

TheAnalyticsTeam

# Sprocket Central Pty Ltd

Data analytics approach

Yini Lai

# Agenda

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

# Introduction

## Identify top 1000 targeted customers

### Introduction & Tasks

#### Introduction:

Sprocket Central Pty Ltd is a company that specialized in high-quality bikes and accessible cycling accessories to riders.

#### Tasks:

The aim of this project is to analyze the existing customer's data to gain some customer insights. Thus, determining customer trends and behavior to help optimize resource allocation for targeted marketing.

### Data Analysis Process

- Distribution of job industry category
- Distribution of age by wealth segment
- Distribution of customers in different state by car owning
- Distribution of past 3 years related purchased by gender
- RFM Analysis

# Data Exploration

## Data Quality Assessment

### Data Quality Overview

Tables & Issues	Accuracy	Completeness	Consistency	Currency	Relevancy	Validity
Transactions	✓	✓			✓	✓
CustomerDemographic	✓	✓	✓	✓	✓	✓
CustomerAddress		✓	✓			

Notes:

- Accuracy:** Correct Values

**Completeness:** Data Fields with Values

**Consistency:** Value Free from Contradiction
- Currency:** Values up to Date

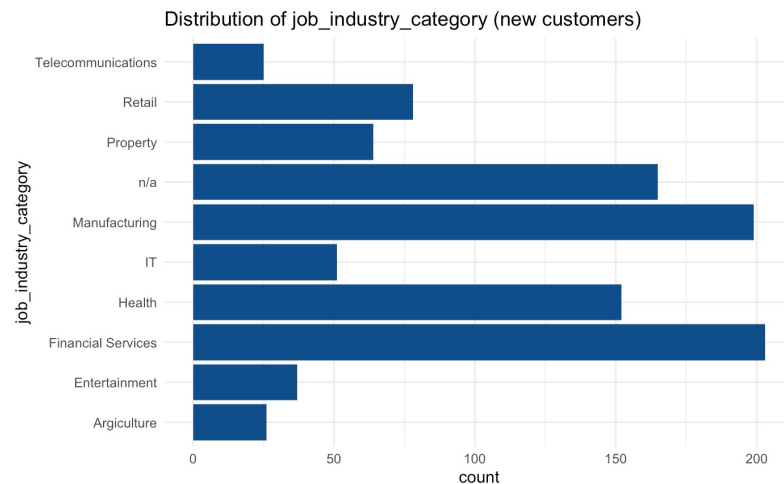
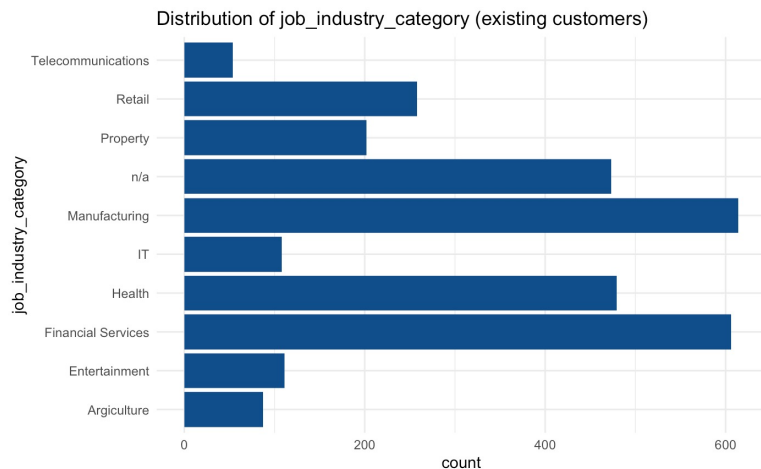
**Relevancy:** Data Items with Value meta-data

**Validity:** Data Containing Allowable Values

Details have been included in the email

# Data Exploration

## Distribution of Job Industry Category

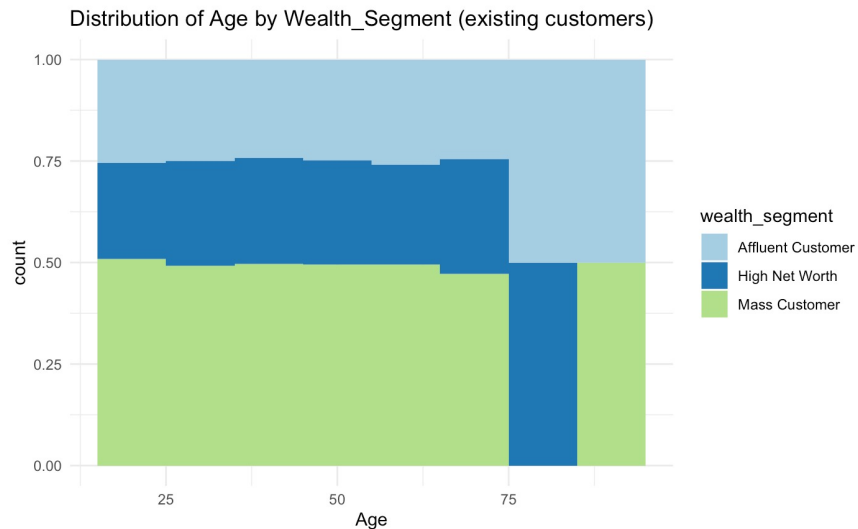


- Most of the customers from the existing customer list are from manufacturing, financial services and health industry, with around 620, 610 and 480, respectively.
- There are just few customers working in telecommunications, agriculture, entertainment and IT industry.

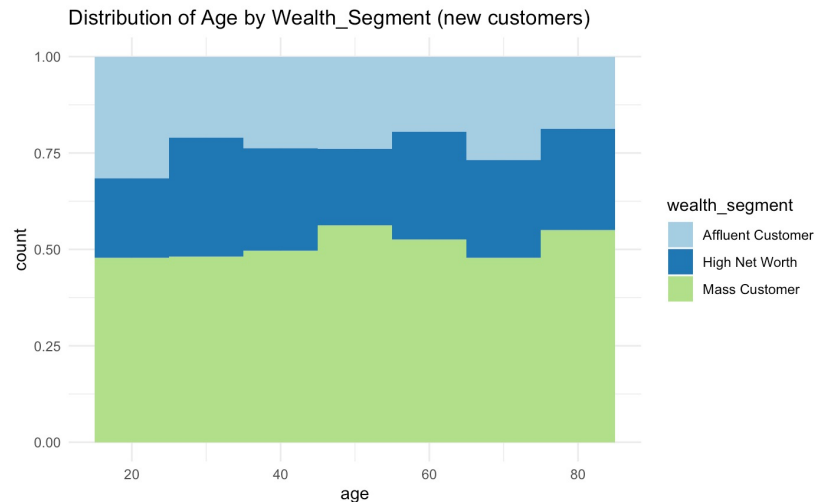
Similar distribution can be seen from the potential customer list.

# Data Exploration

## Distribution of Age by Wealth Segment



- Except customers aged between 75 and 87.5 are high net worth, nearly 50% of the existing customers are mass customer.
- For customers aged between 75 and 87.5, half of them is high net worth, and the rest are affluent customer.

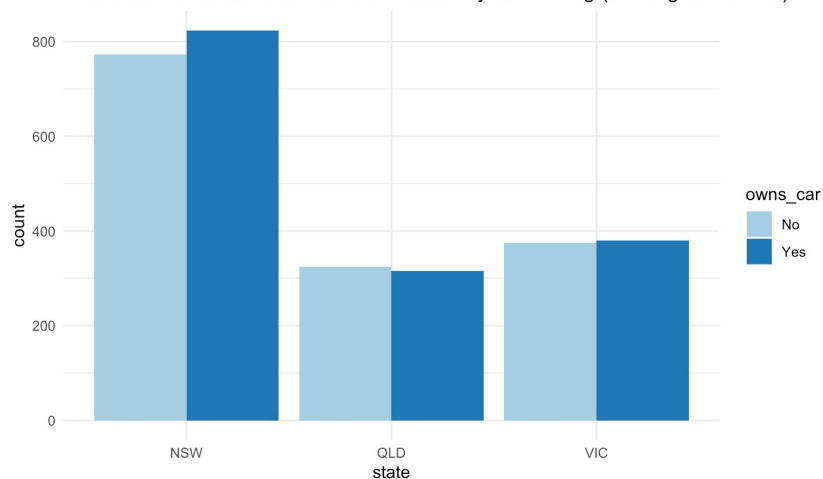


- Similar to the existing customers, nearly half of the customers in all age groups is mass customer.
- For customers aged around 20 and aged around 70, affluent customer outperforms High net customer.

# Data Exploration

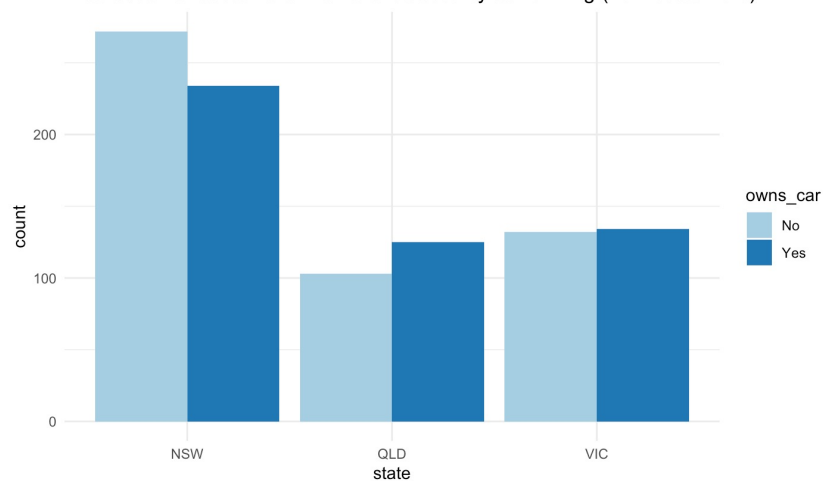
## Distribution of Customers in different states by Car Owning

Distribution of customers in different states by car owning (existing customers)



- Most of the existing customers are from New South Wales. For the existing customers, majority of the NSW customer owns car.
- For the existing customers from QLD and VIC, the number of the customers own car and not distributed quite evenly.

Distribution of customers in different states by car owning (new customers)

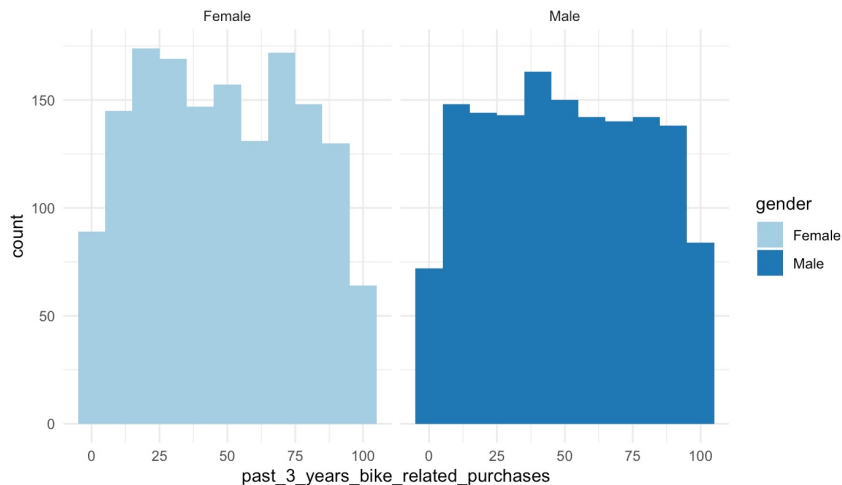


- Most of the new customers from NSW do not own a car, which is a bit different from the existing customers.
- Majority of the customers from QLD own car. The number of the customers own car and not distributed quite evenly in VIC.

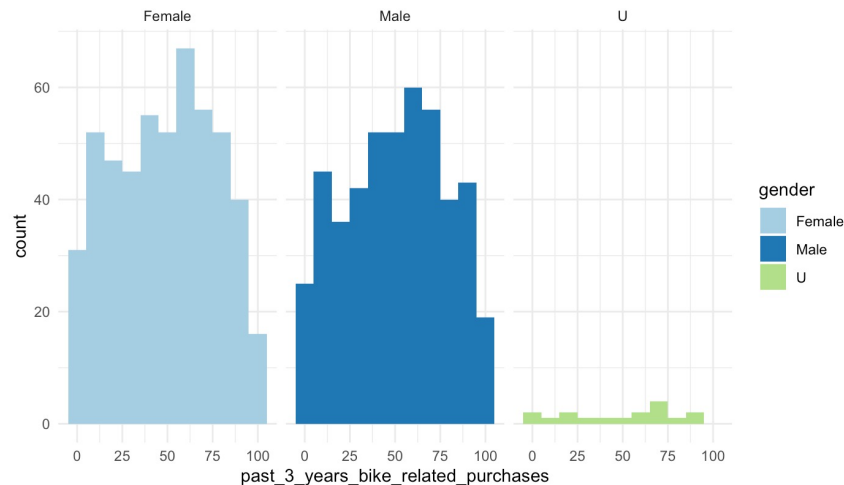
# Data Exploration

## Distribution of past 3 years related purchased by Gender

Distribution of past\_3\_years\_bike\_related\_purchases by Gender (existing)



Distribution of past\_3\_years\_bike\_related\_purchases by Gender (new)



- From the graph, we can see that in the past 3 years, female purchased a little bit more than male.
- In terms of the number of bikes purchased, male had more even distribution than female.
- For the new customers, the pattern that the past 3 years bike related purchases distribution is quite similar between male and female.



# Model Development

## RFM Analysis

### Introduction:

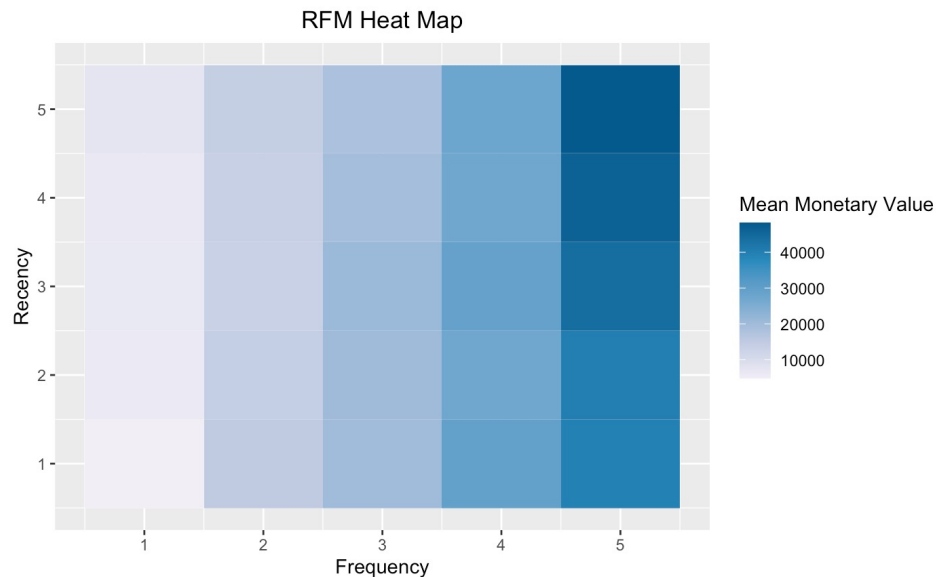
RFM (recency, frequency, monetary) is a method used for analyzing customer value. It measured three dimensions:

- **Recency:** How recently did the customer purchase?
- **Frequency:** How often do they purchase?
- **Monetary Value:** How many do they spend?

We assigned scores (R\_score, F\_score, M\_score) to the customers based on their behaviors in the past. These scores were computed by binning the values in each variable into different categories.

Finally, we calculated the RFM\_score based on the following formula:

$$R\_score * 100 + F\_score * 10 + M\_score$$

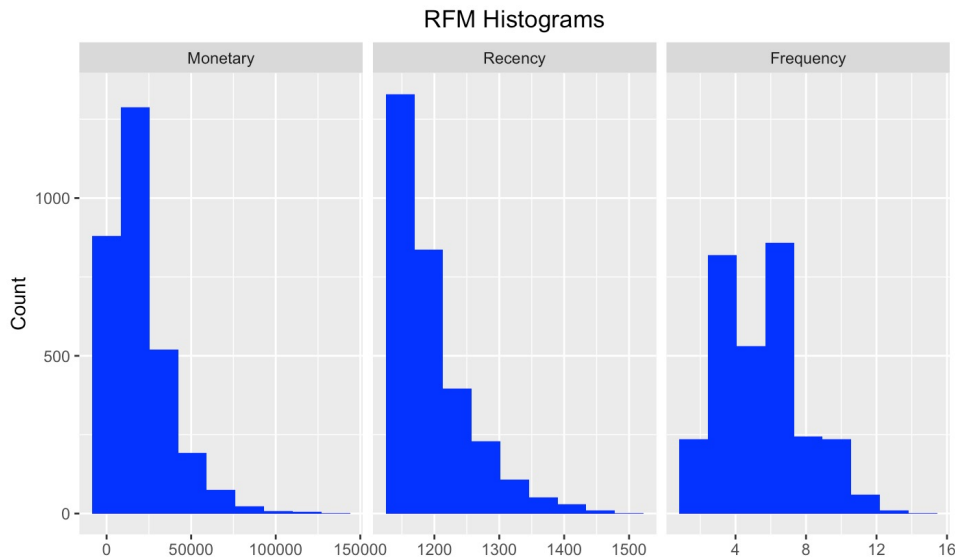


From the heat map, we can see that customer who purchased more frequently and purchased more recently are those who spend the most in purchasing products from Sprocket Central.

# Model Development

## RFM Analysis

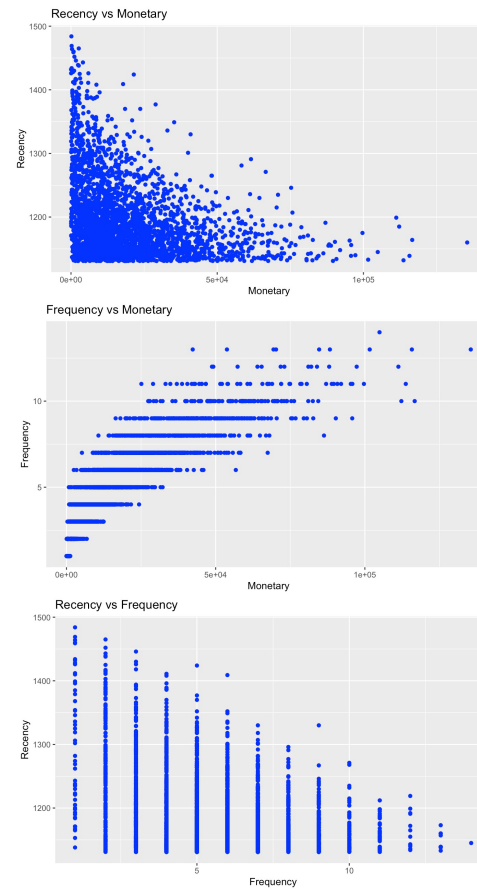
- The graph on the right is the distributions of the values of the three dimensions.
- Most of the customers spend less than 25000.
- Majority of the customers purchased products within recent 1200 days.
- Most of the customers had purchased in Sprocket Central for 3 to 7 times.



# Model Development

## RFM Analysis

- Customers who purchased more recently tend to bring more revenue to the company. Those who purchased 1300 ago spent less than \$25000.
- Customers who purchased more frequently tend to bring more revenue to the company. Those who visited less than 10 times spent less than \$50000.
- Customers who purchased more recently tend to purchase in higher frequency. Those who visited 1300 days ago purchased less than 5 times.



# Model Development

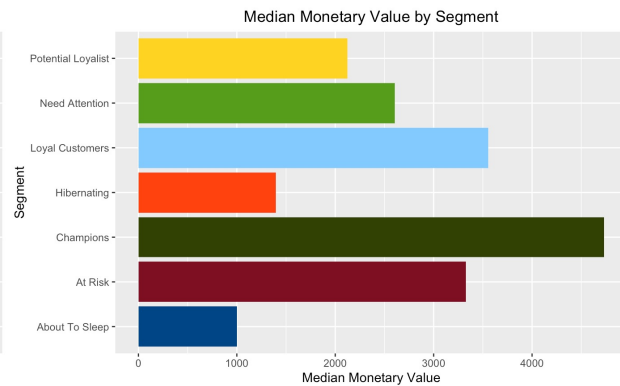
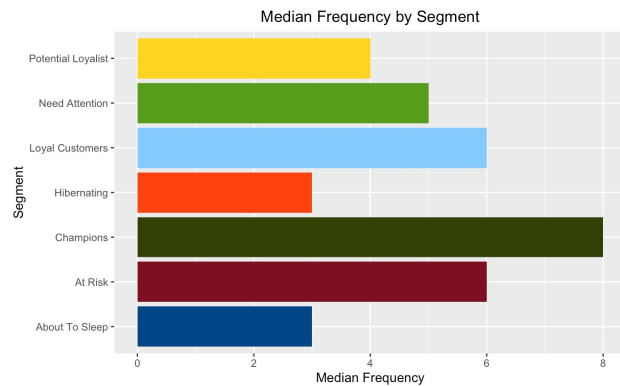
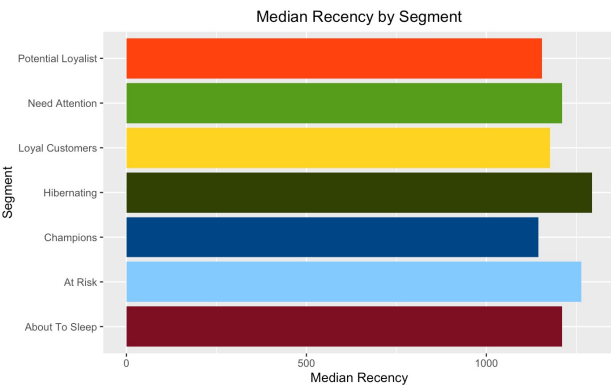
## RFM Analysis

Classify the customers based on their R\_score, F\_score and M\_score.

Segment	Description	R	F	M
Champions	Bought recently, buy often and spend the most	4 - 5	4 - 5	4 - 5
Loyal Customers	Spend good money. Responsive to promotions	3 - 5	3 - 5	2 - 5
Potential Loyalist	Recent customers, spent good amount, bought more than once	3 - 5	1 - 3	1 - 4
New Customers	Bought more recently, but not often	4 - 5	<= 1	<= 1
Promising	Recent shoppers, but haven't spent much	3 - 4	<= 1	<= 1
Need Attention	Above average recency, frequency & monetary values	2 - 3	1 - 3	2 - 4
About To Sleep	Below average recency, frequency & monetary values	2 - 3	<= 2	<= 2
At Risk	Spent big money, purchased often but long time ago	<= 2	2 - 5	1 - 5
Can't Lose Them	Made big purchases and often, but long time ago	<= 1	4 - 5	4 - 5
Hibernating	Low spenders, low frequency, purchased long time ago	1 - 2	1 - 3	1 - 3
Lost	Lowest recency, frequency & monetary scores	<= 2	<= 2	<= 2

# Model Development

## RFM Analysis – customer segmentations



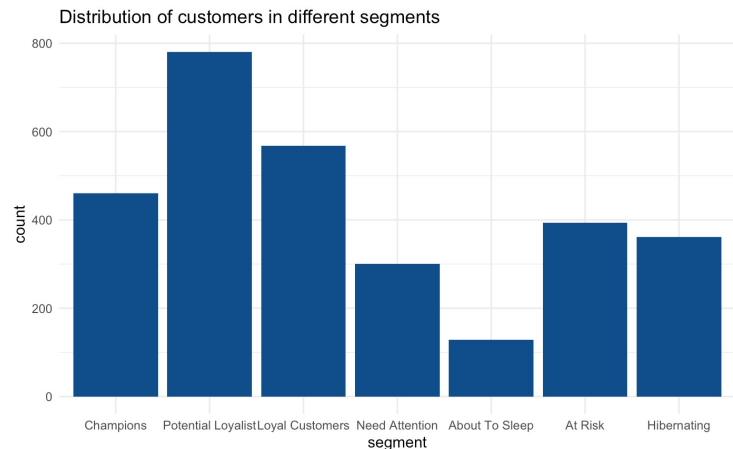
- The median recencies are quite similar between different segments. It is no doubt that “Champions” has the smallest recency values. Also, “Hibernating” and “At Risk” have the largest recency values.

- The median frequency values vary between different segments. Among them, it is obvious that “Champions” has the highest frequency, and “About To Sleep” and “Hibernating” have relatively small frequency.

The median monetary values vary between different segments as well. Among them, “Champions” has the highest amount of median monetary value. “About To Sleep” has relatively smaller monetary value.

# Interpretation

## Summary the top 1000 targeted customers



segment	n
<chr>	<int>
About To Sleep	128
At Risk	393
Champions	461
Hibernating	361
Loyal Customers	568
Need Attention	300
Potential Loyalist	781

7 rows

Based on the segmentation standard showing before and the distribution of the segments showing above, we should pay more attentions to the “Champions” and “Loyal Customers” groups. This means that we should pay more attention to the customers who purchase recently, frequently, tend to spend more money and respond to promotion.