Development

ANOVA Analysis of Profit and Own car

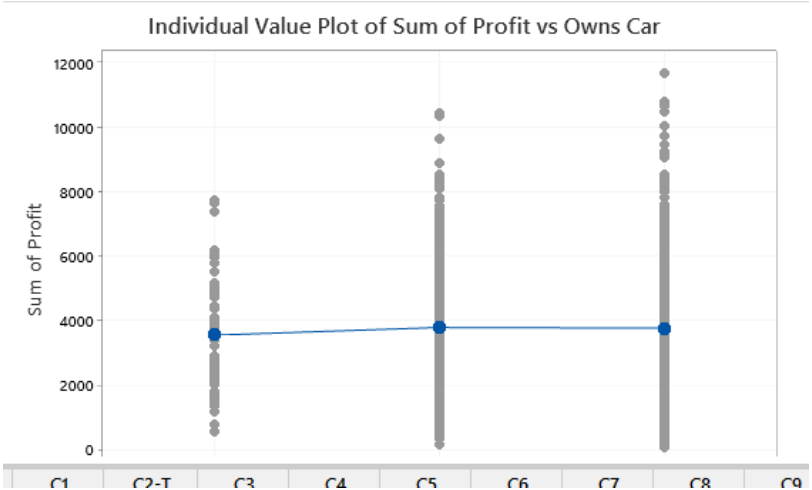
Factor	Levels	Values
Owns Car	3	0, No, Yes

Welch's Test

Source	DF Num	DF Den	F-Value	P-Value
Owns Car	2	157.354	0.49	0.611

Model Summary

R-sq	R-sq(adj)	R-sq(pred)
0.06%	0.00%	0.00%



We can also see that owning a car or not do not significantly affect the profit made. And looking at the individual value plot we see that it's almost horizontal showing that the means from the two groups are the same.

Model Development

Regression Analysis on Profit and Tenure

Regression Analysis: Sum of Profit versus Tenure

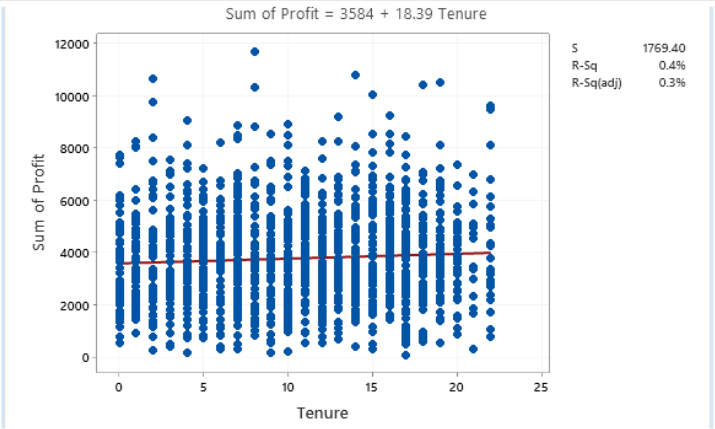
The regression equation is
Sum of Profit = 3584 + 18.39 Tenure

Model Summary

S	R-sq	R-sq(adj)
1769.40	0.38%	0.31%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	17898855	17898855	5.72	0.017
Error	1496	4683618299	3130761		
Total	1497	4701517154			



We can see that the analysis gave a significant p-value of 0.017, which means that tenure may have a significant effect on profit. But from the very low R2, we see that the model did not explain most of the variation. The fitted line plot is also shown with the model equation.

Model Development

Regression Analysis on Tenure and Age

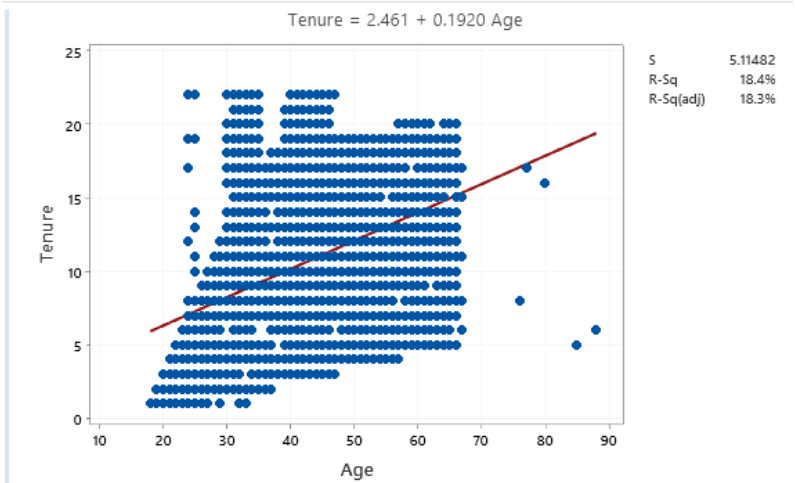
The regression equation is
Tenure = 2.461 + 0.1920 Age

Model Summary

S	R-sq	R-sq(adj)
5.11482	18.35%	18.33%

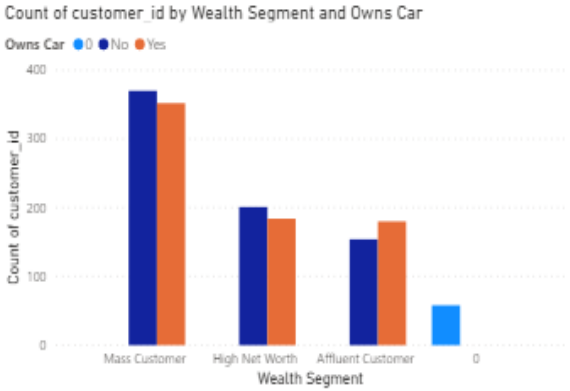
Analysis of Variance

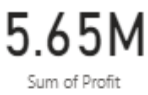
Source	DF	SS	MS	F	P
Regression	1	22979	22979.3	878.37	0.000
Error	3908	102239	26.2		
Total	3909	125218			

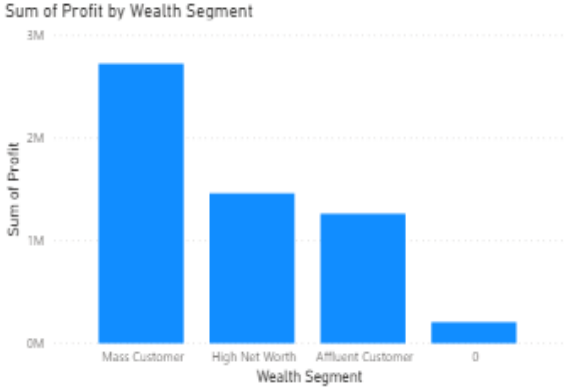


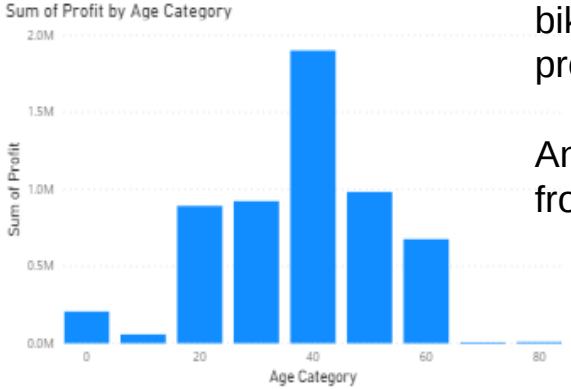
We see from the analysis that age has a significant effect on tenure and is positively correlated with tenure. Which means the older age category will have a higher tenure. The R2 is low but still explain some variations. The fitted plot and model equation is also shown

Interpretation







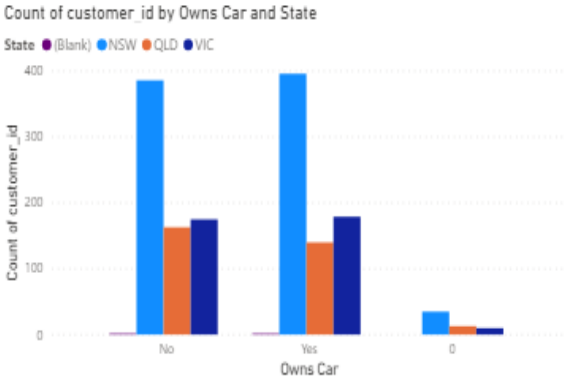


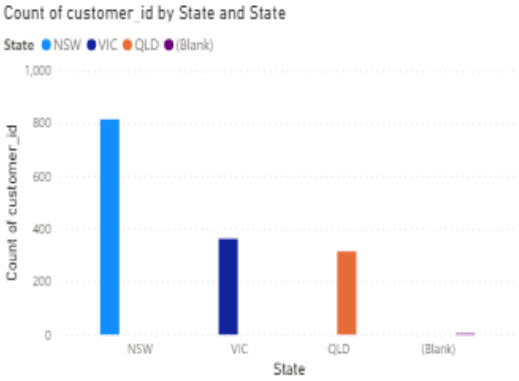
From these visualizations, we can see that we have more of Mass customers and they also contribute more to the profit owing to their number.

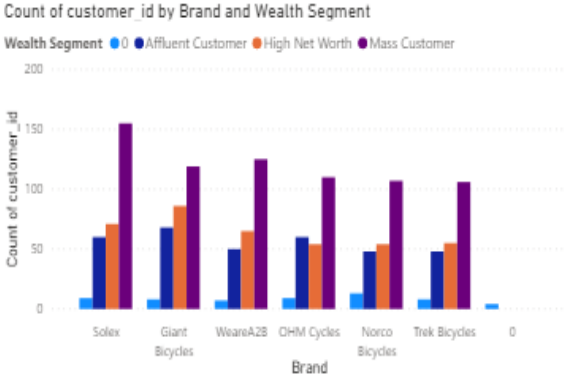
We also see that customers of the age category from 30-50 tend to buy more bikes and also contribute more to the profit.

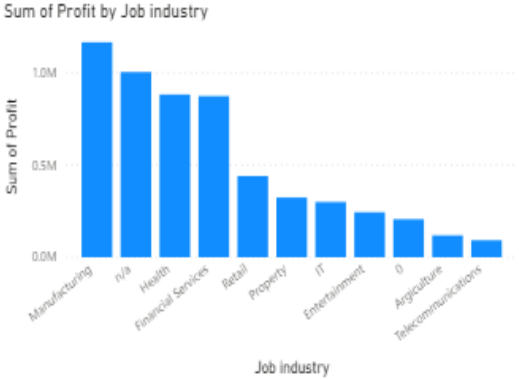
And we can see that the sum of total profit from the store is 5.65M.

Interpretation









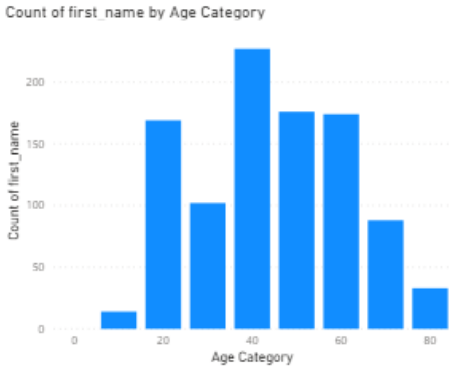
We can see from here that most of the customers live in NSW hence that should be the location of focus.

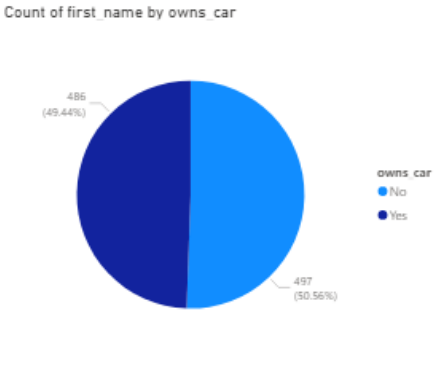
We also see the top industries that contribute more to the profit which are Manufacturing, Health, Financial services and more from an unknown classification.

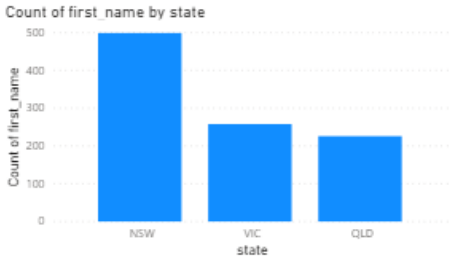
Also the brands that are most purchased include Solex and Giants.

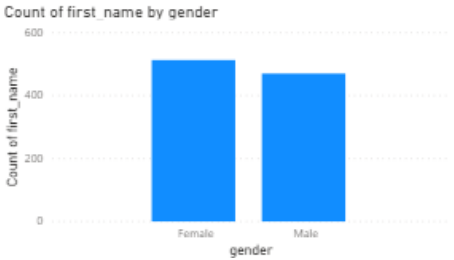
Interpretation

Demographics of New Customers









We see that unlike the old customers we have lesser age 30 category from the new customers but we still have more from 40 and 50. We also see a high number of age category 20.

We also see that the number of customers that own a car is almost equal those that do not.

And also we have more customers coming from our focus location NSW.

Interpretation

Selection from New Customers And Recomendation

first_name	last_name	property_valuation	state
Abbie	Oldman	8	NSW
Adria	Van den Velde	7	NSW
Ajay	Worham	6	NSW
Aldric	Birney	6	NSW
Aleece	Feige	10	NSW
Alexina	Mabley	9	NSW
Andriana	Gosnoll	10	NSW
Aridatha	Sephton	1	NSW
Arty	Strudwick	10	NSW
Aurie	Rhead	2	NSW
Bengt	Bilson	10	NSW
Bevvy	Openshaw	3	NSW
Brendis	Pineaux	4	NSW
Brigitte	Whellams	8	NSW
Brod	Attrey	8	NSW
Calida	Schaben	9	NSW
Carr	Hopkynson	6	NSW
Cecil	Gant	8	NSW
Chanda	Mensler	10	NSW
Cherye	Stanfield	9	NSW
Chryste	Oddboy	9	NSW
Cicily	Hast	3	NSW
Cirillo	Frossell	12	NSW
Clevey	Aisthorpe	9	NSW
Collete	Dory	9	NSW
Conroy	Rappaport	9	NSW
Corinna	Suggey	8	NSW
Darlleen	Shakras	10	NSW
David	Napoleon	11	NSW
Denny	Spleving	8	NSW
Denys	Minshall	8	NSW
Donn	MacGregor	10	NSW

Studying the trend from the old customers data, we know the age category that buy more bikes, the industry which the top buyers come from and also the location that had more valued customers. Hence the new customers data was filtered to these preferences.

I recommend that you focus more on these set of customers selected from the new 1000 customers as they would be drivers of profits and business growth.

A dashboard will be sent to you summarizing all the findings with the full list of selected customers.

Appendix