Geological Based Customer Analysis
Mahindra First Choice
Capstone Project

Presented by

Mr. Sushant S. Bahekar

(B.E., M. Tech.)

Content

- Business requirement
- EDA
- Modeling (Market segmentation approach)

  a. RFM

  b. CLTV
- Evaluate model
- Conclusion & Future scope
- References

Business requirement

- Problem statement 1:

Identifying the ownership patterns of cars throughout the country

- Problem statement 2:

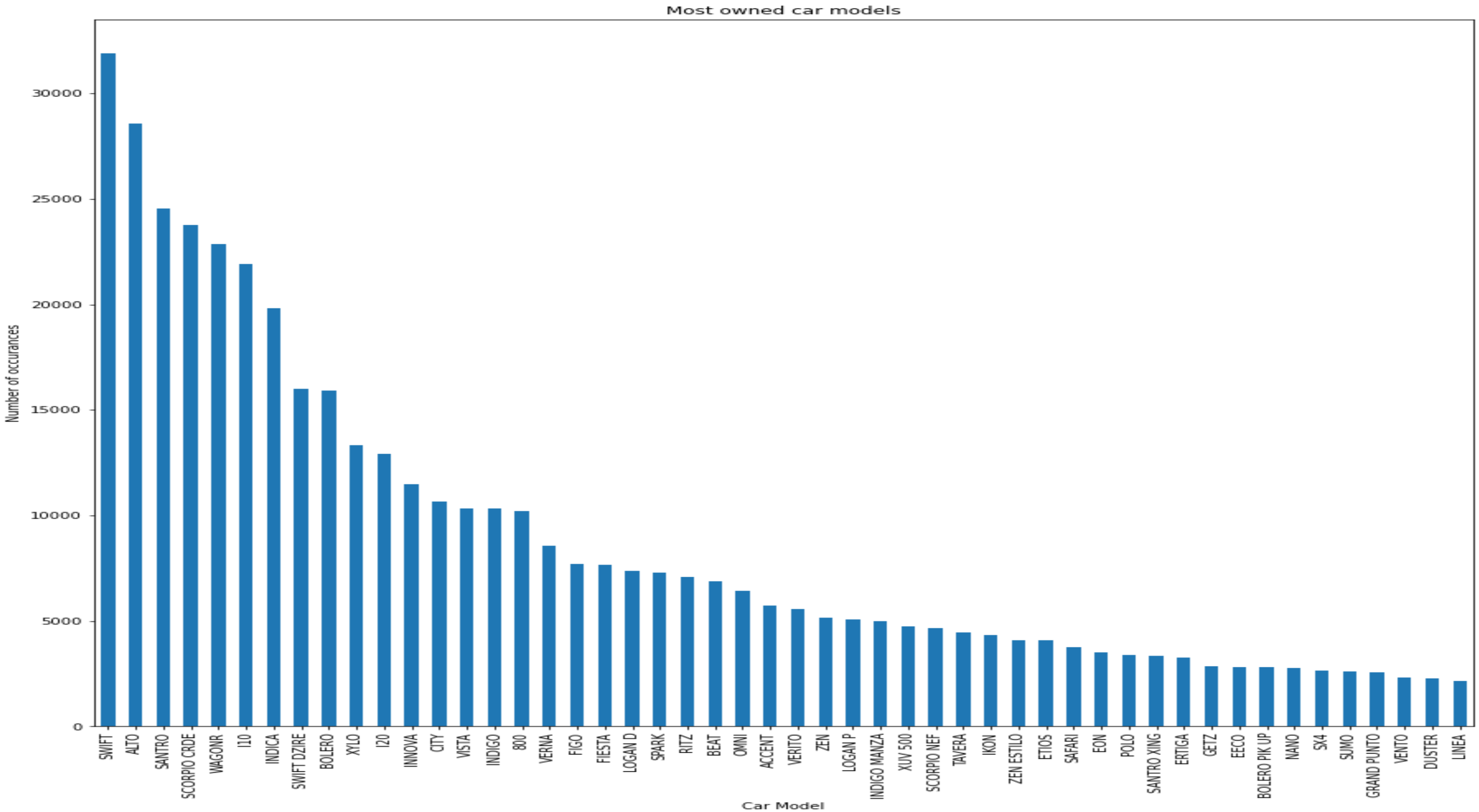
Identifying the type of order (Geography based business overview)

- Problem statement 3:

Market segmentation (Customer lifetime value prediction)

Problem statement 1: Ownership pattern

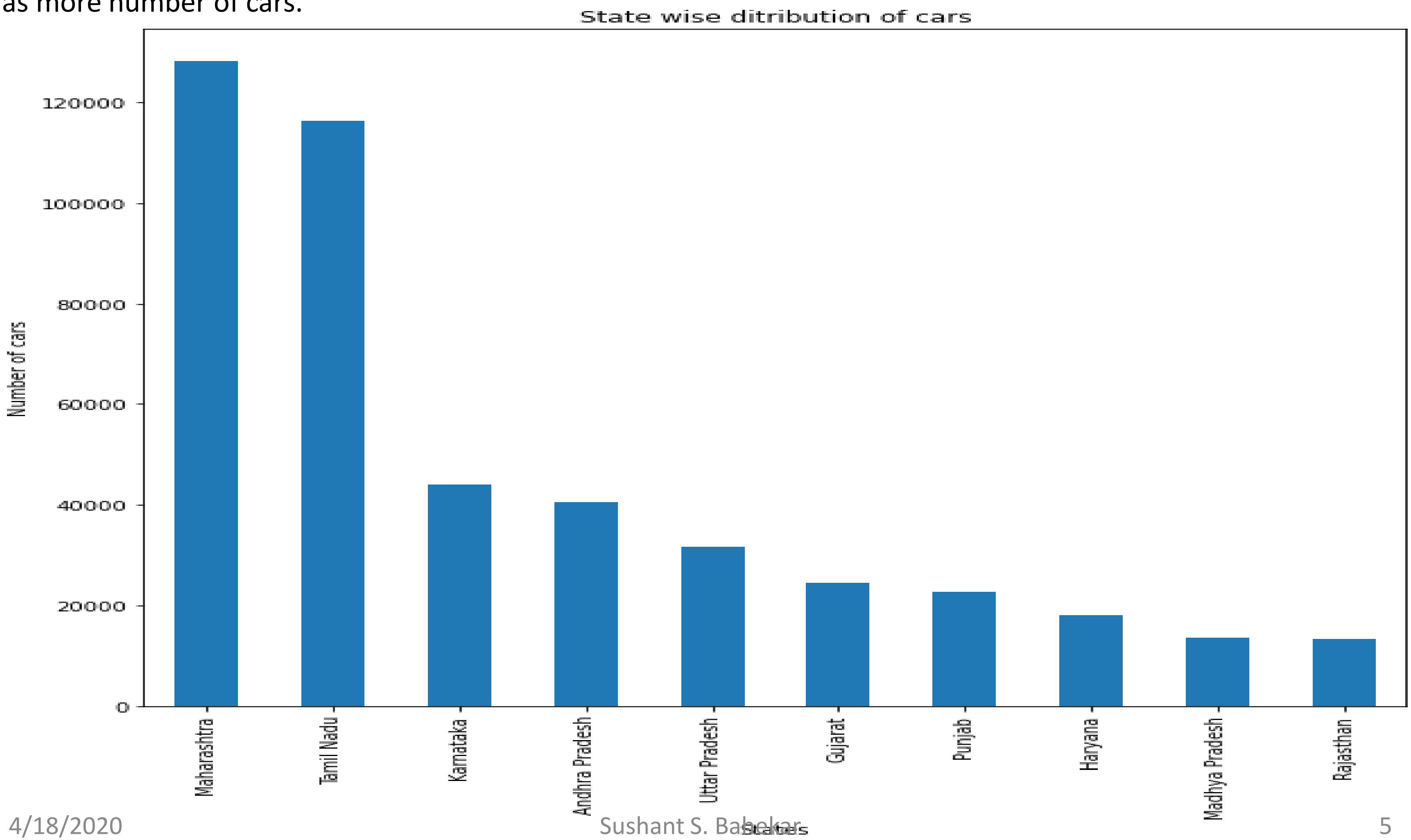
- This chart depicts the top 50 most owned car models throughout the country.
- Pattern suggested that economy priced hatchbacks are the most selling models and value conscious customers.
- We have analysed lowest owned models. Some of them include: GYPSY 1000, GYPSY 1300, PETRA, SIERRA, etc



Problem statement 1: Ownership pattern

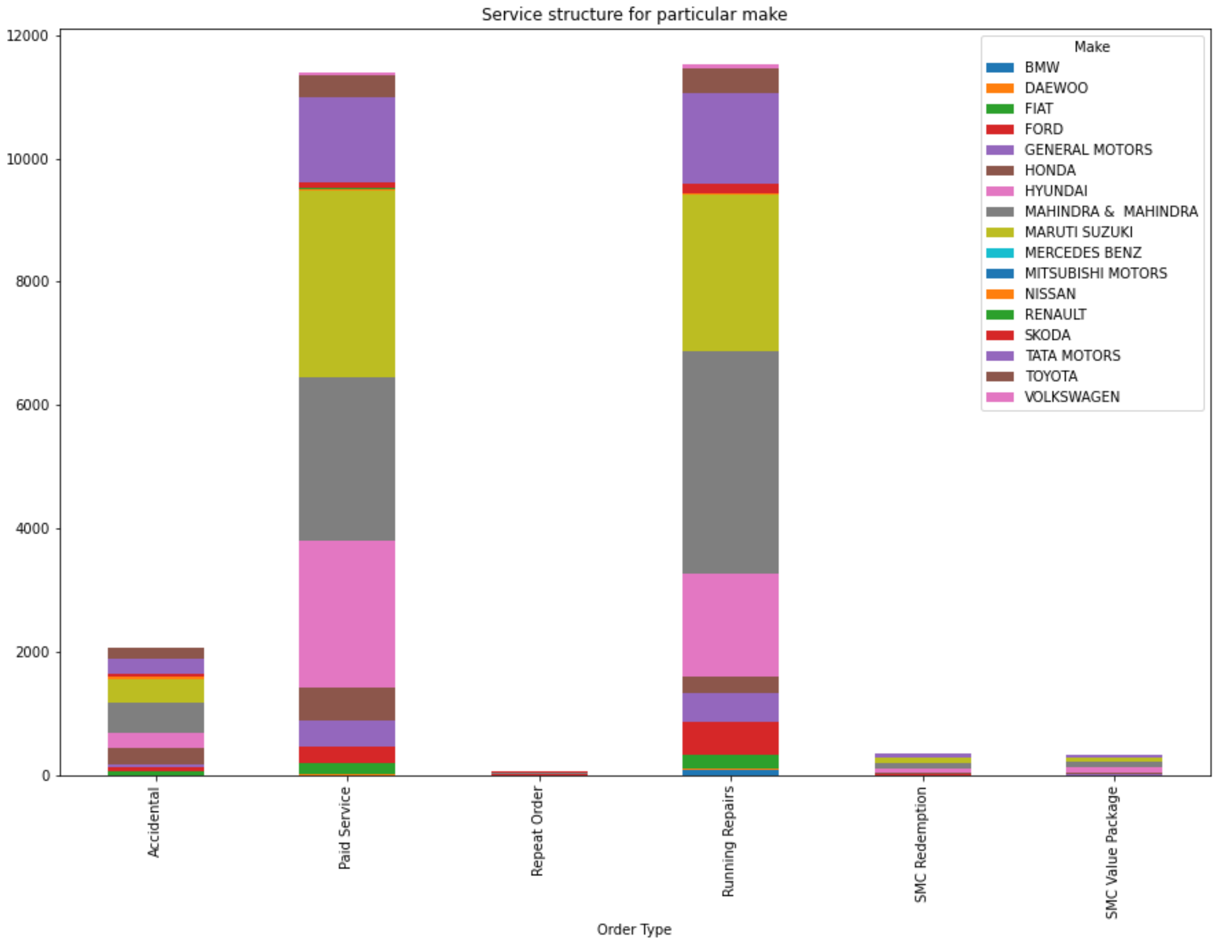
Which area has most cars?

Below chart depicts statewise wise distribution of cars which shows that Maharashtra, Tamilnadu, Karnataka state has more number of cars.



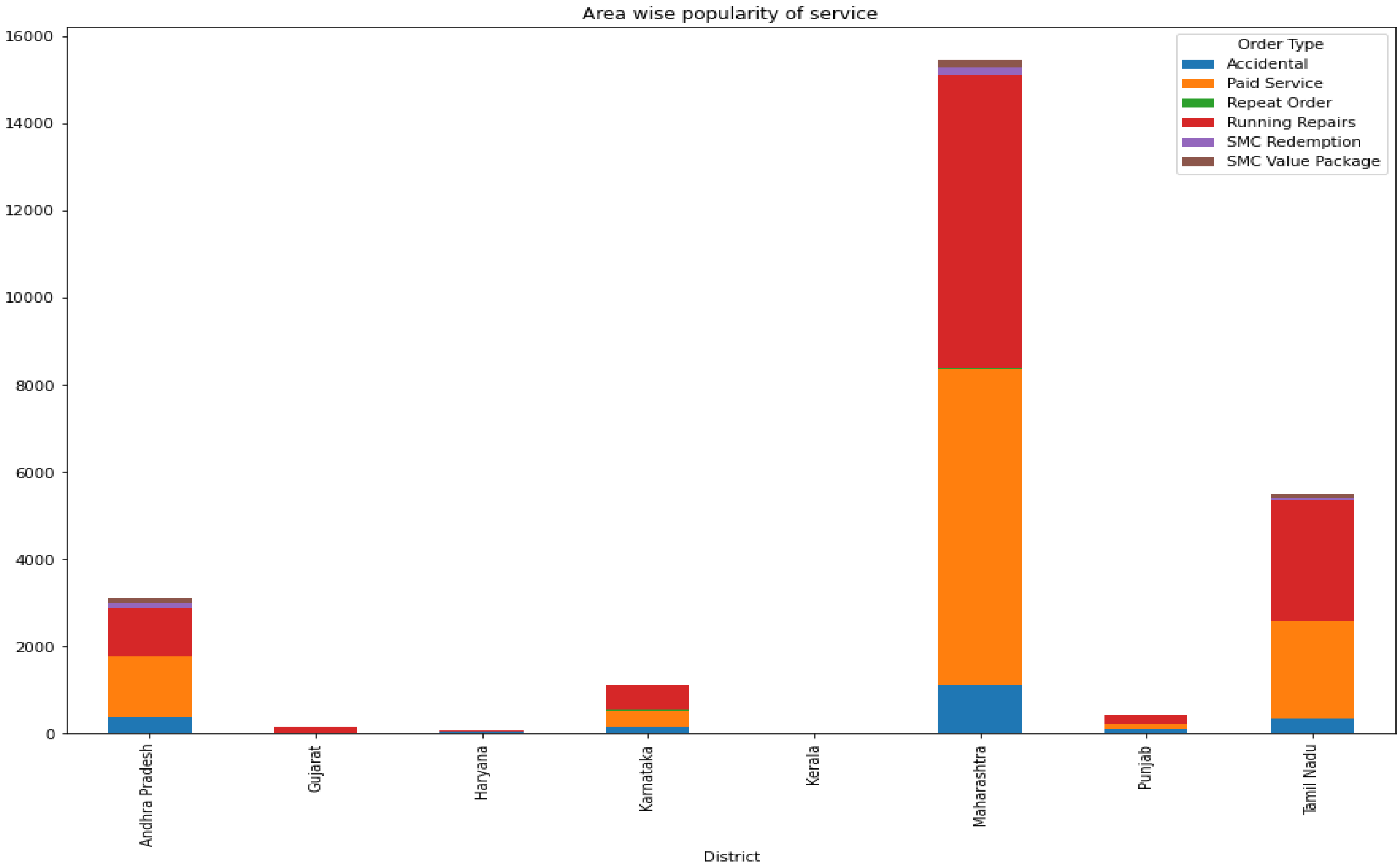
Problem statement 1: Ownership pattern

What is service structure for particular car/make?



Problem statement 1: Ownership pattern

Which type of service popular in certain area?



Problem statement 2: Geography based business overview)

Number of orders generated by top Districts

Maharashtra	128119
Tamil Nadu	116309
Karnataka	44135
Andhra Pradesh	40604
Uttar Pradesh	31534

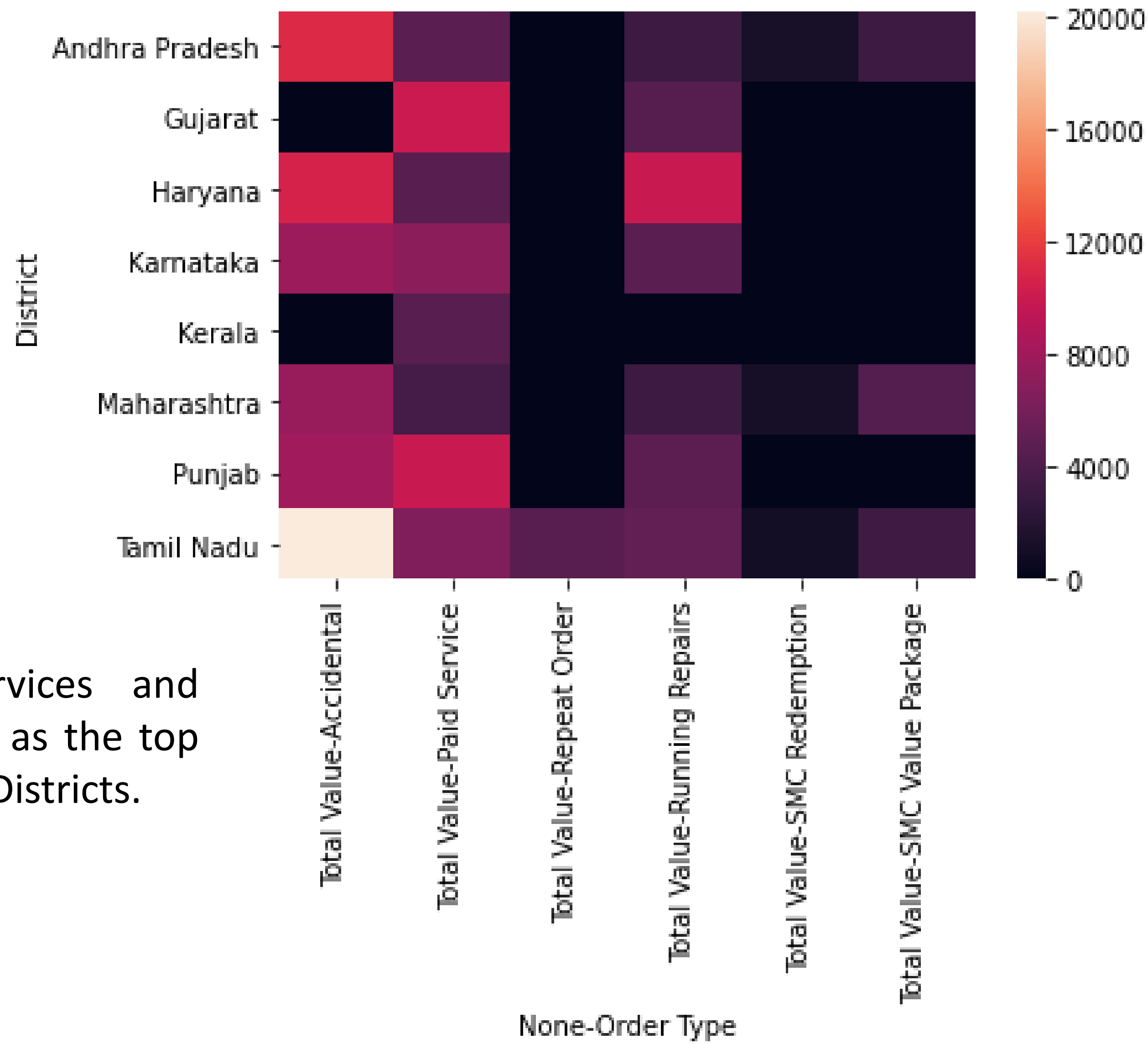
Ranking of Districts by Order Type orders

	Order Type	District
0	Accidental	Tamil Nadu
1	Mechanical	Tamil Nadu
2	Paid Service	Maharashtra
3	Repeat Order	Maharashtra
4	Running Repairs	Tamil Nadu
5	SMC Redemption	Maharashtra
6	SMC Value Package	Maharashtra
7	WBW Order	Punjab
8	Workshop Damage	Maharashtra

Problem statement 2: Geography based business overview)

- Heat Map allows us to visualize and rank the top Districts in terms of Revenue as follow:

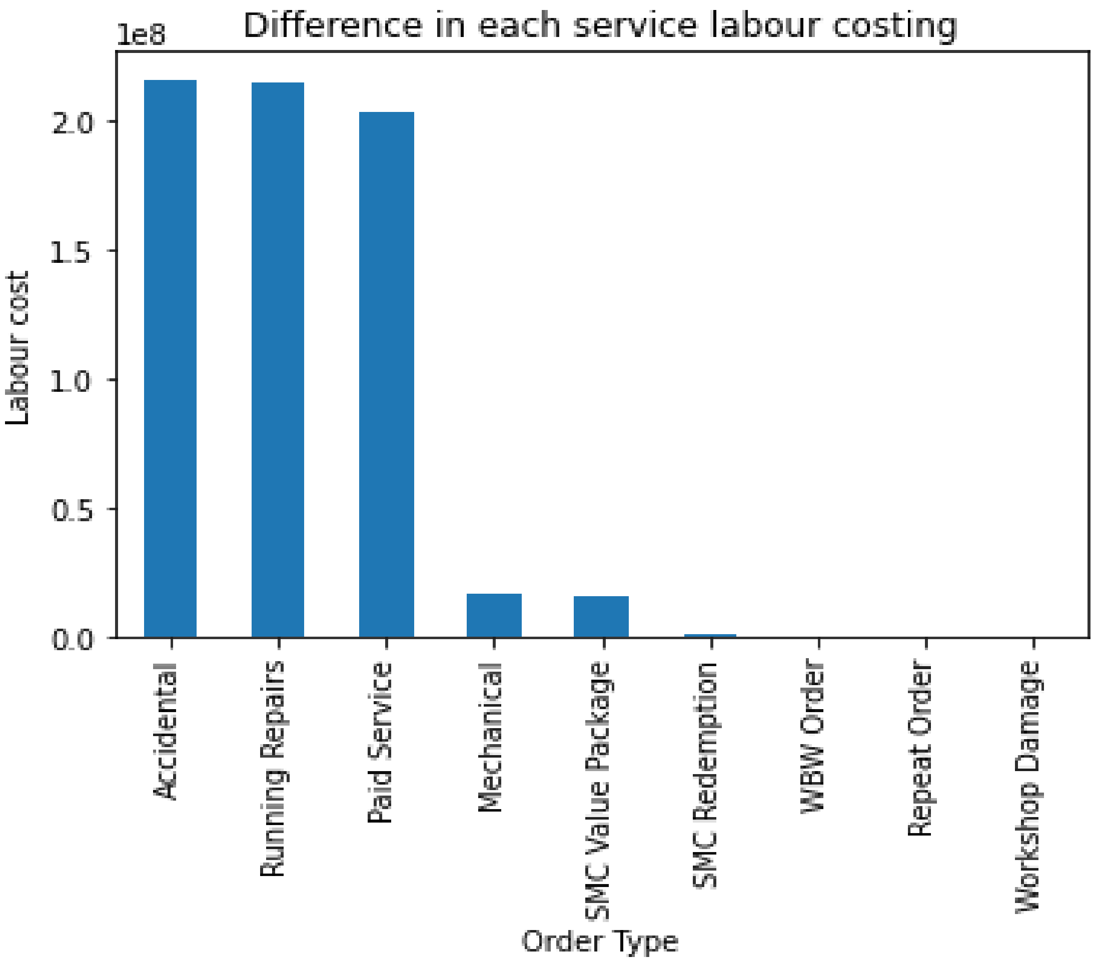
1. Tamil Nadu
2. Maharashtra
3. Karnataka
4. Andhra Pradesh



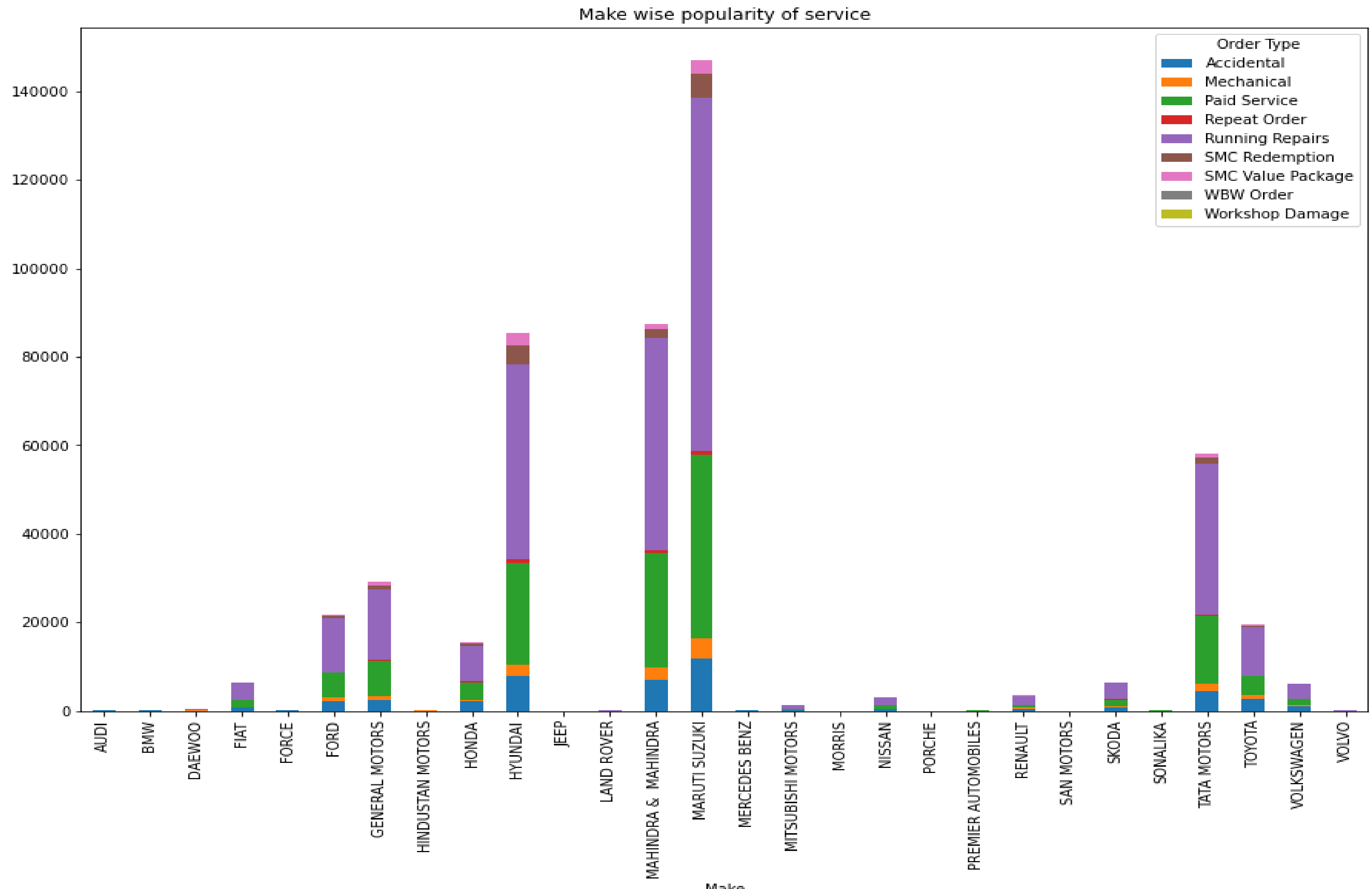
- Accidental, Paid Services and Running repairs ranks as the top order types across all Districts.

Problem statement 2: Geography based business overview)

Order Type	Labour Total
Accidental	215715368
Mechanical	16390034
Paid Service	203348433
Running Repairs	214775678
SMC Redemption	716801
SMC Value Package	15275125
WBW Order	26983



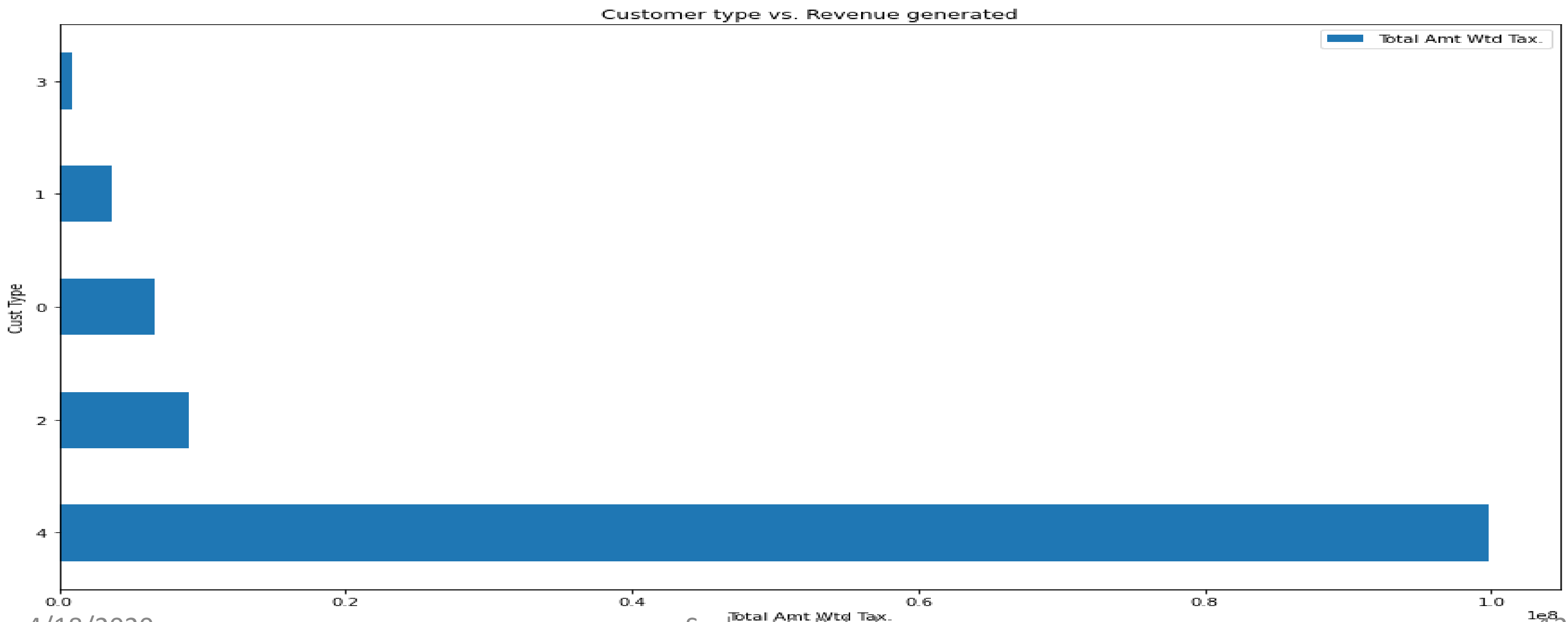
Problem statement 2: Geography based business overview)



Order Type and Customer Type Overview

-Although “Retail” Customer type is the main revenue source.

Cust Type	Total Amt Wtd Tax.
Retail	9.978584e+07
Fleets	9.053690e+06
Corporate others	6.676231e+06
Corporate- M&M	3.722937e+06
MFCWL	9.072157e+05



Modeling
(Market segmentation approach)

1. RFM Score Calculations

- RECENCY (R): Days since last purchase
- FREQUENCY (F): Total number of purchases
- MONETARY VALUE (M): Total money this customer spent

2. Customer Life Time Value (CLTV)

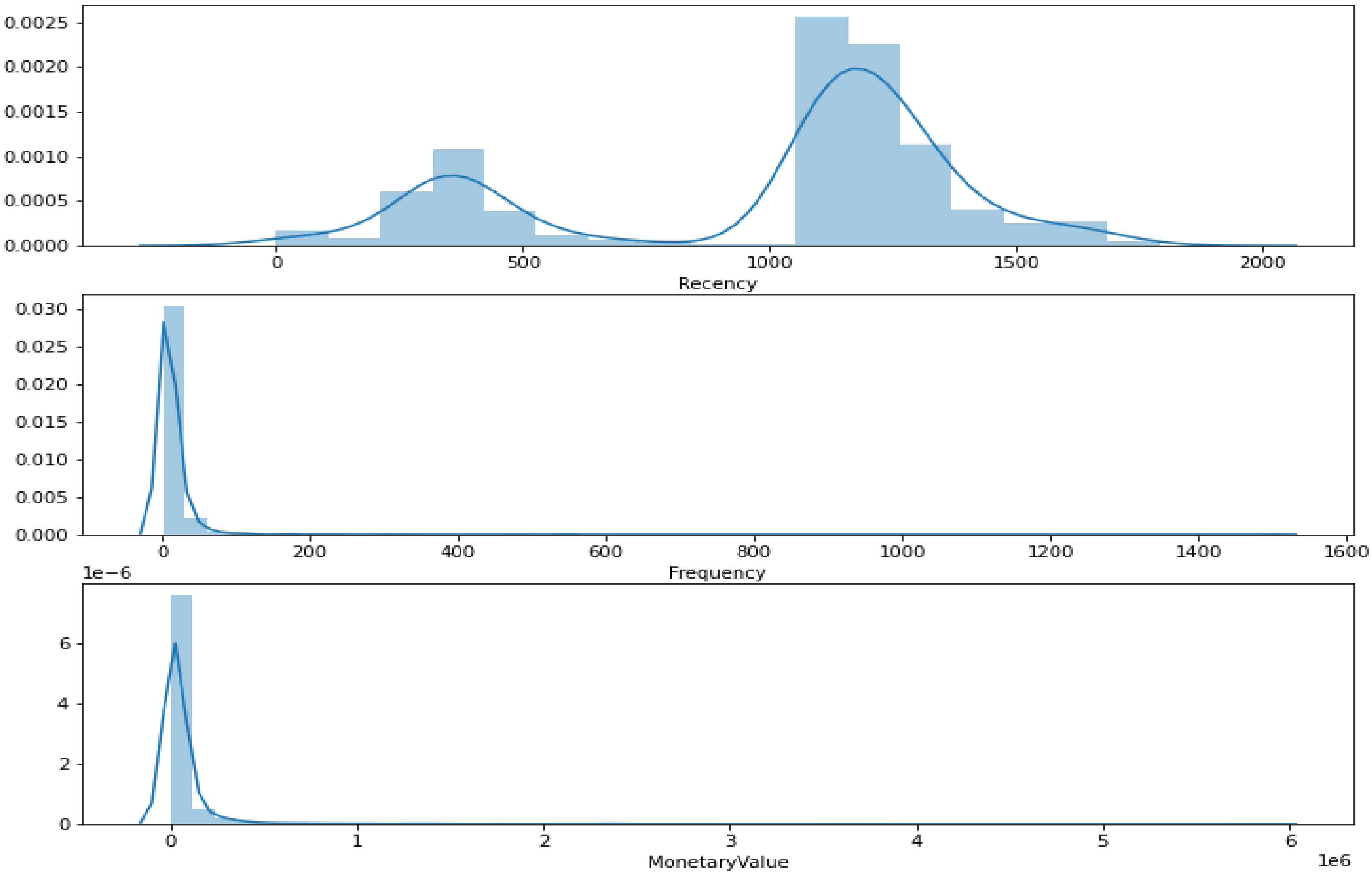
RFM Score Calculations

The first we need to do is to sort customers based on recency, frequency, and monetary values

Customer No.	Recency	Frequency	MonetaryValue
E10007	1336	3	3251.31
E10013	1085	3	3003.87
E10014	1323	1	785.21
E10023	376	4	597.56
E10052	372	43	141718.15

let’s examine the distribution of our Recency, Frequency, and Monetary.

RFM Score Calculations



RFM Score Calculations

We create a 4 labels for our f_labels, where 4 is the "best" quantile. We do the same for our f_label. We then create new columns "R" and "F" and assign the r_group and f_group values to them respectively.

	Recency	Frequency	MonetaryValue	R	F
Customer No.					
E10007	1336	3	3251.31	1	1
E10013	1085	3	3003.87	3	1
E10014	1323	1	785.21	1	1
E10023	376	4	597.56	4	2
E10052	372	43	141718.15	4	4

Next, we do the same for our monetary value by grouping the values into 4 quantiles using .qcut() method. Finally, with these 3 scores in place, R, F, and M, we can create our first RFM segment by concatenating the values together below.

	Recency	Frequency	MonetaryValue	R	F	M	RFM_Segment_Concat
Customer No.							
E10007	1336	3	3251.31	1	1	1	111
E10013	1085	3	3003.87	3	1	1	311
E10014	1323	1	785.21	1	1	1	111
E10023	376	4	597.56	4	2	1	421
E10052	372	43	141718.15	4	4	4	444

RFM Score Calculations

Summing the Score & find out RFM level

One of the most straightforward methods is to sum our scores to a single number and define RFM levels for each score range.

We can get creative and hypothesize about what each score range entails & we can determine RFM level

	Recency	Frequency	MonetaryValue	R	F	M	RFM_Segment_Concat	RFM_Score	RFM_Level
Customer No.									
E10007	1336	3	3251.31	1	1	1	111	3.0	Require Activation
E10013	1085	3	3003.87	3	1	1	311	5.0	Promising
E10014	1323	1	785.21	1	1	1	111	3.0	Require Activation
E10023	376	4	597.56	4	2	1	421	7.0	Loyal
E10052	372	43	141718.15	4	4	4	444	12.0	Can't Loose Them

RFM Score Calculations

Finally, we can then group our customers by their RFM level.

	Recency mean	Frequency mean	Monetary Value mean	count
RFM_Level				
Can't Loose Them	**730.8**	**30.2**	**157157.7**	**660**
Champions	**1005.5**	**9.2**	**31536.6**	**207**
Loyal	**1049.7**	**7.5**	**20757.5**	**210**
Needs Attention	1268.7	2.0	4178.7	191
Potential	1057.1	4.7	11554.0	223
Promising	1226.0	3.2	7159.5	230
Require Activation	1365.4	1.4	2119.3	111

From here, we can see that a large percentage (~60%) of our customers are in the top tier RFM levels. The store must be doing something right to be maintaining their loyalty!

The other 40% will need some work. Let’s explore using some ads to re-target them:

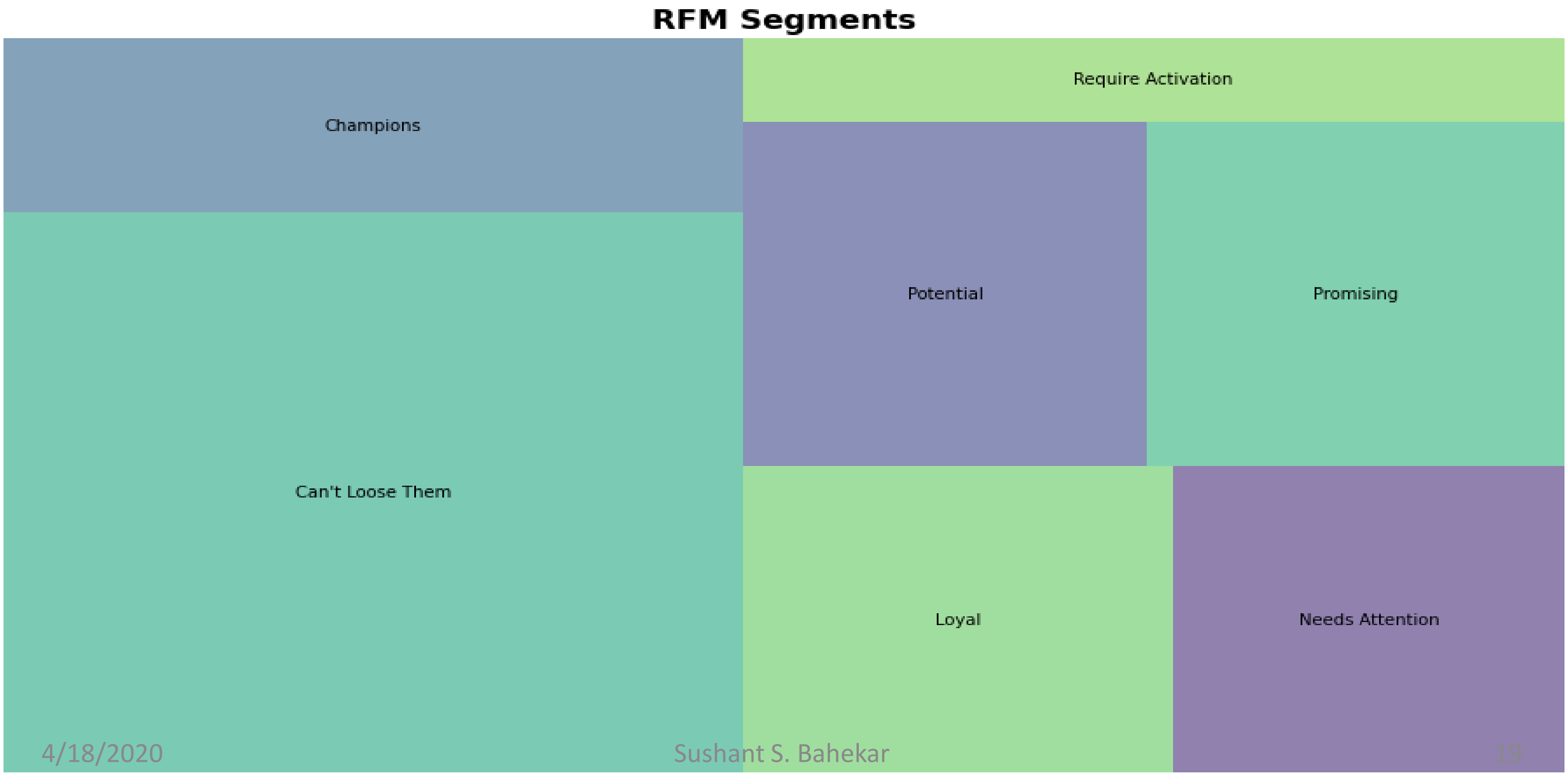
Potential : High potential to enter our loyal customer segments, why not throw in some freebies on their next purchase to show that you value them!

Promising : Showing promising signs with quantity and value of their purchase but it has been a while since they last bought sometime from you. Let’s target them with their wish list items and a limited time offer discount.

RFM Score Calculations

Needs Attention : Made some initial purchase but have not seen them since. Was it a bad customer experience? Or product-market fit? Let's spend some resource build our brand awareness with them.

Require Activation : Poorest performers of our RFM model. They might have went with our competitors for now and will require a different activation strategy to win them back.



Customer Life Time Value (CLTV)

- CLTV = ((Average Order Value * Purchase Frequency)/Churn Rate) * Profit margin.
- Customer Value = Average Order Value (AOV)* Purchase Frequency
- AOV = Total Revenue / Total Number of Orders.
- Purchase Frequency = Total Number of Orders / Total Number of Customers.
- Churn Rate: Churn Rate is the % of customers who have not ordered again.
- Customer Lifetime = 1/ churn rate
- Churn Rate= 1-Repeat Rate

Customer Life Time Value (CLTV)

Let us see,

(a) the number of days between the present date and the date of last purchase from each customer,

(b) the number of orders for each customer

(c) the sum of purchase price for each customer.

	num_days	**num_transactions**	**spent_money**
Customer No.			
E10007	1792	3	3251.31
E10013	1792	3	3003.87
E10014	1792	1	785.21
E10023	1792	4	597.56
E10052	1792	43	141718.15

Customer Life Time Value (CLTV)

Now, we shall compute CLTV using the formula mentioned in the beginning.

CLTV = [AOV * Purchase Frequency/Churn Rate] Profit margin.

Customer Value = Average Order Value * Purchase Frequency

Purchase Frequency: 14.055131004366812

repeat_rate: 0.8395196506550219

churn_rate 0.16048034934497812

Profit Margin & CLTV

Let's assume that business making 25% profit.

	num_days	num_transactions	spent_ money	avg_order value	profit_margin	CLTV	cust_lifetime_ value
Customer No.							
E10007	1792	3	3251.31	1083.77000	812.8275	94918.346020	7.715224e+07
E10013	1792	3	3003.87	1001.29000	750.9675	87694.612959	6.585580e+07
E10014	1792	1	785.21	785.21000	196.3025	68769.973776	1.349972e+07
E10023	1792	4	597.56	149.39000	149.3900	13083.820102	1.954592e+06
E10052	1792	43	141718.15	3295.77093	35429.5375	288648.99892	1.022670e+10

Prediction Model for CLTV

Now we predict CLTV using Linear Regression Model.

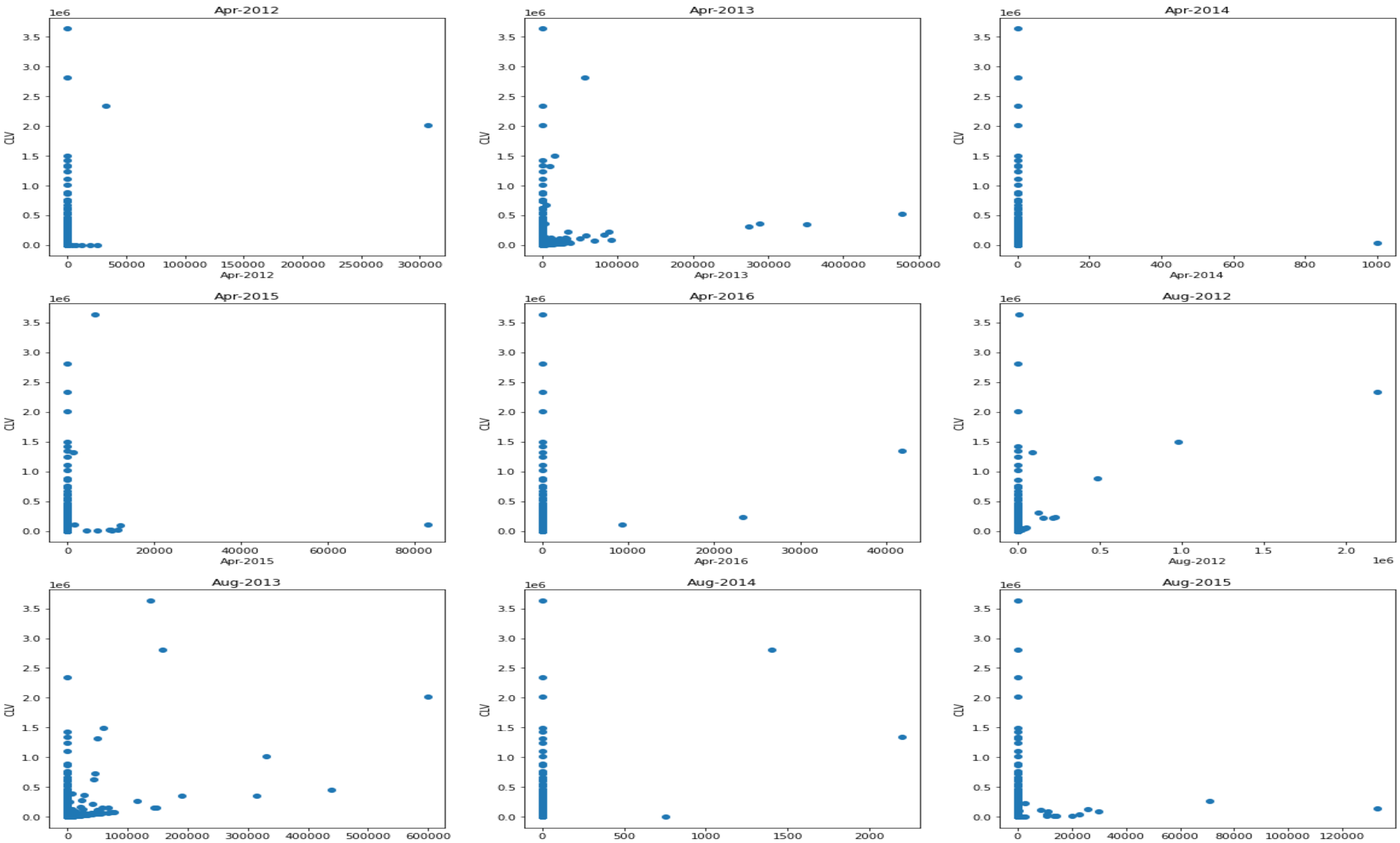
We build a regression model for existing customers. We need a set of dependent (y) and independent variables (X). So, we took last five years data as independent variables and total revenue over five years as a dependent variable followed by usual ML process of splitting training & test data.

Code:

```
X = sale.drop(['Customer No.', 'CLV'], axis = 1)

y = sale['CLV']

X_train, X_test, y_train, y_test = train_test_split(X, y,test_size = 0.3,random_state=6)

cols = X_train.columns
```

Prediction Model for CLTV

Let's check the scatter_plot for different features vs target variable CLV



Model evaluation

mean_squared_error: 28861.934336179806

r2_score: 0.9999996635779342

mean_absolute_error: 9.490854546364975

root_mean_squared_error: 169.88800527459202

Conclusion & Future work

When the model is deployed, the ML algorithm helps to understand the patterns, it will also categorize the customers according to their CLTV predictions. The marketing strategy is important here considering that CLTV can figure out most profitable customers, but how we are going to make profit from them, will depend on the adopted marketing strategy. Moreover loyalty programmes can be formulated based on the insights.

The RFM was close to perfect here predicting what an average customer will spend during their lifetime. CLTV helps to design an effective business plan and also provide a chance to scale the business. However, as already discussed, a lot will depend on marketing strategy to extract profit.

References

- Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. Journal of interactive marketing, 23(1), 61–69.
- Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). “Counting your customers” the easy way: An alternative to the Pareto/NBD model. Marketing science, 24(2), 275–284.
- https://towardsdatascience.com/recency-frequency-monetary-model-with-python-and-how-sephora-uses-it-to-optimize-their-google-d6a0707c5f17
- https://towardsdatascience.com/measuring-users-with-customer-lifetime-value-cltv-94fccb4e532e
- https://medium.com/@henryfeng/know-users-behaviors-better-with-cohort-analysis-in-python-6c0dfc373963
- https://www.kaggle.com/vjchoudhary7/kmeans-clustering-in-customer-segmentation