TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Jialu Wang

Agenda

1. Introduction
2. Data Exploration
3. Model Development: RFM Analysis
4. Interpretation

Introduction

Identify Top 1000 old Customer based on RFM analysis analyze their demographic features

Outline of Problem

- Sprocket Central Pty Ltd want to find high value customers from a list of 1000 potential customers
- The data about new customers don't have transaction history but with demographics and attributes
- Sprocket have a dataset with existing customers' transaction history with demographics and attributes

Analysis Approach

- Use dataset about existing customers to get insights about demographics and attributes of high value customers based on RFM analysis
- Encode the new customers' demographics and attributes to predict their value

Data Exploration

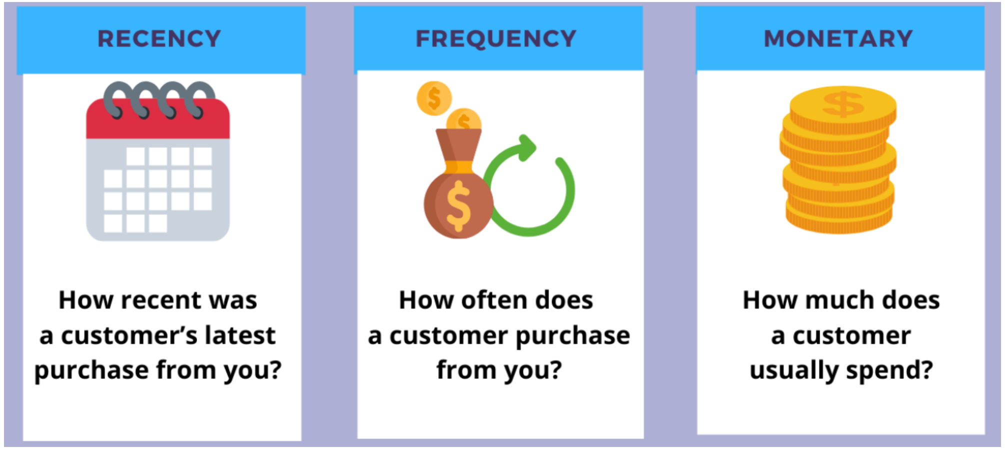
Data Quality Assessment and Data Cleaning

Consistency	• Inconsistent values for the same attribute • Inconsistent data type for the same attribute	• In CustomerDemographic, gender have 6 unique values, change F & Femal to Female, M to Male. Additionally, gender records where 'U' have been replaced based on the distribution from the training dataset. • In CustomerAddress, state has different rules, some are full spelt, some are abbrev, change all to abbrev. • Some column have wrong data types, such as product_first_sold_date and list_price in transactions, one is date not number, the other is currency not number. • Default column in CustomerDemographic has Mojibake, don't know what's the right decode rule;
Completeness	• Every sheet has missing values	• If only a small number of rows are empty, we will filter out the record entirely from the training set for prediction, like less 1% transactions have missing fields. But if it is core field or the missing rate is high (like 12.5% missing rate of job_title in CustomerDemographic, we will impute the missing fields based on distribution in the training dataset.
Currency	• customer_id's maximal in CustomerDemographic is 4000 , but three values of customer_id in Transactions is 5034	• ensure that all tables are from the same period. Only customers in CustomerDemographic will be used as a training set for our model. it's better to merge data in both CustomerDemographic and NewCustomerList into one sheet, adding customer_id by DOB.

Model Development

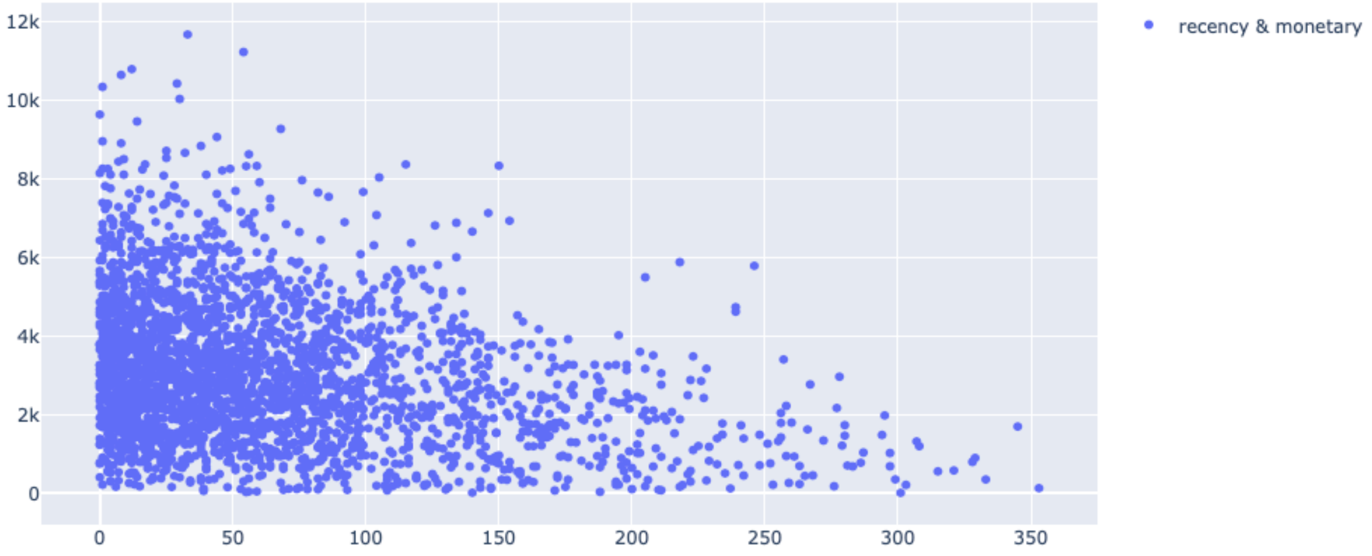
RFM Analysis and Customer Segmentation

- RFM stands for Recency, Frequency, and Monetary
- Often used for reactivation campaigns, high values customer programs etc.



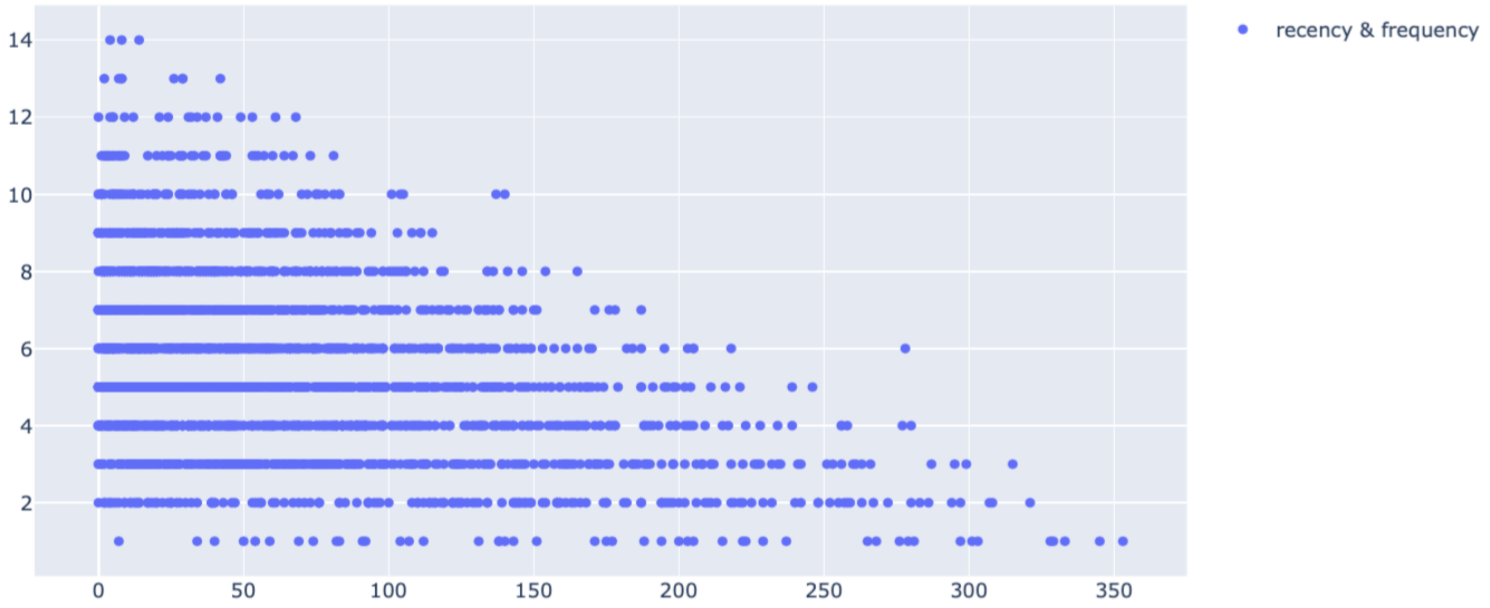
Model Development

Recency vs Monetary



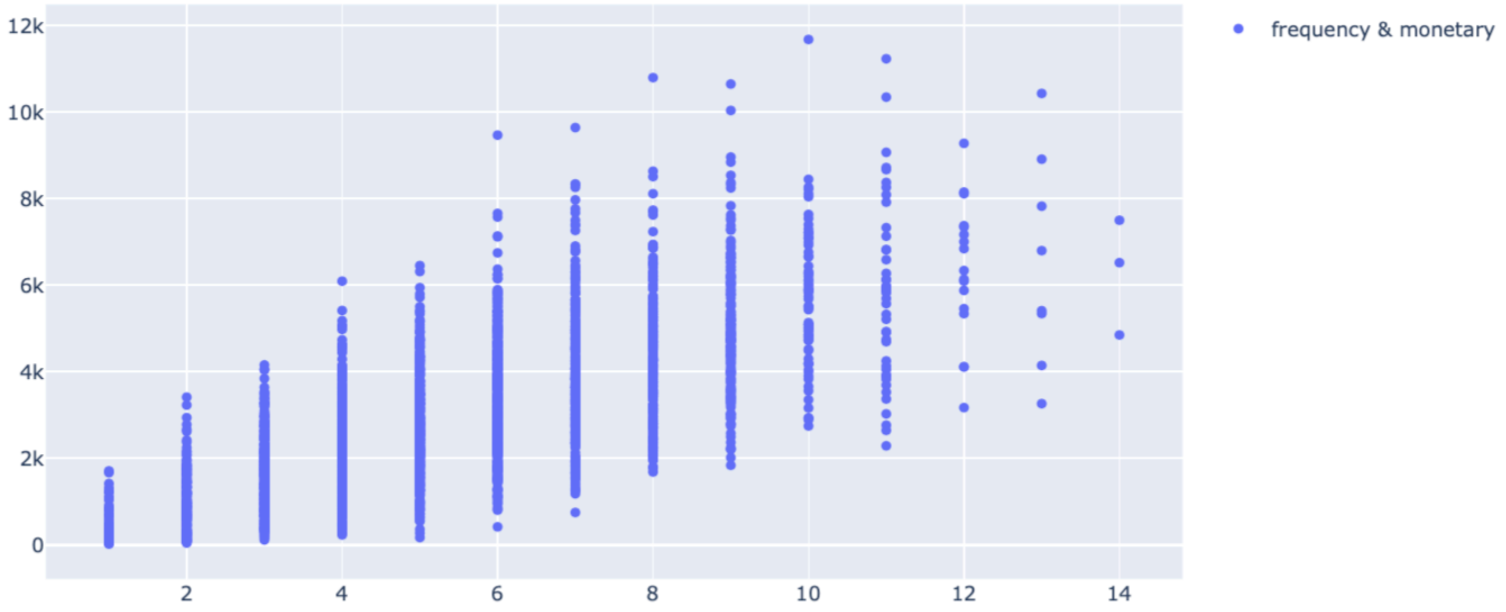
Model Development

Recency vs Frequency



Model Development

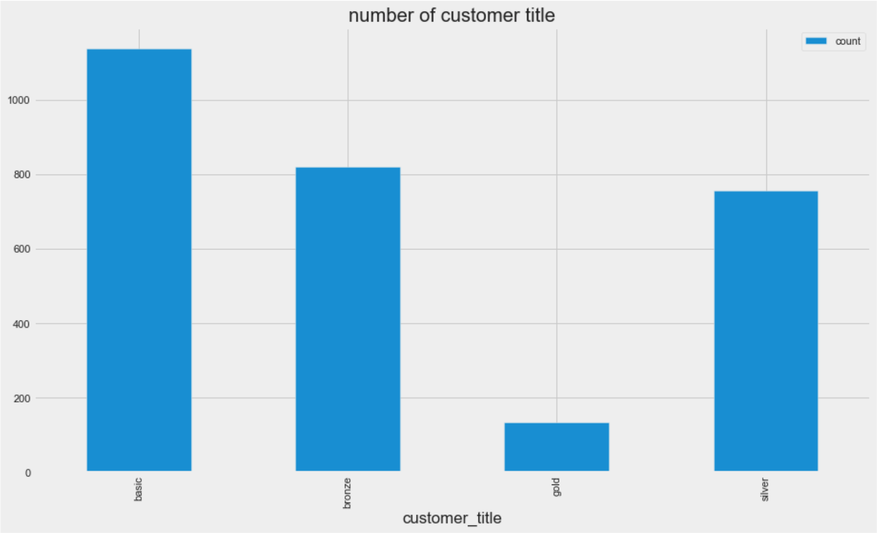
Frequency vs Monetary



Model Development

Existing Customer Segmentation

- The more recent a customer buy
- The monetary value a customer pay
- The more times a customer purchase
- The higher value a customer is

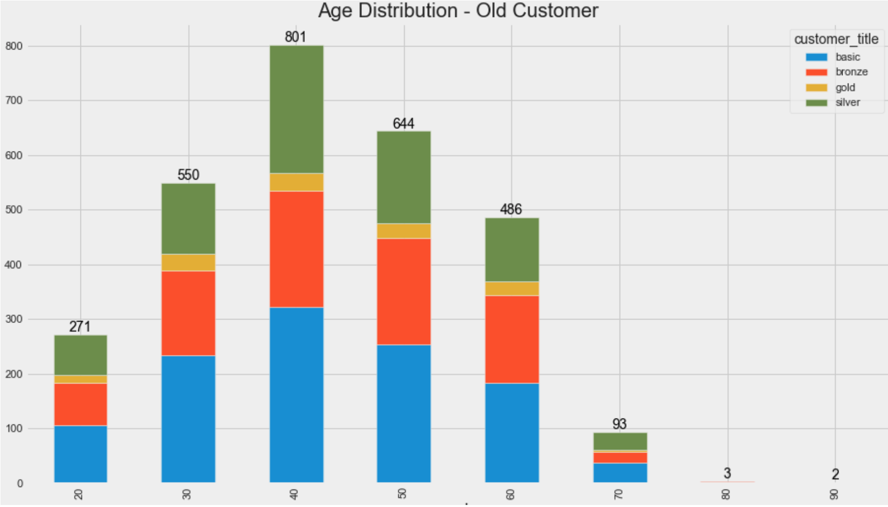


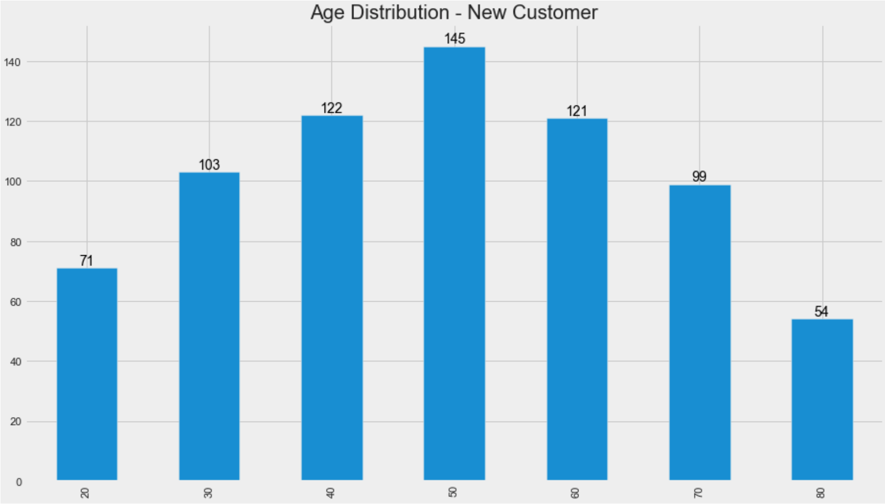
Gold: total score <= 3,
Silver: total score <= 6, 25%
Bronze: total score <= 8, 50%
Basic: total score > 8

RFM Class	Basic	Bronze	Gold	Silver
Number	1138	821	134	757

Interpretation

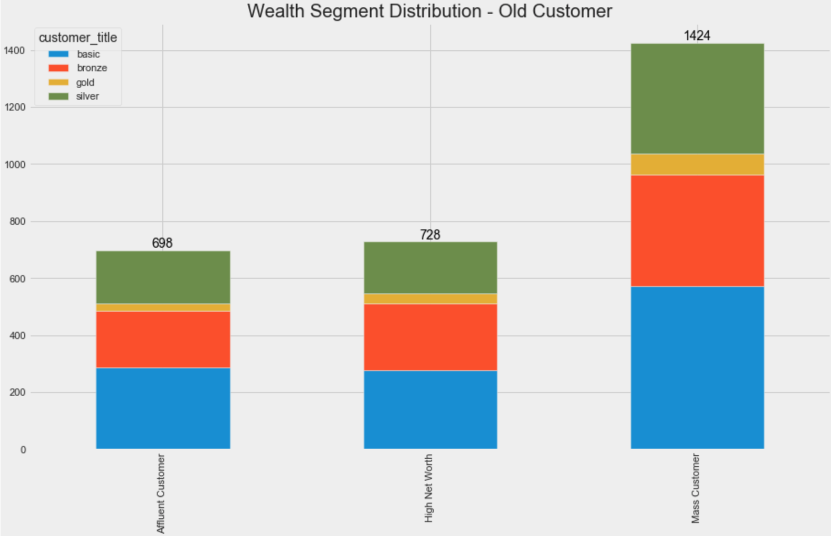
Age Distribution

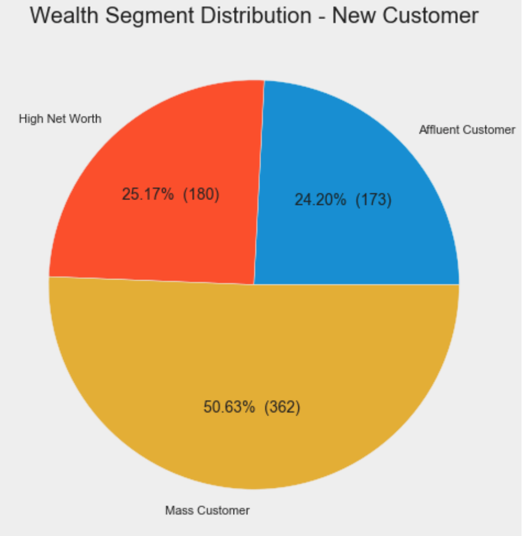




Interpretation

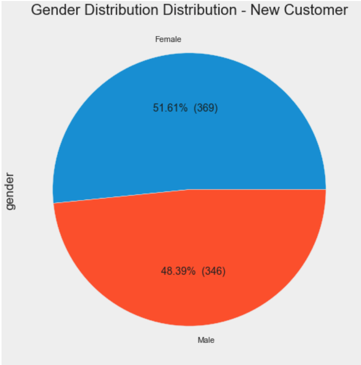
Wealth Segment Distribution

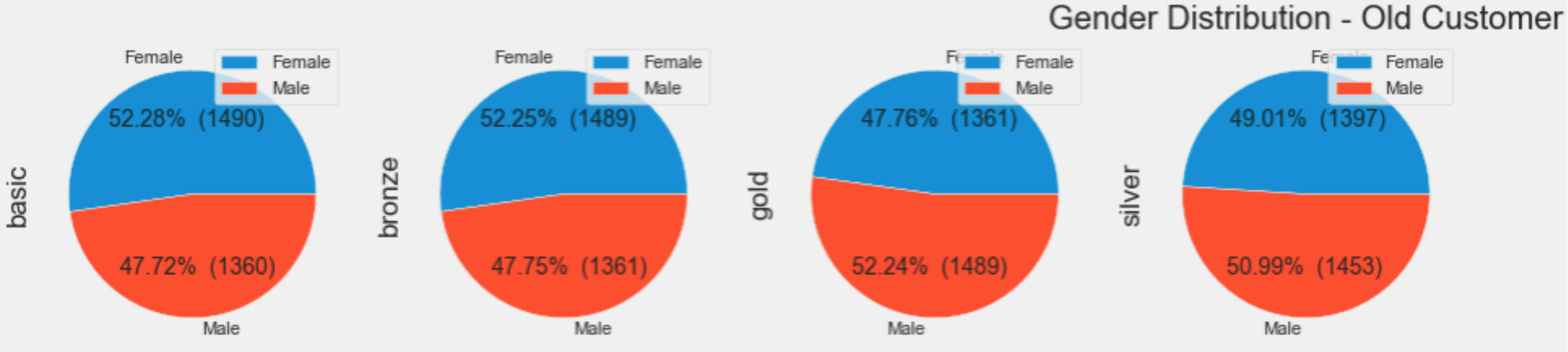




Interpretation

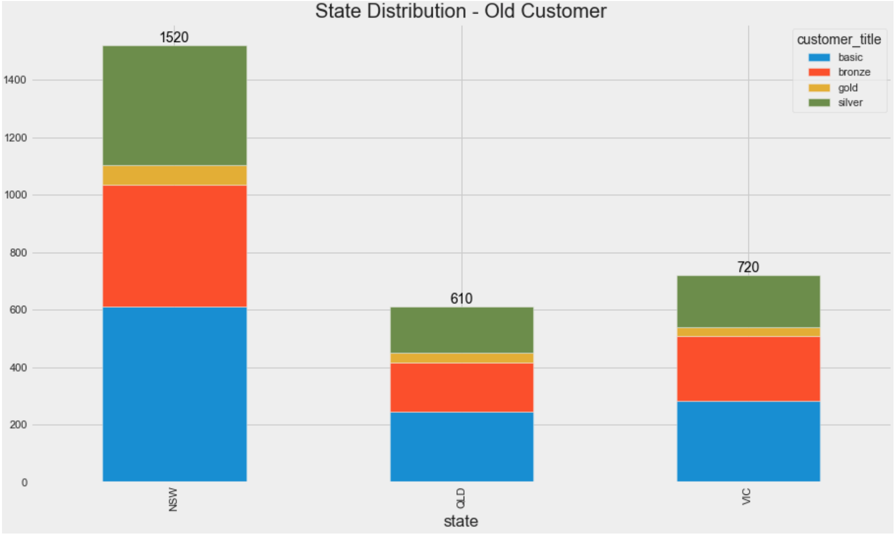
Gender Distribution

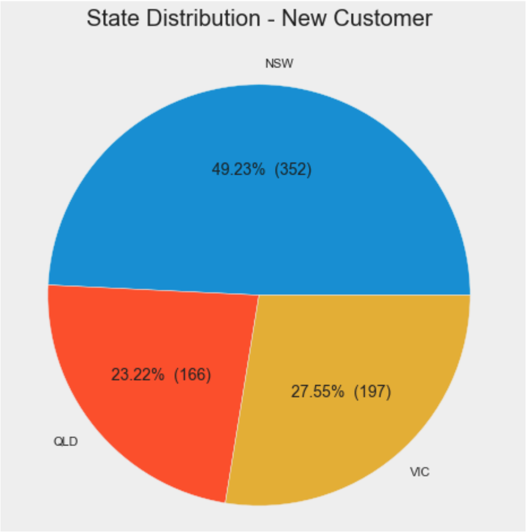




Interpretation

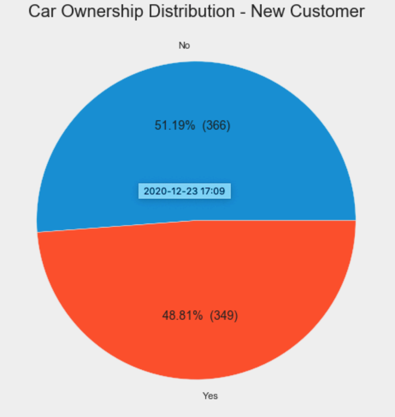
State Distribution

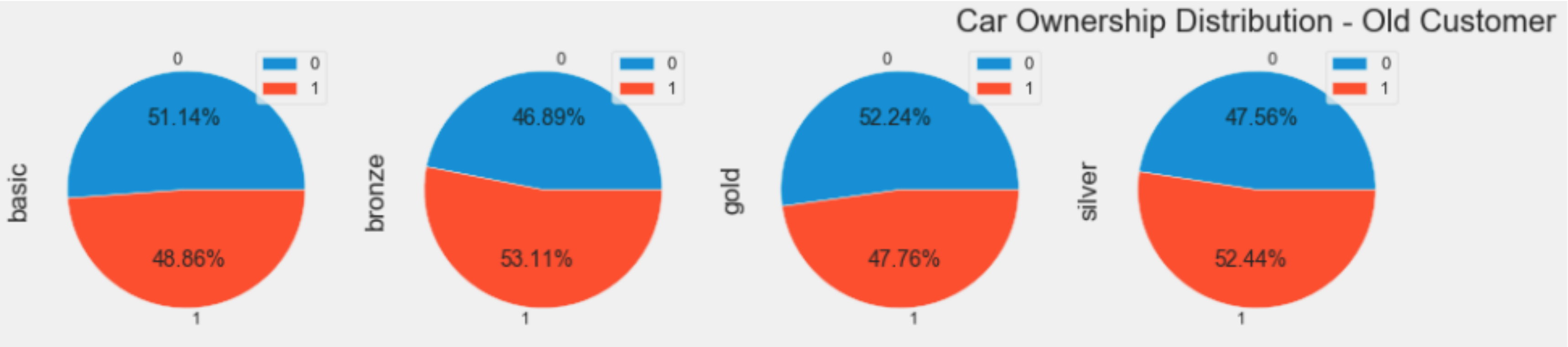




Interpretation

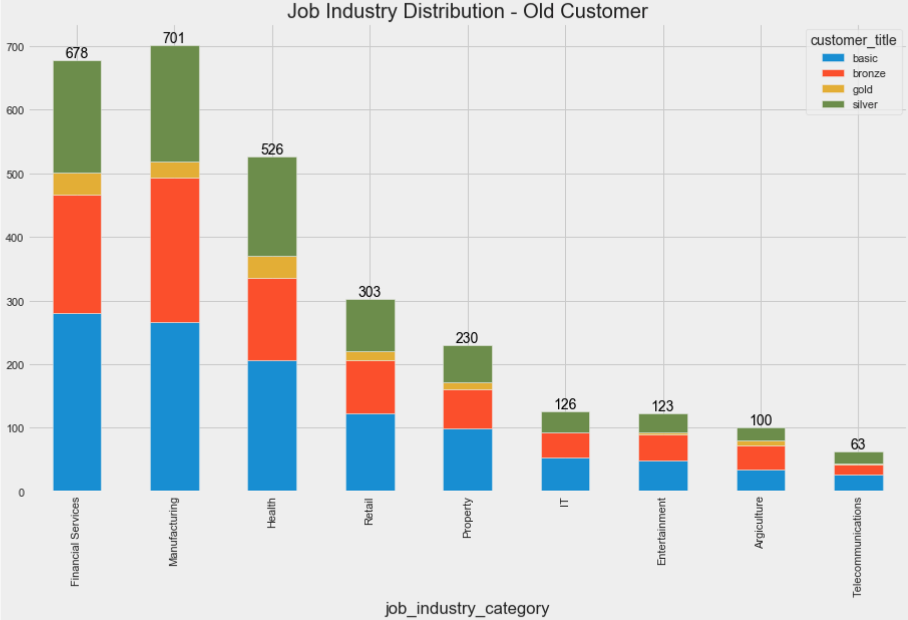
Car Ownership Distribution

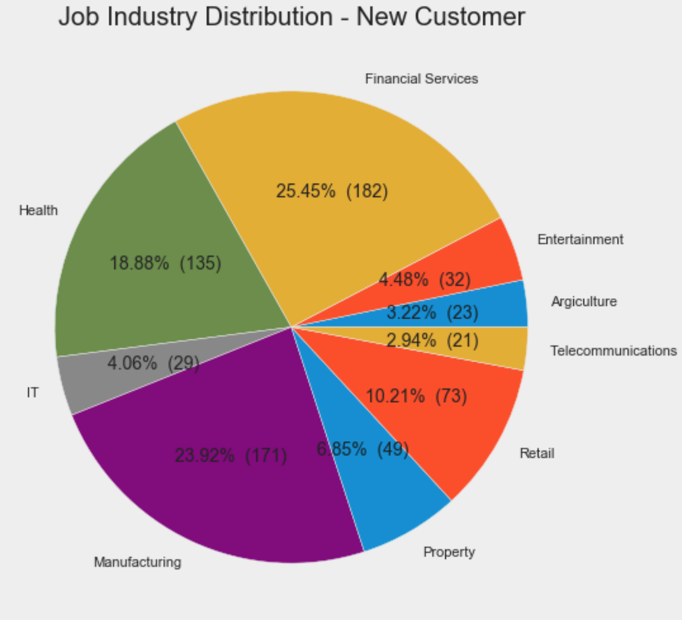




Interpretation

Job Industry Distribution

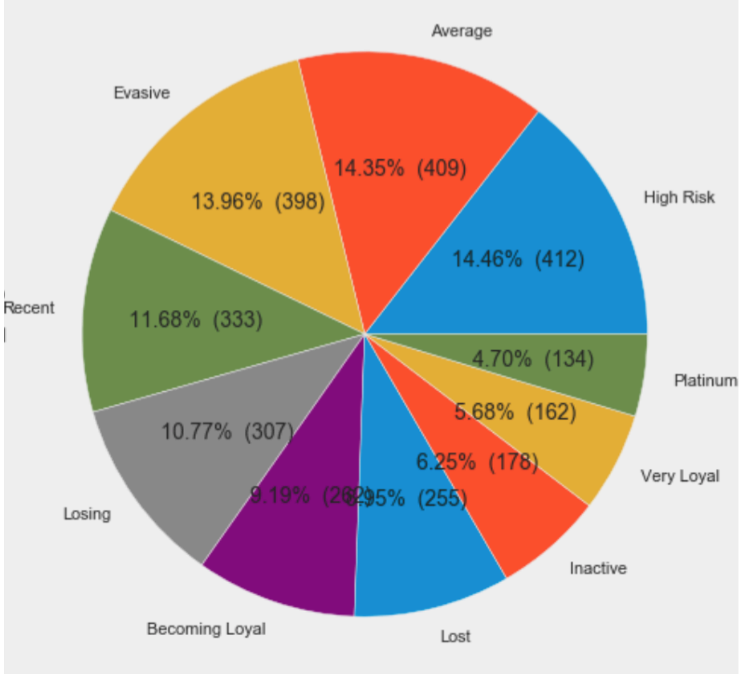




Interpretation

Customer Segmentation Standard

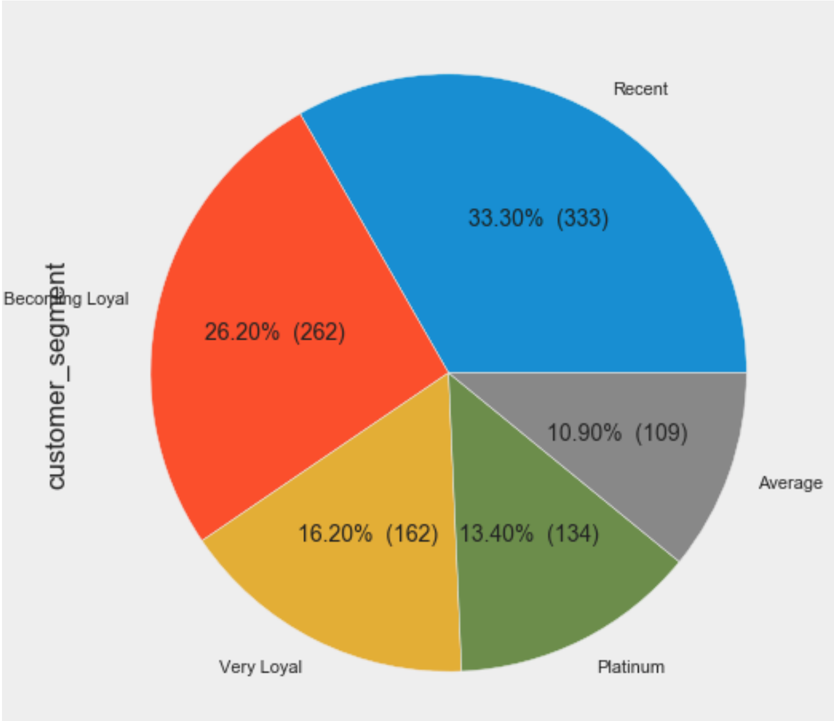
Rank	Class Name	Counts
1	Platium Customer	134
2	Very Loyal	162
3	Becoming Loyal	262
4	Recent Customer	333
5	Average Customer	409
6	High Risk	412
7	Evasive	398
8	Losing	307
9	Inactive	178
10	Lost	255



Interpretation

Top 1000 Customers Segmentation

Rank	Class Name	Counts
1	Platium Customer	134
2	Very Loyal	162
3	Becoming Loyal	262
4	Recent Customer	333
5	Average Customer	109

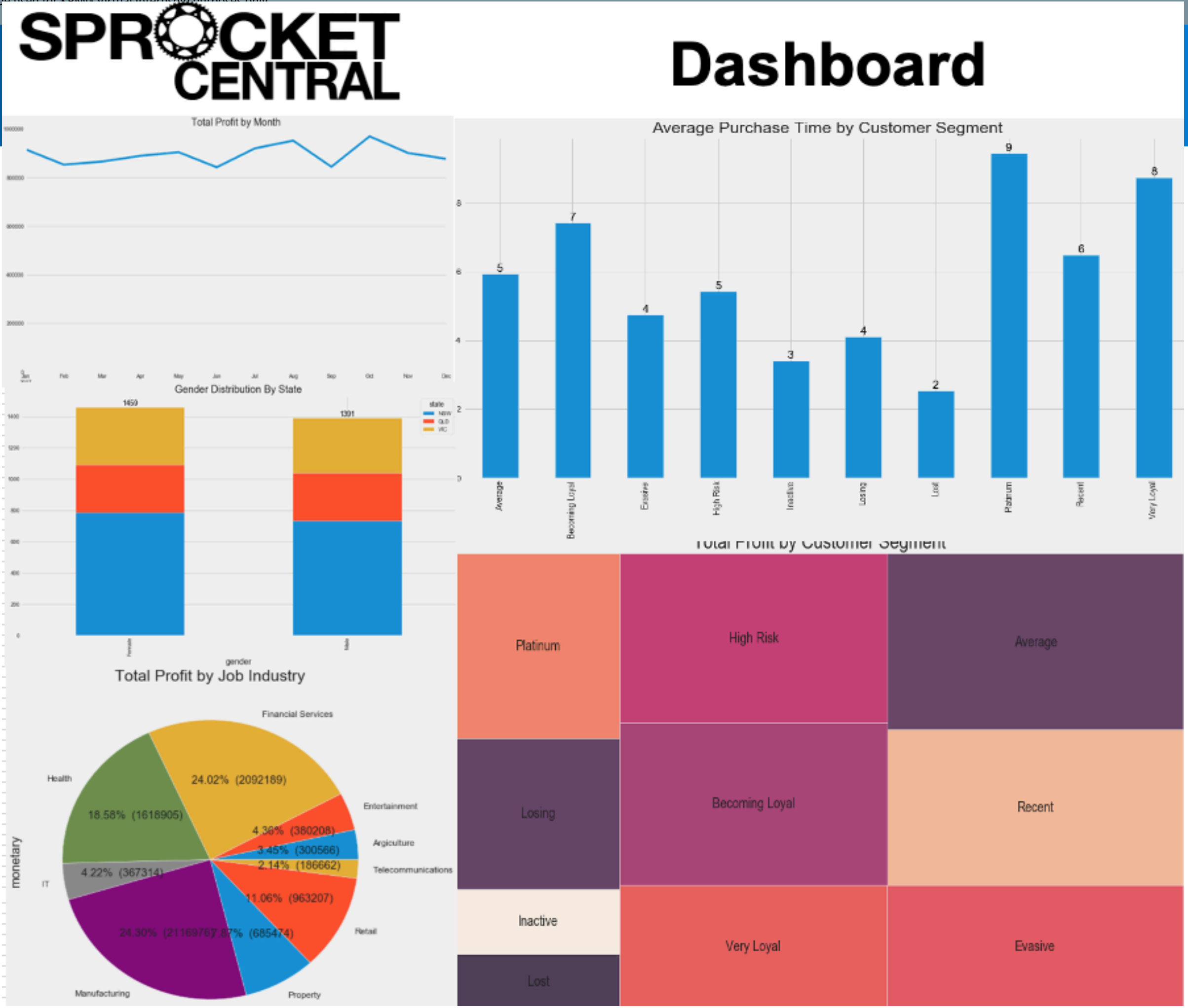


Interpretation

Top 1000 Customers Demographics

customer_id	gender	DOB	job_industry_category	wealth_segment	owns_car	tenure	state	customer_segment
2103	Male	1975/9/22	Financial Services	Affluent Customer	0	18	NSW	Platinum
3470	Female	1967/10/1	Health	Affluent Customer	1	6	VIC	Platinum
725	Male	1965/8/27	Health	High Net Worth	1	19	QLD	Platinum
2476	Male	1956/9/25	Property	High Net Worth	0	17	QLD	Platinum
902	Female	1989/7/26	Retail	Mass Customer	0	18	NSW	Platinum
...	...	...	...	...	...	...	...	...
1763	Female	1994/10/30	Manufacturing	Affluent Customer	1	7	VIC	Average
1776	Male	1978/8/26	Financial Services	Affluent Customer	0	10	QLD	Average
1760	Female	1966/4/27	Health	High Net Worth	0	15	VIC	Average
1759	Male	1969/6/2	Financial Services	High Net Worth	0	15	NSW	Average
2354	Female	1958/12/19	Retail	Mass Customer	0	17	VIC	Average

Dashboard

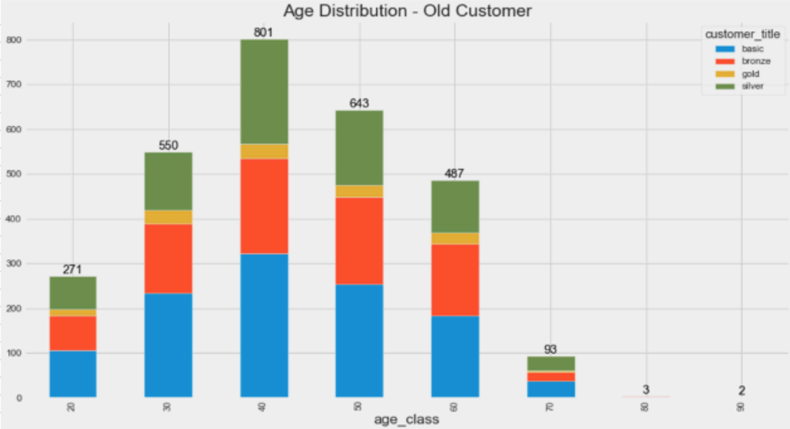

SPROCKET CENTRAL
Dashboard
Total Profit by Month
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Average Purchase Time by Customer Segment
5
7
4
5
3
4
2
9
6
8
Average
Becoming Loyal
Evasive
High Risk
Inactive
Losing
Lost
Platinum
Recent
Very Loyal
Gender Distribution By State
1459
1391
state
NSW
QLD
VIC
gender
Total Profit by Customer Segment
Platinum
High Risk
Average
Losing
Becoming Loyal
Recent
Inactive
Lost
Very Loyal
Evasive
Total Profit by Job Industry
monetary
Financial Services
24.02% (2092189)
Health
18.58% (1618905)
Entertainment
4.36% (380208)
Argiculture
3.45% (300566)
Telecommunications
2.14% (186662)
IT
4.22% (367314)
Retail
11.06% (963207)
Property
7.87% (685474)
Manufacturing
24.30% (2116976)

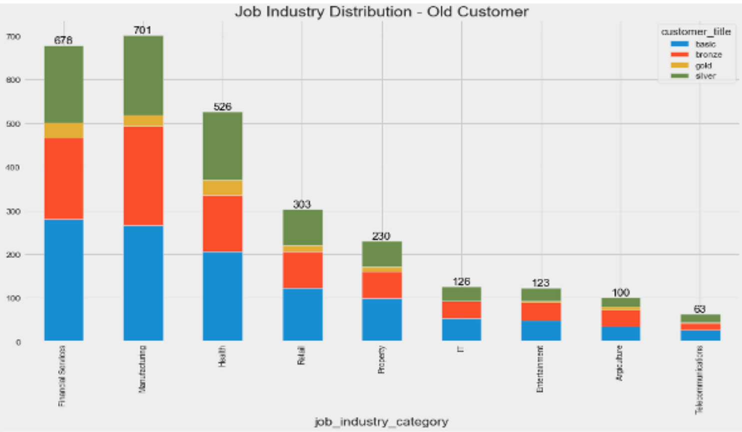
Note: The data and information in this document is reflective of a hypothetical situation and client. This document is to be used for KPMG Virtual Internship purposes only.

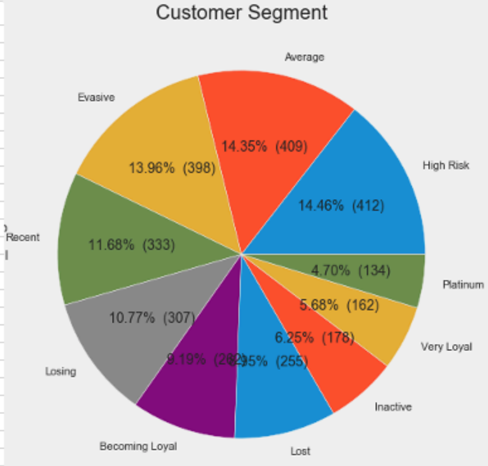
Dashboard

SPROCKET CENTRAL

Dashboard







Top 1000 Customers

2103	Male	1975-09-22	Financial Services	Affluent Customer	0	18.0	NSW	Platinum
3470	Female	1967-10-01	Health	Affluent Customer	1	6.0	VIC	Platinum
725	Male	1965-08-27	Health	High Net Worth	1	19.0	QLD	Platinum
2476	Male	1956-09-25	Property	High Net Worth	0	17.0	QLD	Platinum
902	Female	1989-07-26	Retail	Mass Customer	0	18.0	NSW	Platinum
...	...	...	...	...	...	...	...	...
1763	Female	1994-10-30	Manufacturing	Affluent Customer	1	7.0	VIC	Average
1776	Male	1978-08-26	Financial Services	Affluent Customer	0	10.0	QLD	Average
1760	Female	1966-04-27	Health	High Net Worth	0	15.0	VIC	Average
1759	Male	1969-06-02	Financial Services	High Net Worth	0	15.0	NSW	Average
2354	Female	1958-12-19	Retail	Mass Customer	0	17.0	VIC	Average

Appendix

Appendix

Python code about this work: https://github.com/jarrywangcn/KPMG-Data-Analytic