



# Sprocket Central Pty Ltd

Data Analytics Approach

Amish Verma

Data Analytics Consulting Virtual Intern

# Agenda

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

# Introduction

## Data Quality Assessment

	Customer Demographic	Customer Addresses	Transaction Data
Accuracy	DOB: Inaccuracy Job Industry Category: Misspelling		
Completeness	DOB: Blanks Job Title: Blanks Job Industry Category: Blanks Tenure: Blanks Customer IDs: Not in Sync	Customer IDs: Not in Sync	Standard Cost: Blanks Brand: Blanks Product Line: Blanks Product Class: Blanks Product Size: Blanks Product First Sold Date: Blanks Customer IDs: Not in Sync
Consistency	Gender: Inconsistency	States: Inconsistency	
Currency	Deceased Customers: Filter Out		
Relevancy	Default: Exclude Field		Order Status: Exclude Cancelled
Validity			Product First Sold Date: Format

# Introduction

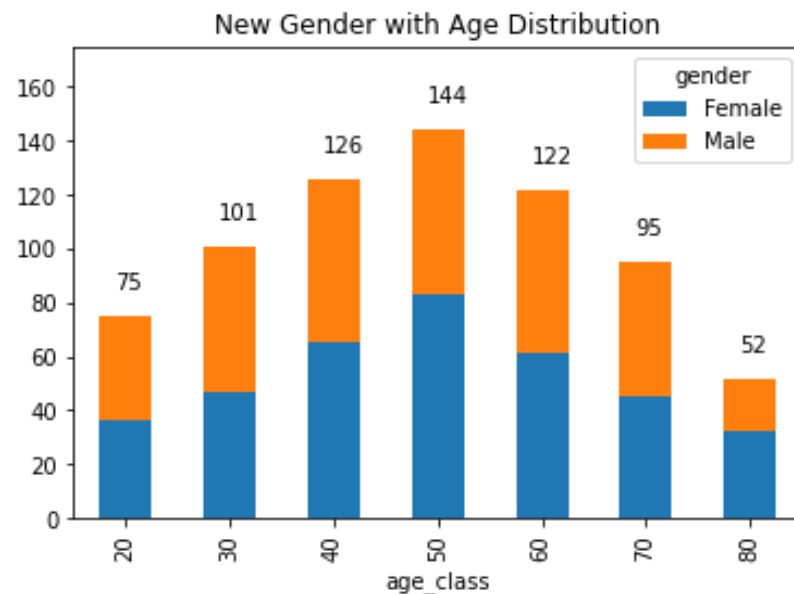
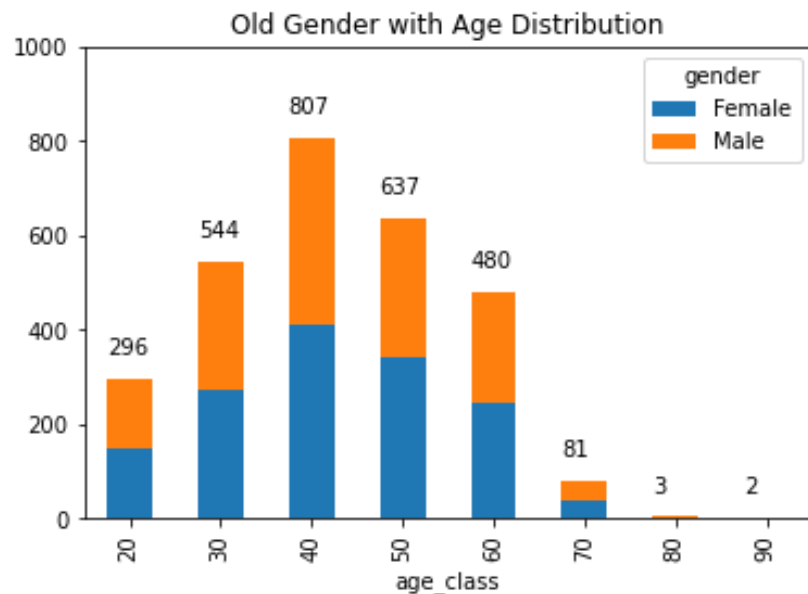
## Data Cleaning

- Records with Missing Fields were Dropped.
- Join Keys between Tables were considered and conflicting Records were Dropped.
- Age, Last Purchase (Days Ago) and Profit Fields were Added.
- Records pertaining to Deceased Customers were Dropped.
- Transactions more than a year old were Dropped.

Distinct Customer IDs As Received	4,000
Distinct Customer IDs after Data Cleaning	3,492

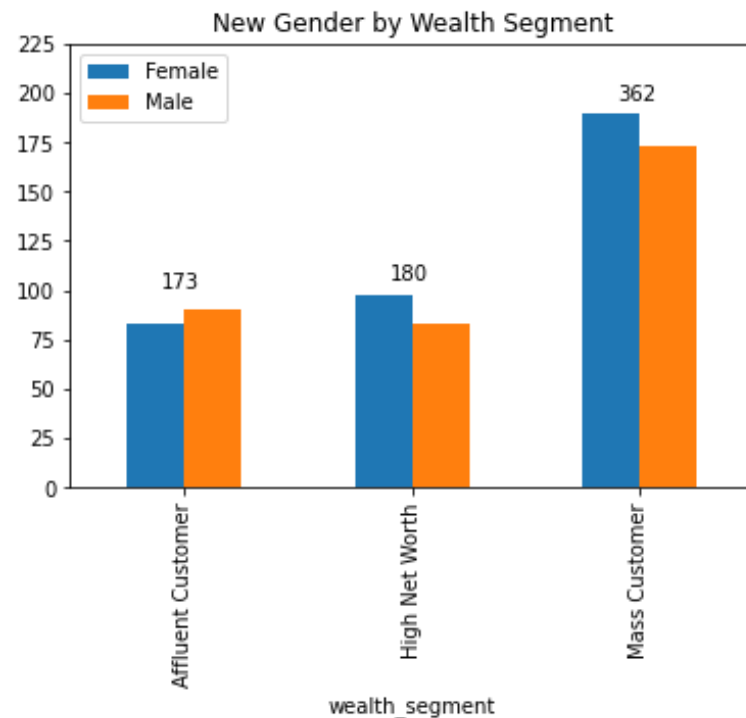
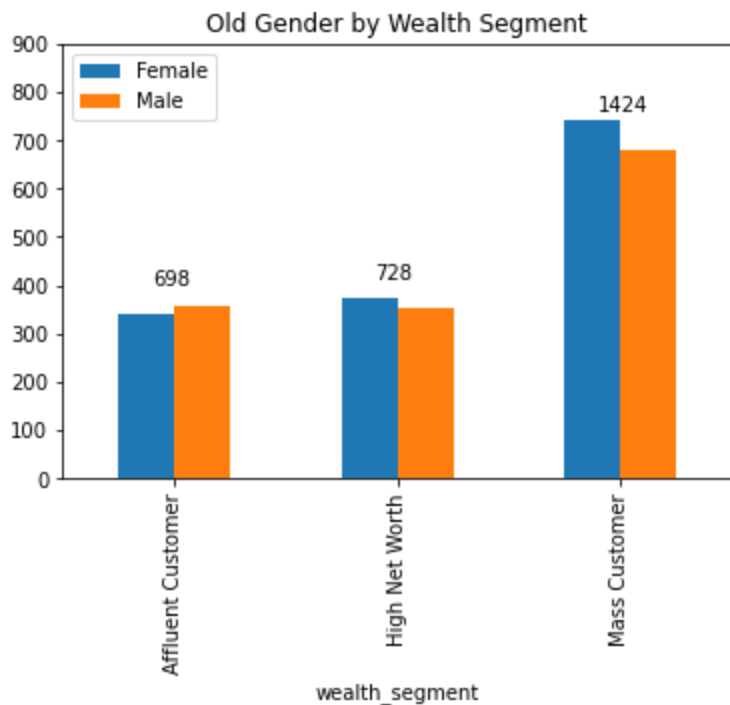
# Data Exploration

## Gender with Age Distribution



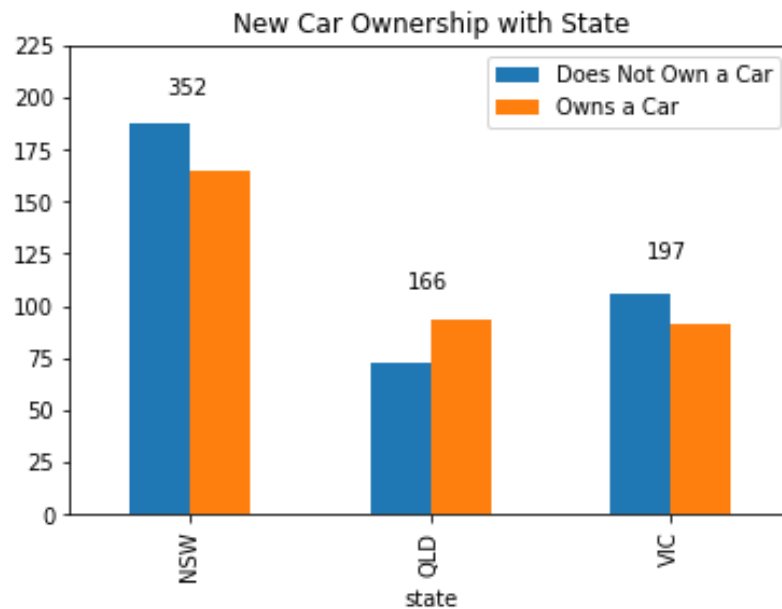
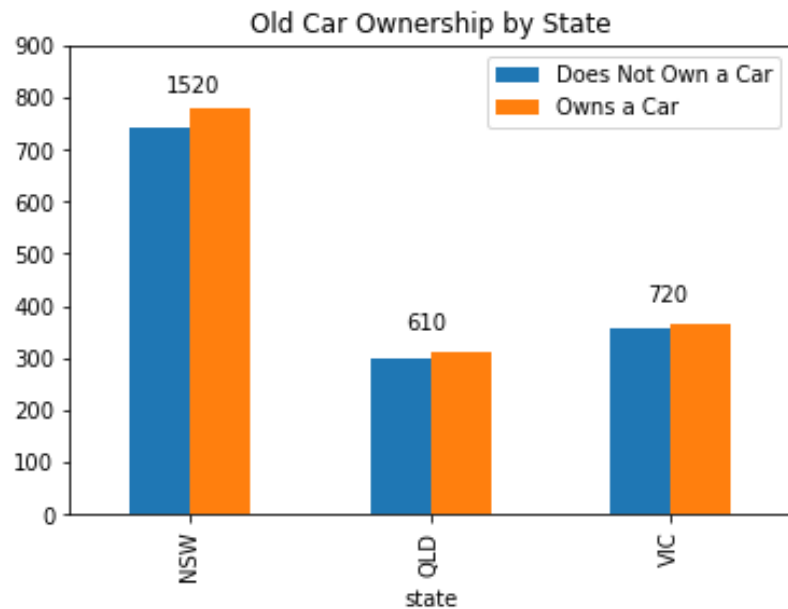
# Data Exploration

## Gender with Wealth Segment



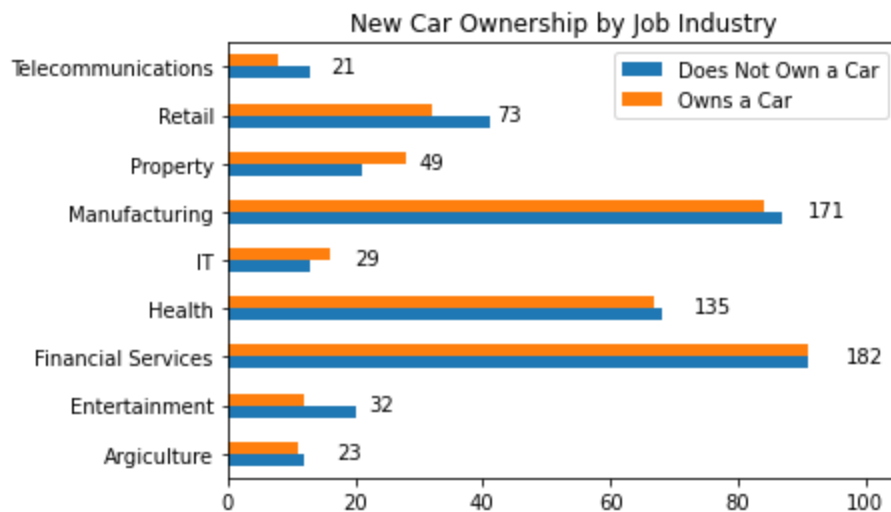
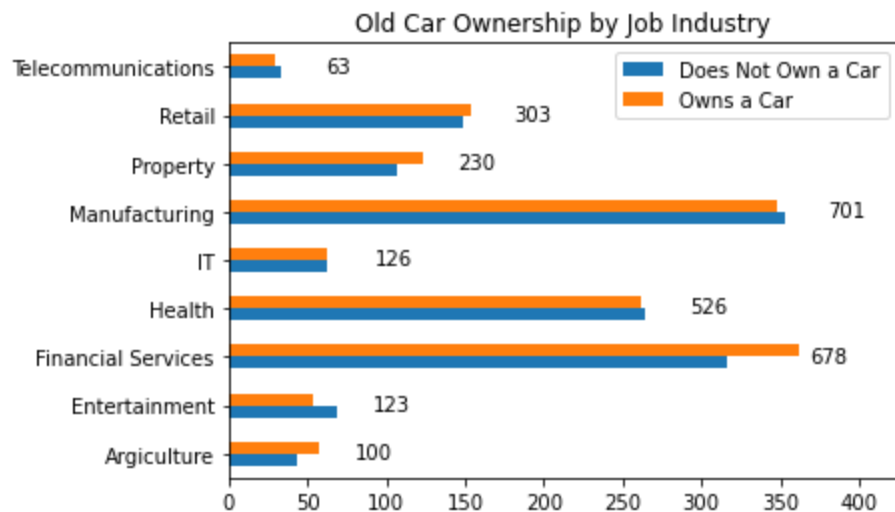
# Data Exploration

## Car Ownership with State



# Data Exploration

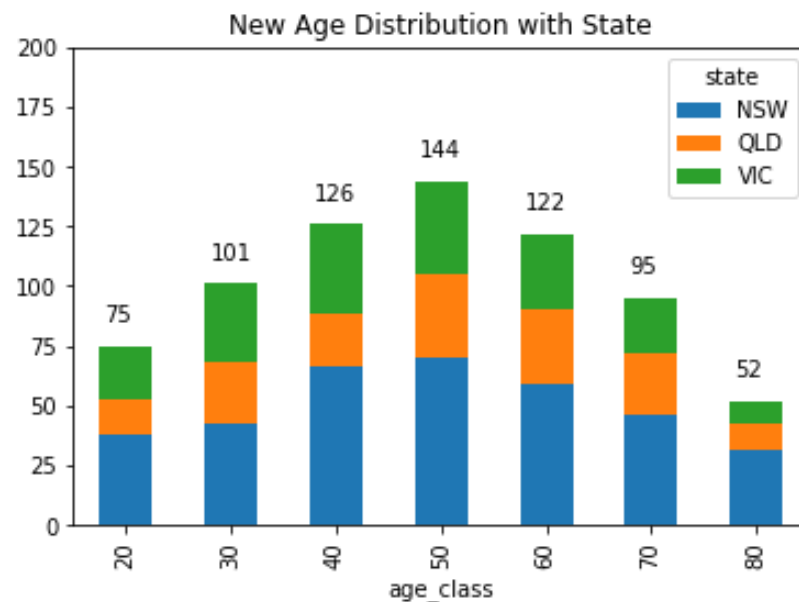
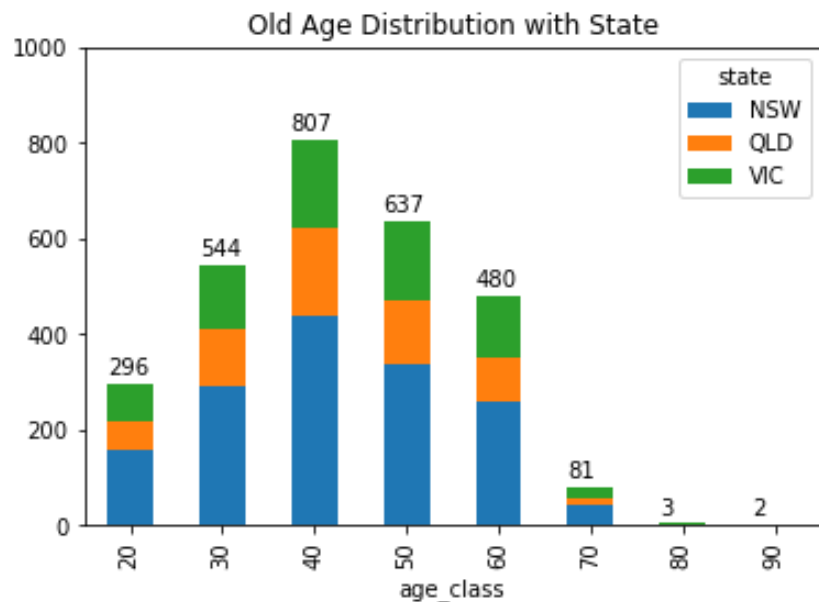
## Car Ownership with Job Industry





# Data Exploration

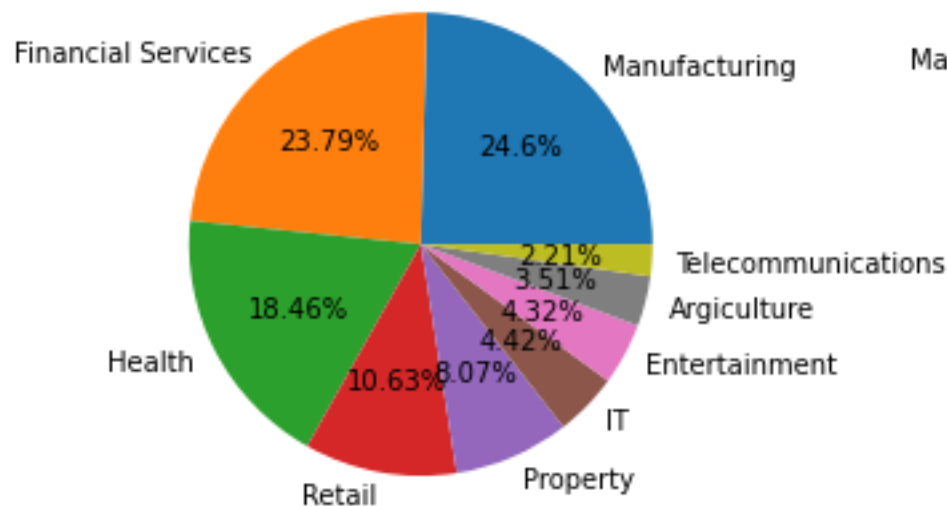
## Age Distribution with State



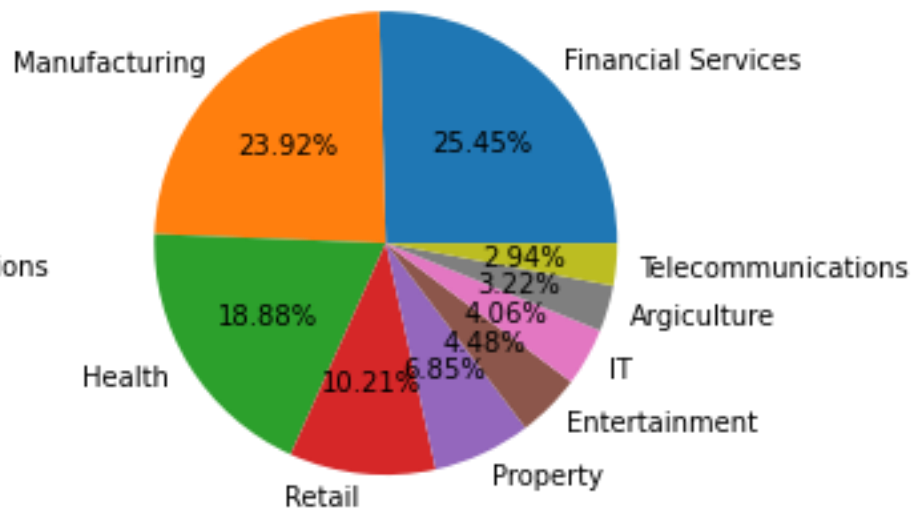
# Data Exploration

## Job Industry Diversity

Old Job Industries

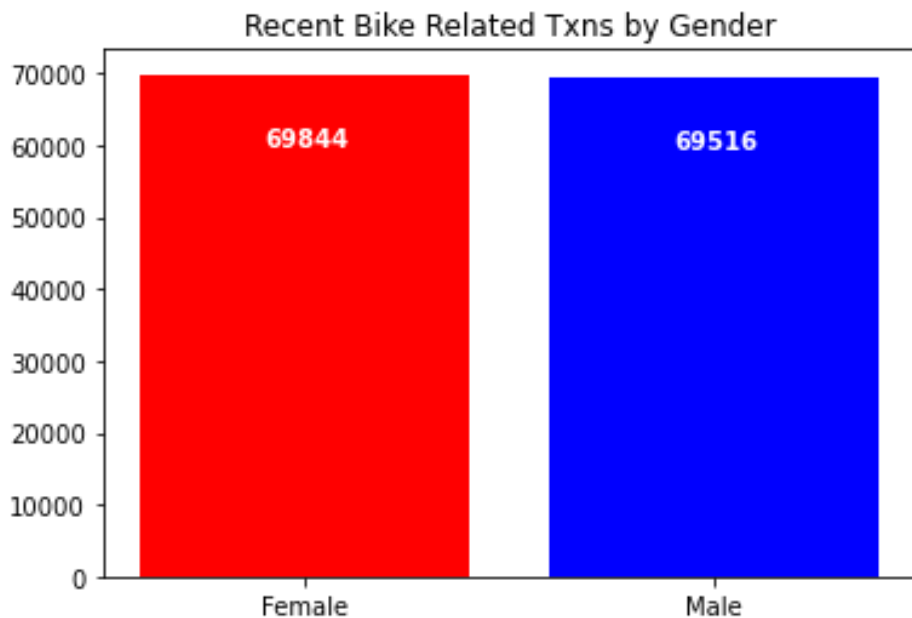


New Job Industries



# Data Exploration

## Recent Transactions



# Model Development

## RFM Analysis

- **Recency**

- The last day on which a customer performed a transaction was taken as the recency parameter.
- Customers were divided into 4 quartiles and given a R\_Score.

- **Frequency**

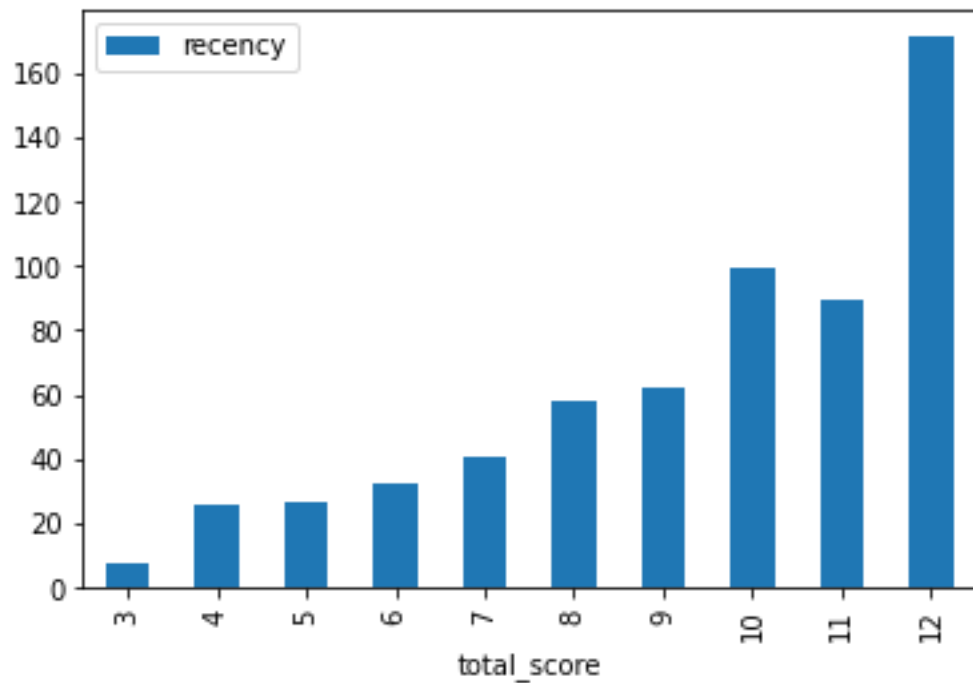
- The frequency of transactions done by a particular customer was taken as the frequency parameter.
- Customers were divided into 4 quartiles and given a F\_Score.

- **Monetary Value**

- The average profit per customer was taken as the monetary value parameter.
- Customers were divided into 4 quartiles and given a M\_Score.

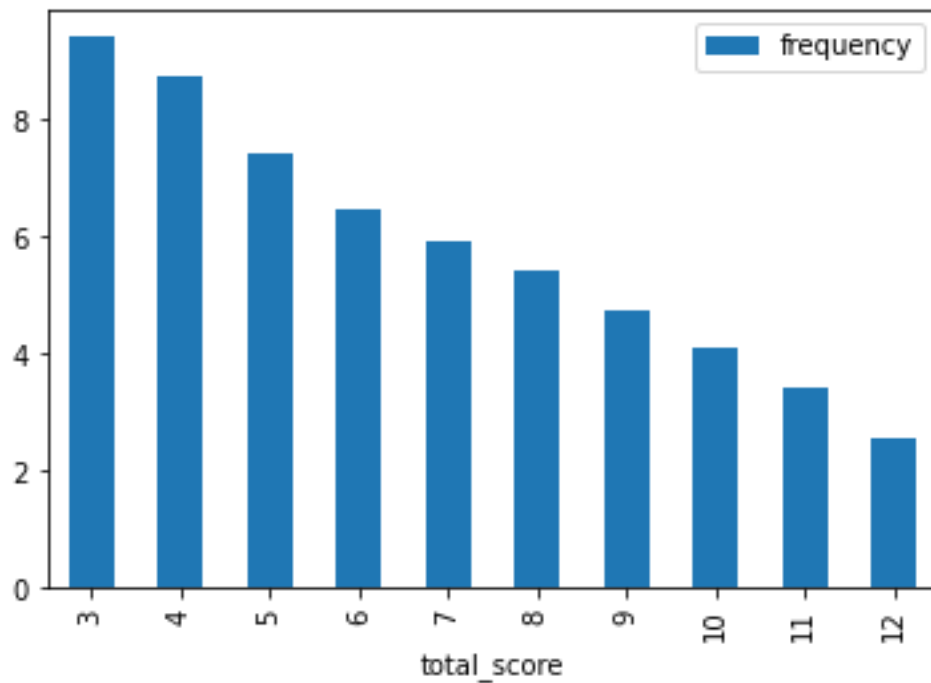
# Model Development

## RFM Analysis



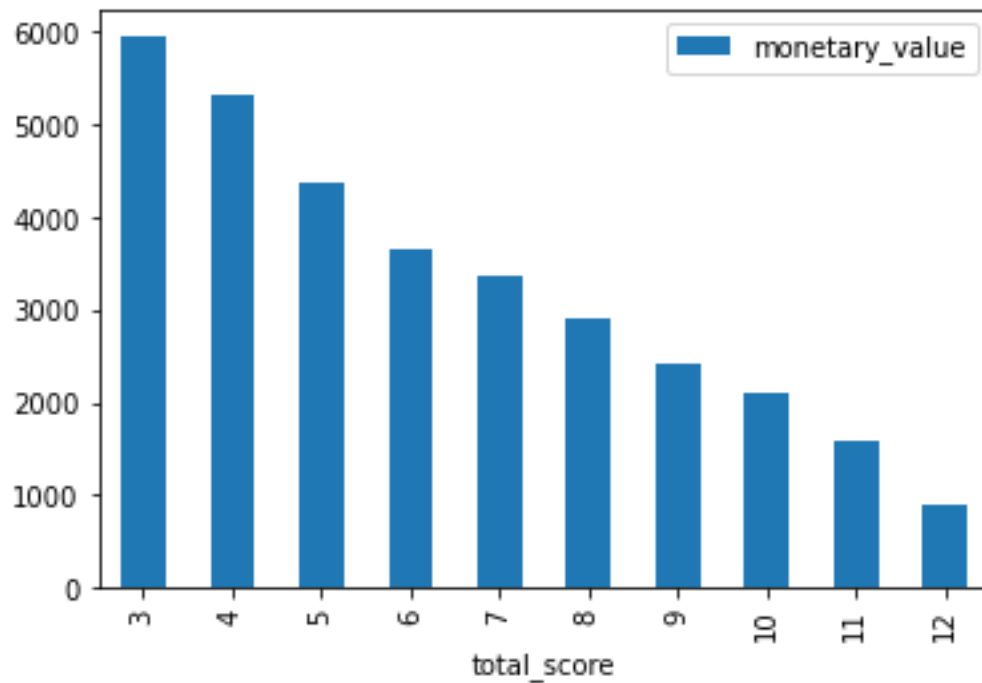
# Model Development

## RFM Analysis



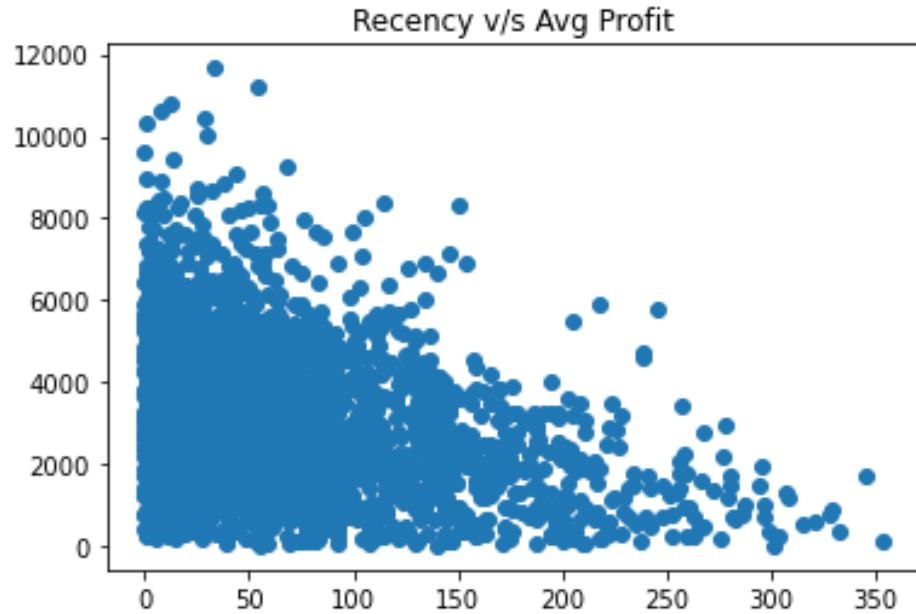
# Model Development

## RFM Analysis



# Model Development

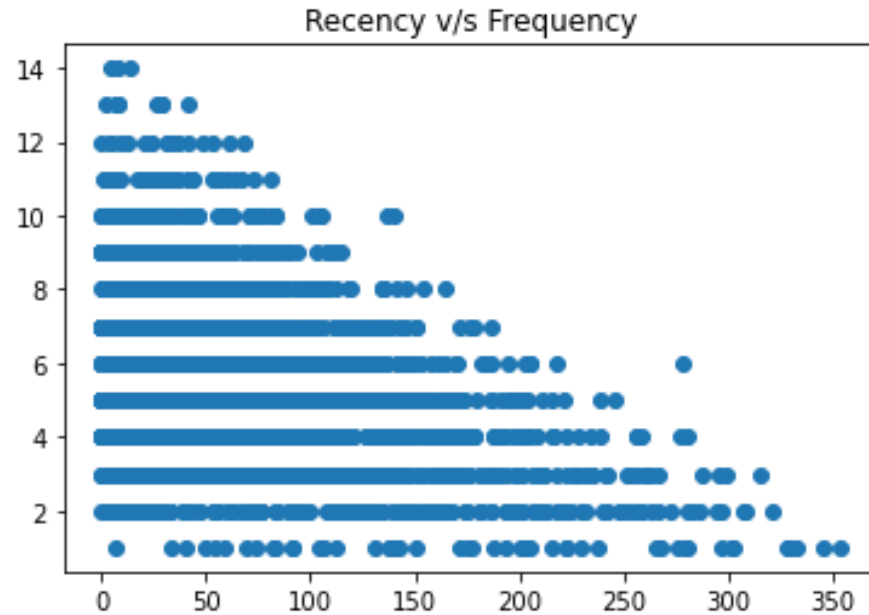
## RFM Analysis





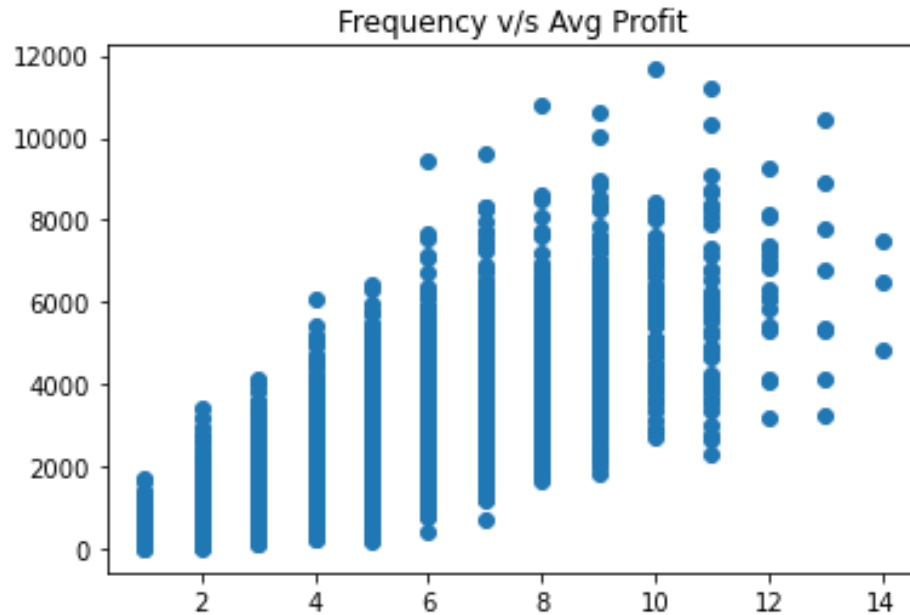
# Model Development

## RFM Analysis



# Model Development

## RFM Analysis

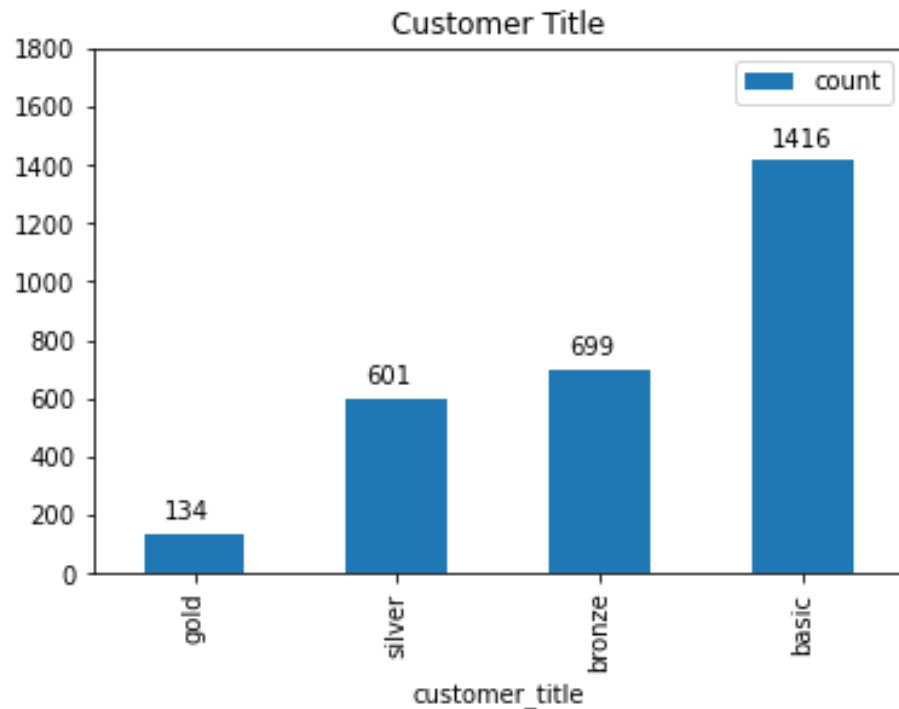


# Model Development

## RFM Analysis

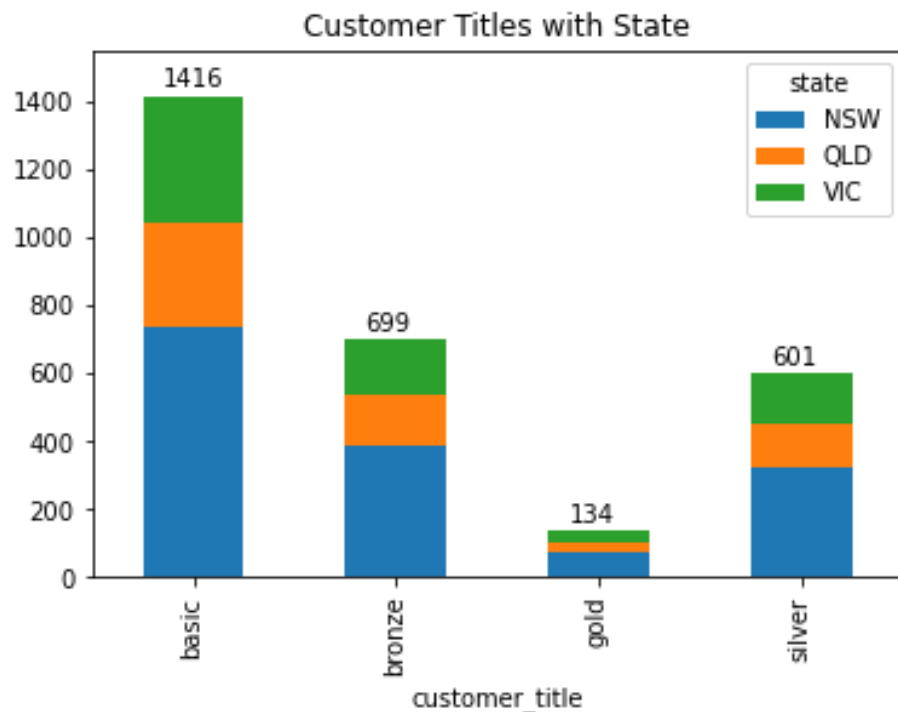
Based on the RFM Class, four customer tiers were identified:

1. Gold Class: These customers have recently made a purchase, are frequent and are most profitable.
2. Silver Class
3. Bronze Class
4. Basic Class: These customers have not made any recent purchase, are not frequent and do not contribute majorly.



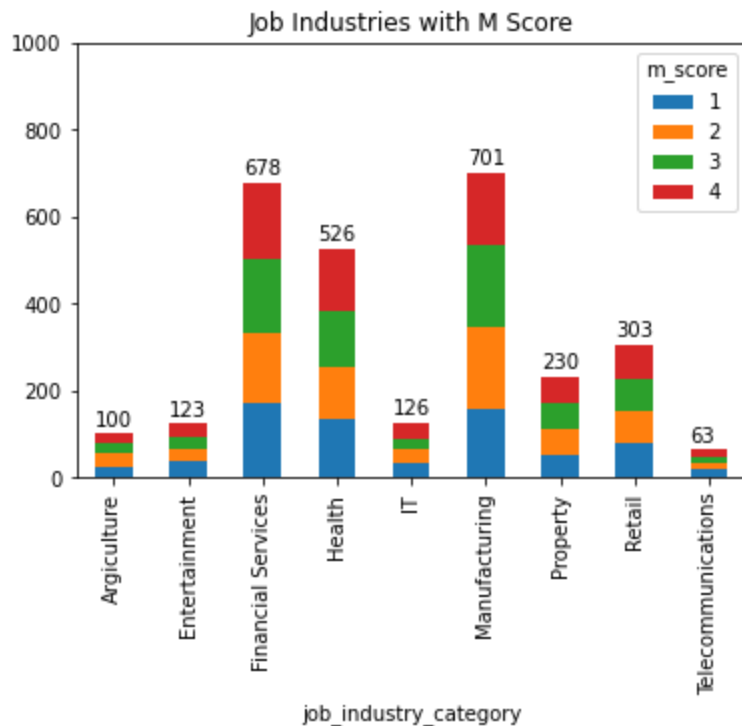
# Interpretation

## Customer Tiers by State



# Interpretation

## Profitable Job Industries



# Interpretation

## Customer Segments

Segment	RFM Score	Cumulative
Platinum	3	134
Very Loyal	4	296
Becoming Loyal	5	558
Recent	6	891
Average	7	1300
High Risk	8	1712
Evasive	9	2110
Losing	10	2417
Inactive	11	2595
Lost	12	2850

## Targeting Methodology

- Customers having high RFM Scores can be filtered and targeted to boost .
- The customers have made recent purchases, are frequent, and provide the most profits.