TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Data Analytics Team] - [Yadvendar Singh]

Agenda

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

Introduction

1000 New Customers to Target from the Datasets

Problem Outline:

- Sprocket Central Pty Ltd is our valuable client and their marketing team is looking to boost their business.
- The team need recommendation of 1000 new customers that should be targeted to drive the most value for the organization.

Proposed Solution:

- The Project has been divided into 3 phases as follows – Data Exploration, Model Development and Interpretation to build the recommendation.

- Our **Data Analysis** will compose of:

1. Job industry, wealth segmentation by age.
2. Product purchases by age, and gender etc.
3. Loyal customers classification.
4. RFM analysis and customer classification.
5. Forecasting of Revenue.

Data Exploration

Data Quality Assessment

Data Quality Dimensions

Assessment was made on the following mentioned dimension and has been sent to you via an email.

Correct Values	Accuracy
Data Fields with Values	Completeness
Values up to Date	Currency
Values Free from Contradiction	Consistency
Data Items with Value Meta-data	Relevancy
Records that are Duplicated	Uniqueness
Allowable Values	Validity

Data Exploration

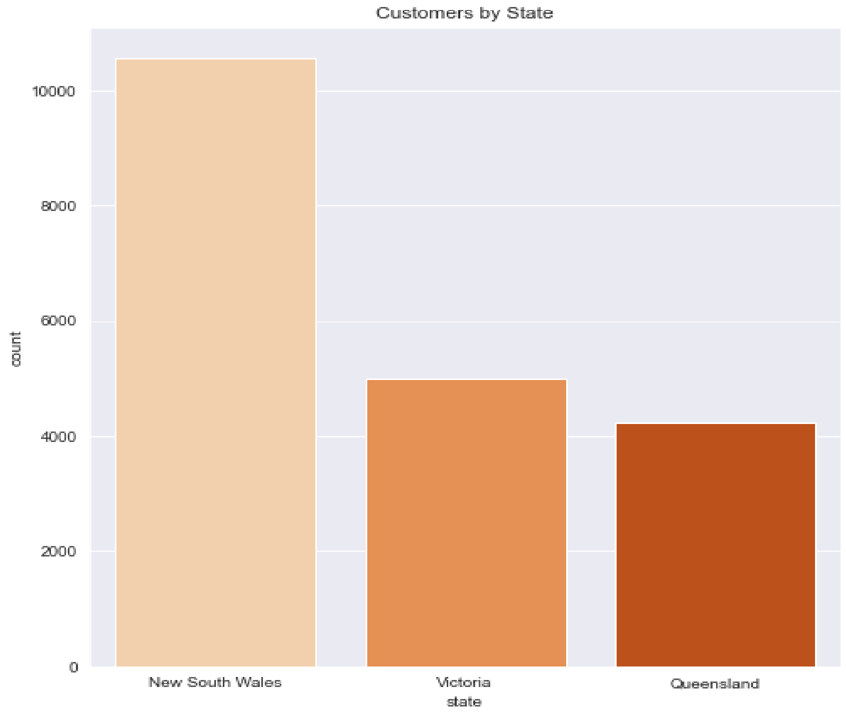
Data Quality Assessment Summary:

Dimensions	Transactions	Customer Demographics	Customer Address
Accuracy		1. default	
Completeness	1. online_order 2. brand 3. product_line 4. product_class 5. product_size 6. product_first_sold_date	1. last_name 2. DOB 3. Completeness 4. Completeness 5. default 6. tenure	
Currency		1. default	
Consistency		1. DOB 2. default	1. states
Relevancy		1. default	
Uniqueness		1. default	
Validity	1. product_first_sold_date	1. gender_dimension 2. default	

Data Exploration

Customer By State

- New South Wales has highest number of customers, followed by Victoria and Queensland.

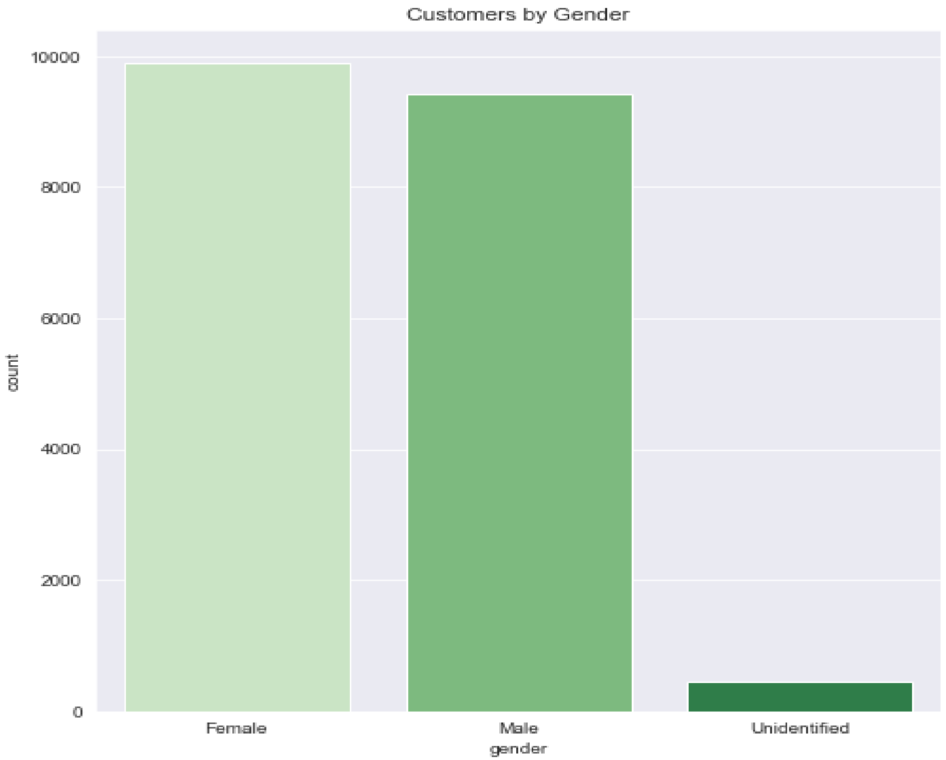


- ❑ X axis -> state
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customer By Gender

- Females have topped our customers by gender chart, followed by Males.
- Unidentified gender are those people who didn't identify themselves as either "Male" or "Female".

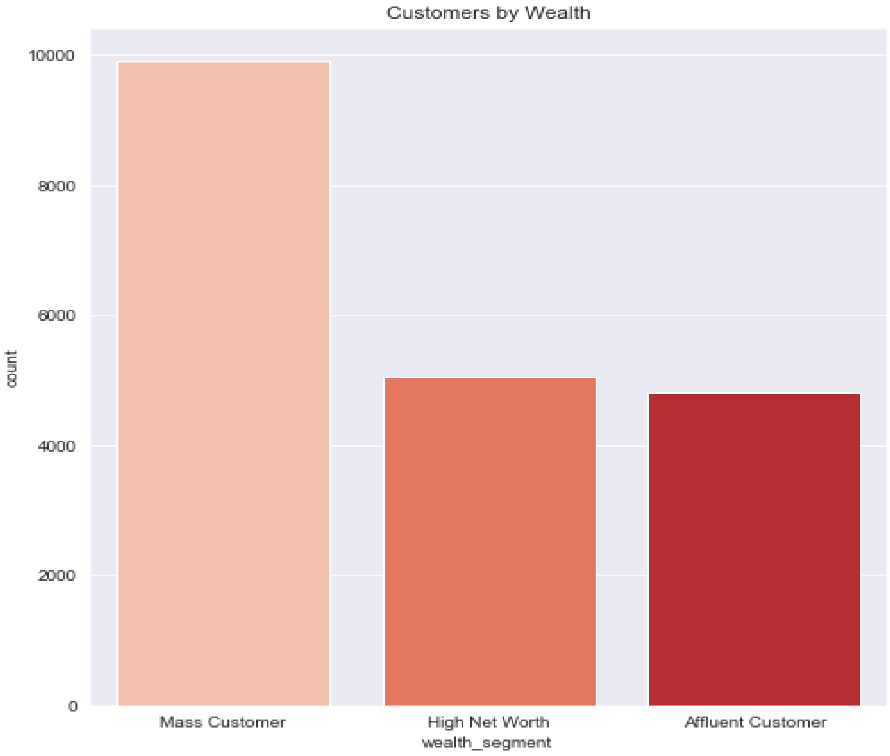


- ❑ X axis -> gender
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customer By Wealth Segment

- There are almost 2x more Mass Customers than Affluent Customers.
- Ratio of High Net Worth and Affluent Customers is ~1.

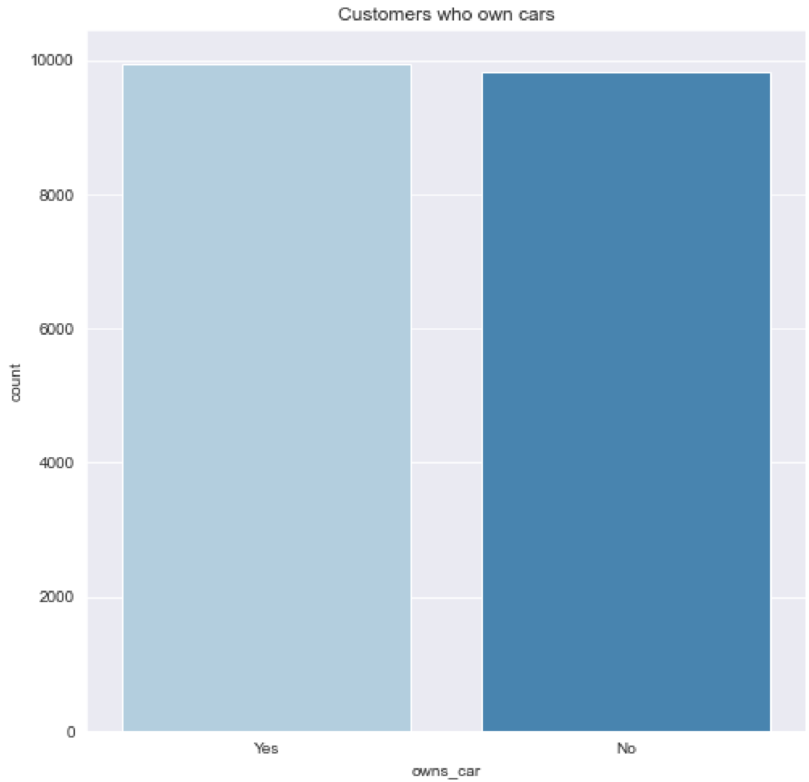


- ❑ X axis -> Wealth Segment
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers who OWN Cars

- There is no significant difference between the customers who owns the car to those who doesn't.
- This signifies that our Bicycle sales is independent of car variable.

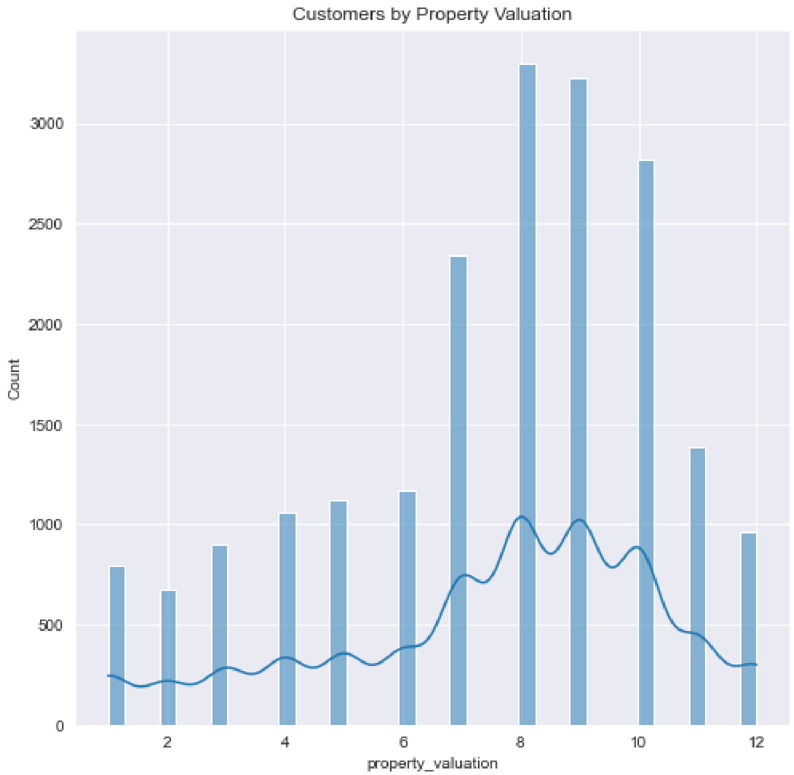


- ❑ X axis -> Owns Cars -> Category (Yes or No)
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers by Property Valuation

- The highest number of customers are those whose property has been evaluated between (8,10).
- The distribution of customers between (1,6) and (11,12) is same.

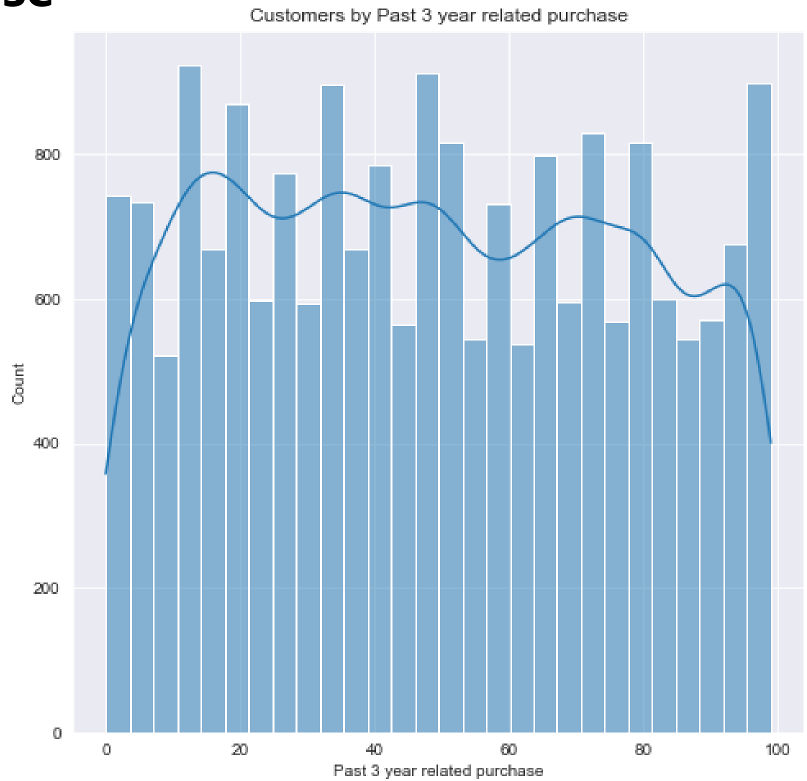


- ❑ X axis -> Property_valuation -> Rank [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers by Past 3-year related purchase

- We have customers who made 0 purchase to the customers who made around 100 purchase.
- Majority of our customers have bought more than 20 bicycles.

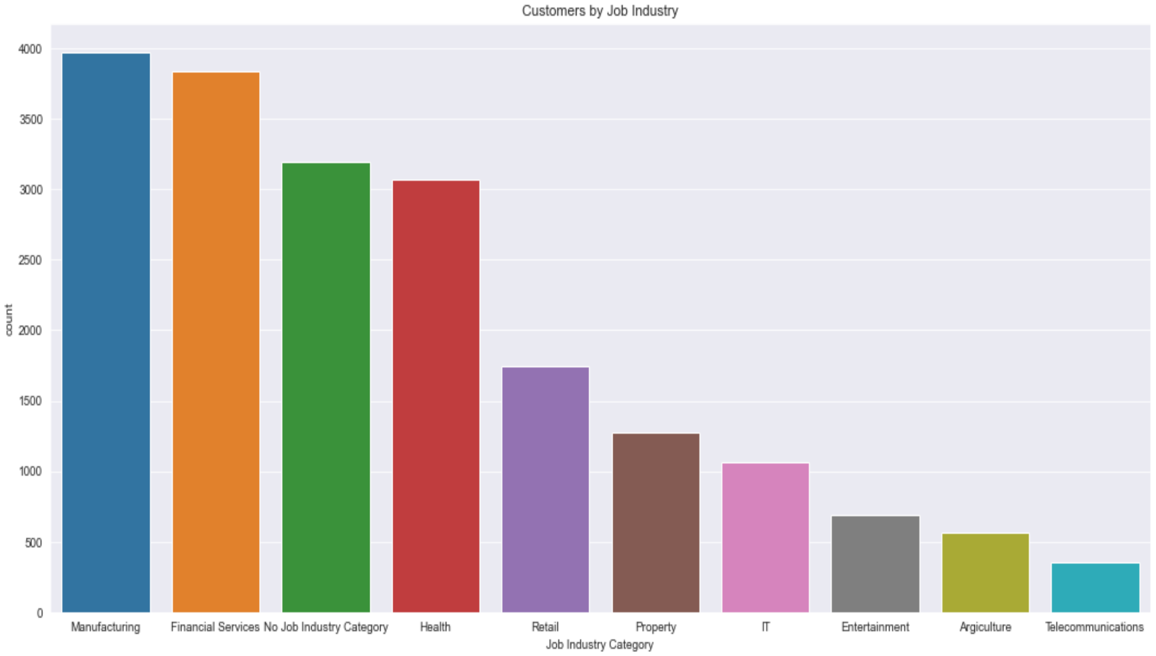


- ❑ X axis -> Past 3-year bike related purchase
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers by Job Industry

- Financial and Manufacturing services marks up highest customers.
- Customers in Entertainment, Agriculture, and Telecommunication are the lowest.
- Moderate number of customers come from retail and property sectors.

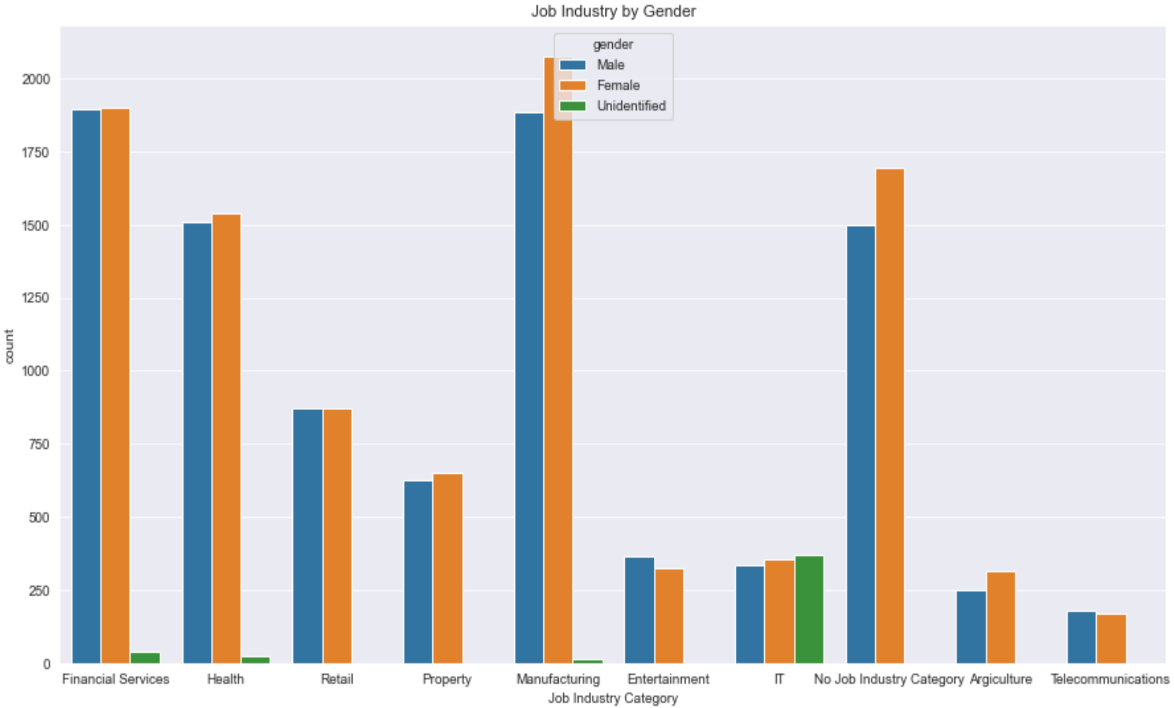


- ❑ X axis -> Job Industry Category
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers by Job Industry and Gender

- There are a greater number of Females in almost all Industries.
- There are slightly more Males in Entertainment and Telecommunication Category Only.

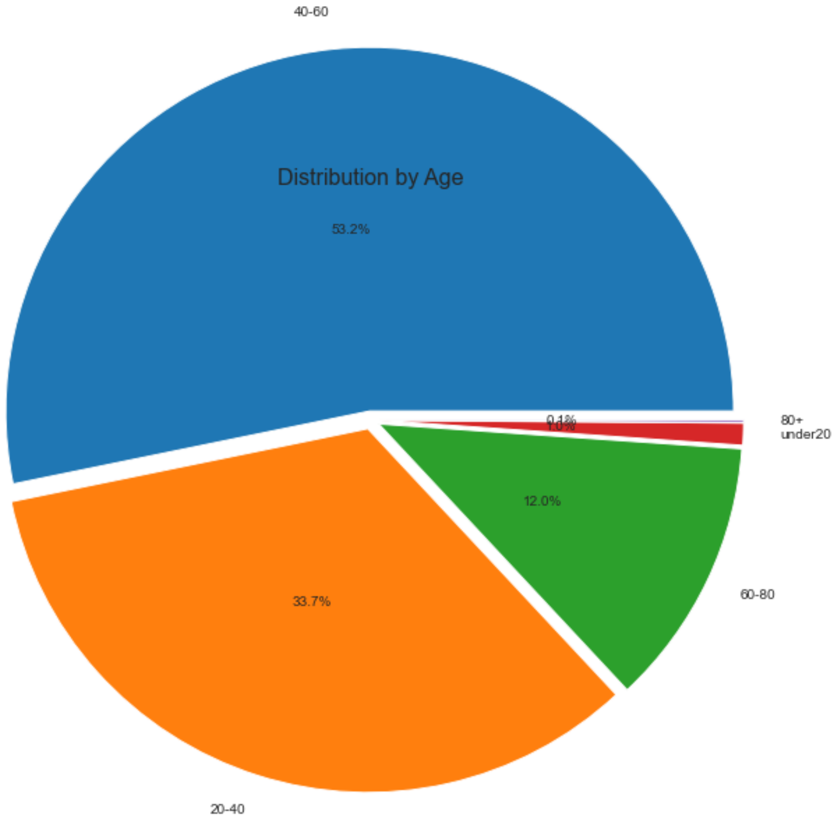


- ❑ X axis -> Job Industry Category
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Customers Distribution by Age

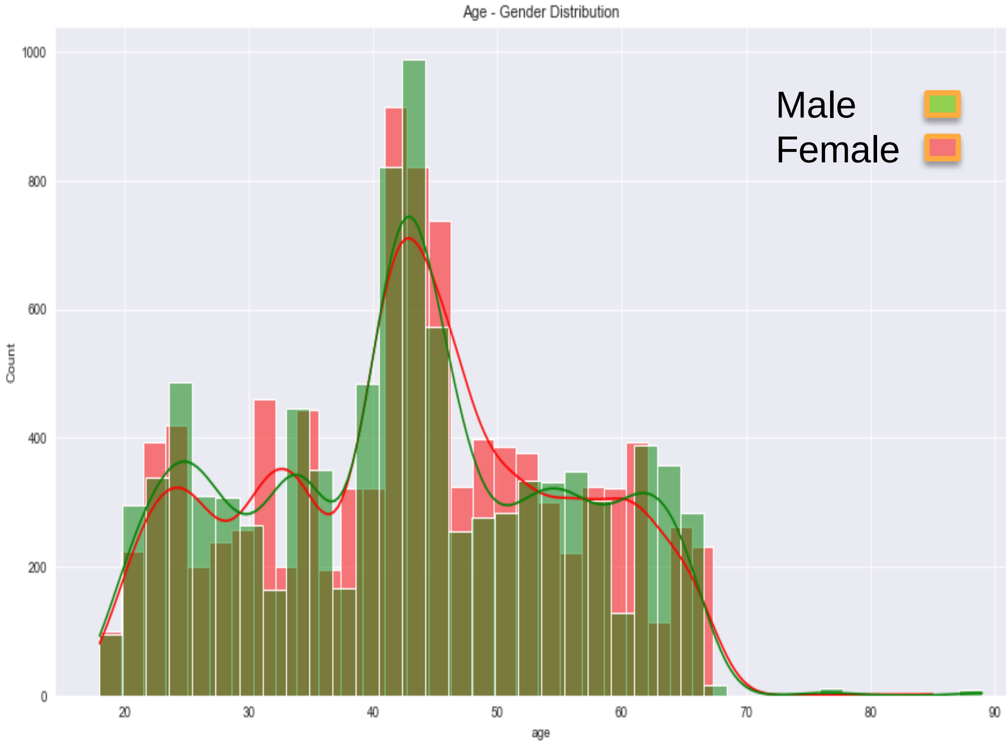
- Highest number of customers are from age-group [40-60], (Mid Age Group).
- 0.1% customers are above 80 years of age.
- 1.0% customers are from under20 age group.
- More than 1/4th of the customers are in the age range [20-40].
- 12.0% customers are in the age range (60-80)



Data Exploration

Customers Distribution by Age and Gender

- Overall, More Females makes up Bike related Transactions.
- There is not much variation in the Male Female Customers.

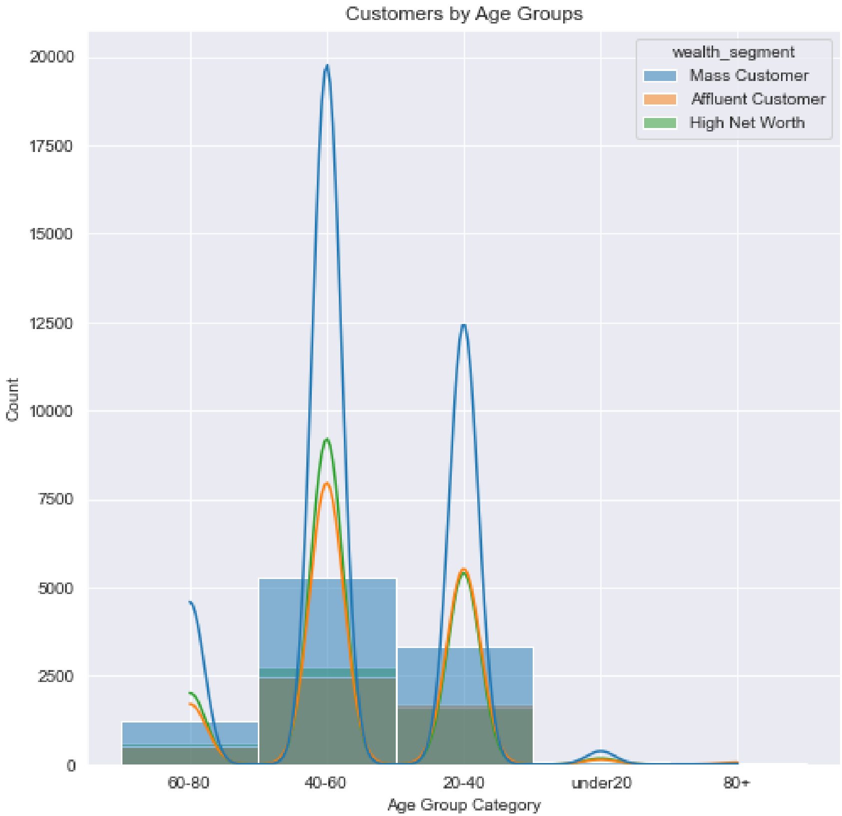


- ❑ X axis -> Age
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Age Group Distribution by Wealth Segment

- The count of Mass Customers is nearly equal to the sum of both Affluent Customers and High Net Worth Customers.
- Mass Customer >High Net Worth > Affluent Customer

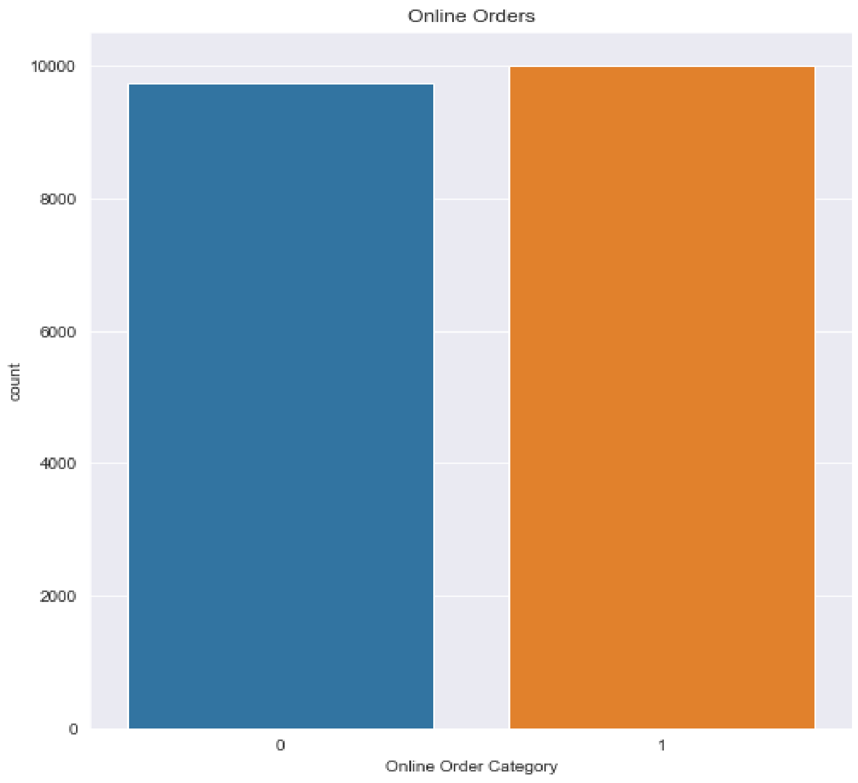


❑ X axis -> Age Group Category
❑ Y axis -> Count/Number of Customers

Data Exploration

Order Segmentation

- Offline and Online order shows the similar trend.
- Customers have equal preference for both online and in-store purchases.

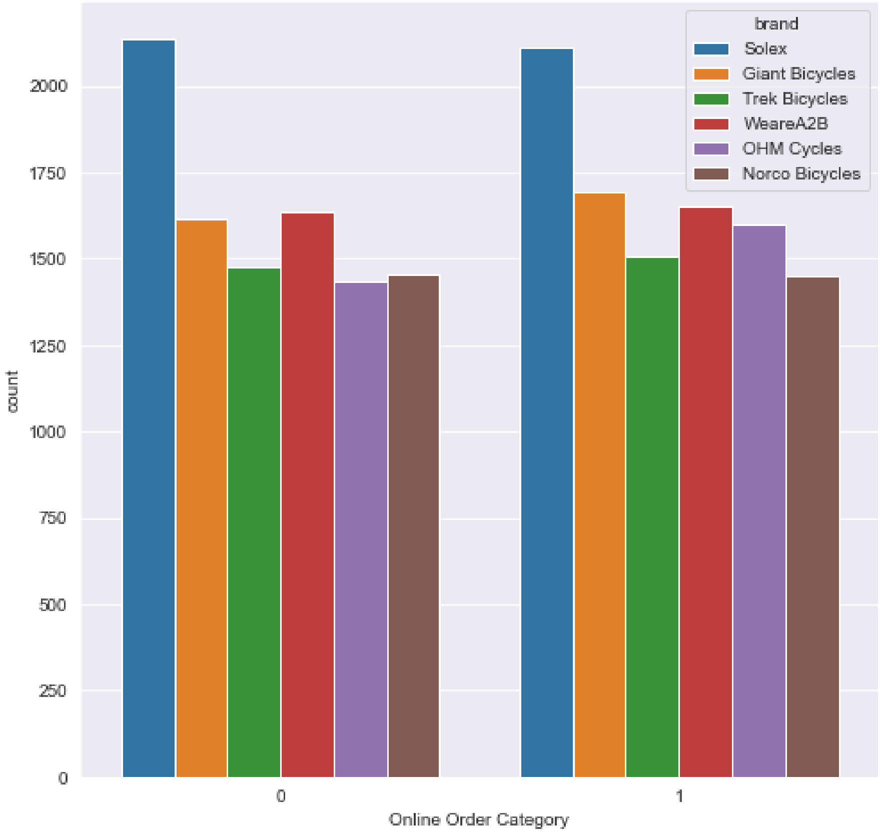


- ❑ X axis -> Online Order Category
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Order Segmentation by Brand

- Offline and Online order for all brands shows similar kind of trend.
- There is not much variation in online and offline orders.

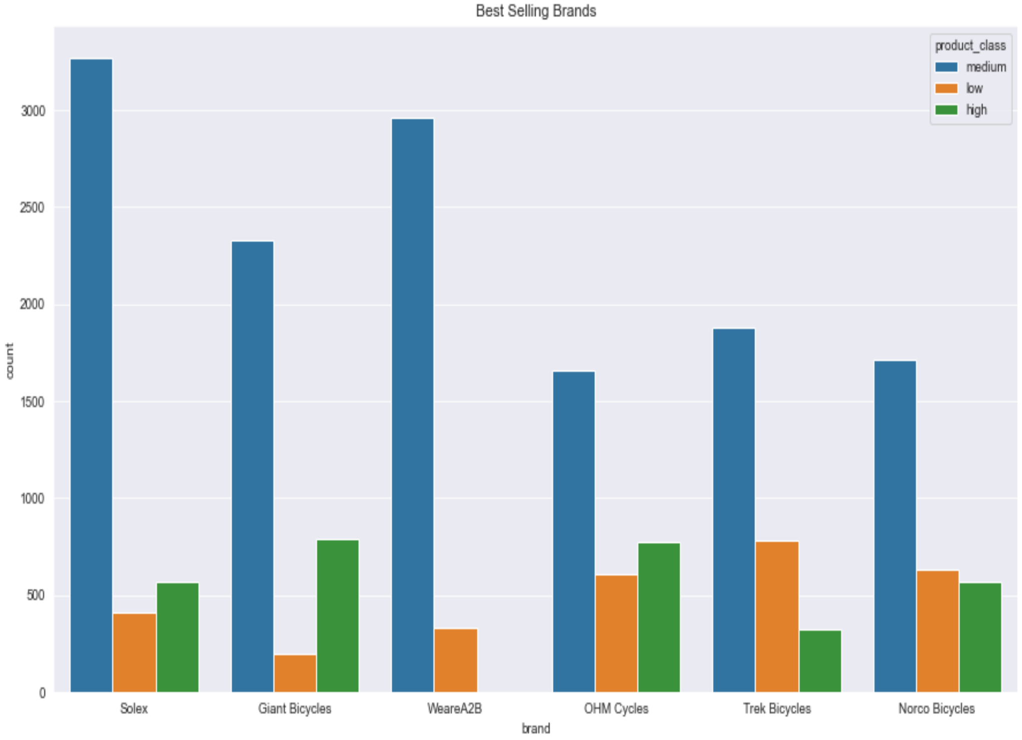


- ❑ X axis -> Online Order Category
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Best Selling Brands by Product Class

- Best Selling Products are from Solex , followed by WeareA2B.
- Trek Bicycles and Norco Bicycles have some of the lowest sales.
- Medium Product Class has high number of sales in each Brands.
- WeareA2B have no to less sales related to high product class.

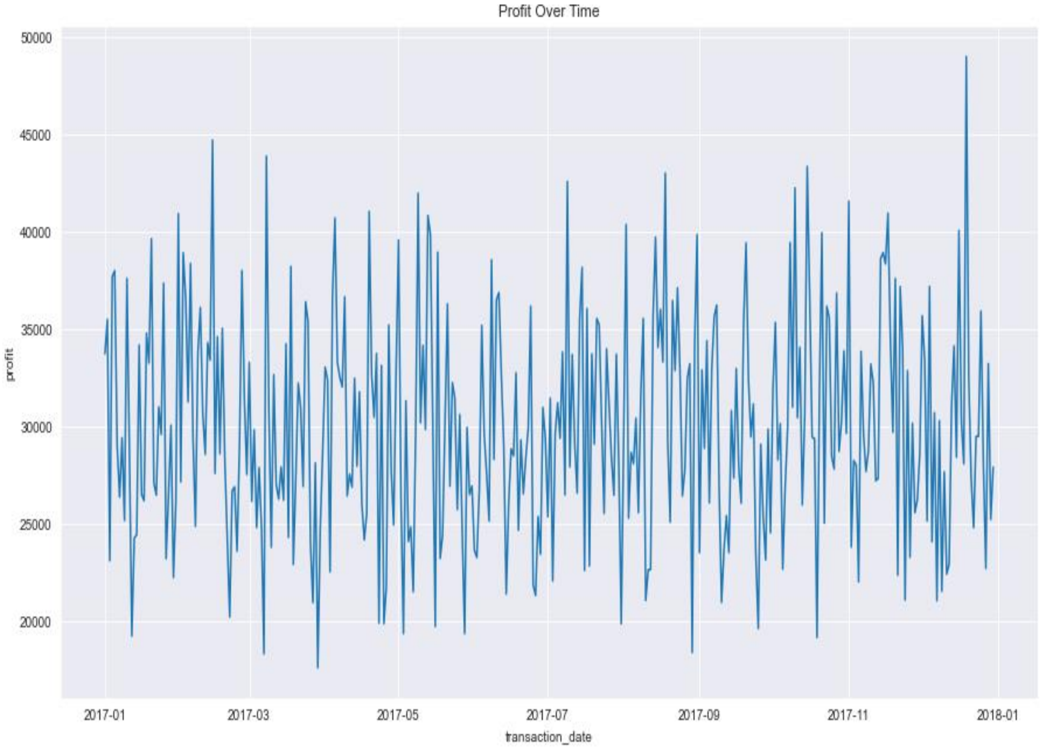


- ❑ X axis -> Brand
- ❑ Y axis -> Count/Number of Customers

Data Exploration

Profit Over Time

- Profit is constant over the entire year.
- There is a Big Spike in Profit in the month of December.

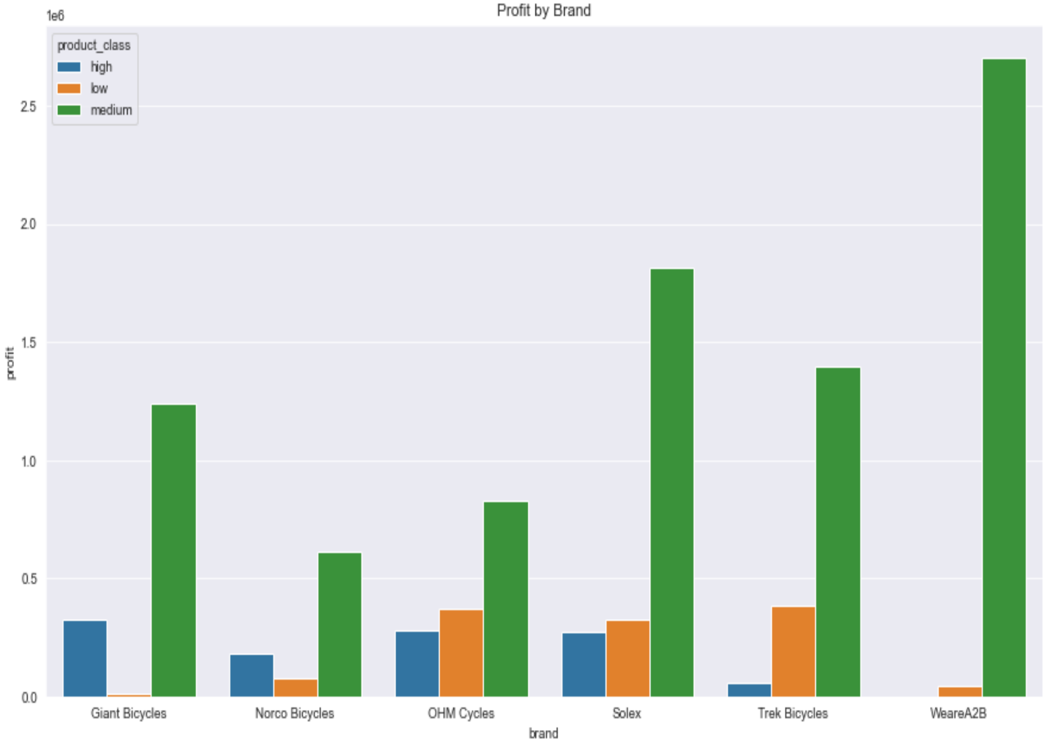


- ❑ X axis -> Transaction Date
- ❑ Y axis -> Profit

Data Exploration

Profit By Brand and Product Class

- Highest Profit grossed (>2.5 million) is from WeareA2B brand and product class Medium.
- This can be attributed to the fact that best sold bicycles are from Medium product class.
- Medium product class has made more than 0.5 million for every brand.



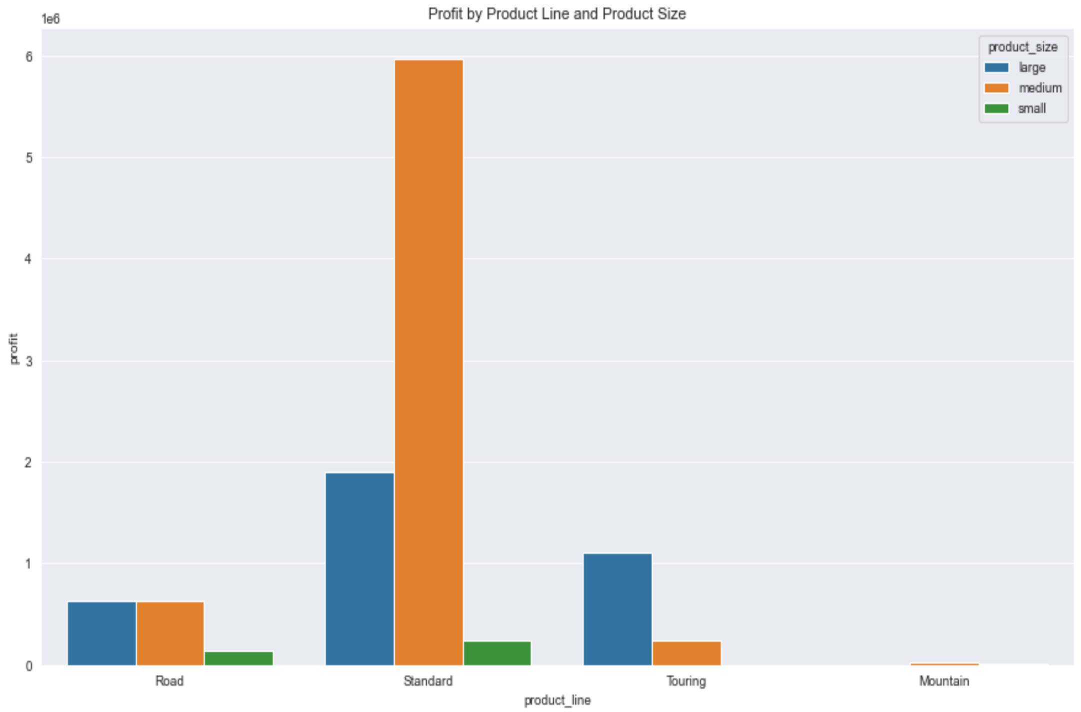
❑ X axis -> brand
❑ Y axis -> Profit in Millions

Data Exploration

Profit By Product Line and Product Size

- Highest Profit comes from Standard product line with Medium Product Size.
- This can also be attributed to a greater number of sales towards standard product line.
- Standard product line is more versatile, because of that it registered high number of sales.

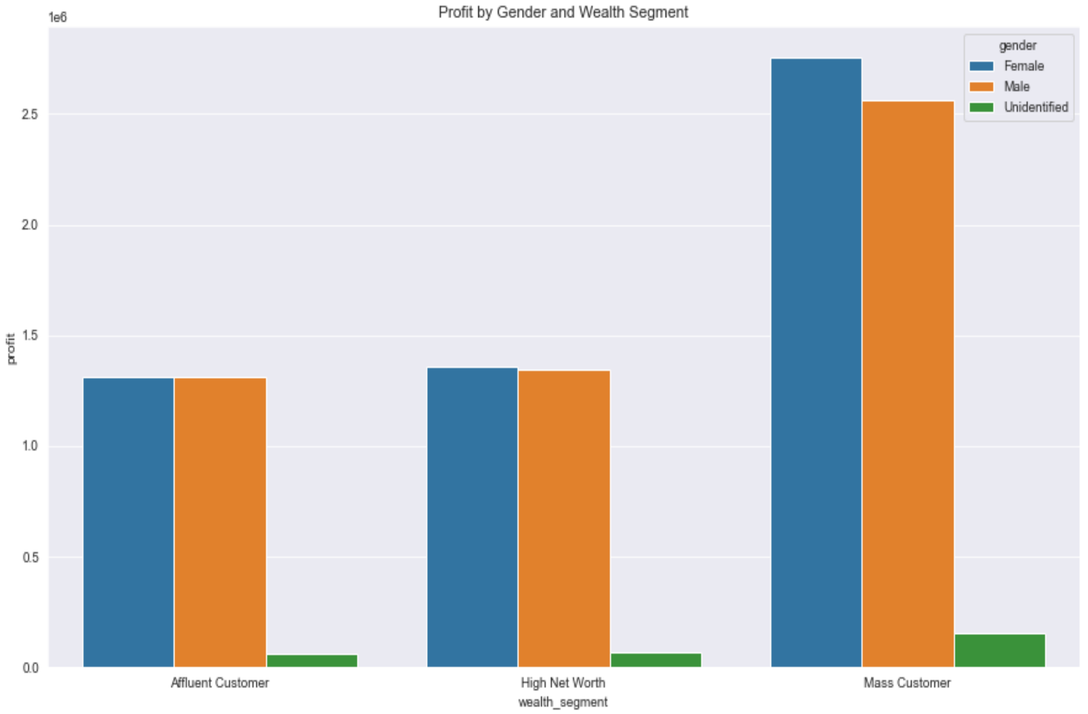
❑ X axis -> Product Line
❑ Y axis -> Profit in Millions



Data Exploration

Profit By Gender and Wealth Segment

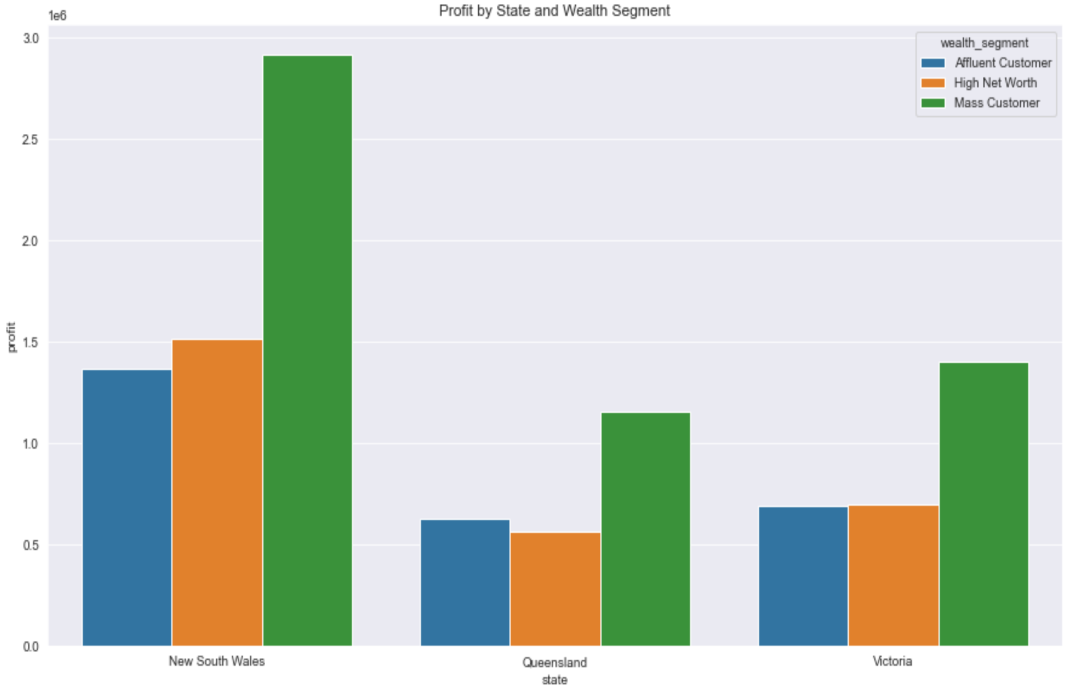
- Mass Customers generated more revenue than both Affluent and High Net Worth Customers.
- Female made more purchase in each category.



❑ X axis -> Wealth Segments
❑ Y axis -> Profit in Millions

Data Exploration

Profit By State and Wealth Segment



- New South Wales has highest number of customers in all three wealth segments.
- Only in Queensland we have low High Net Worth customers as compared to other states.

- ❑ X axis -> Wealth Segments
- ❑ Y axis -> Profit in Millions

Model Development

Customer Classification based on RFM Analysis.

- RFM stands for Recency, Frequency, and Monetary value, each corresponding to some key customer trait. Frequency and Monetary value affects a customer's lifetime value, and recency affects retention, a measure of engagement.

customer_id	recency	frequency	monetary	R	F	M	RFM_Segment_Concat	RFM_Score	RFM_Level
1	8	11	3018.09	4	4	3	443	11	Top Most Priority
2	129	3	2226.26	1	1	2	112	4	Needs Attention
4	196	2	220.57	1	1	1	111	3	Require Activation
5	17	6	2394.94	4	2	2	422	8	Champions
6	65	5	3946.55	2	2	3	223	7	Loyal

Model Development

Customer Description based on RFM Value

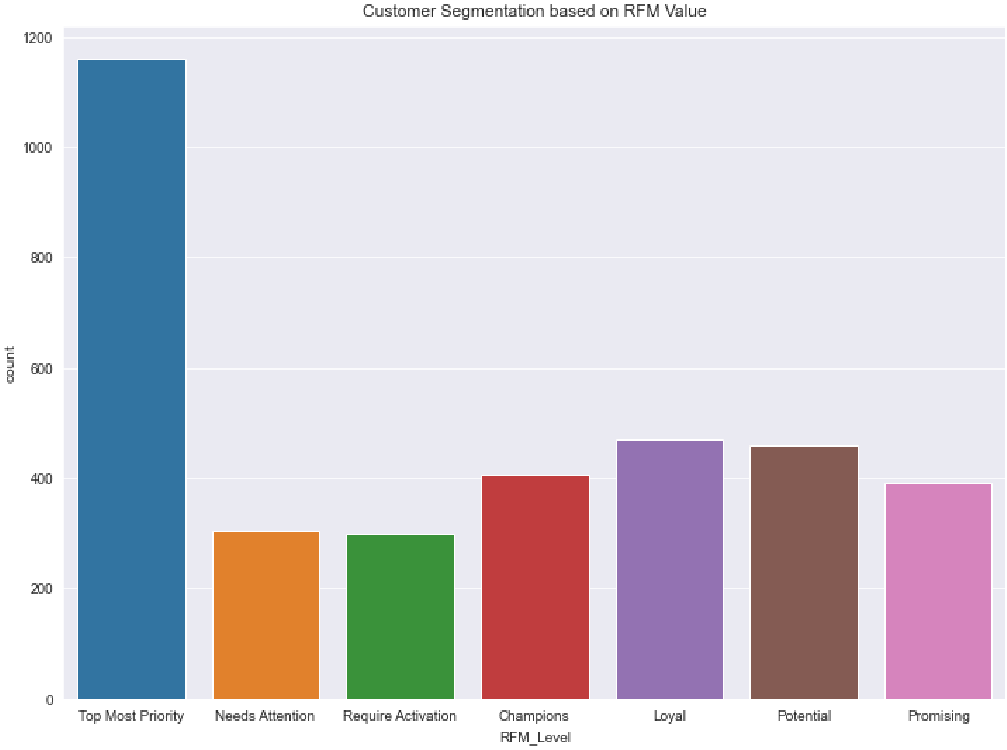
Rank	Description	RFM Score	RFM Level
1	Customers who used to visit and purchase quite often but haven't visited quite recently. We can bring them back with some sort of promotion strategies	9	Top Most Priority
2	Customers who bought recently, with highest total purchase. So, we can create strategies to give them priority	(8,9]	Champions
3	Customers who visited more often than "Potential" customers but are not so good with their purchase. So, we can give more promotion on our product for them.	(7,8]	Loyal
4	Customers who doesn't visit often but are good with their purchase if compared to "Loyal" customers. So, we can reach the potential customers give more promotion, recommended product in our retail, and give them membership.	(6,7]	Potential
5	Promising because we hope they would be our loyal customers. So, we can reach them with information about membership program and benefits. They visit more often than "Needs Attention" customers.	(5,6]	Promising
6	Customers whose visit history and purchase are very low as compared to other customers. They require direct approach from our team. Maybe we can start with sending product catalog, daily product or most recent purchase from another customer.	(4,5]	Needs Attention
7	This segment is very hard to reach as they might be the customers who visited our chain, or online store but never comeback.	Less than 4	Require Activation

Model Development

Customer Classification based on RFM Analysis.

- Customers belonging to Top Most Priority Category are highest.
- There are almost similar number of customers belonging to Loyal and Potential type.
- Require Activation and Needs Attention have the lowest number of customers.
- We need to convert all these low revenue generating customer types to our Top Most Priority by the approach mentioned above.

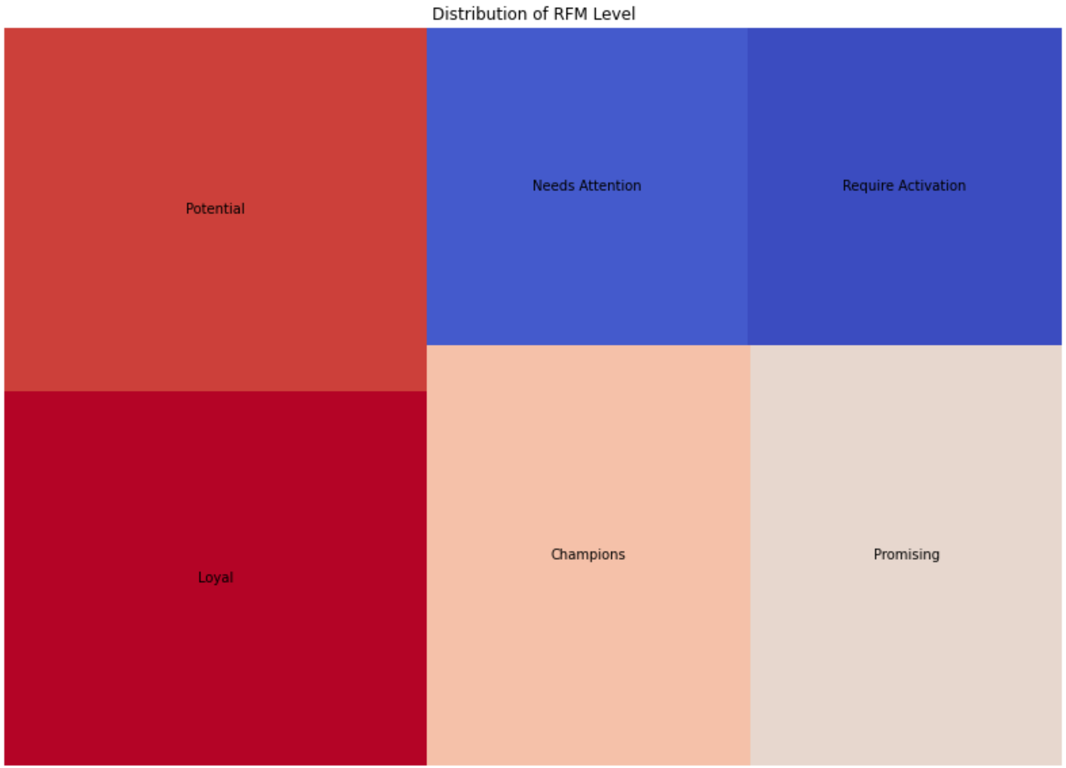
❑ X axis -> RFM Level
❑ Y axis -> Count of Customers.



Model Development

Customer Classification based on RFM Analysis.

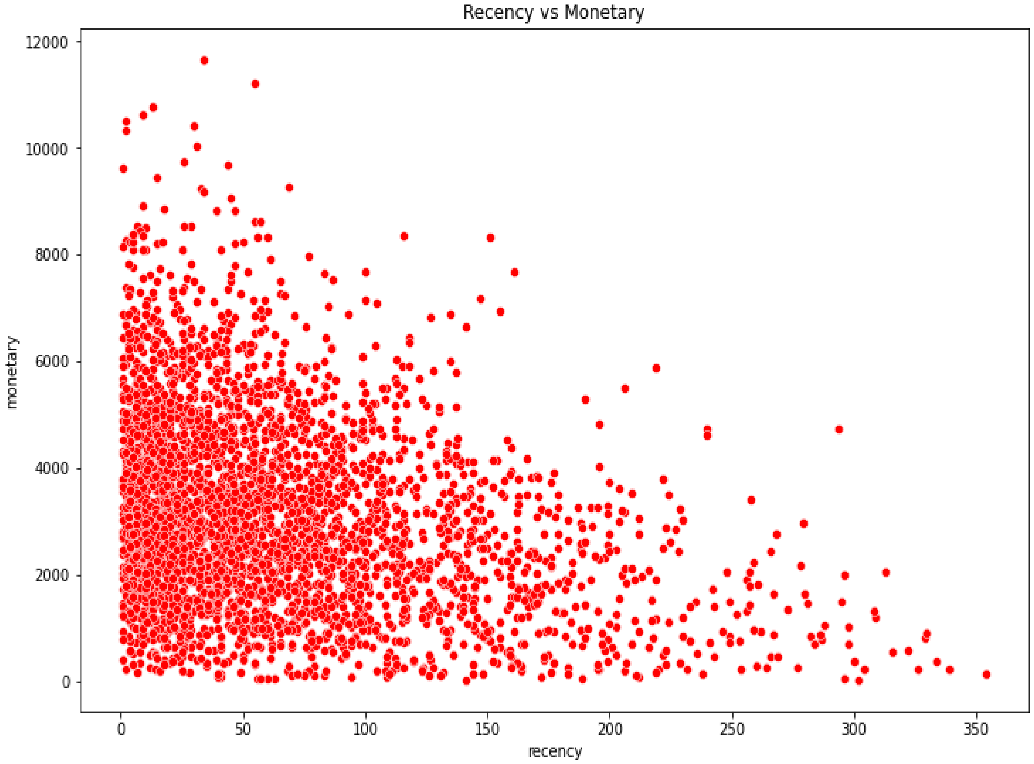
- Squarify plot gives us a clear picture of how our customers are segmented based on RFM Level.



Model Development

Recency Vs Monetary

- Customers who purchased more recently have generated more revenue than the customers who visited earlier.
- Customers who have made purchase between 2 to 5 months have generated a moderate revenue.
- Those with purchase history more than 7 months old generated low revenue.

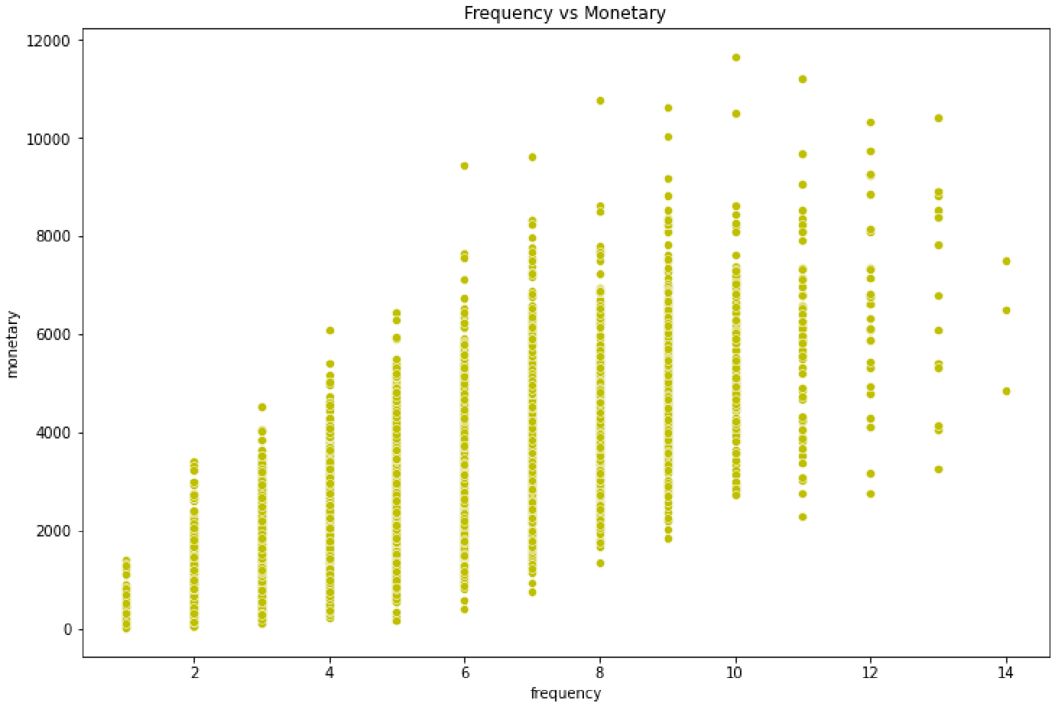


- ❑ X axis -> Recency
- ❑ Y axis -> Monetary

Model Development

Frequency Vs Monetary

- Frequency and Monetary are directly proportional to each other.
- As Frequency increases revenue also increases.
- Customers with frequent transactions made highest revenue.

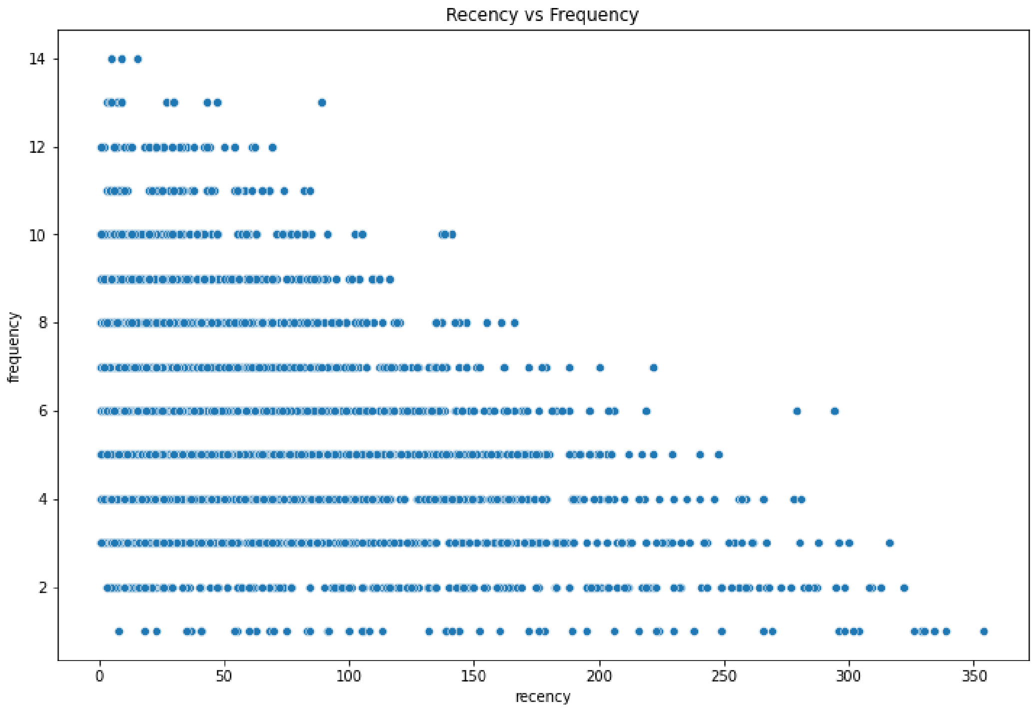


- ❑ X axis -> Frequency
- ❑ Y axis -> Monetary

Model Development

Recency Vs Frequency

- As recency increases frequency decreases. i.e., Customers who purchased more than 250 days ago are not visiting website frequently.
- Customers who have visited more recently have higher chance of visiting more frequently and make purchase.

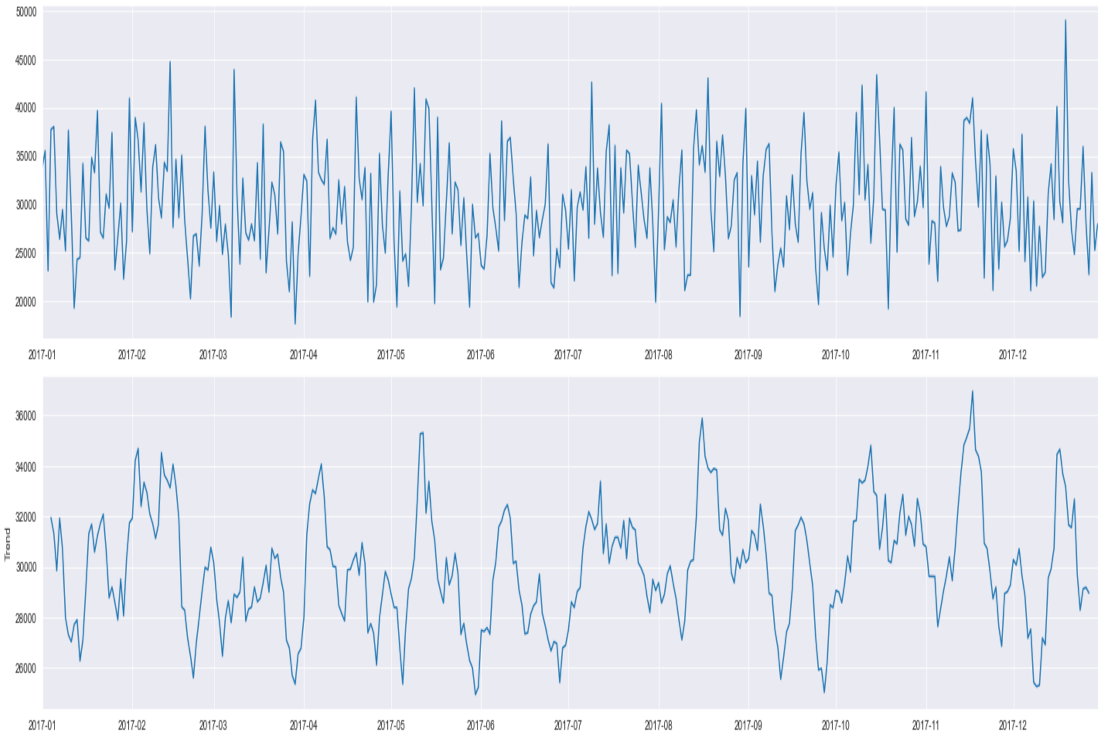


- ❑ X axis -> Recency
- ❑ Y axis -> Frequency

Model Development

Profit Analysis (Additive Model)

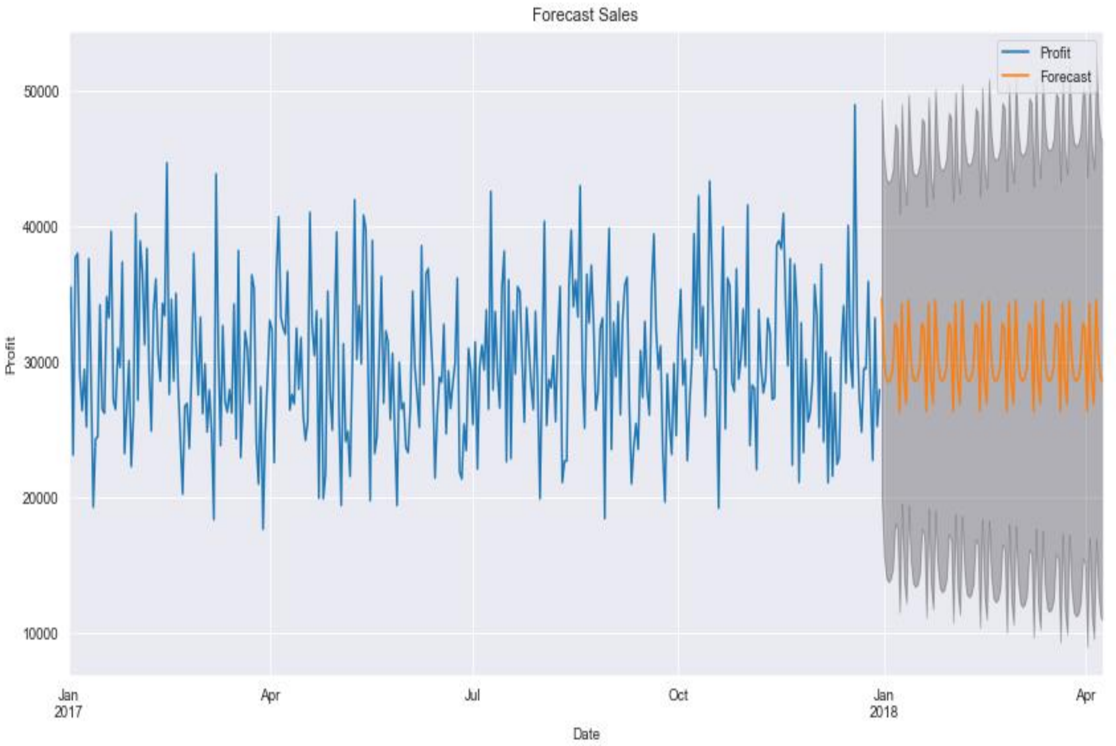
- This model gives us variation and Trend in a Time series related data.
- Sales were high in mid of every month.
- March, June and July recoded some of the lowest sales.
- November marked highest sale of all months.



Model Development

Profit Forecasting (ARIMA Model)

- This model helps us in predicting future profit based on past sales record.
- The model predicts stable profit based on previous sales record.
- Confidence interval of our prediction for revenue lies between 10,000 to 50,000 $.



- ❑ X axis -> Date
- ❑ Y axis -> Forecasted Profit

Interpretation

Top 1000 Customers to Target

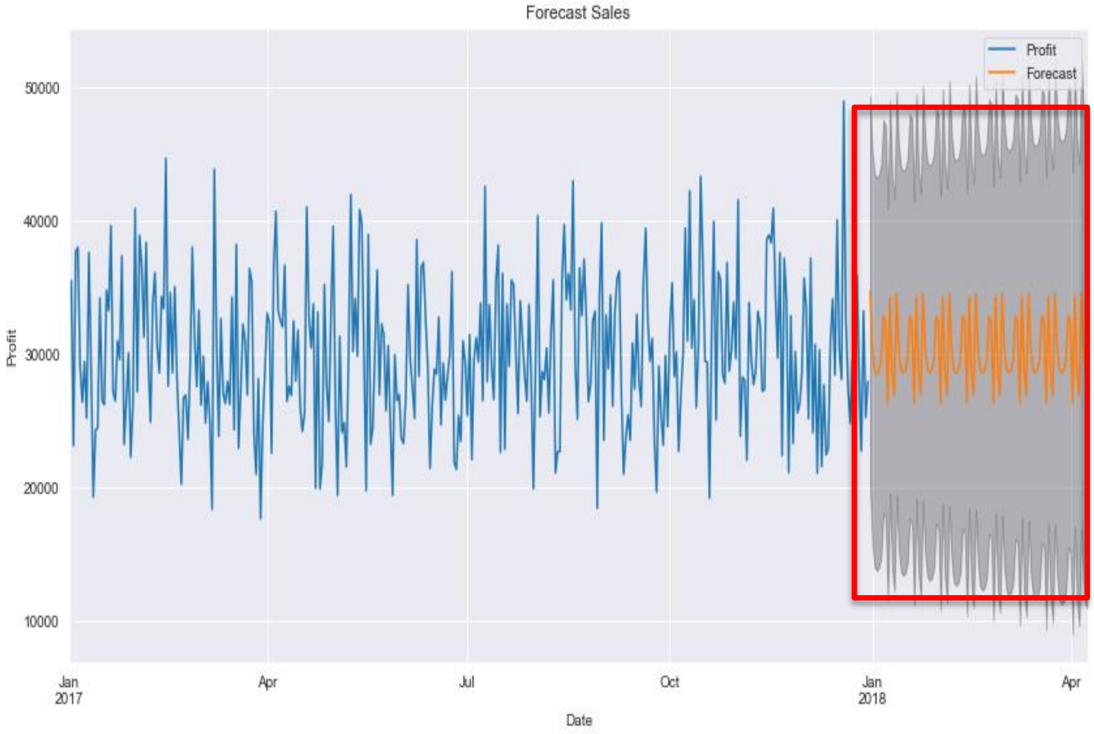
- The customers marked in the red box are the customers that should be targeted in the new customer list.
- Top thousand customers can be filtered using the conditions discussed above.
- These 1000 customers will have low recency, high frequency and high monetary.
- These customers will generate highest amount of the profit.

Rank	RFM Level	Total Customers	Selection of Customers
1	Top Most Priority	1161	1161
2	Champions	407	407
3	Loyal	470	470
4	Potential	458	458
5	Promising	391	391
6	Needs Attention	304	0
7	Require Activation	298	0

Interpretation

Future Profit Prediction

- By following and targeting our customers based on above mentioned approach we can maximize our profit.
- The above-mentioned approach ensures that our profit remains in high confidence interval limit.



- ❑ X axis -> Date
- ❑ Y axis -> Forecasted Profit

Appendix

Appendix

References

- Squarify (https://github.com/laserson/squarify)
- ARIMA Model (https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/)
- All the Analysis is done in Python (https://www.python.org/)