

TheAnalyticsTeam

# Sprocket Central Pty Ltd

Data analytics approach

Ting-Lun, Fan

# Agenda

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

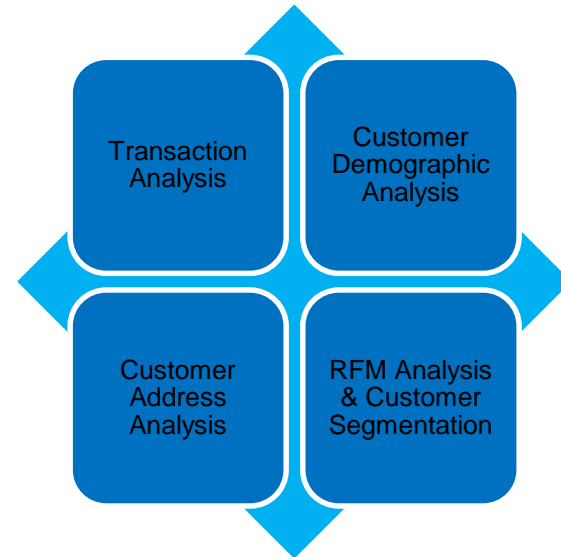
# Introduction

## Identified Top 1000 Customers From Datasets as Targeted Customer

### Background

- Sprocket Central Pty Ltd is a medium size bikes & cycling accessories organization.
- The organization has a large dataset relating to its customers, but their team is unsure how to effectively analyze it to help optimize its marketing strategy.
- Their marketing team is looking for boosting business sales.
- Sprocket Central has provided 3 datasets. Our job is to analyze the datasets and recommend top 1000 customers that they need to focus on to generate higher business value

### Analysis Content



# Data Exploration

## Data Quality Assessment & Data Cleaning

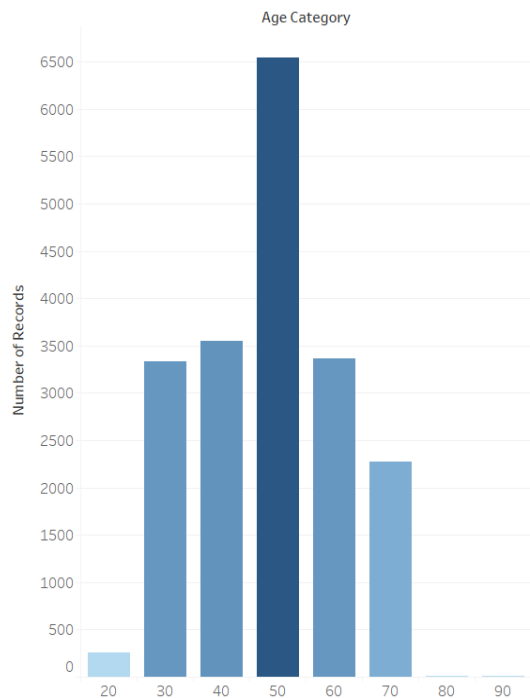
### Key Issues:

- 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

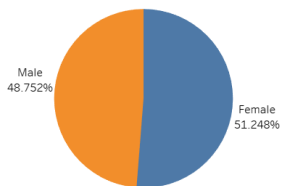
	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity	Uniqueness
Customer Demographic	DOB: Inaccurate Age: Missing attribute	last_name, DOB, job_title, job_industry_category: Missing value	Gender: Inconsistency		Default: Deleted		
Customer Address			States: Inconsistency				Country: All the same
Transaction	Profit: Missing attribute	Online order, brand, product_line, class, size and cost: Missing value			Cancelled status order: Filtered out		List price format Product sold date format

# Data Exploration

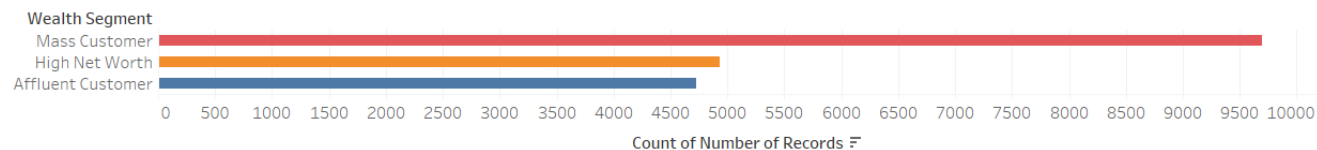
## Customer Demographic Analysis



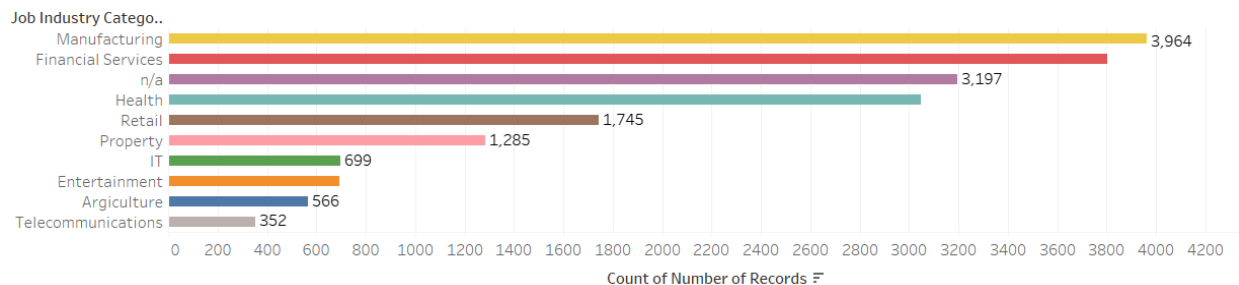
20 => age between 10~20



### Wealth Segment Category



### Job Industry Category



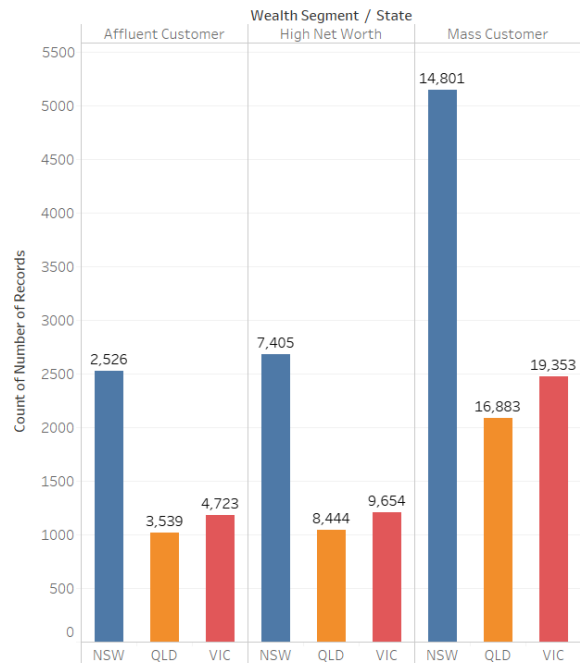
## Insights

- Most of our clients are around 40~50 years old, very less teenager.
- Female are slightly higher than male
- Most customer are from manufacturing and financial service, while telecommunication holds the lowest.

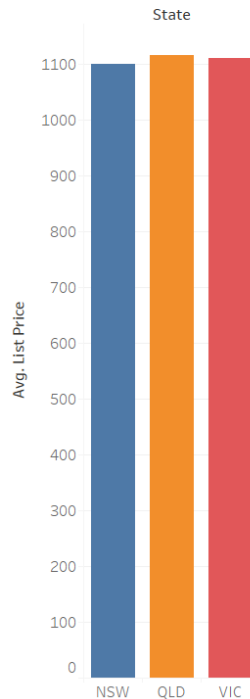
# Data Exploration

## Customer Address Analysis

Wealth Segment in State Category



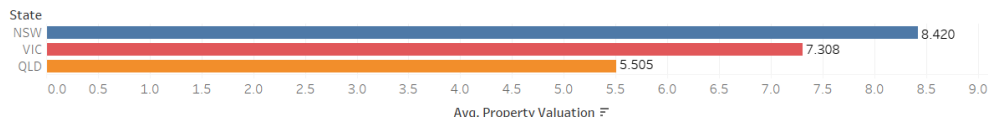
Average Purchase Spend by States



### Insights

- As we can see, New South Wales has the highest number of customer, which also holds the highest proportion in each wealth segment category.
- The average purchase spend in each state is similar.
- However, New South Wales has relatively high average property valuation, while QLD holds the lowest.

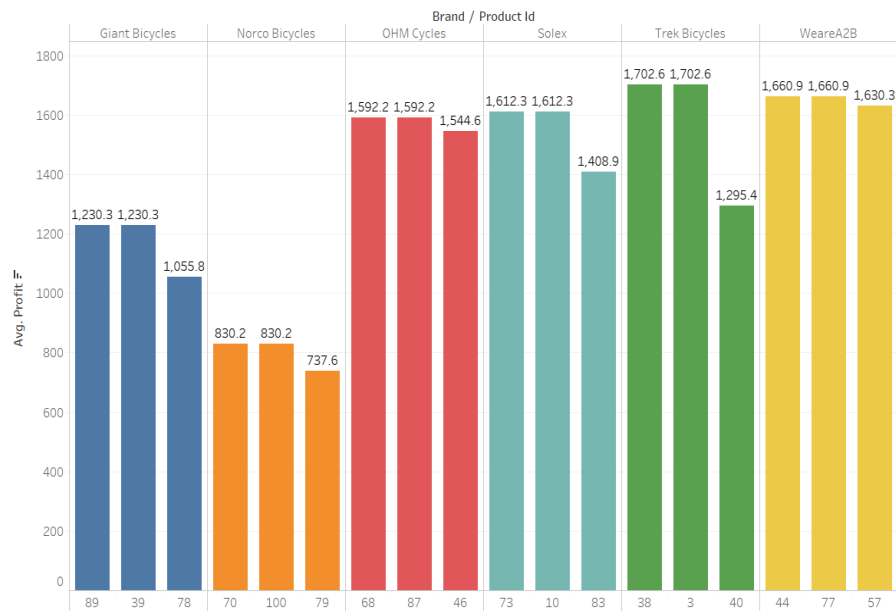
Average Property Valuation by States



# Data Exploration

## Transaction Analysis

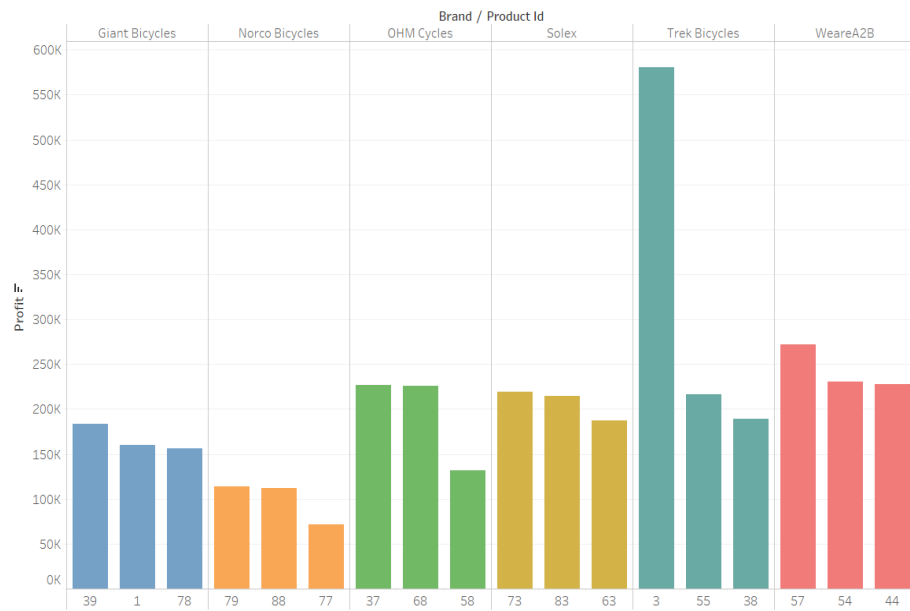
Top 3 Highest Average Profit Product ID in Each Brand



## Insights:

- In this two charts we can see the top 3 highest average and total profit product id in each brand. This can help us understand what kind of product can generate best value for us.

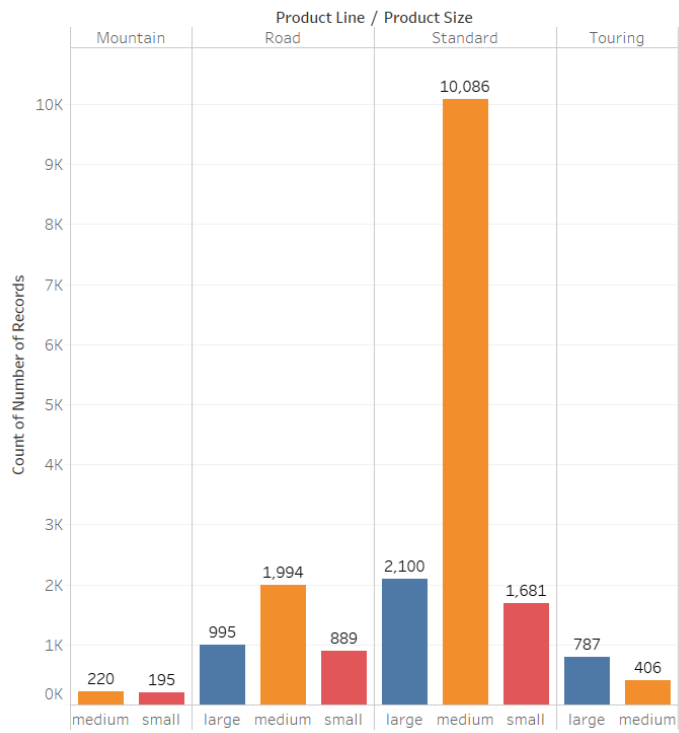
Top 3 Highest Total Profit Product ID in Each Brand



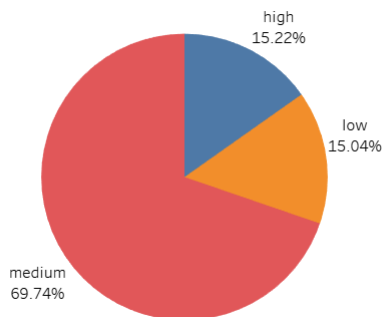
# Data Exploration

## Transaction Analysis

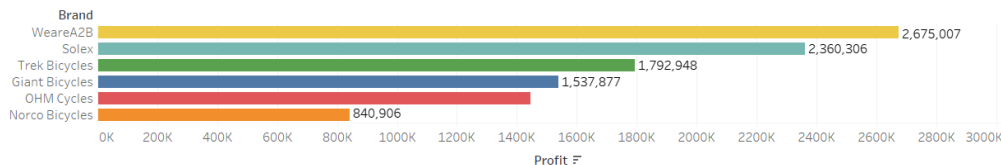
Number of Product Size in Each Product Line



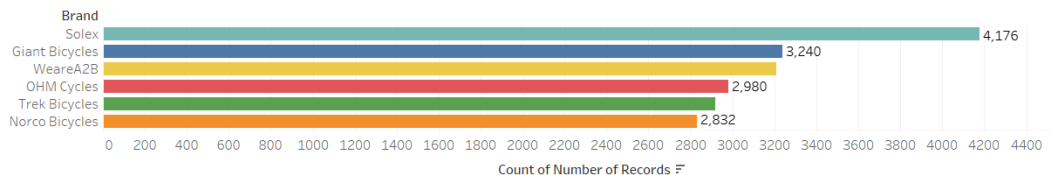
Percentage of Product Class



Total Profits by Brands



Total Number of Transactions by Brands



## Insights

- From first graph, we can see that Standard has the highest transaction number in product line. Medium is usually the most popular size, while in Touring type large size is more welcome than medium.
- Medium holds more than half in Product class
- From the last two graph, we can see that Solex holds the highest number of transaction but WeareA2B has the highest sum profits. We can say WeareA2B might has higher unit price.



# Model Development

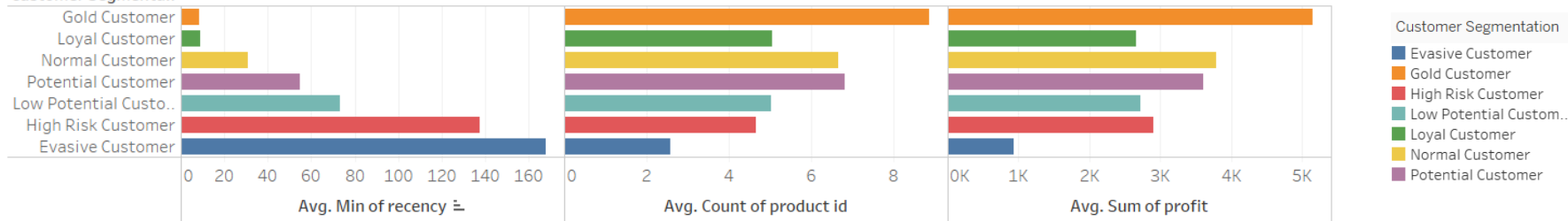
## RFM Analysis & Customer Segmentation

Score Range from 1~4	Weight
R_socre (Recency)	100
F_score (Frequency)	10
M_score (Monetary)	1

- RFM is a method used for analyzing customer value.
- RFM stands for Recency, Frequency, and Monetary value. These RFM metrics are important indicators of a customer's behavior and can display different levels of business engagement.
- We have segmented 7 different customer groups based on different level of RFM metrics.
- As we can see, Gold Customer, which is our top customer group, has the lowest Min of recency and highest Count of product id and Sum of profit. On the contrary, Evasive Customer group , which is the worst one, has the totally opposite metrics.

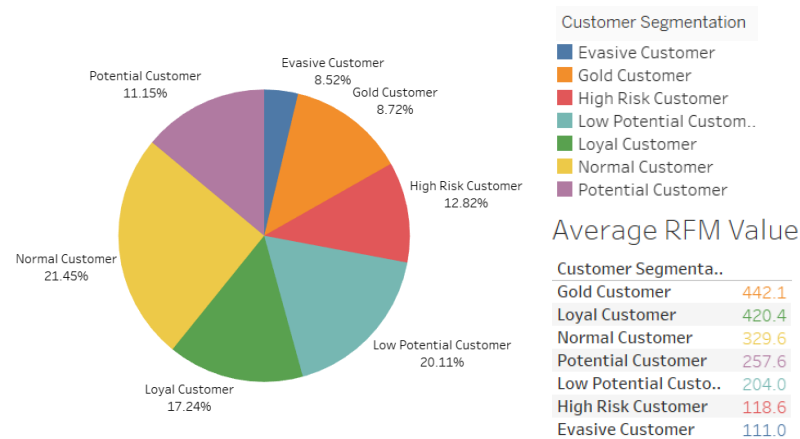
### RFM Analysis

#### Customer Segmenta..



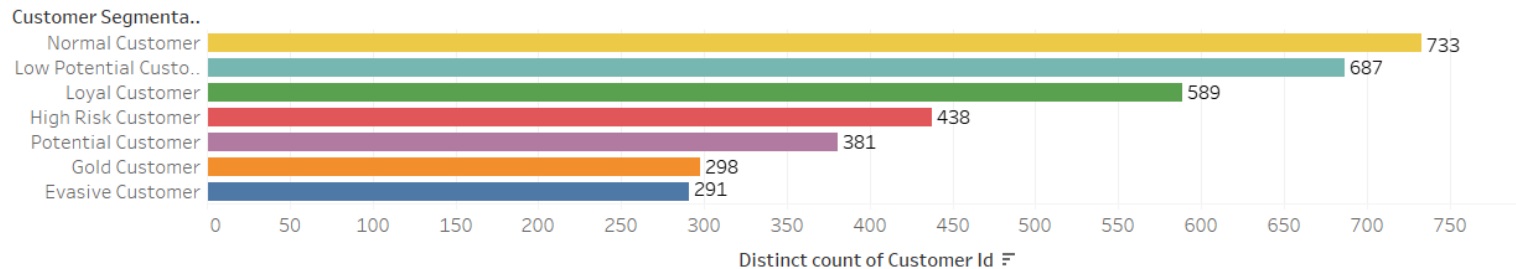
# Model Development

## RFM Analysis & Customer Segmentation



- Normal customer group has highest proportion, while Gold and Evasive have the two lowest proportion.
- RFM range from 444 to 111, Gold Customer has the highest average 442.1 RFM value, Evasive Customer has the lowest one with 111 only.

### Numbers of Customer Segmentation



# Interpretation

## Summary Table of Top 1000 Customer to target

Customer Segmentation	Number of customers	Cumulative	Customer Selection
Gold Customer	298	298	298
Loyal Customer	589	887	702
Normal Customer	733	1620	0
Potential Customer	381	2001	0
Low Potential Customer	687	2688	0
High Risk Customer	438	3126	0
Evasive Customer	291	3417	0

- Those top 1000 customers have all bought recently, more frequent and tend to spend more than others.

