

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Sudhanshu Gosain

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 the provided datasets.
- Using the 3 datasets provides, the aim is to analyse and recommend 1000 customers that Sprocket Central should target to drive higher value to the company.

This will done in 3 phases:

Contents of Data Analysis

- ☐ “New” and “Old” customer age distribution
- ☐ Getting Customer status by “RFM” Approach
- ☐ Job – Industry distribution
- ☐ Customer status by age and wealth segment category
- ☐ Profit by wealth segment and gender category
- ☐ Car owners by genders
- ☐ Last-3 Year purchase by Age and Gender

1. **Data Exploration**
2. **Model Development**
3. **Interpretation**

Data Exploration

Data Quality Assessment and Clean up

Summary Table

Data Quality issue

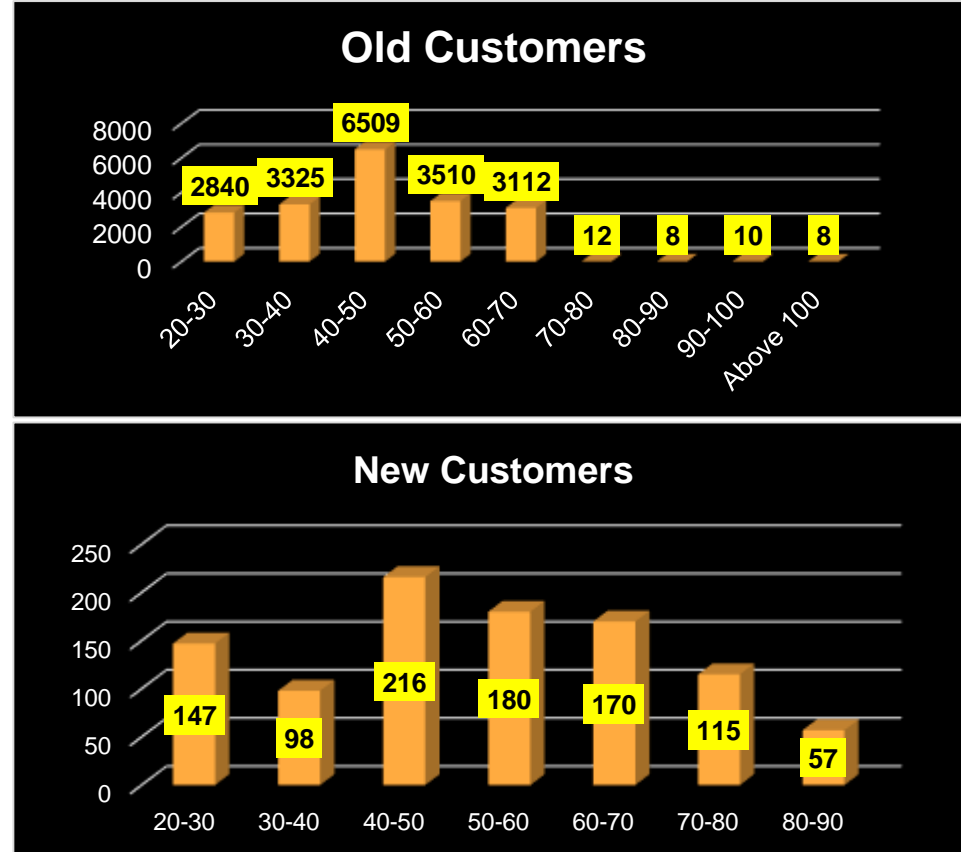
- **Accuracy** : Correct Values
- **Completeness**: Data Filled with Values
- **Consistency**: Values Free from Contradiction
- **Currency**: Values up to date
- **Relevancy**: Data items with value Meta-data
- **Validity**: Data Containing allowable values
- **Uniqueness**: Records that are duplicated

	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
<i>Customer Demographic</i>	DOB: Inaccurate Age: Missing	Job title: Blanks Last Name: Blanks DOB: Blanks Job Industry: N/A	Gender: Inconsistency	Deceased Customers: Filter out	Default column: Irrelevant	Customer id: format DOB: Format Tenure: Format Past_3_year: Format
<i>Customer Address</i>		Customer id: Incomplete	State: Inconsistency			Post code: Format Property Valuation: Format
<i>Transactions</i>	Profit: Missing	Customer id: Incomplete Product line: Blanks Product Class: Blanks Product sold date: Blanks. Product size: Blanks Online Order: Blanks Brand: Blanks	Standard cost: inconsistent			Customer id: Format Transaction date: Format List Price: Format Product sold date: format
<i>New Customer list</i>	DOB: inaccurate Age: Missing	Job title: Blanks Last Name: Blanks DOB: Blanks			4 unknown columns: Irrelevant	Post code: Format Property Valuation: Format Past_3_year: Format

Data Exploration

Age Distribution

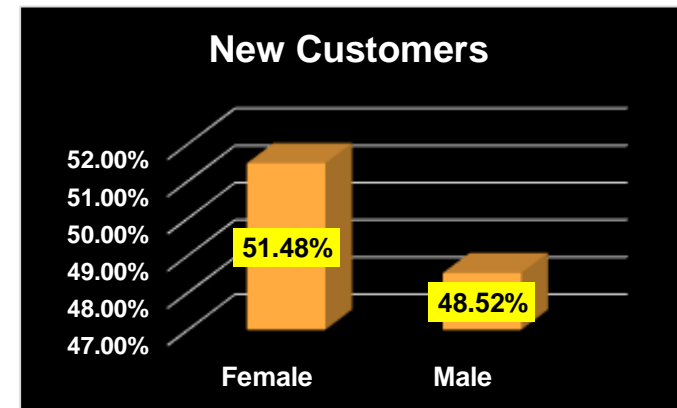
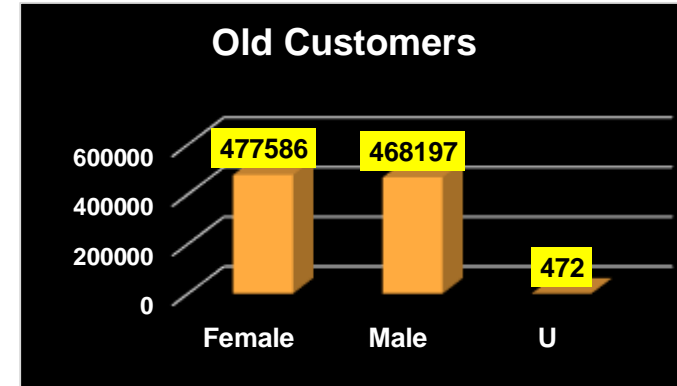
- Most customers are aged between 40-49 in “Old Customers” as well as in “New Customers”.
- Least number of customers are 70+ in the “Old Customers” and 80+ in the “New Customer”.
- “Old Customers” has majority of customers from 20-70 years.
- 40-69 is the most popular age group range in “New Customers”
- “New Customers” sees a steep drop of customers in the 30 – 39 age group.



Data Exploration

Bike Related purchases over last 3 years by gender

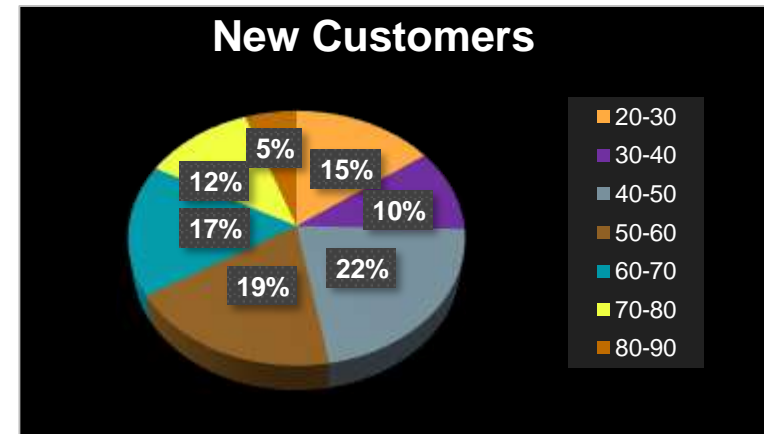
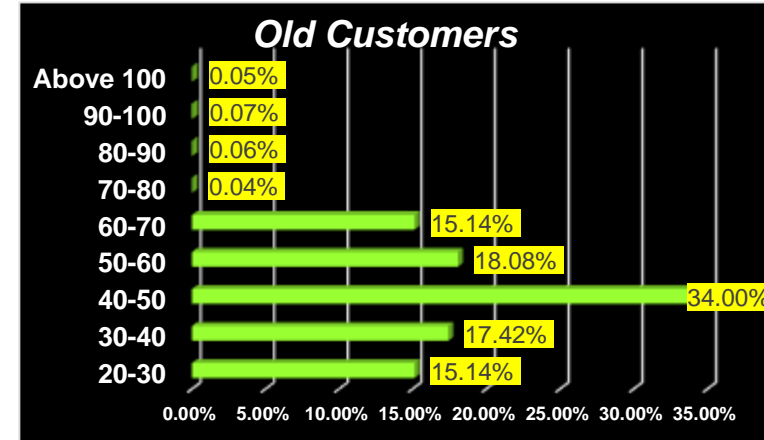
- In the last three years around 50% of the bike related purchases were made by the females in the “Old customers”
- 472 purchases in the “Old Customers” is done by unknown gender.
- In the “New Customers” the female percentage is 51.48%, whereas the 48.52% purchases is done by males.



Data Exploration

Bike Related purchases over last 3 years by AGE

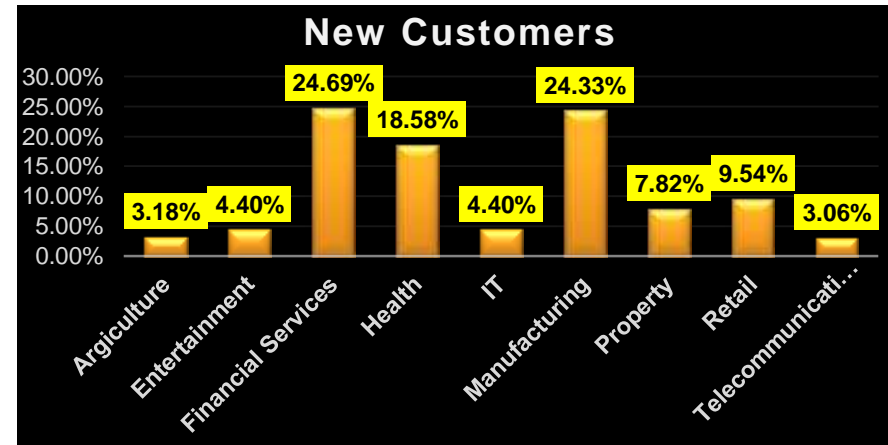
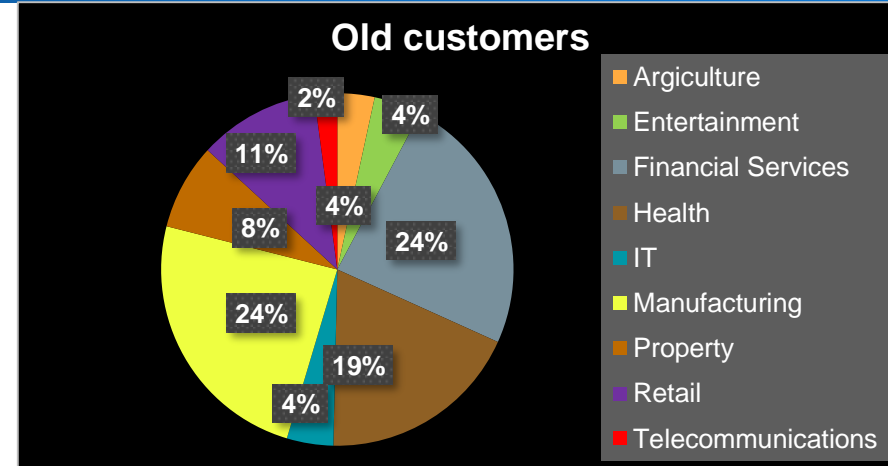
- In “Old customers” majority of purchases is done by customers from age group 30-59.
- In “Old Customers” 40-49 age group has done highest 34% purchases in the last 3 years.
- In the “New Customers” highest 22% is made by the customers belonging to 40-49 age group.
- 80+ customers purchased only 5% bike related produced in the last 3 years from the “New Customers”



Data Exploration

Job Industry Distribution

- In “Old customers” 48% of customers are in Manufacturing and Financial Services industry.
- This become 49.03% when we analyse the customers in “New Customers”
- In both the group the smallest numbers of customers are from Agriculture and Telecommunication industries.
- Only 2% customers are from Telecommunication in “Old Customers” and 3.06% in the “New Customers”.

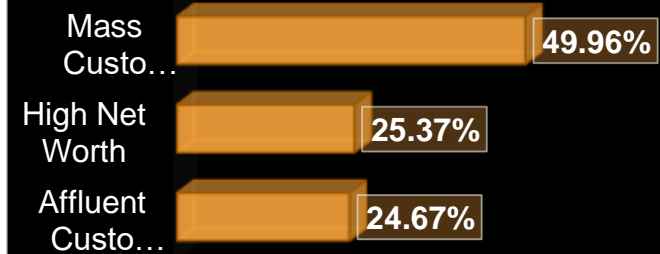


Data Exploration

Wealth Segment Analysis

- In “Old customers” 49.96% of customers are in Mass Customer Segment.
- In “New Customers” Highest 499 customers are in Mass Customer Segment.
- In both the group the smallest numbers of customers are in Affluent Customers Segment

OLD CUSTOMERS



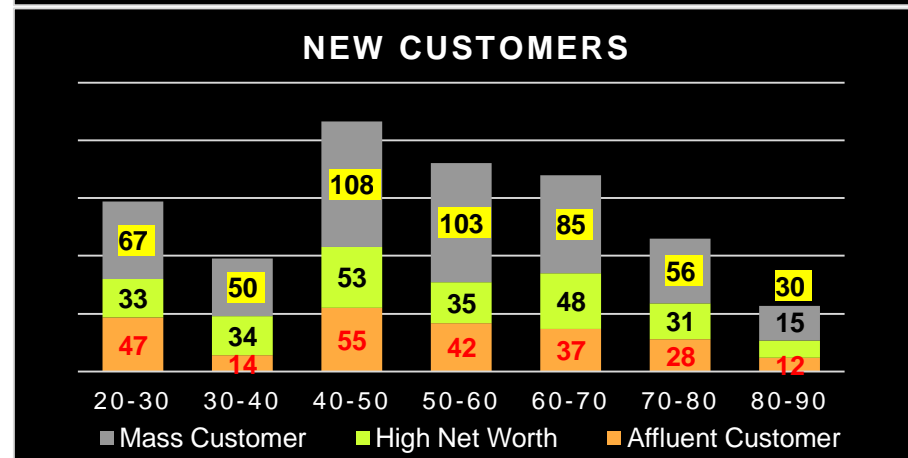
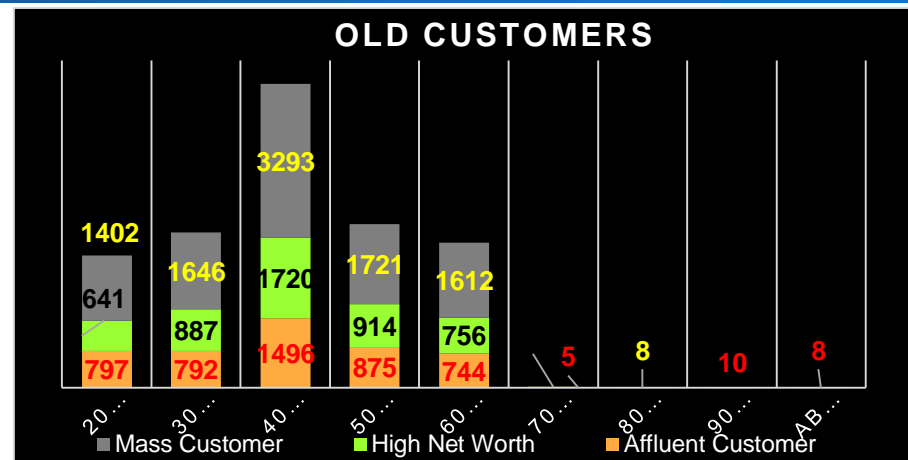
New Customers



Data Exploration

Wealth Segment Analysis by age group

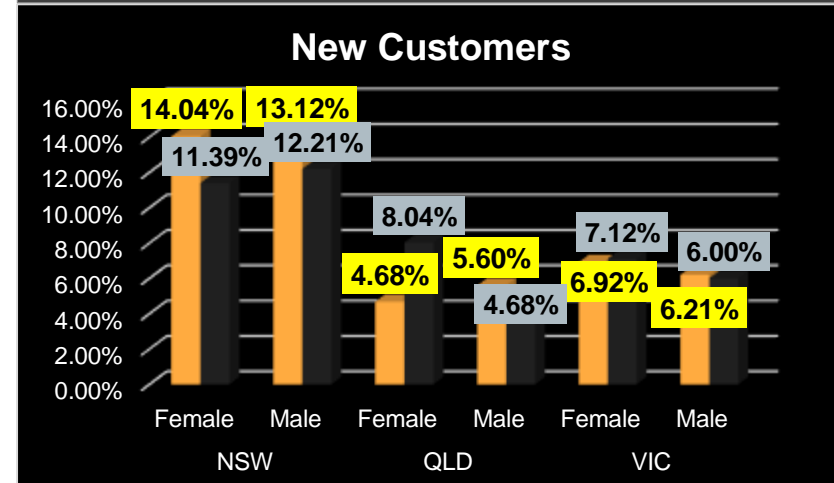
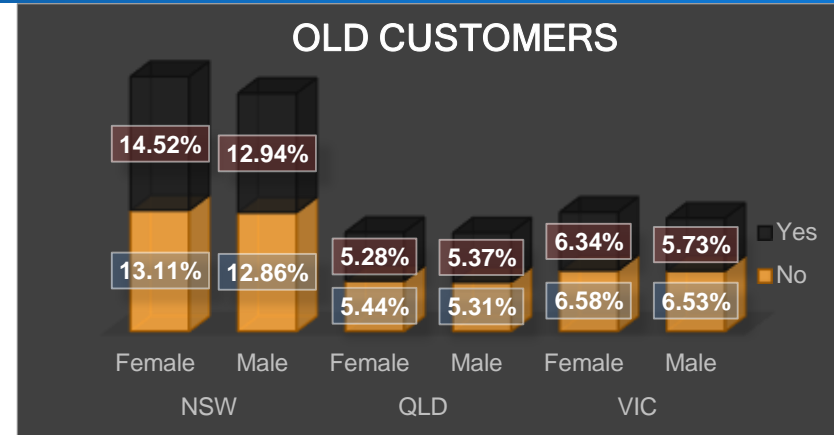
- In “Old customers” highest 3293 are from Mass Customers from age group 40 – 49.
- In “New Customers” Highest 108 are from Mass Customers from age group 40 – 49.
- In both the group the numbers of customers are start decreasing from the age 50 onwards.
- In “Old Customers” only 24 customers are 70+ age.



Data Exploration

Car Ownership analysis

- In “Old customers” 13.11% females and 12.86% males in NSW owns a car.
- In “New Customers 14.04% females and 13.12% males in NSW owns a car.
- In both the group the highest numbers of car owners are in New South Wales.
- Victoria splits quite evenly as the male to female car ownership ratio is looking similar.
- Least number of customers are from QLD and in “New Customers ” 8.04% of female do not have a car which is double as compared to the female car owner in the city.



Data Exploration

RFM Analysis and Customer classification

- RFM analysis is used to determine which customers a business should target to increase its revenue and values.
- The RFM (REGENCY, FREQUENCY, and MONETARY) model shows customers that have displayed high levels of engagements in 3 categories mentioned.
- We have 5322 “Premium Customers”, 5168 “Gold Customers”, 4705 “Silver Customers”, Least 4139 “ Bronze Customers”.

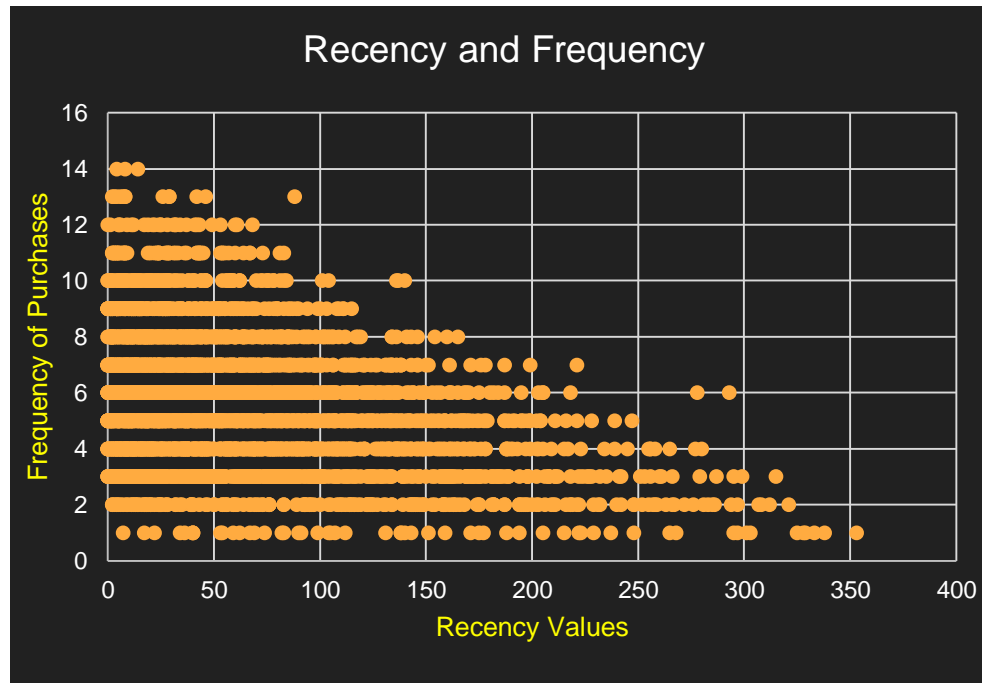


RFM Values and Customers Status				
min	111	→ Brone Customers		
q1	211	→ Silver Customers		
median	311			
q3	411	→ Gold Customers		
max	444	→ Premiun Customers		

Data Exploration

RFM Analysis

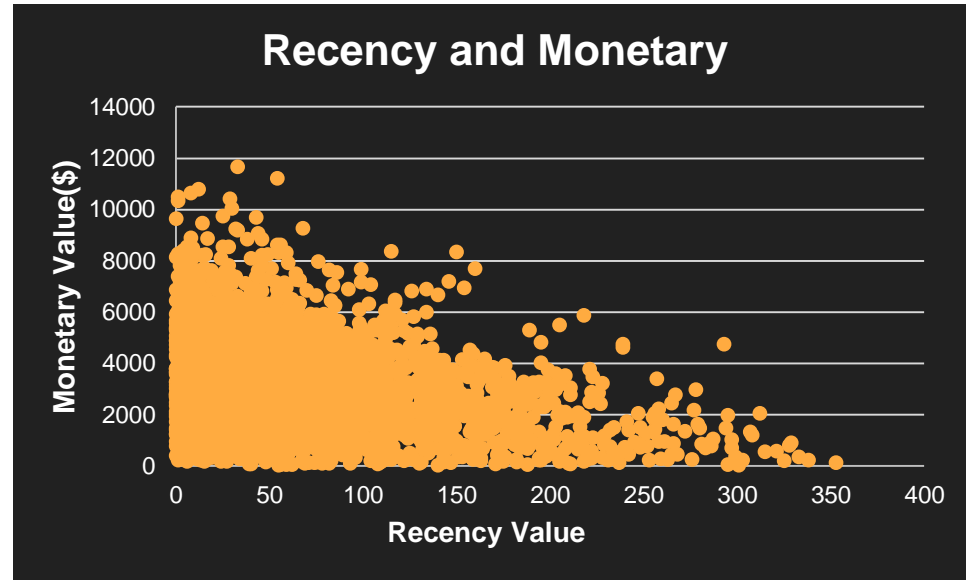
- Very low frequency of 0-2 correlated with high recency values. i.e. more than 250 days
- Customers who visited recently (0-100) days have a higher frequencies i.e. up to 12 times and they are visiting frequently .
- Higher frequency has a inverse relationship with frequency as the recent customers are the less frequent customers



Data Exploration

RFM Analysis

- The chart shows that customers who purchased more recently have generated more revenue, than the customers who visited a while ago.
- Customers who visited recently (80-150) days also show to generate a moderate amount of revenue .
- Customers visited more than 200 days ago generate low revenue



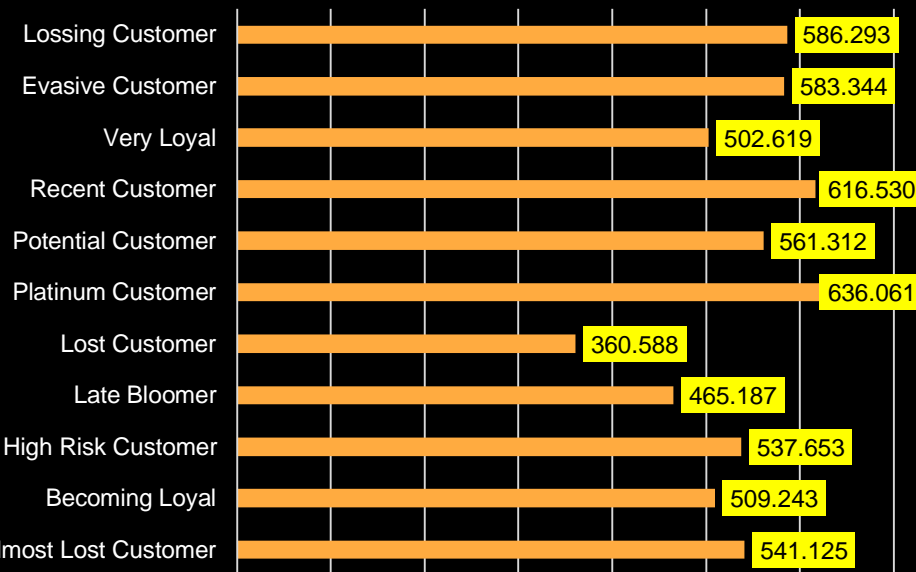
Model Development

Customer Status table with RFM values

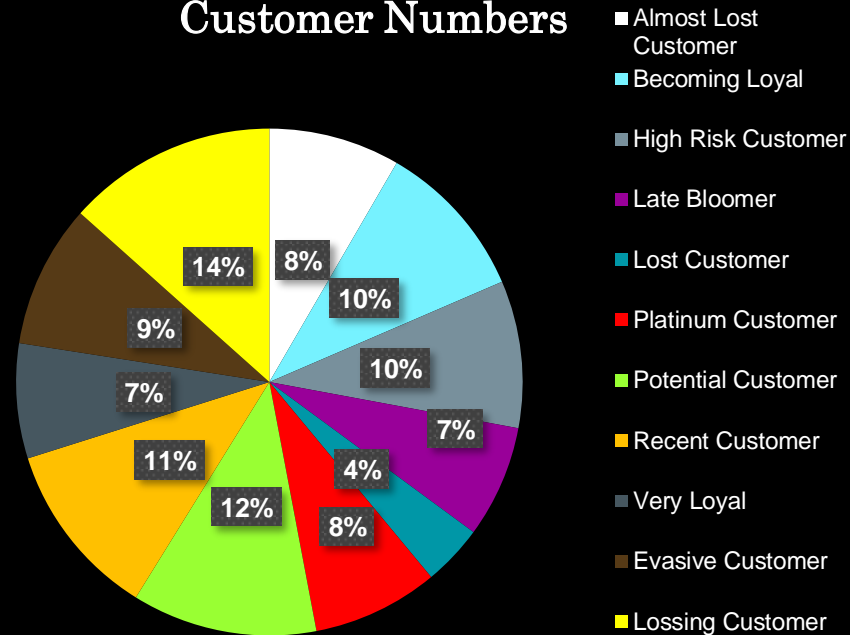
	Customer Status Analysis	
Customer Title	Description	RFM Value
Platinum Customer	Most recent buy, buys often, most spent	444
Very Loyal	Most recent, buys often, spends large amount of money	433
Becoming Loyal	Relatively recent, bought more than once, spends large amount of money	421
Recent Customer	Bought recently, not very often, average money spent	344
Potential Customer	Bought recently, never bought before, spent small amount	323
Late bloomer	No purchases recently, but RFM value is larger than average	311
Losing Customer	Purchases was a while ago, below average RFM value	224
High Risk Customer	Purchase was long time ago, frequency is quite high, amount spent is high	212
Almost Lost Customer	Very low recency, low frequency, but high amount spent	124
Evasive Customer	Very low recency, Very low frequency, small amount spent	112
Lest Customer	Very Low RFM	111

Model Development

Average Profit

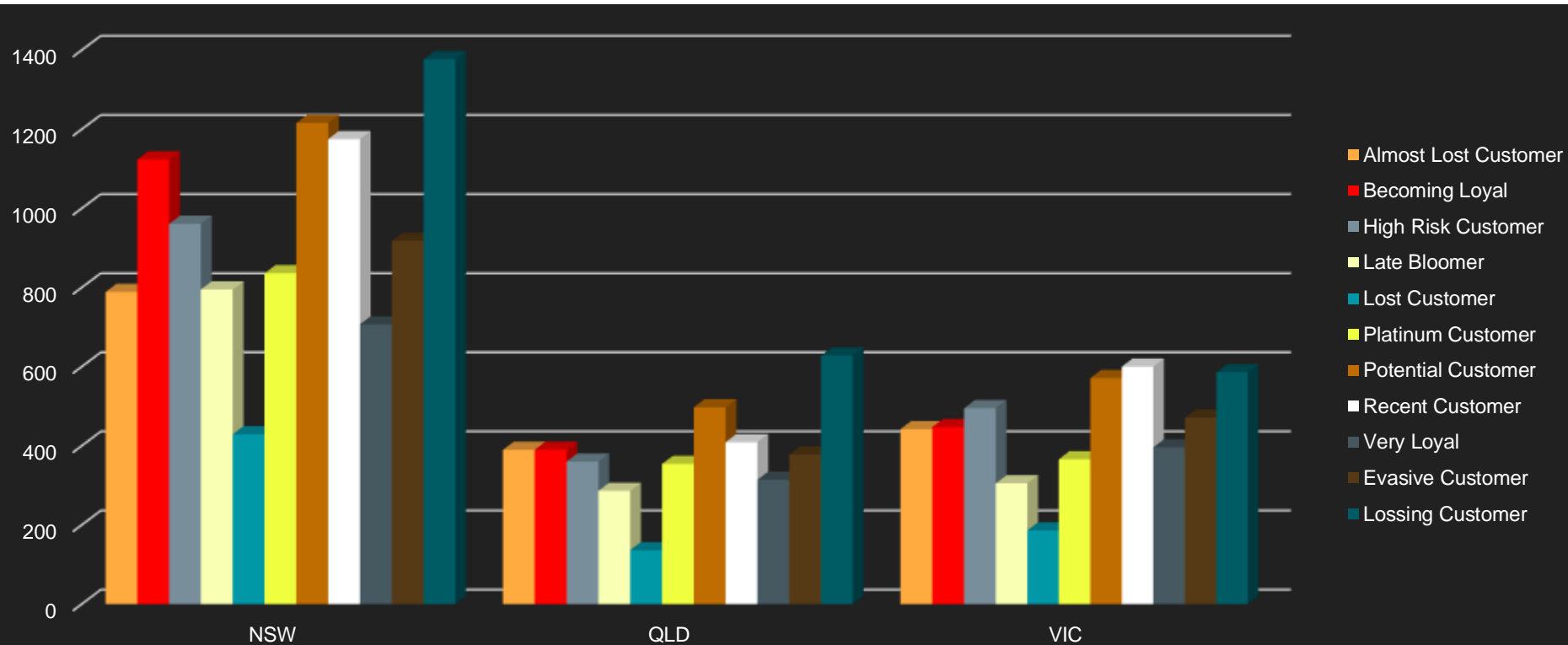


Customer Numbers



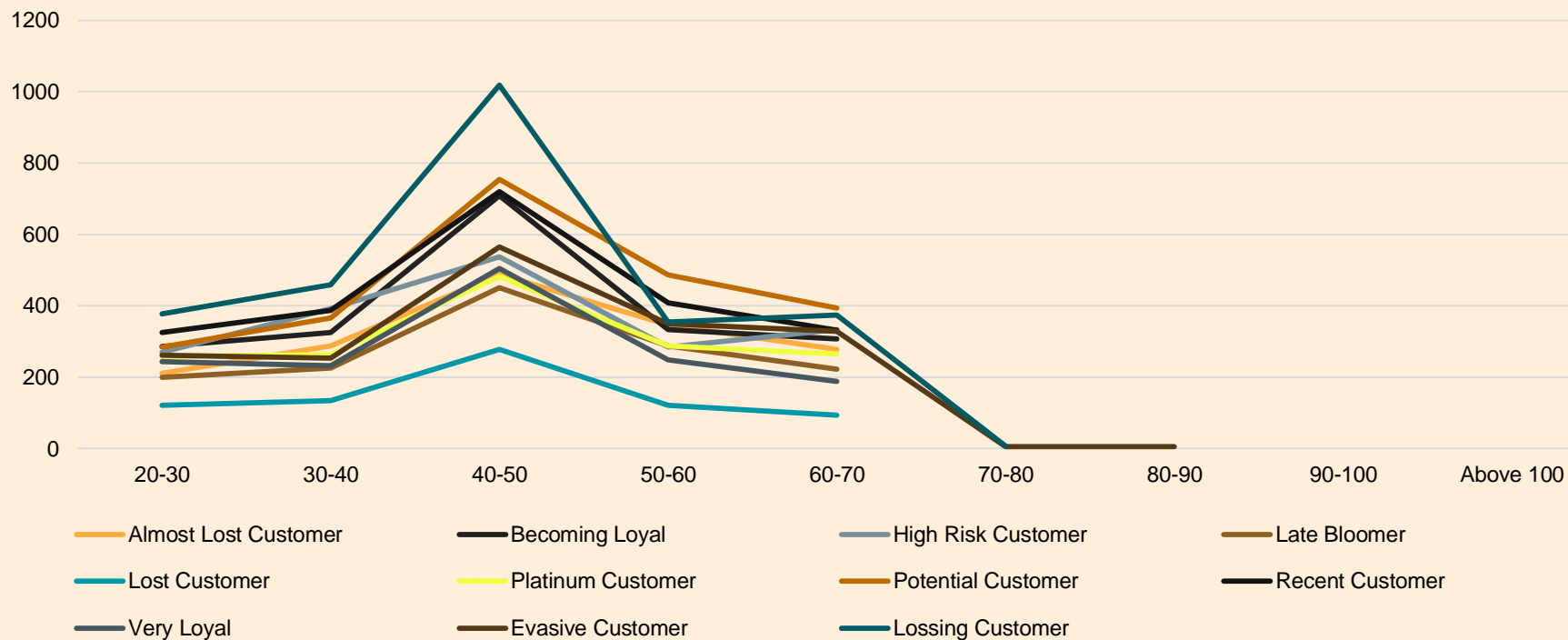
Model Development

Customers status state wise



Model Development

Age distribution of Customer status



Data Exploration

- These are the top 10 Customers who provides best business to the company.
- 6 out of 10 are from NSW , 2 customers are from QLD & VIC.
- 11669 is the revenue provide by “TYE” from QID which is highest .
- Customers visited more than 200 days ago generate low revenue

TOP 10 CUSTOMERS			
Tye	941	QLD	11669
Marcile	2637	NSW	11222.7
Morley	1460	NSW	10787.6
Oberon	1558	NSW	10640.3
Ammamaria	729	VIC	10497.8
Wes	3326	NSW	10422
De	322	QLD	10341.6
Olvan	2770	NSW	10028.8
Jeffry	1597	NSW	9739.46
Kynthia	1887	VIC	9695.59

For Better Understanding of RFM Model visit the following links:

<https://www.optimove.com/resources/learning-center/rfm-segmentation>

<https://clevertap.com/blog/rfm-analysis/>



Thank you