

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

Arnab De

Data Analytics Consulting Virtual Intern

Agenda

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

Introduction

Data Quality Assessment

FIELD NAME	ERRORS
<u>DOB</u>	01 record 1843 87 records Blanks
<u>last_name</u>	125 records Blanks
<u>Gender</u>	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
<u>job_title</u>	506 records Blanks
<u>job_industry</u>	656 records mention 'N/A'
<u>Default</u>	3317 records value 'special characters' includes null and Blanks
<u>Tenure</u>	87 records Blanks

Introduction

Data Cleaning

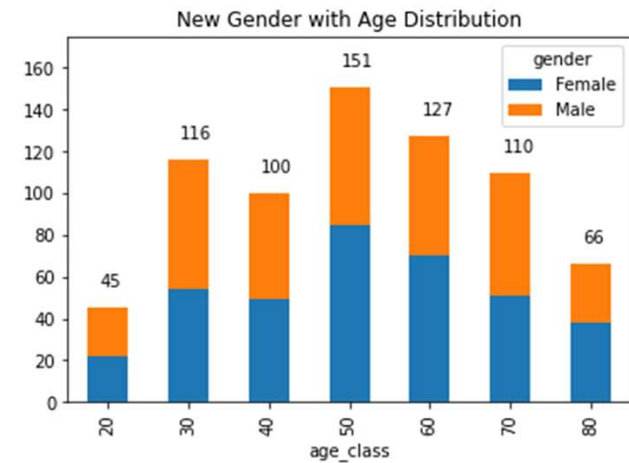
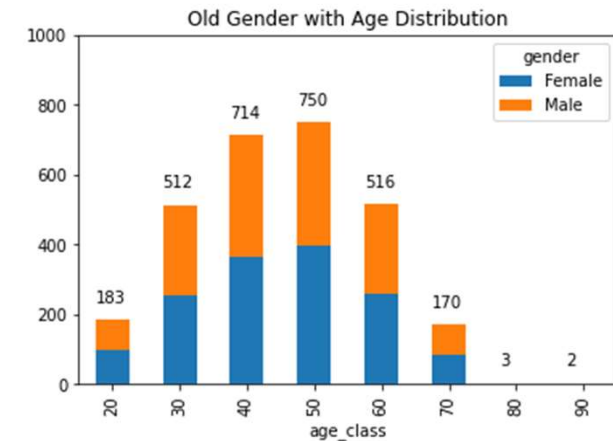
- Records pertaining to Deceased Customers and transactions more than a year old were Dropped.
- Join Keys between tables were considered and conflicting Records were discarded.
- Records with Missing Fields were dropped, as well as transactions that were more than one year old or more than five years old.

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

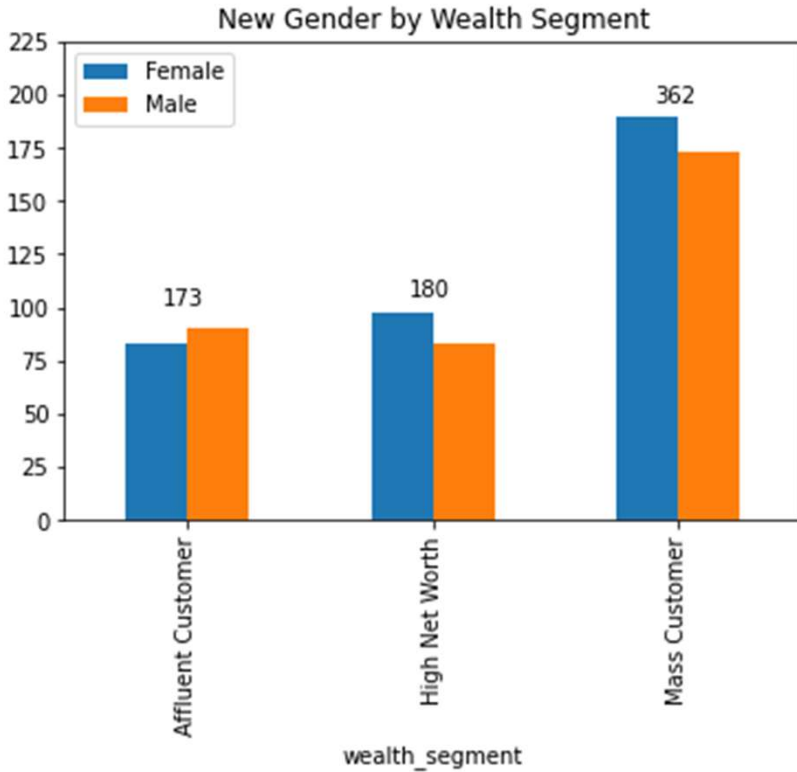
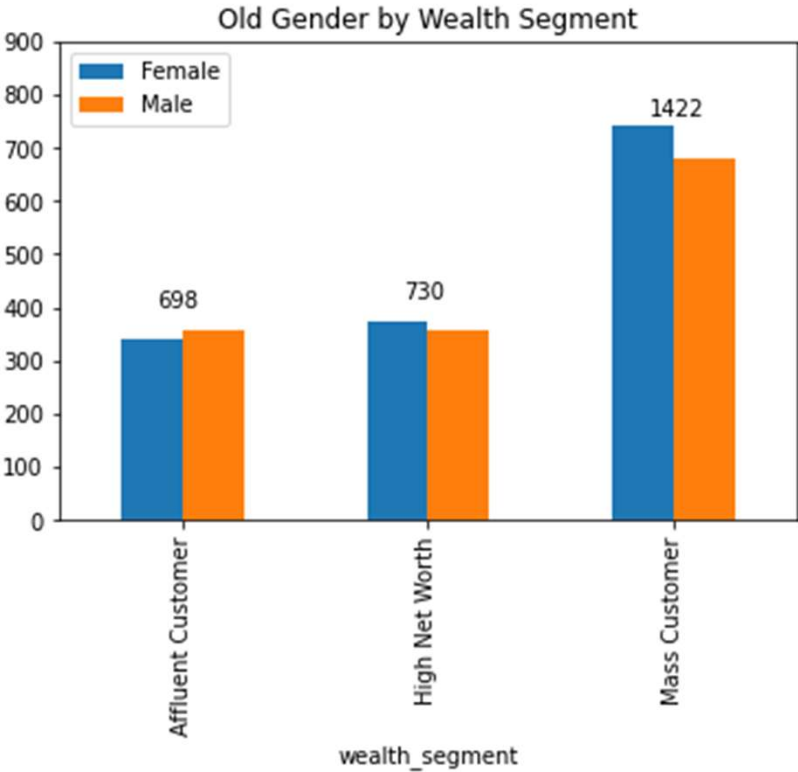
Data Exploration

Gender with Age distribution

- As we can see, mostly our new customers are between 25 to 48 years old.
- Number of customers from 48 to 59 years old has big drops on percentages.
- There is a slightly increase in number of customers over 59 years old in term of percentages
- It looks like the percentages of under 25 years old not really change.

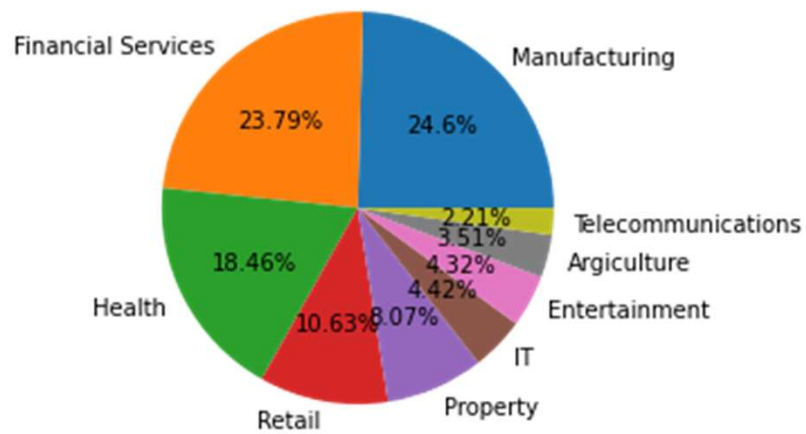


Gender with Wealth Segment

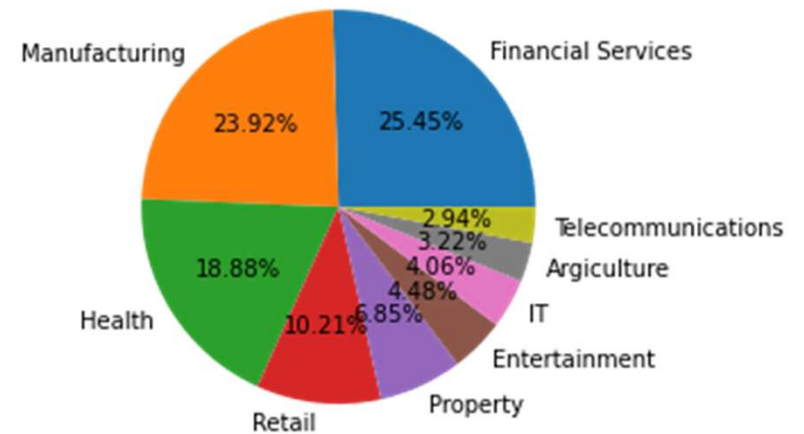


Job Industry Diversity

Old Job Industries



New Job Industries



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.

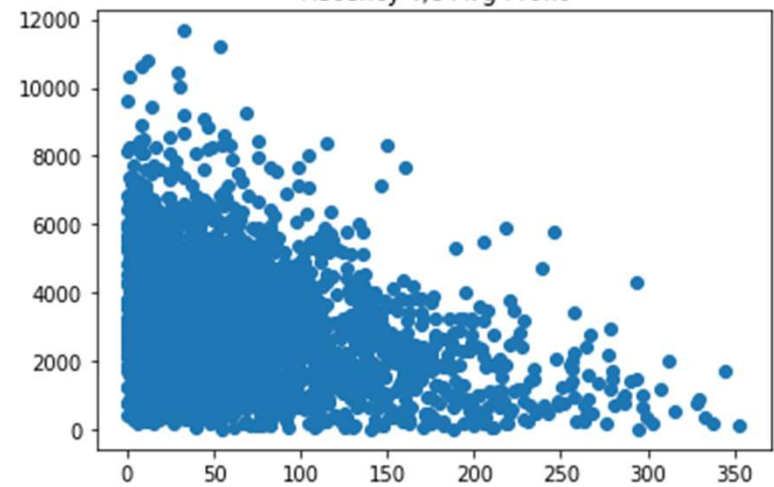
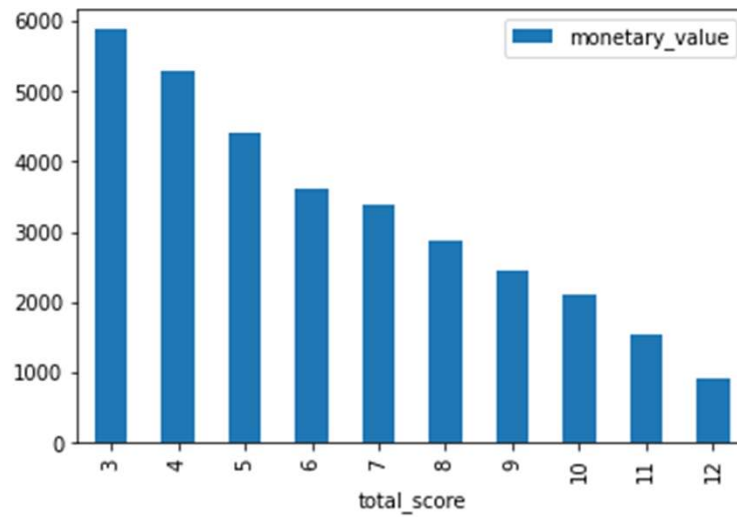
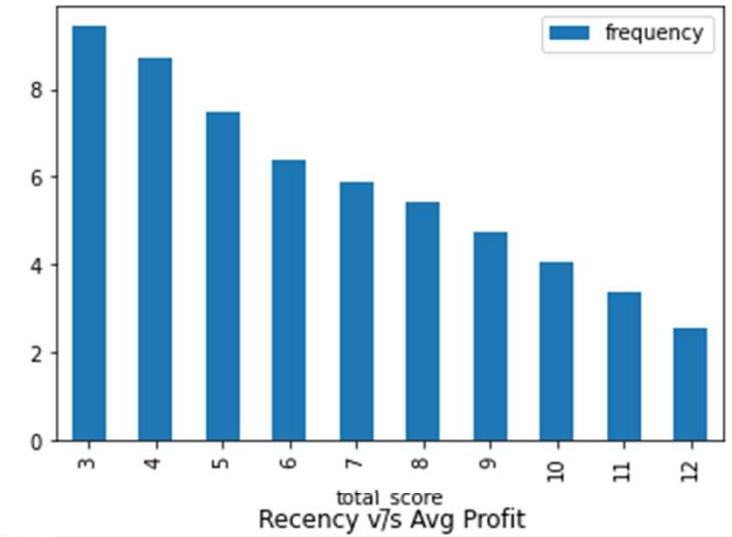
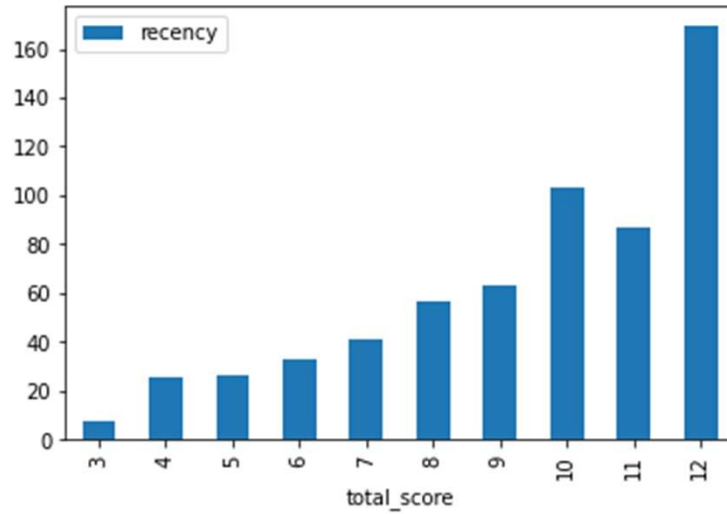
- **Frequency**

- 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.

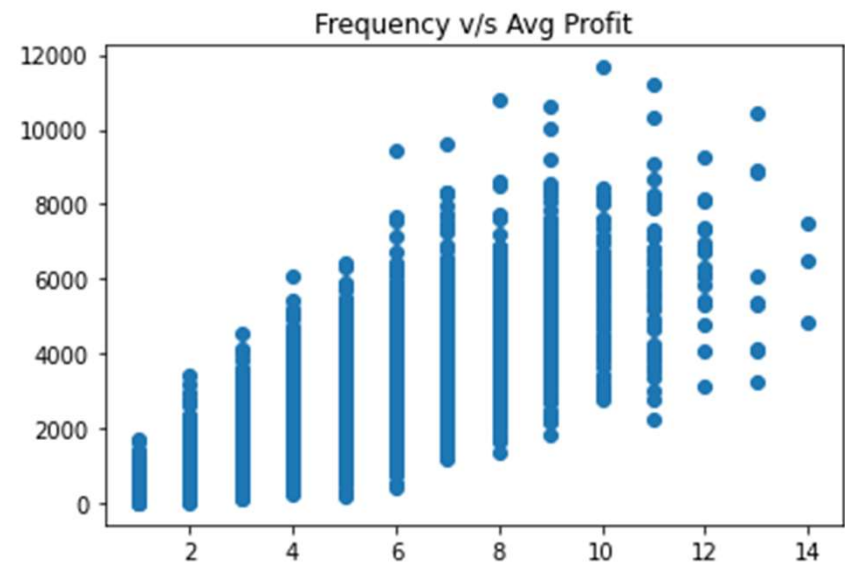
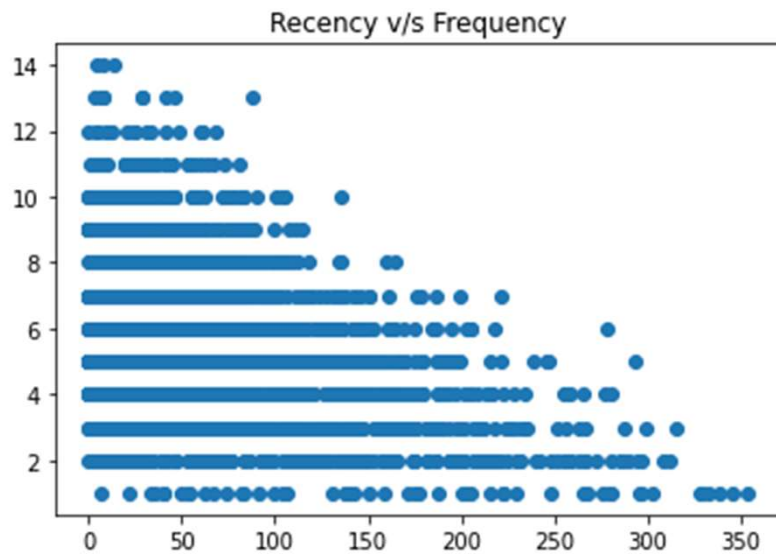
- **Monetary Value**

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

RFM Analysis



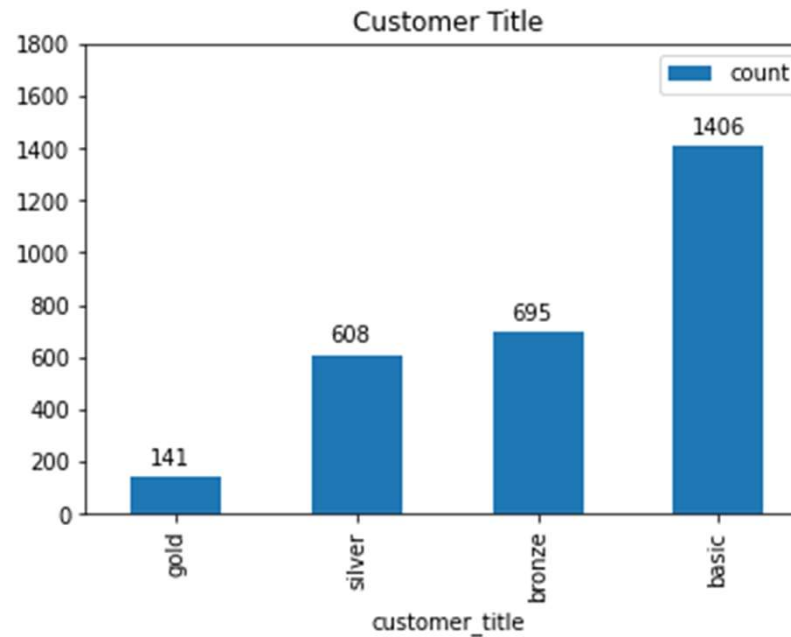
RFM Analysis



RFM Analysis

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

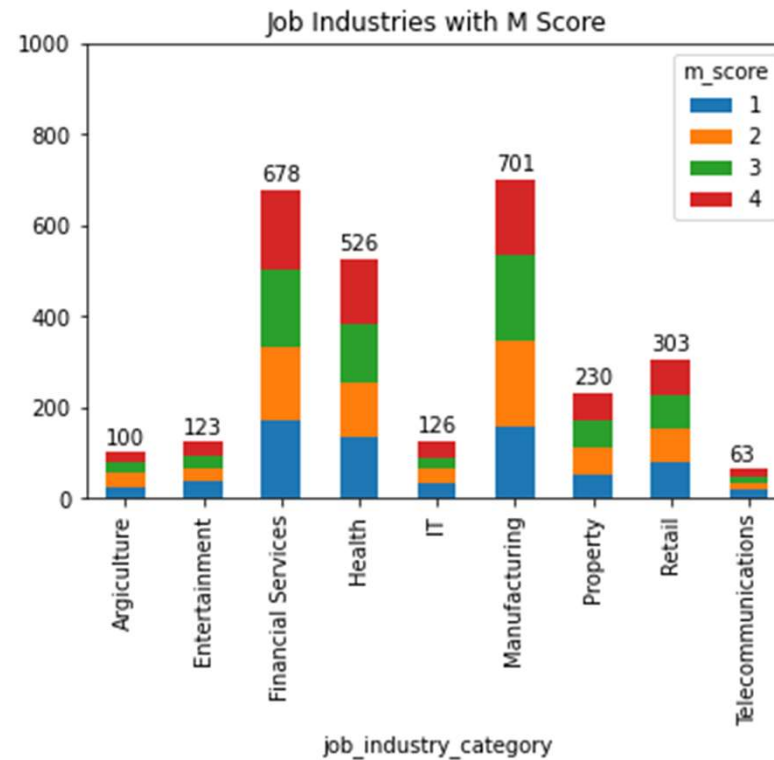
1. Gold Class: These customers have recently made a purchase, are frequent and are most profitable.
2. Silver Class
3. Bronze Class
4. Basic Class: These customers have not made any recent purchase, are not frequent and do not contribute majorly.



Interpretation

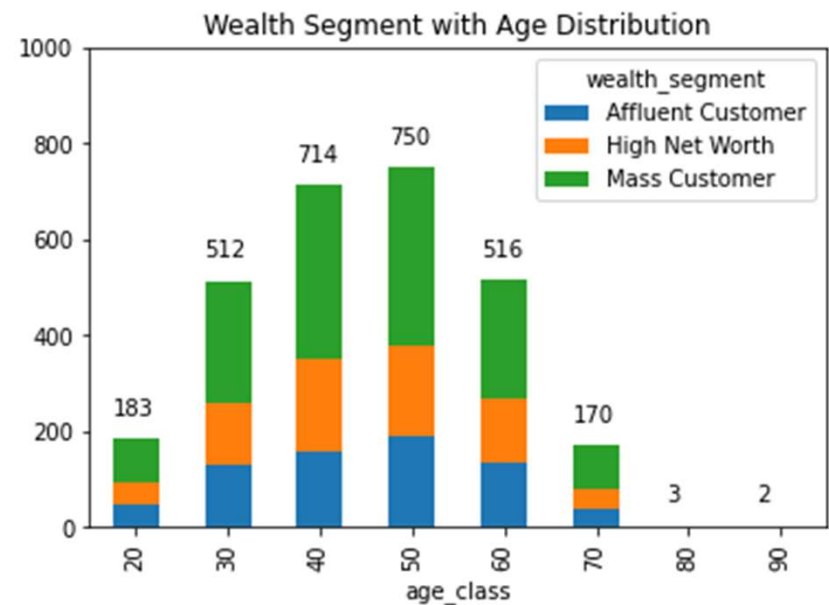
Job industry category

- Mostly our new customers are on Finance industry and our Manufacturing customers are still on top 2.
- The rest industries is still same

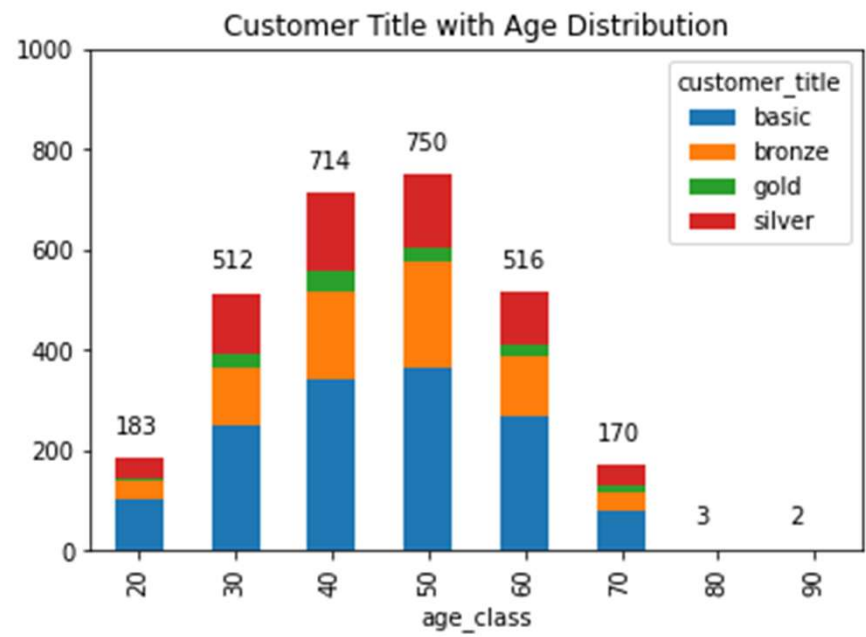
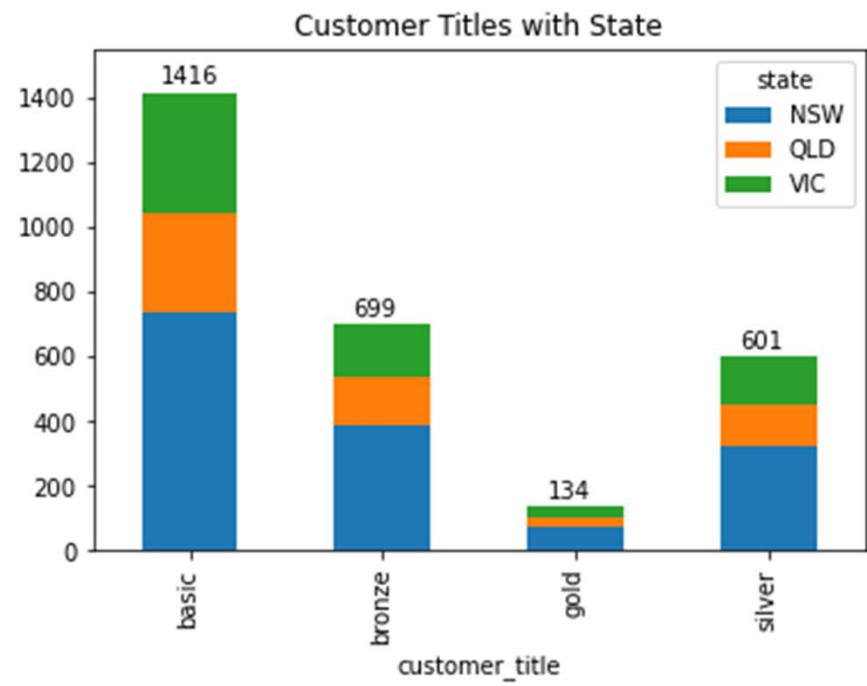


Wealth segments

- In all ages, the number of Mass Customers is the highest so we should focus on this social class.
- After that, we should focus on High Net Customer.
- Then Affluent Customers but mostly second and third quadrant



Customer Tiers by State and age distribution



Appendix

Appendix

S4. SELECTED DISTRIBUTION INDICATORS, Household net worth and gross household income—2011–12

		<i>Household net worth(a)</i>	<i>Gross household income(a)</i>
Ratio of values at top of selected percentiles			
P90/P10	ratio	53.87	9.06
P80/P20	ratio	11.61	4.45
P80/P50	ratio	2.36	1.90
P20/P50	ratio	0.20	0.43
Percentage share received by households in			
Lowest net worth quintile	%	0.9	12.1
Middle net worth quintile	%	12.0	17.4
Highest net worth quintile	%	60.8	31.3
Percentage share received by households in			
Lowest gross income quintile	%	12.4	4.3
Middle gross income quintile	%	16.4	15.7
Highest gross income quintile	%	36.8	46.5

Source: Australian Bureau of Statistics

Thank you