



InsideSherpa

Inspiring and empowering
future professionals

Akshay Kumar

Data Analytics Consulting Virtual Internship

Certificate of Completion
September 7th, 2020

Over the period of September 2020, Akshay Kumar has completed practical task modules in:

Data Quality Assessment
Data Insights
Data Insights and Presentation

Deborah Yates
National Managing
Partner People
Performance and
Culture

Tom Brunskill
CEO, Co-Founder of
InsideSherpa

Enrolment Verification Code n5PhAL2drjtPu3g29 | User Verification Code E2sxtEmA9Mo5c347 | Issued by InsideSherpa



The Analytics Team

Sprocket Central Pvt.Ltd

Data Analytics Approach
By

Akshaya Kumar





AGENDA

Introduction

Data
Exploration

Model
Development

Interpretation

INTRODUCTION

Identify and
Recommend
Top 1000
Customer to
Target from
Datasets



Problem Statement

- Sprocket Company specializes in high quality bikes and cycling accessories
- The marketing team is looking to boost the sales by analysing three given data sets
- The main aim is to analyze 1000 customers that Sprocket should target to drive higher value for the company



Contents of Data Analysis

- New and Old Customer Age Distribution
- Bike related purchase for the past three years
- Job Industry Distribution
- Wealth segment Distribution by Age
- Number of cars owned/not owned by state
- RFM Analysis and Customer Satisfaction



Data Exploration

Data Quality Assessment and Clean Up

Key Issues for
Data Quality
Assessment

Accuracy

Completeness

Consistency

Currency

Relevancy

Validity

	<u>Accuracy</u>	<u>Completeness</u>	<u>Consistency</u>	<u>Currency</u>	<u>Validity</u>	<u>Relevance</u>
<u>Customer Demographic</u>	DOB: Inaccurate Age: Missing	Job Title Customer ID: Incomplete	Gender: Inconsistent	Deceased customers: Filter out	-NA-	Default Column deleted
<u>Customer Address</u>	-NA-	Customer ID: Incomplete	States: Inconsistent	-NA-	-NA-	-NA-
<u>Transactions</u>	Missing profit	Customer ID: Incomplete Online order: Blanks eliminated Brand: blanks	-NA-	-NA-	Product sold date: Formatted List price: Formatted	Cancelled Status order: Filtered out

New and Old Customer Age Distributions

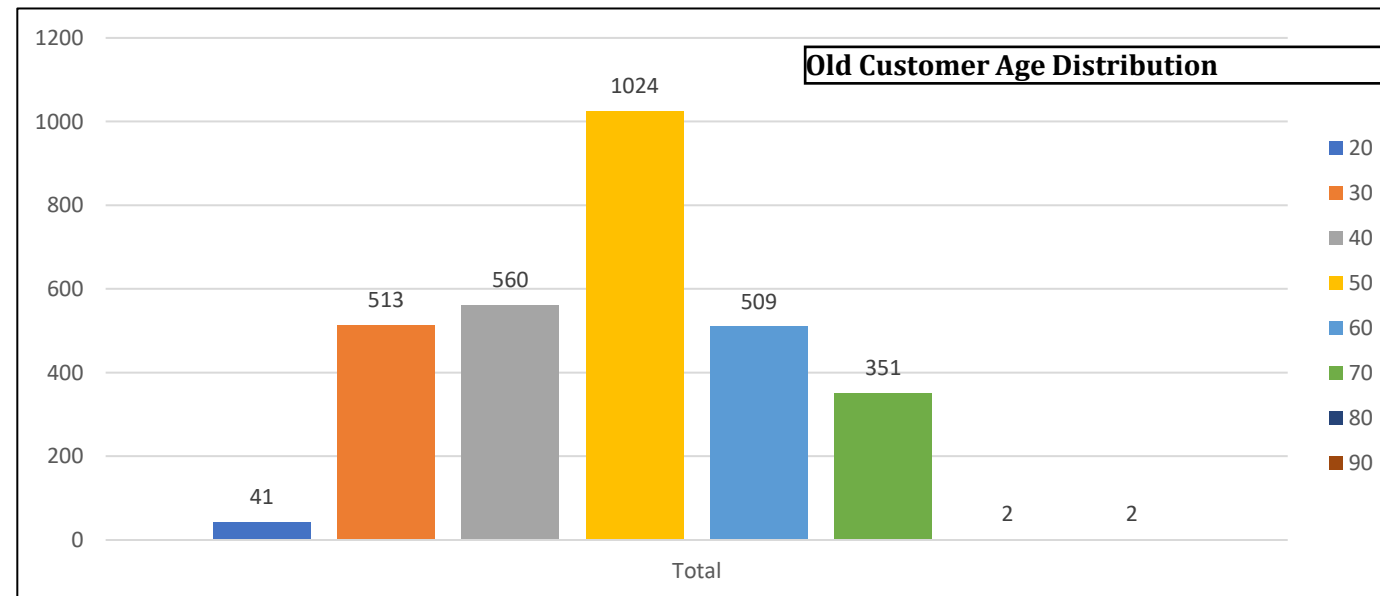
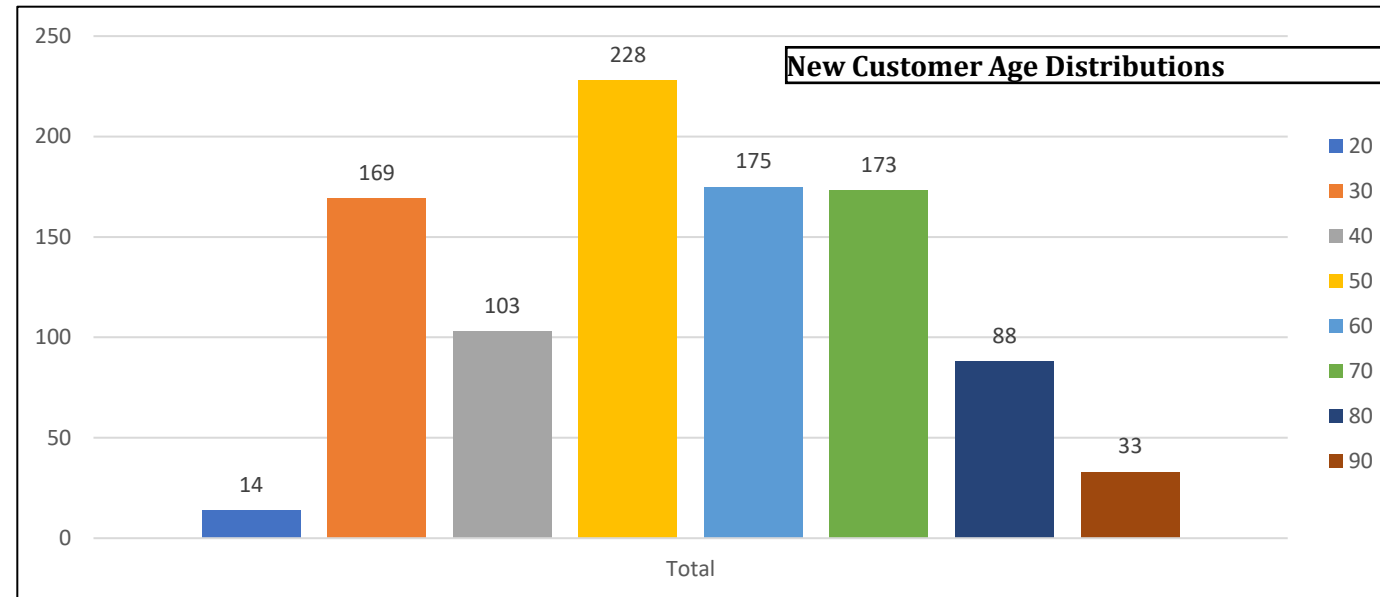
In both the cases it is seen that most of the customers are aged between 40-49

In both the cases the lowest age groups are either below 20 years or above 80 years

The most populated aged for the 'New' list are between 20-39, and 40-69

The 'Old' customer list recommends 20-69

It is confirmed that customer aging between 30-39 in 'New' list have been reduced

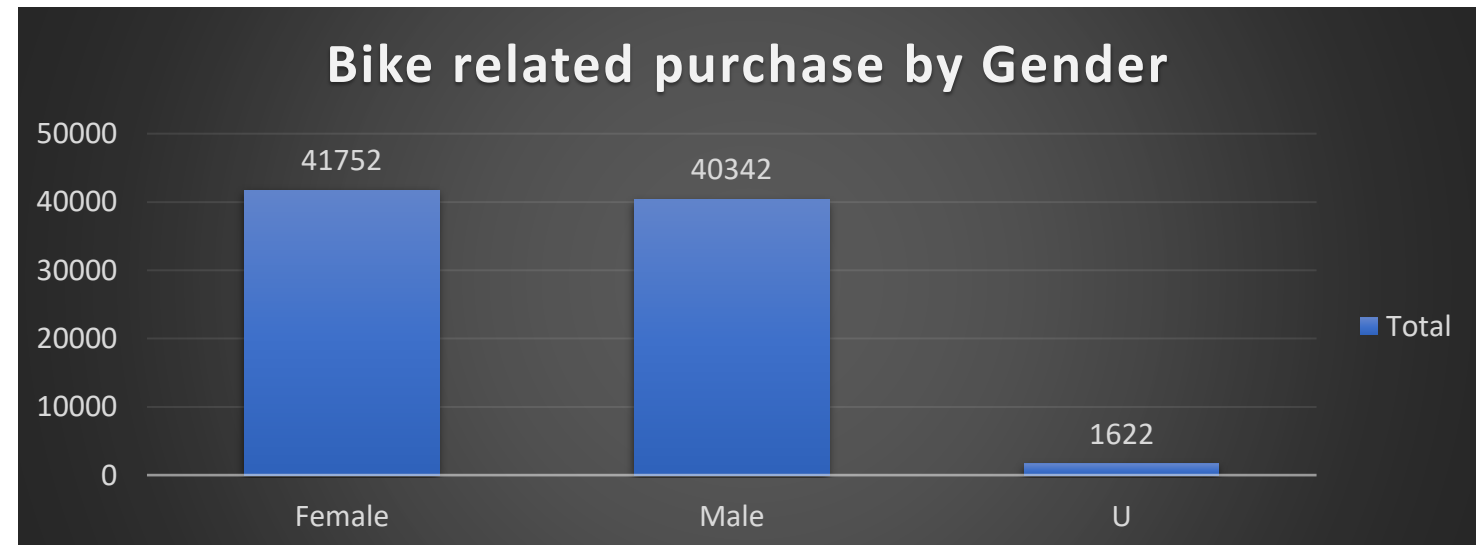
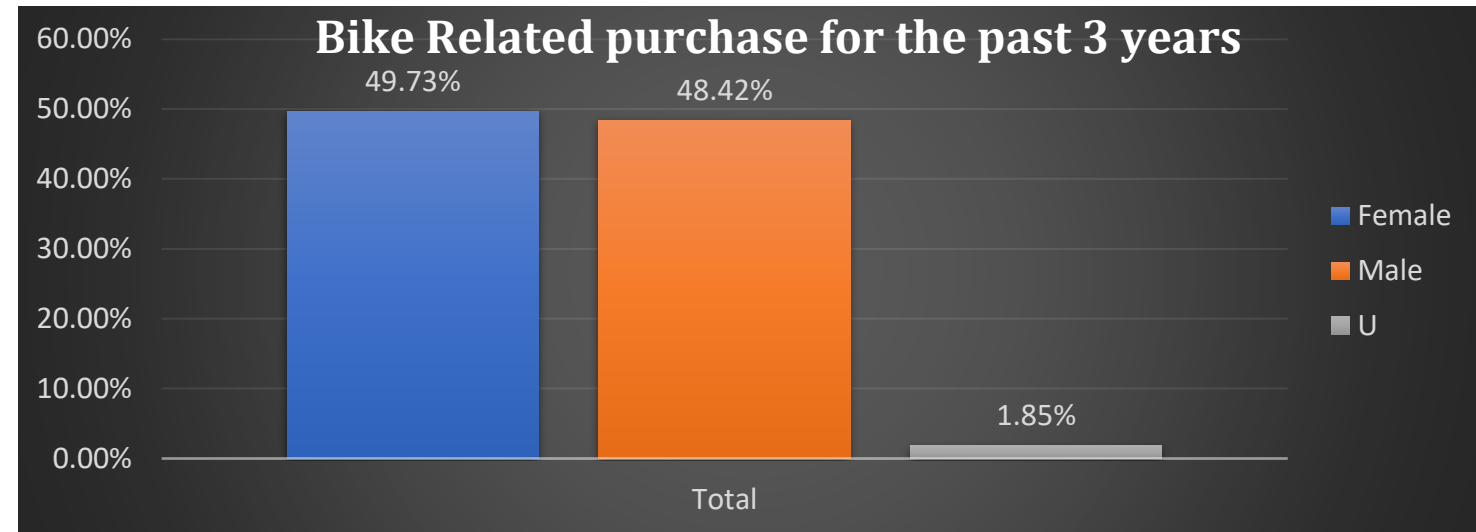


Bike Related Purchase for past 3years

As per the analysis almost 50% purchase of bikes were by Female

Nearly 48% bike related purchases were by Male

Also, Female purchased 41752 biked related accessories in comparison to Male who purchased 40342



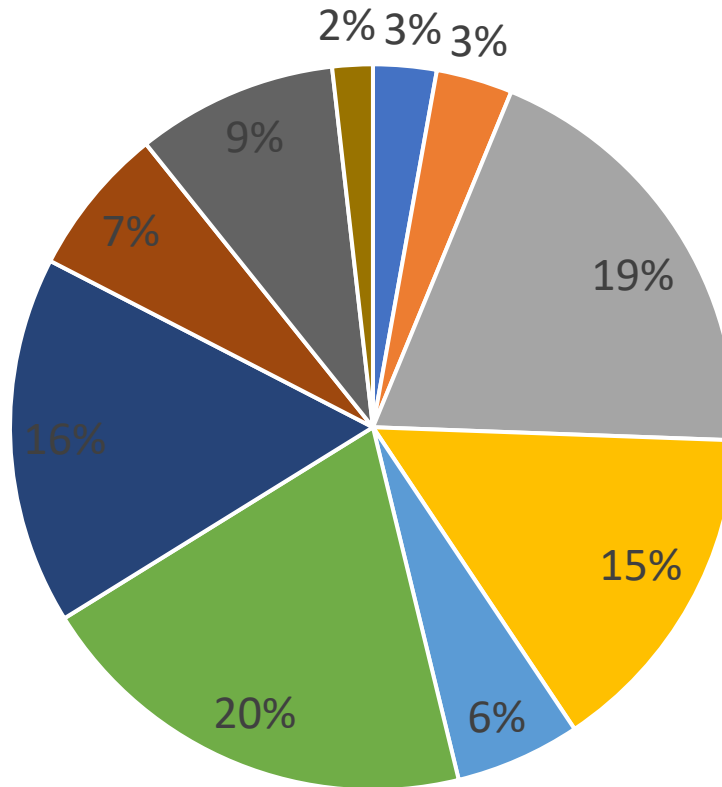
Job Industry Distribution

20% of 'New' customers are in Manufacturing and Financial Services

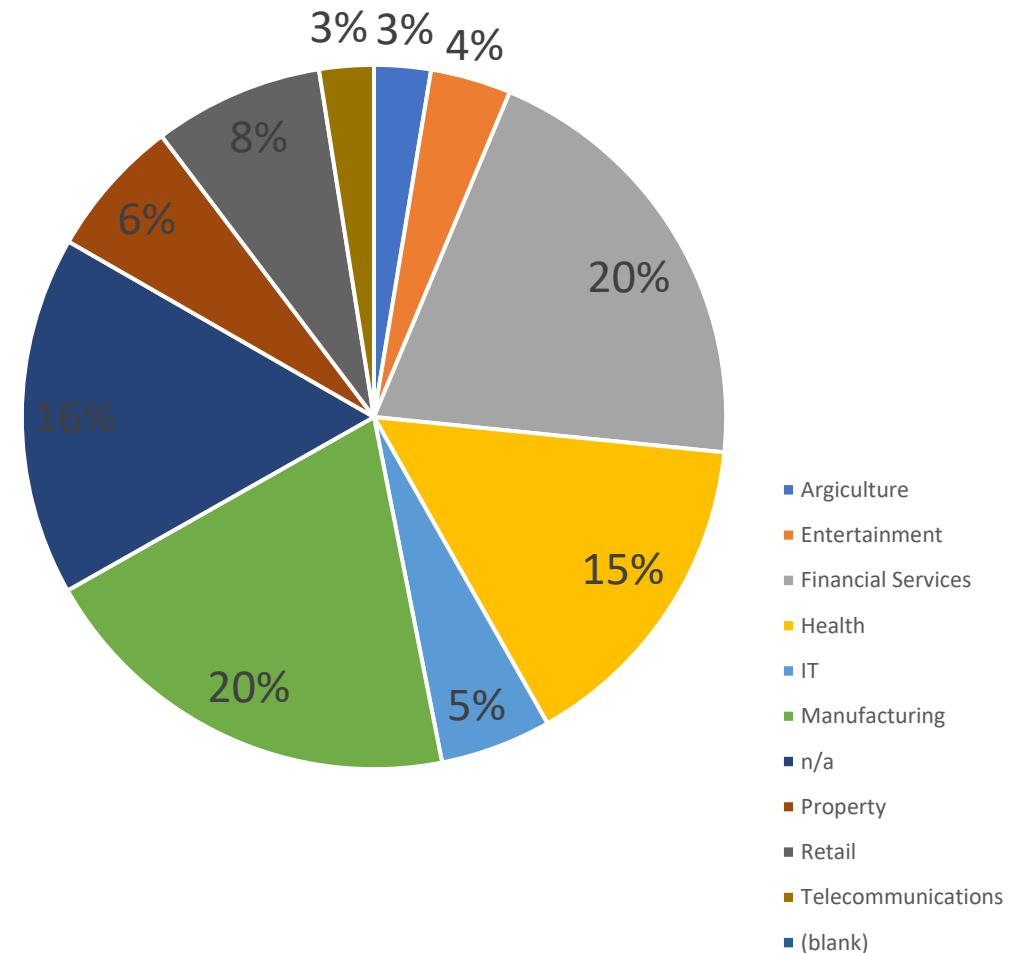
The smallest number of customers are in Agriculture and Telecommunications 3%

In 'Old' Customer list 20% and 19% are assigned under the category Manufacturing and Financial Services

Old Customer Job Industry Distribution



New Customer data by job category



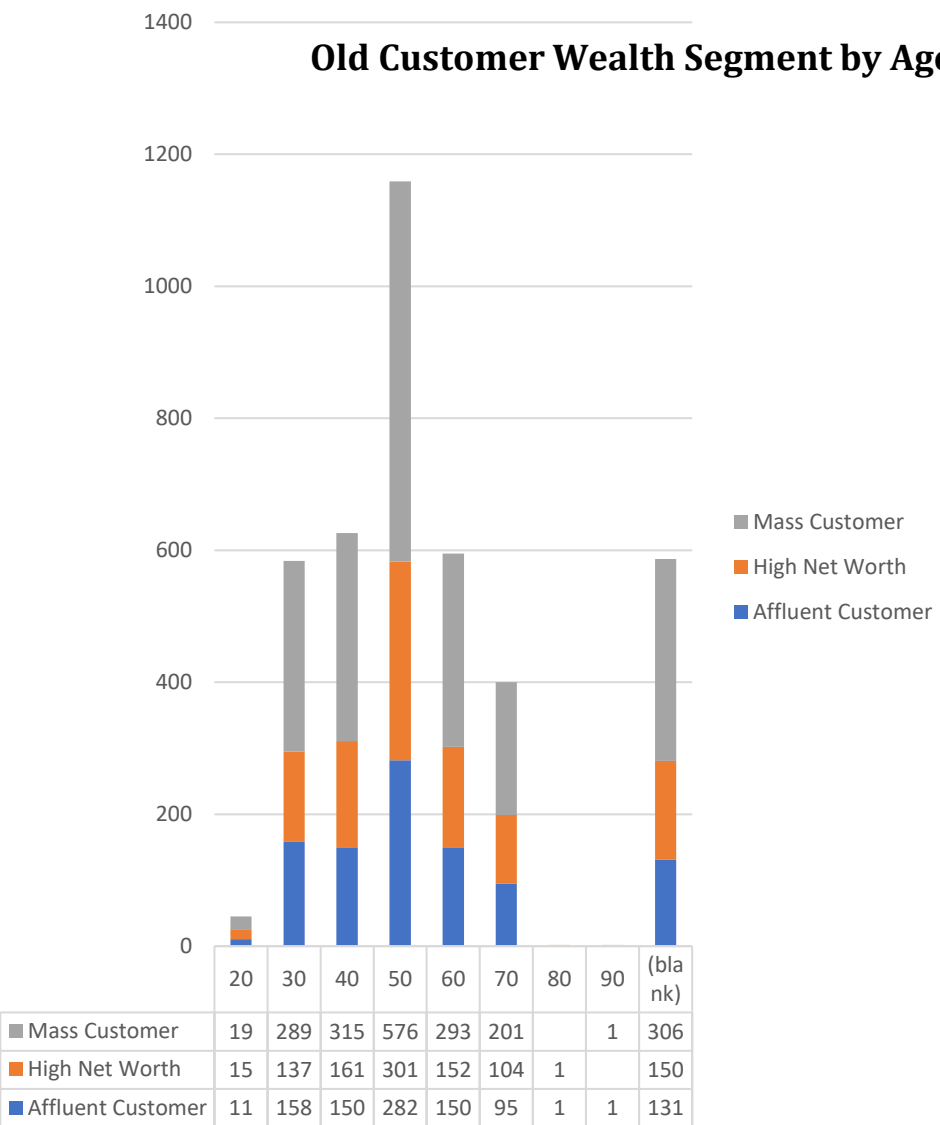
Wealth Segmentation by Age Category

*In all age categories
largest number
of customers
are classified as
'Mass
Customer'*

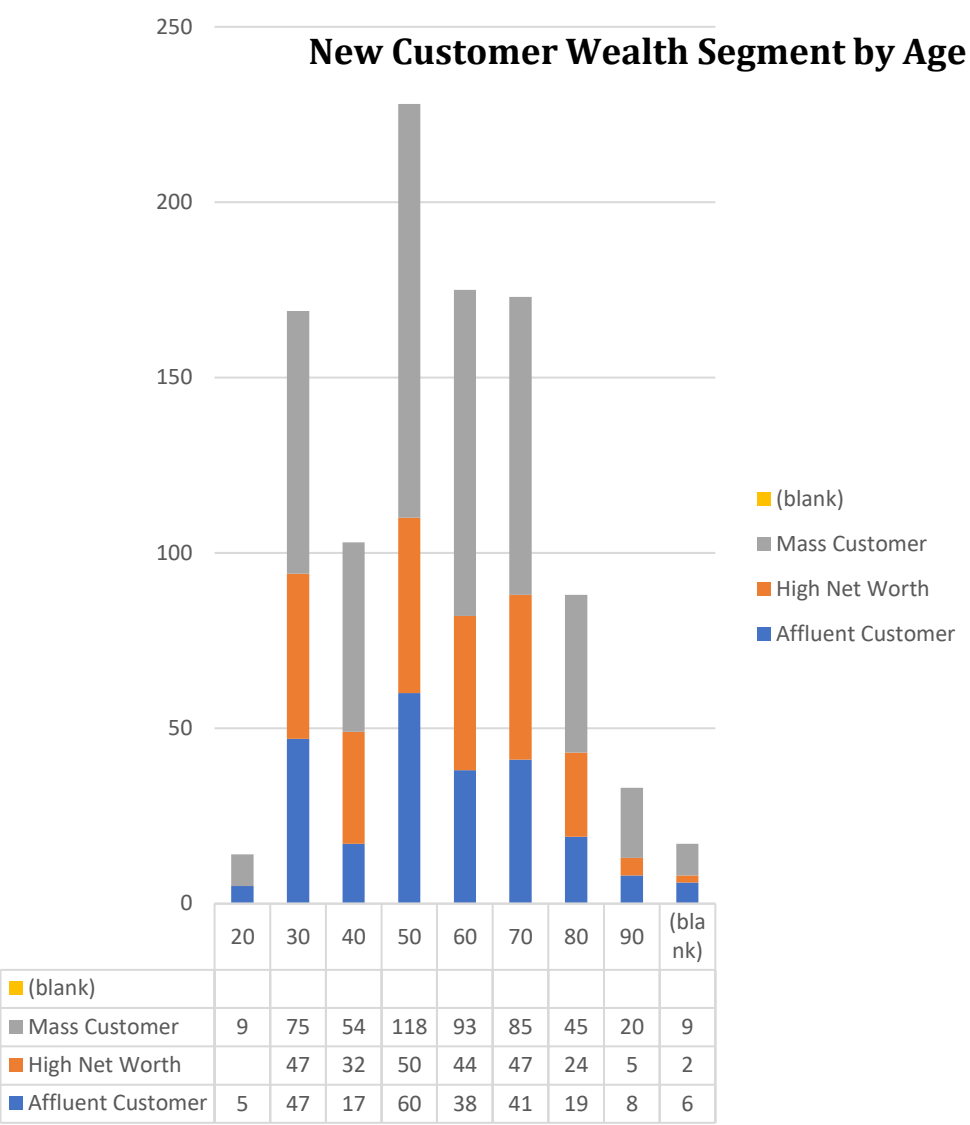
*The second
category is
'High Net
Worth'
customers*

*The 'Affluent
customers
outperforms the
'High Net
Worth' between
40-49 age group*

Old Customer Wealth Segment by Age



New Customer Wealth Segment by Age

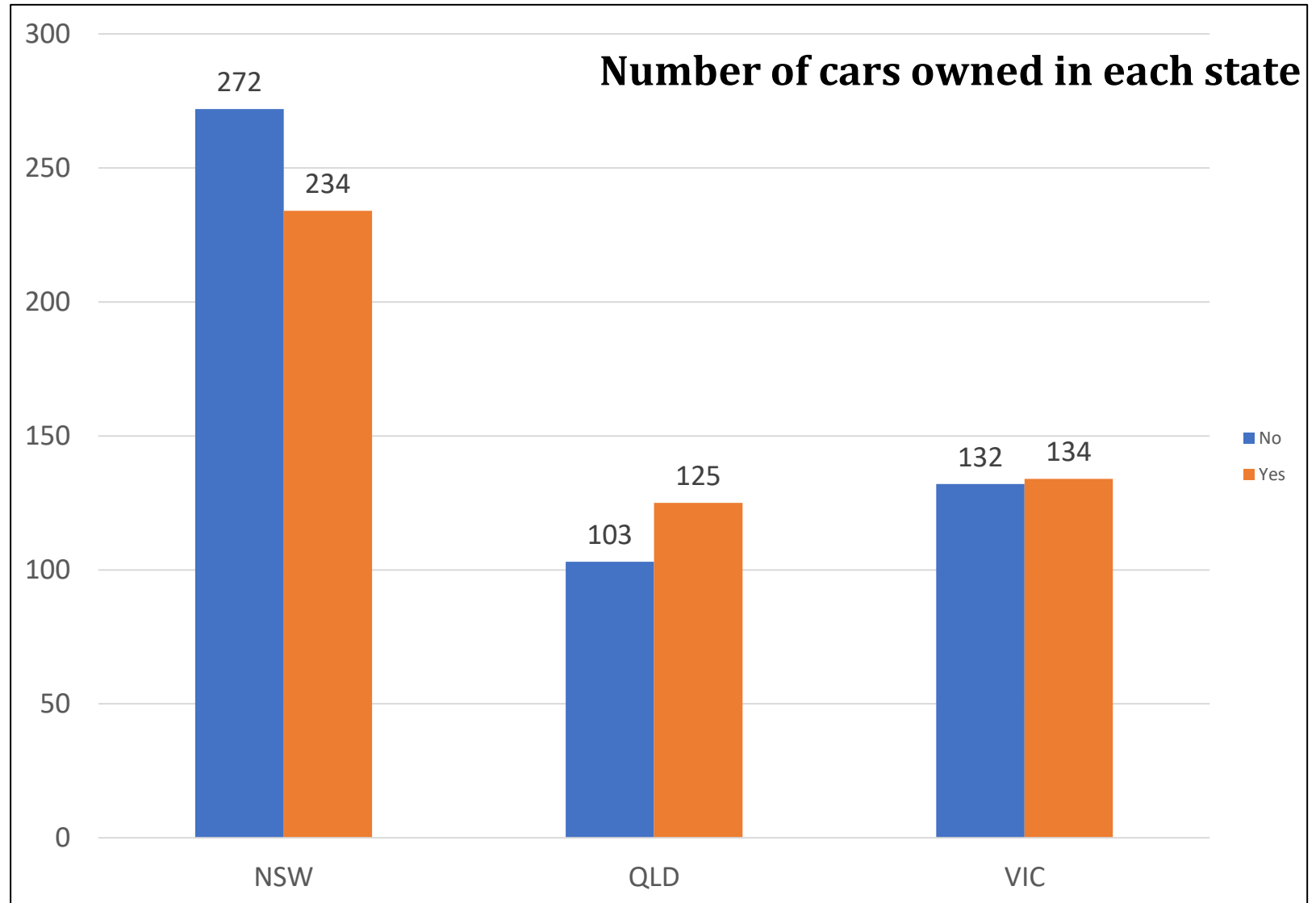


Cars Owned/Not Owned by State

NSW has the largest amount of people that do not own a car

The state VIC is seemed to have split equally

The State QLD is identified having more number of people who own a car

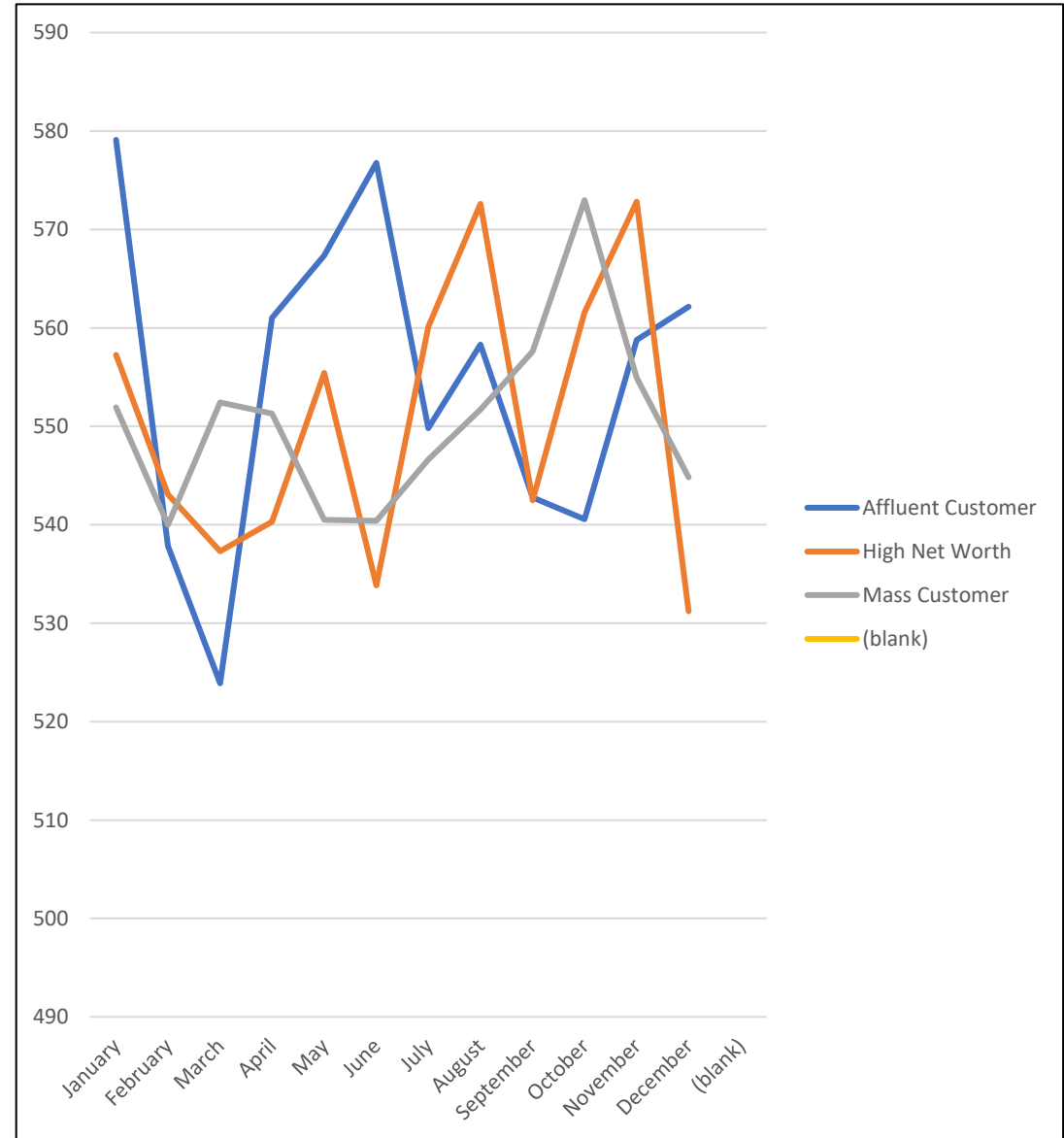
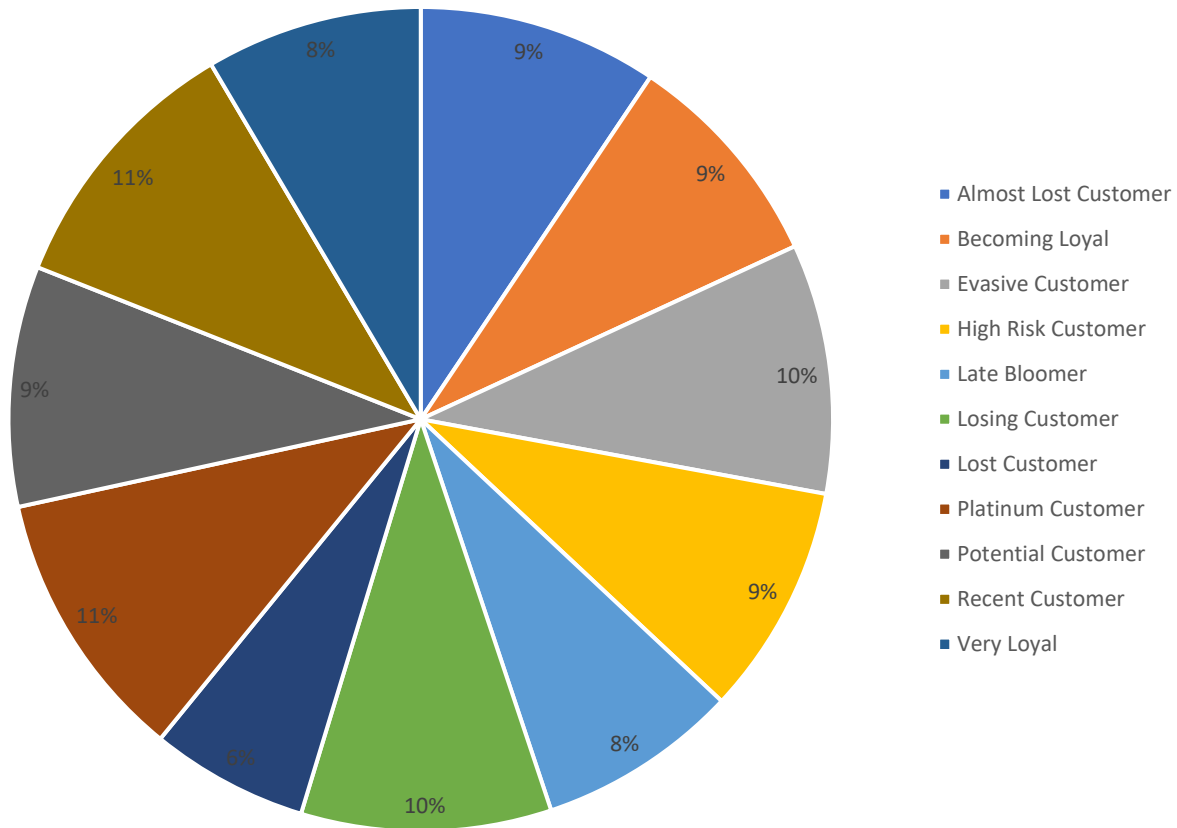


RFM Legend and Description

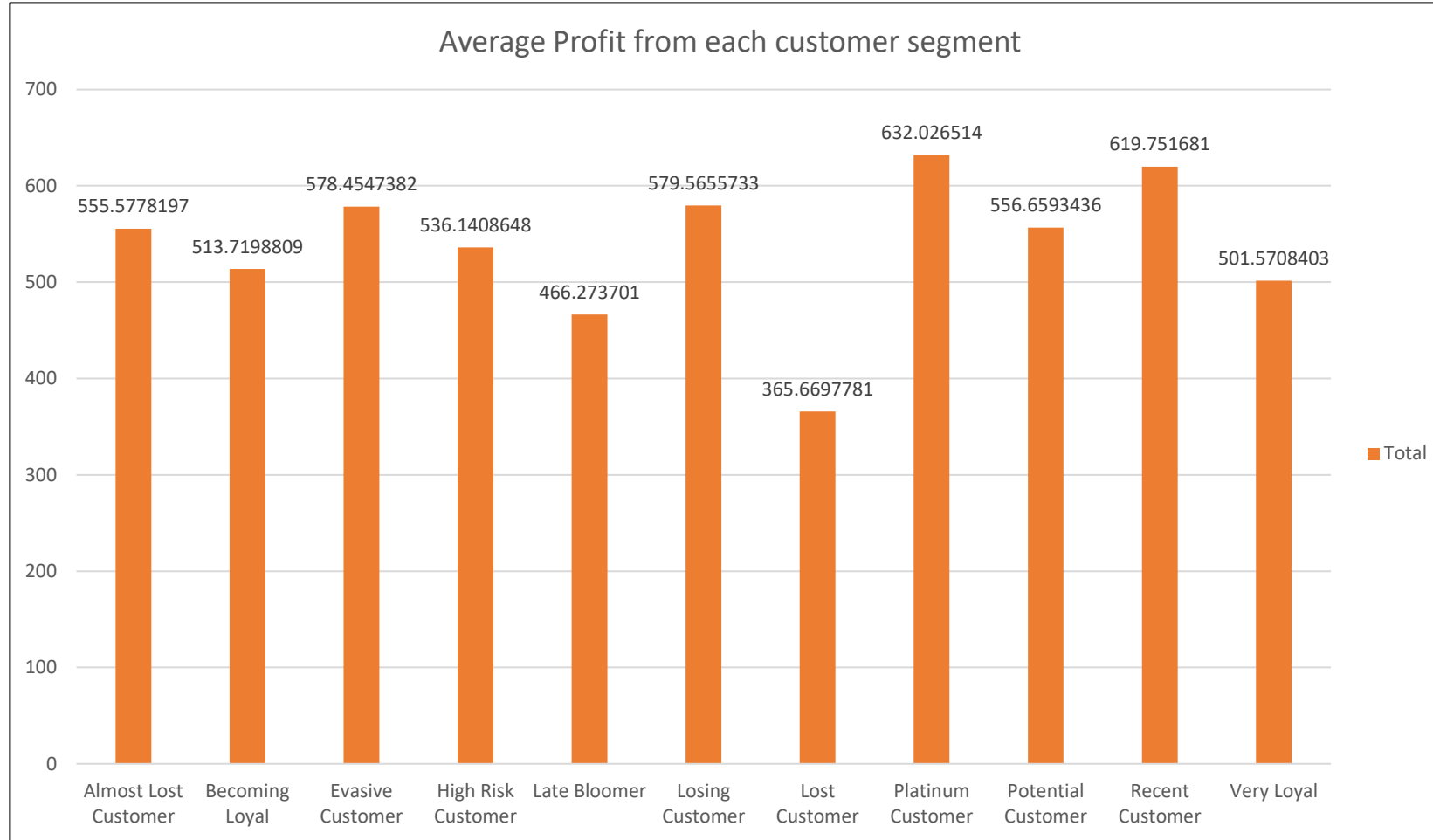
Rank	Customer title	Description	RFM Value
1	Platinum Customer	Most recent buy, buys often, most spent	444
2	Very Loyal	Most recent buy, buys often, spends large amount of money	443
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	421
4	Recent Customer	Bought recently, not very often, average money spent	344
5	Potential Customer	Brought recently, never brought before, spent small amount	323
6	Late bloomer	No purchase 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 better, high amount spent	212
9	Almost lost Customer	low recency, low frequency, but high amount spent	124
10	Evasive Customer	low recency, low frequency, very large amount spent	112
11	Lost Customer	very low RFM value	111

Model Development

Average Profit from Customers

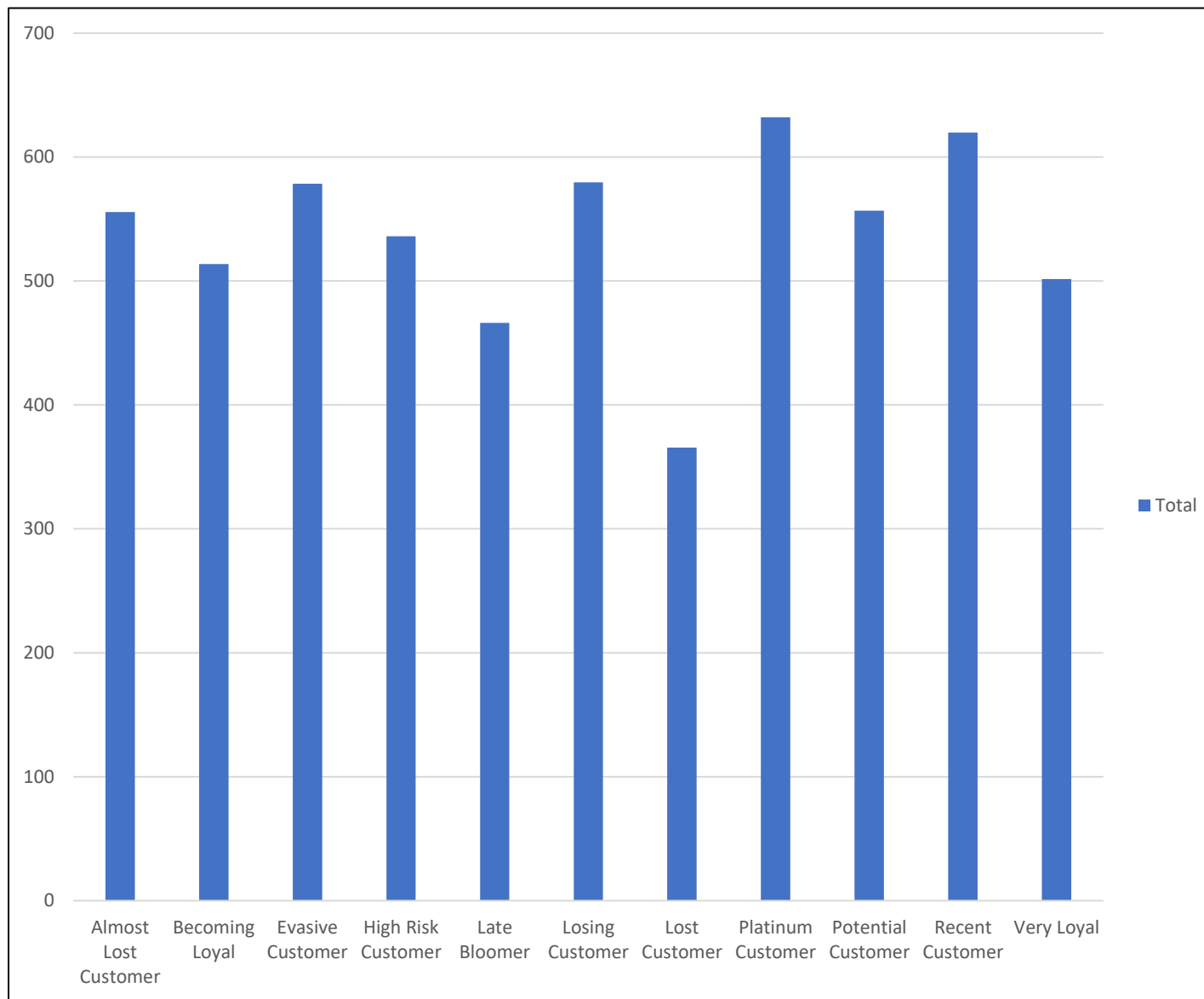


Model Development

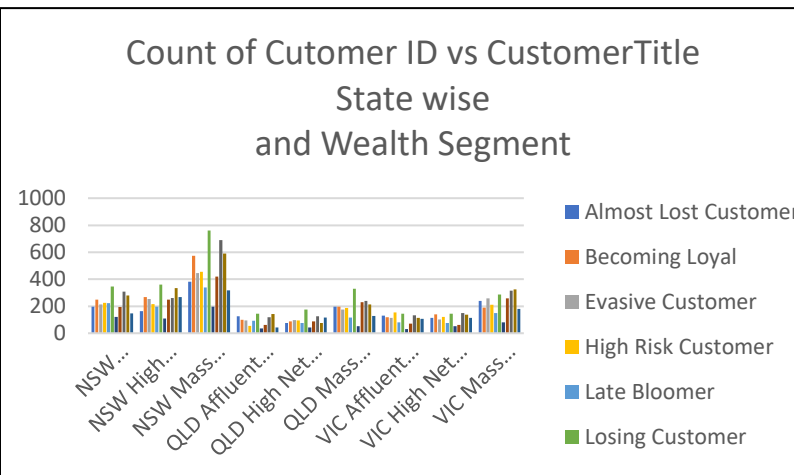
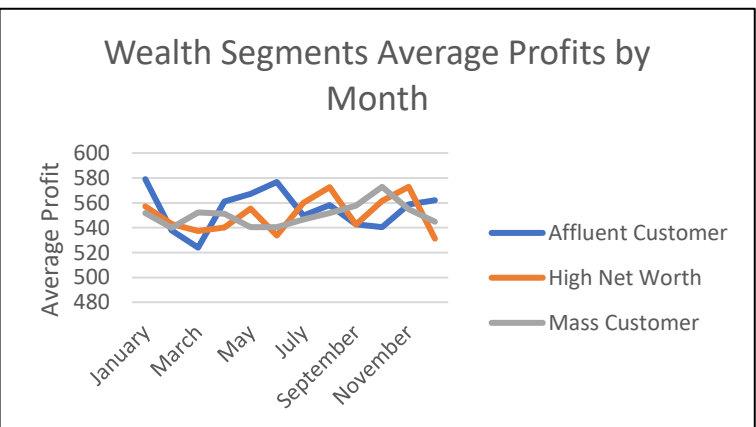
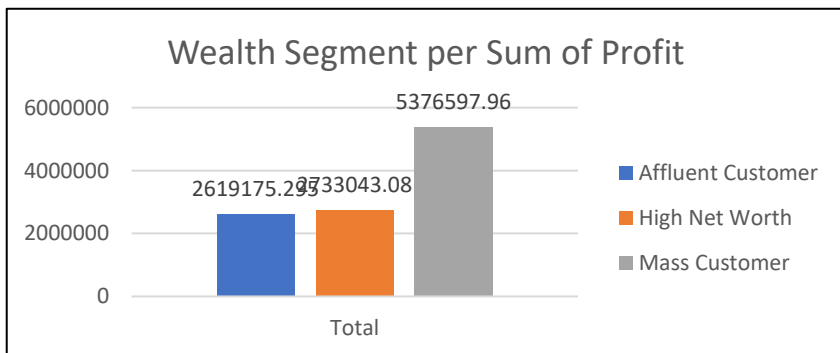
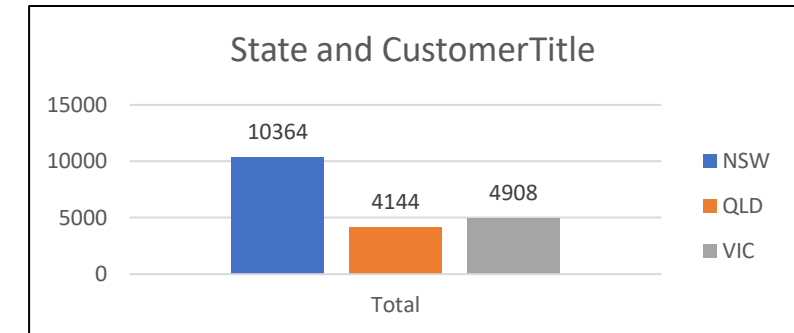
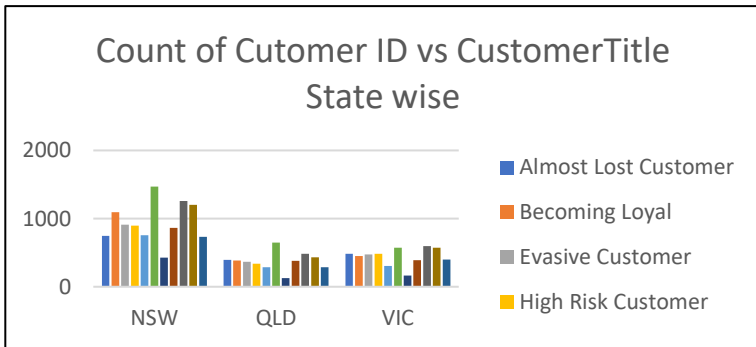
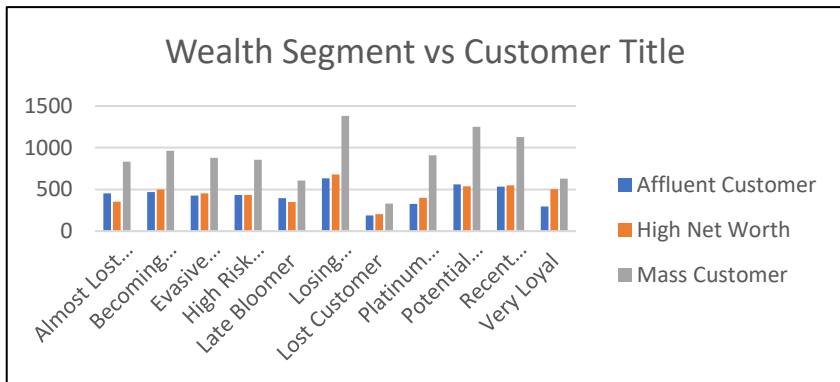
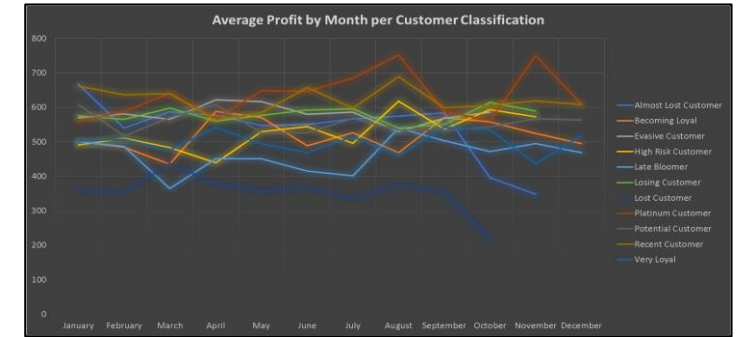
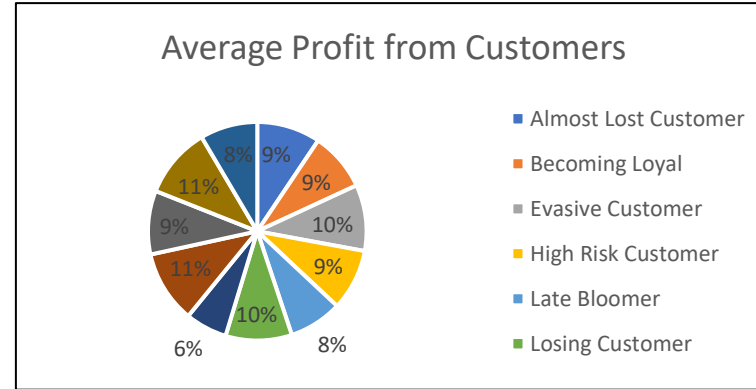
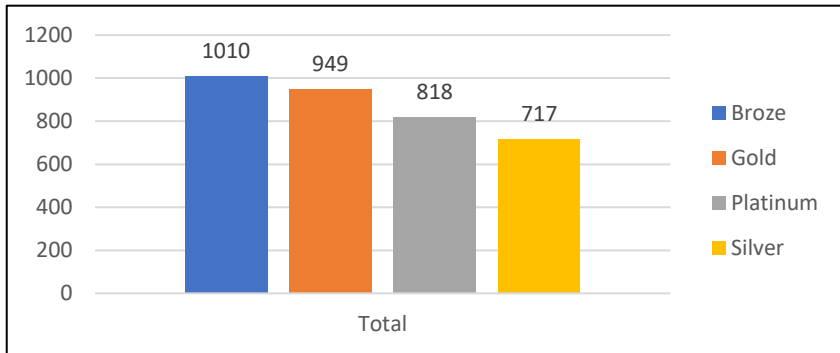


Model Development

Wealth
Segment
Average
Profit



DASHBOARD



Summary Table Top 1000 Customers

RANK	Customer Title	Description	Number of Customers
1	Platinum Customer	Most recent buy, buys often, most spent	177
2	Very Loyal	Most recent buy, buys often, spends large amount of money	187
3	Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	347
4	Recent Customer	Bought recently, not very often, average money spent	301