MAHINDRA FIRST CHOICE SERVICES REPORT

Created by:
Cyberbots

Problem Statement

Geo location Based Analysis

Generate meaningful insights on Geolocation based customer & revenue analysis

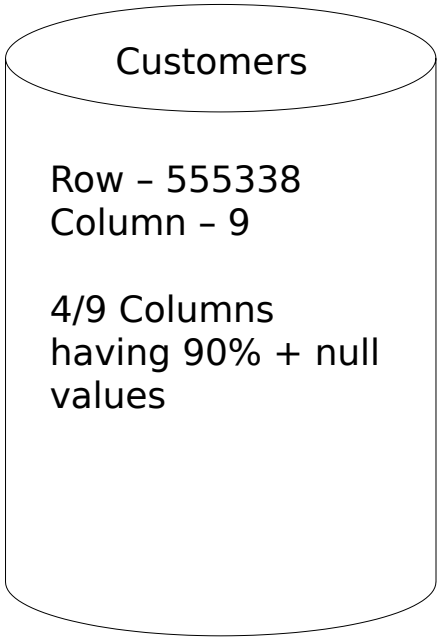
Market Segmentation

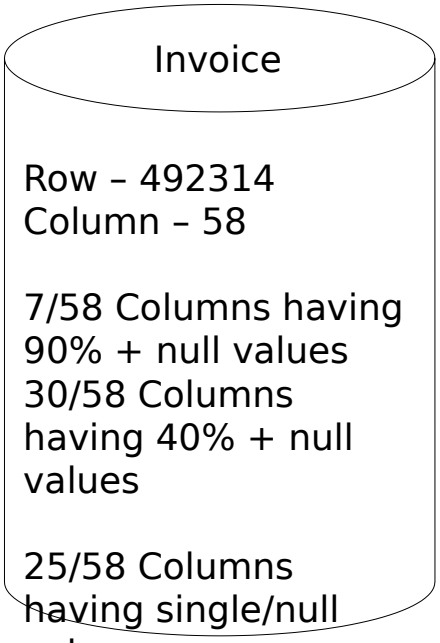
Divide the customer base into segments based on their car type, service type & total spend such that these customers will respond similarly to the different marketing campaigns.

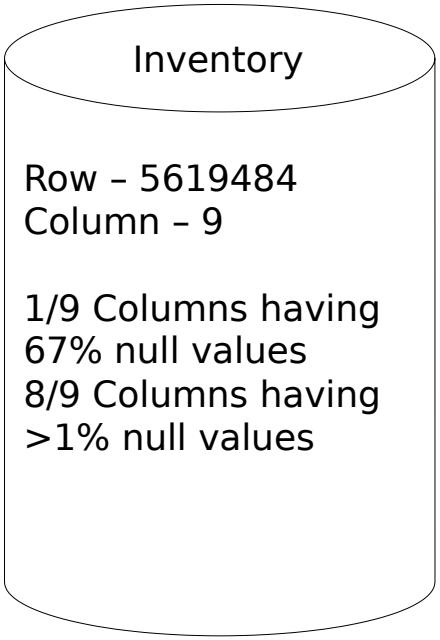
Customer Lifetime Value Prediction

Life time value analysis of customers to maintain relationship with high profit generating customers in future & finding potential customers from the population.

Datasets



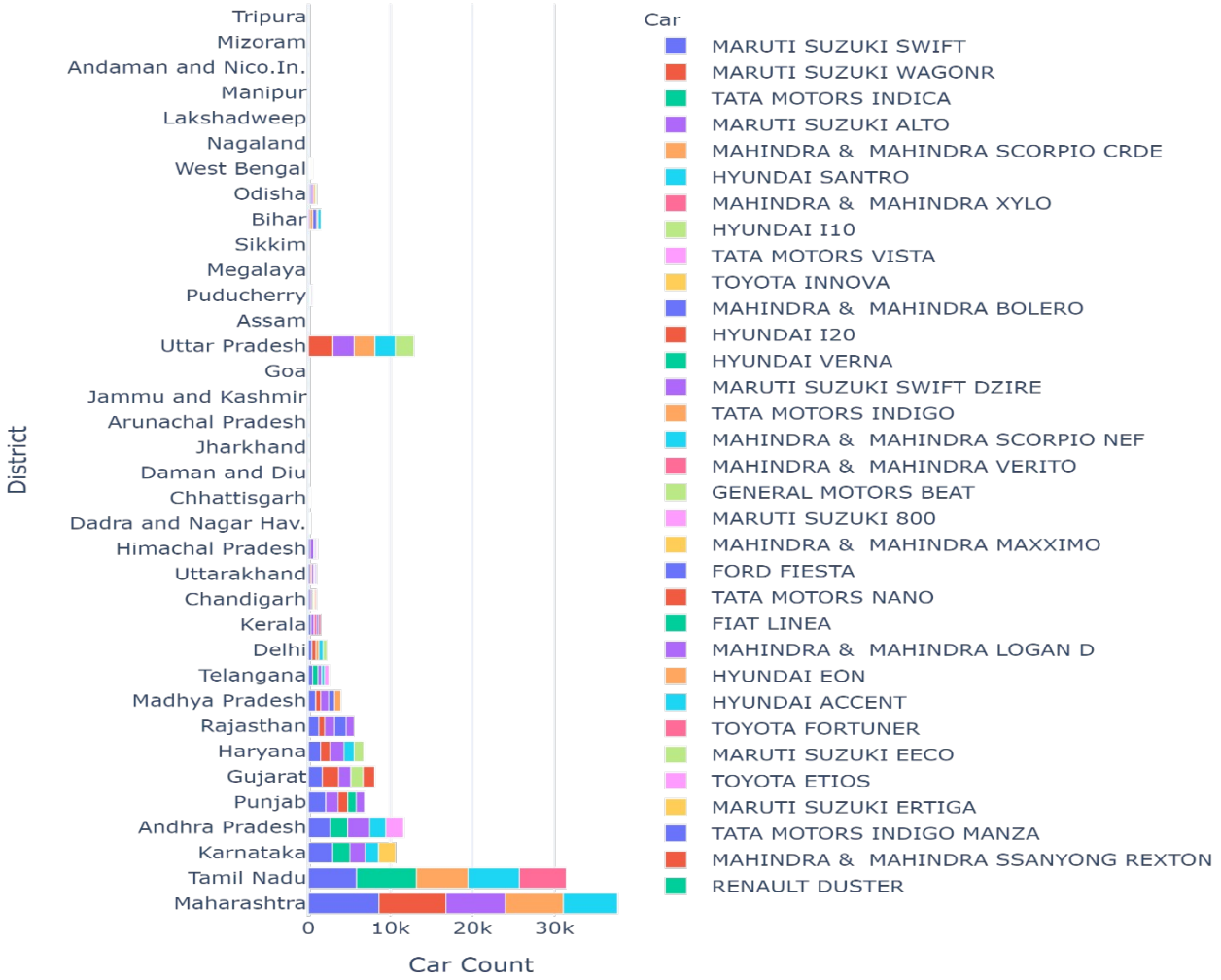




- Dataset is provided in excel and csv format

Problem Statement 1 : Identifying the ownership pattern

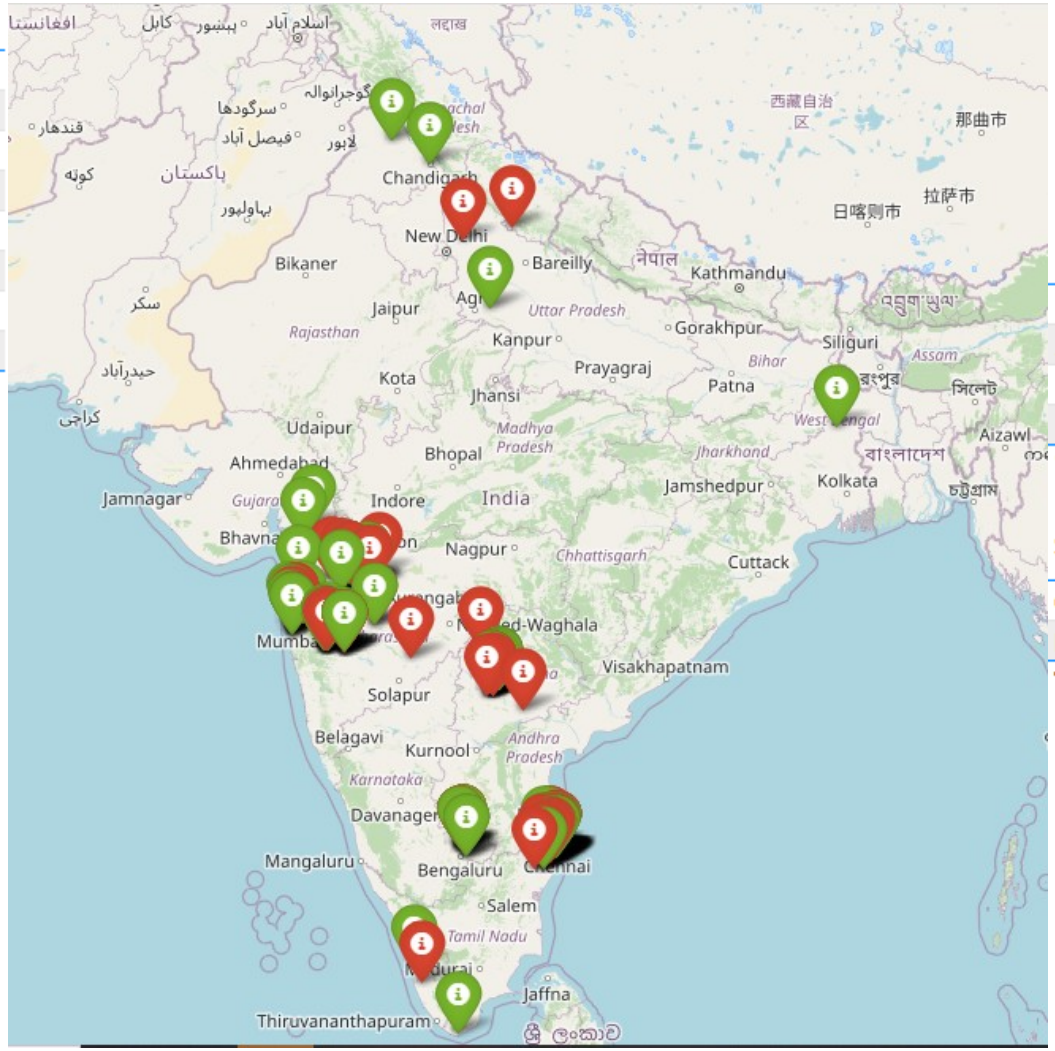
- This chart depicts car count by each state
- 90% of the car belongs to first 10 States
- Chart suggest that economical cars are most popular
- Avigo, Z4, Pushpak, Estilo, Trailblazer are least owned cars at count of 1



Plant Analysis

State	Plant
Chandigarh	1
Delhi	1
Haryana	20
Himachal Pradesh	5
Jammu and Kashmir	2
Punjab	16
Uttar Pradesh	46
Uttarakhand	9
Total	**100**

State	Plant
Goa	1
Gujarat	13
Karnataka	31
Maharashtra	48
Rajasthan	52
Total	**145**



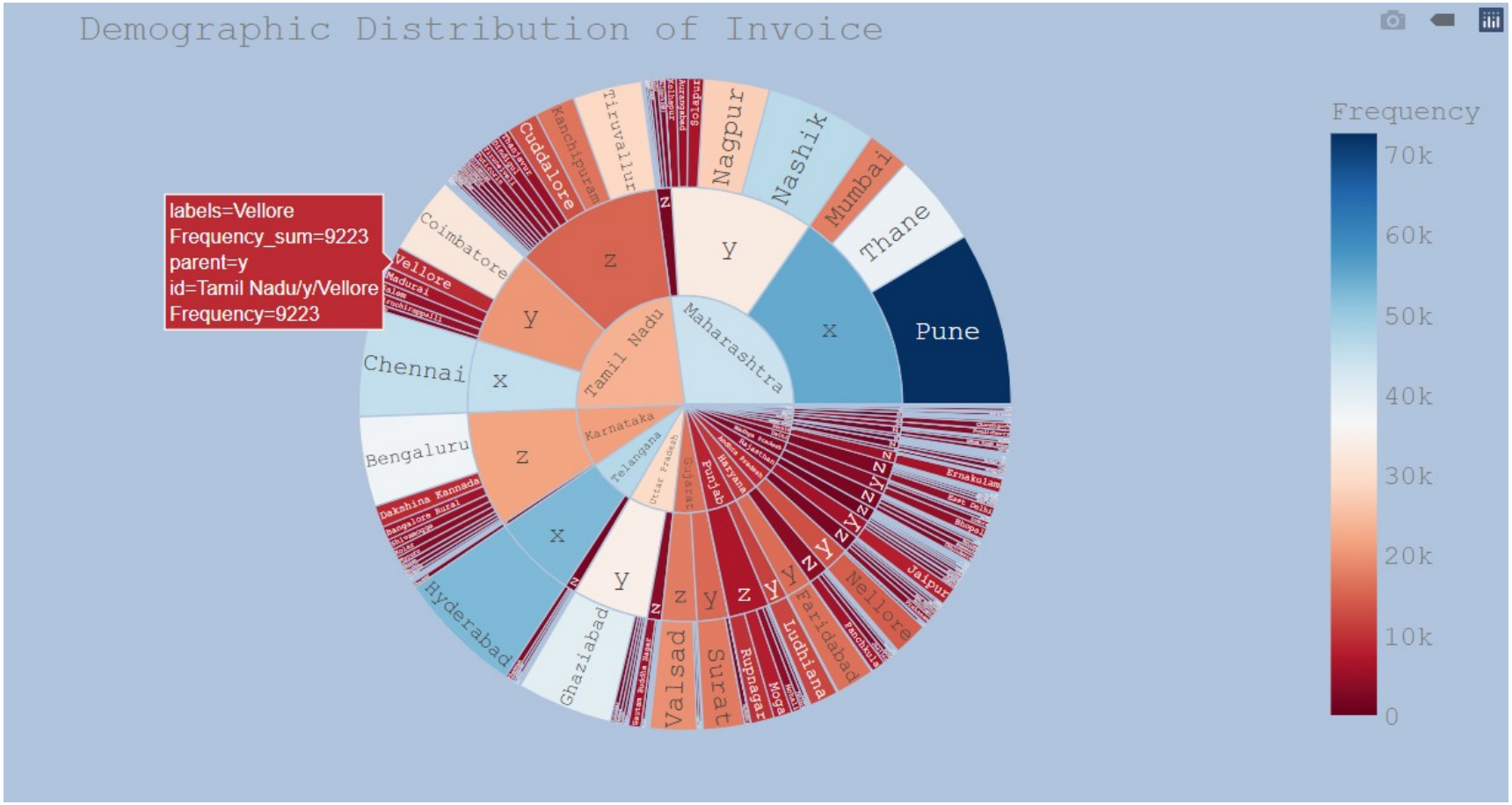
State	Plant
Assam	7
Manipur	1
Nagaland	1
Total	**9**

State	Plant
Bihar	21
Jharkhand	12
Odisha	13
West Bengal	14
Total	**60**

State	Plant
Chhattisgarh	5
Madhya Pradesh	35
Total	**40**

State	Plant
Tamil Nadu	47
Andhra Pradesh	15
Telangana	13
Kerala	6
Puducherry	3
Total	**84**

Customer Analysis

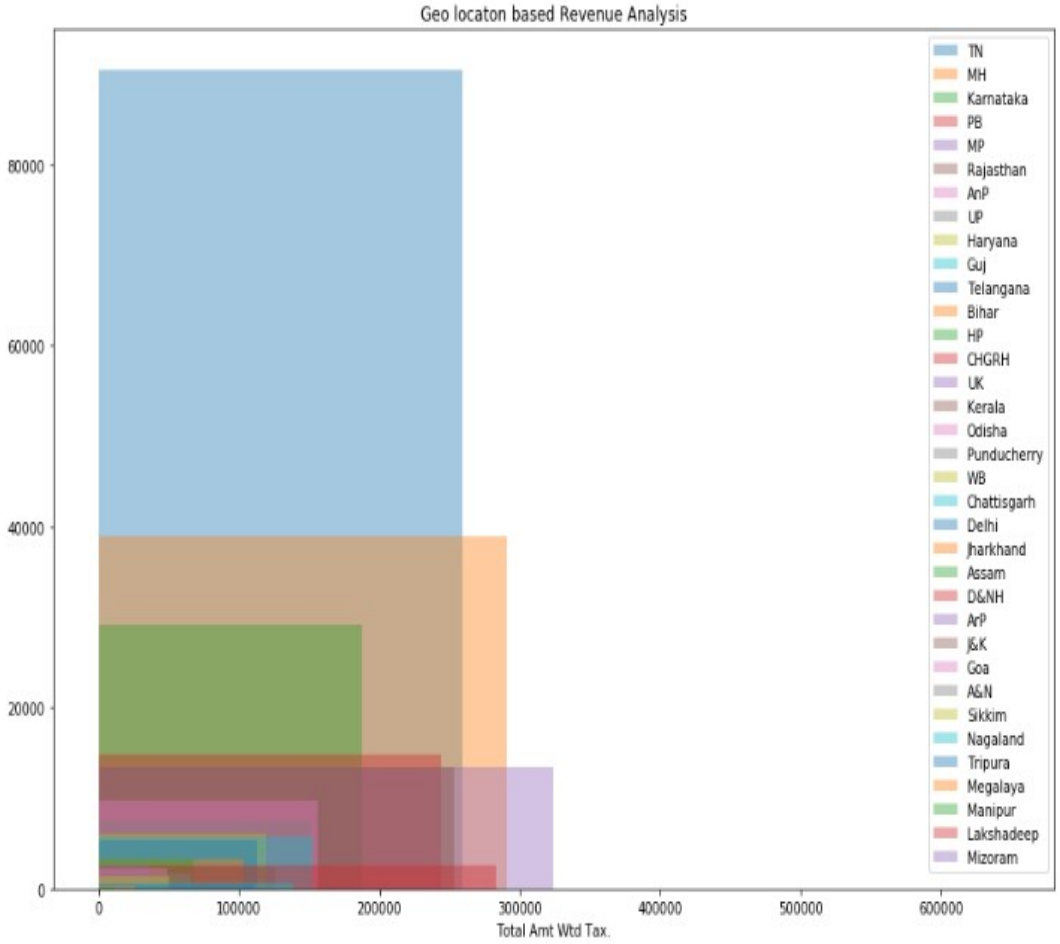


State wise revenue generation

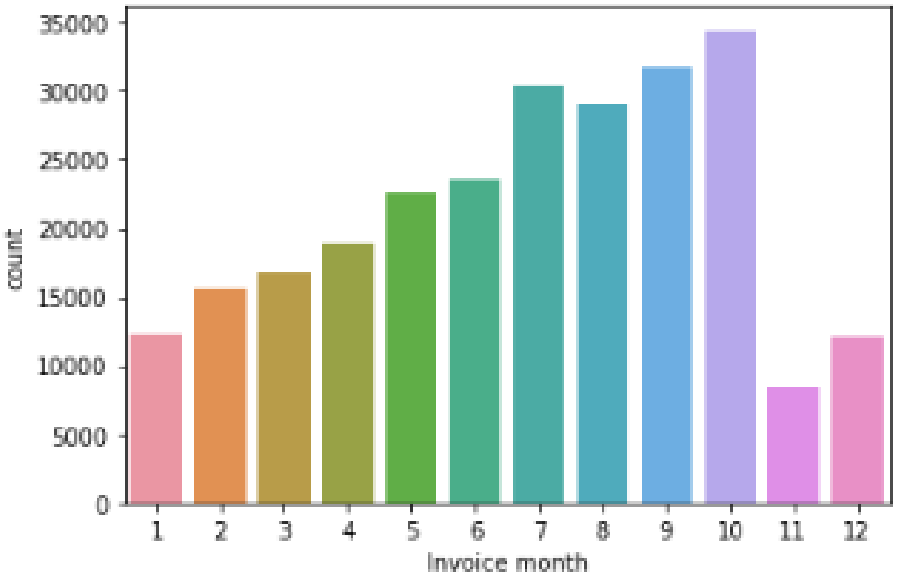
- Tamilnadu has highest customer base followed by Maharashtra.

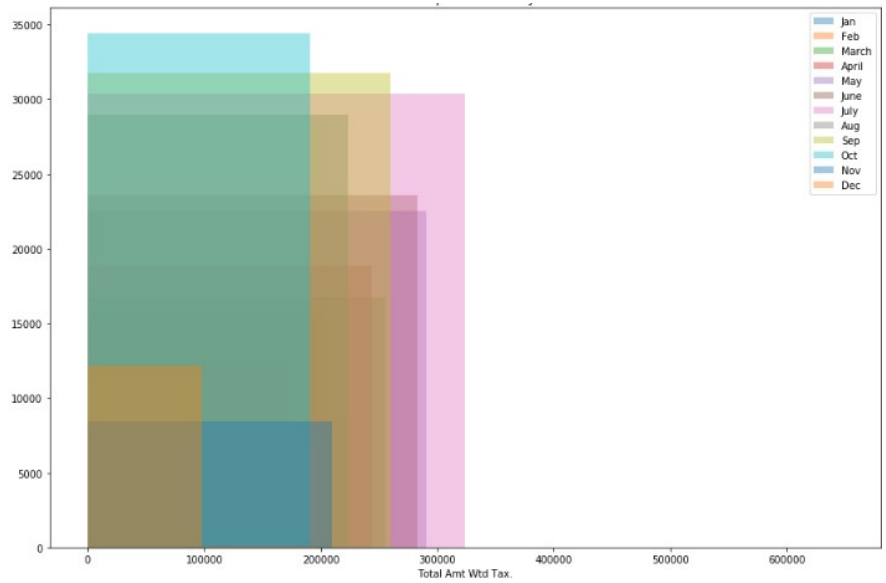
- Top 3 revenue generating states are

1. Pondicherry
2. Maharashtra
3. Tamilnadu



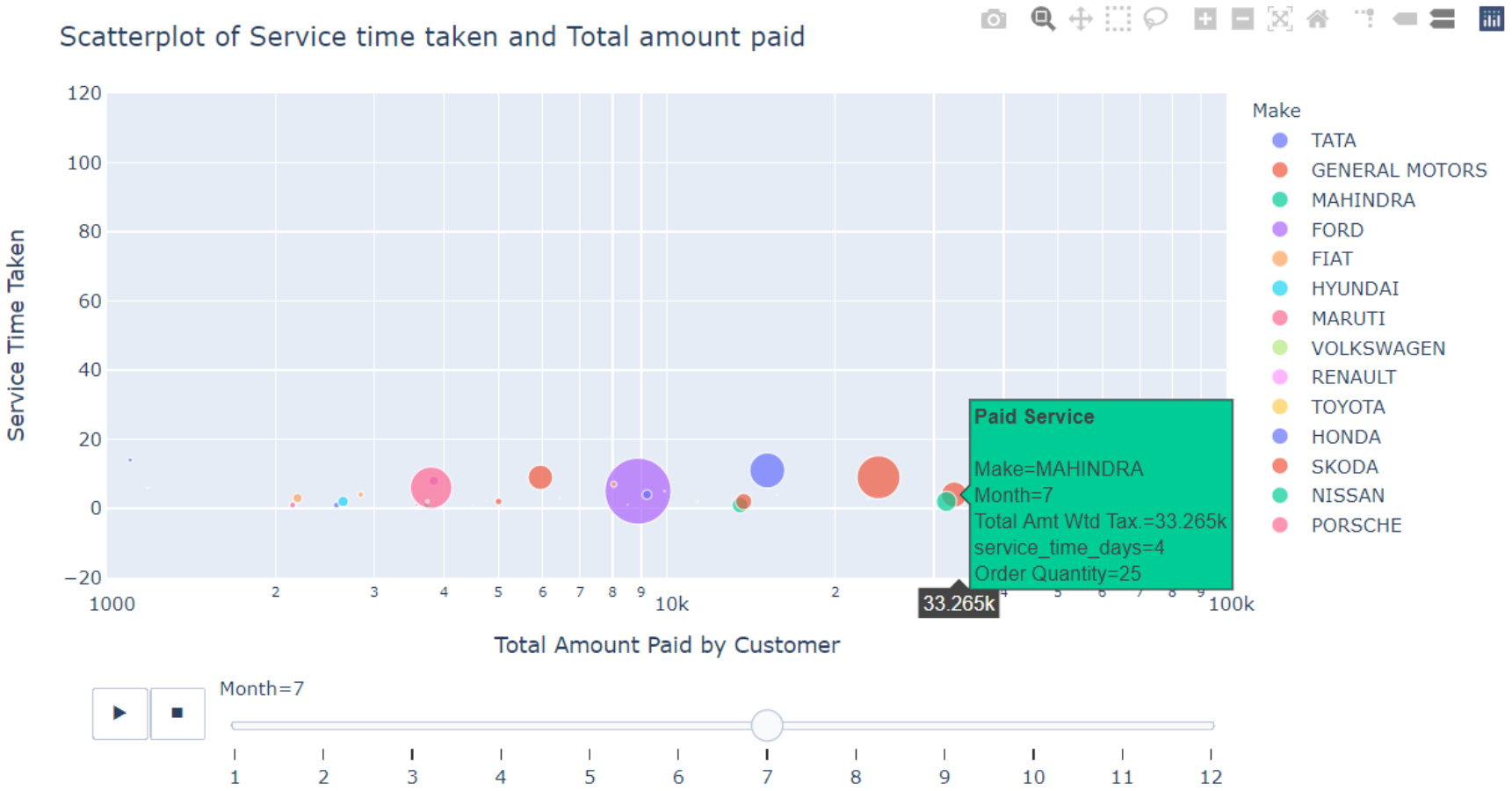
Month wise revenue trend





Observations: Car visits is highest in the month of October though revenue in July month is highest (due to monsoon related servicing)

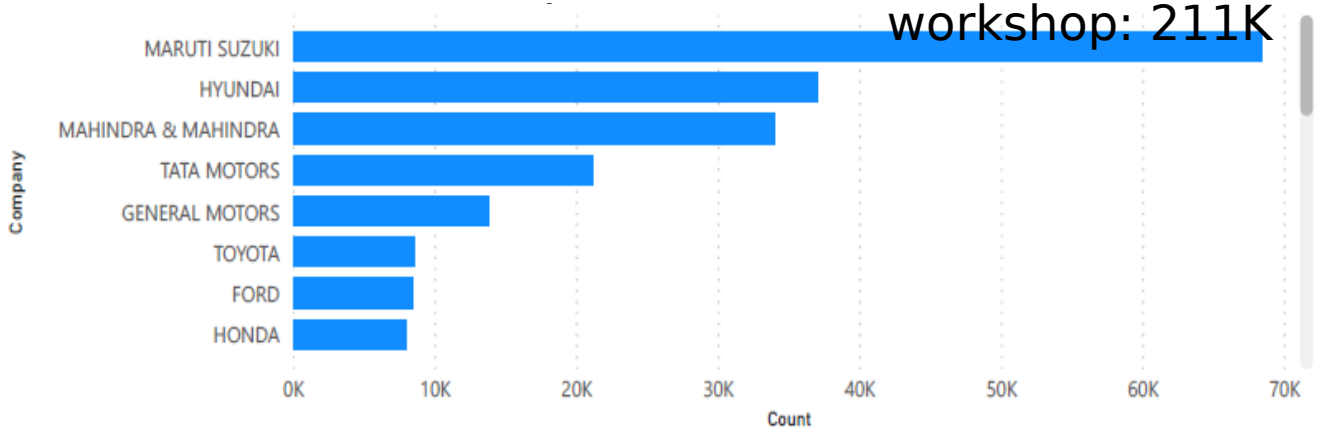
Data Analysis



Most visited Cars Companies in Different Zones

Western Zone : Total Car

Total No. of Car visited to workshop: 211K



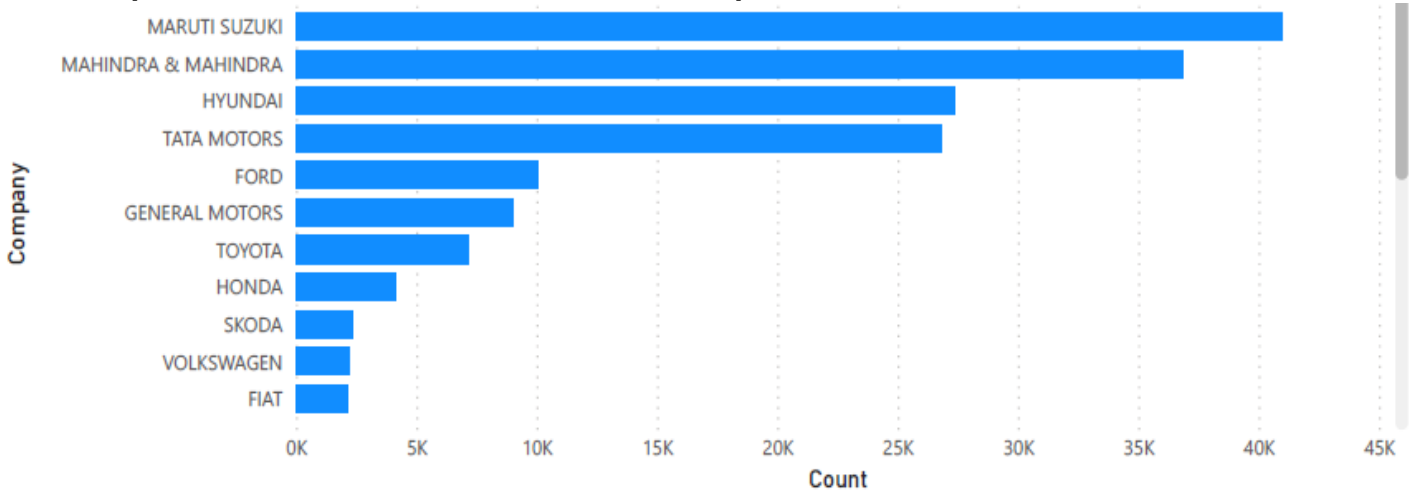
Most visited model

Column1	Model
SWIFT	14619
WAGONR	12078
ALTO	11844
SCORPIO CRDE	9884
SANTRO	9821
I10	9182
INDICA	7572
SWIFT DZIRE	7107
BOLERO	6326
I20	6301
CITY	5627
INNOVA	5037
XYLO	4367

Southern Zone

Total Car Companies: 26

Total No. of Car visited to workshop: 172K

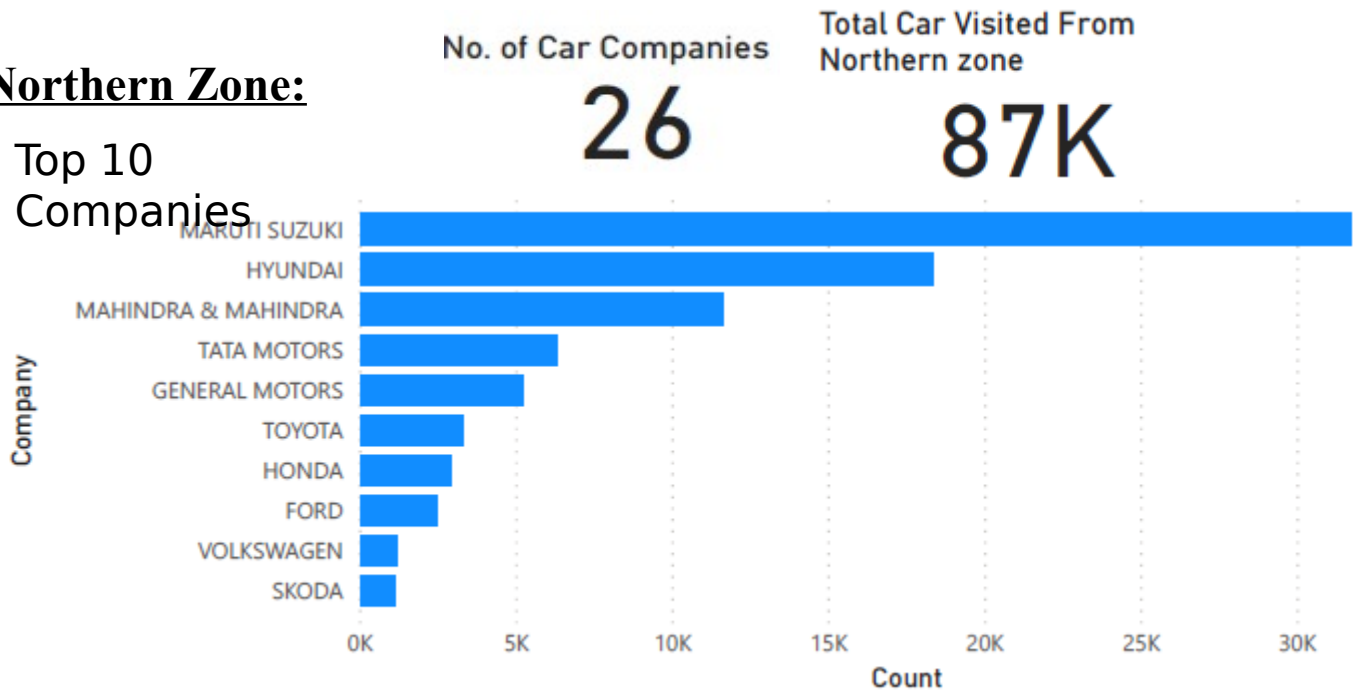


Model visited

Column1	Model
INDICA	10303
SWIFT	9412
SANTRO	8853
SCORPIO CRDE	8490
ALTO	8196
XYLO	7060
I10	6759
BOLERO	6288
VISTA	5167
WAGONR	4419
SWIFT DZIRE	4353
INNOVA	4290
800	4031
FIESTA	3906
INDIGO	3879
LOGAN D	3740
FIGO	3457
I20	3279
LOGAN P	3057
VERITO	2995

Northern Zone:

Top 10 Companies

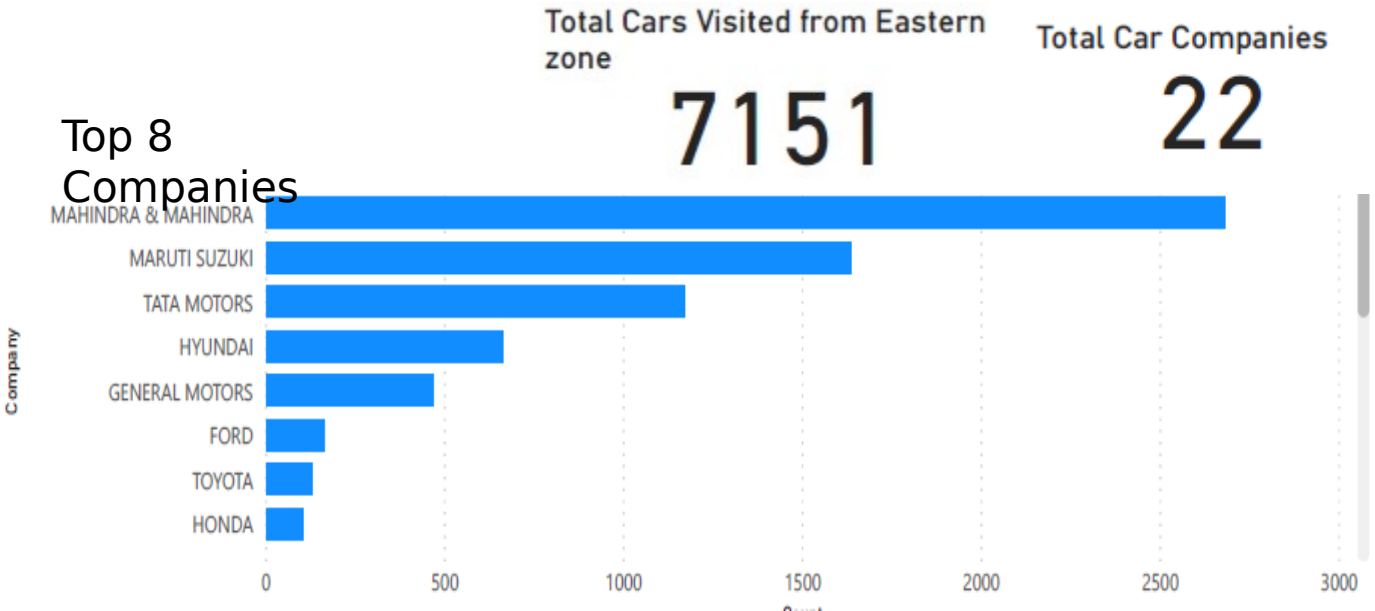


Most visited model

Column1	Model
ALTO	7085
SWIFT	6640
WAGONR	5497
SANTRO	5379
I10	5172
SCORPIO CRDE	4407
SWIFT DZIRE	3771
I20	2826
VERNA	2180
CITY	2041
INNOVA	1858
BOLERO	1801
RITZ	1740
BEAT	1737
XYLO	1538
SPARK	1510
INDIGO	1466
800	1459
INDICA	1316
FIGO	1195
ACCENT	1075
SAFARI	1072
ZEN	1064

Eastern Zone:

Top 8 Companies



Most visited Model

Model	Count
BOLERO	739
SCORPIO NEF	537
SCORPIO CRDE	516
INDIGO	439
ALTO	423
SWIFT DZIRE	339
SWIFT	262
BEAT	259
WAGONR	209
XYLO	196
I10	191
INDICA	185
MAXXIMO	179
BOLERO PIK UP	174

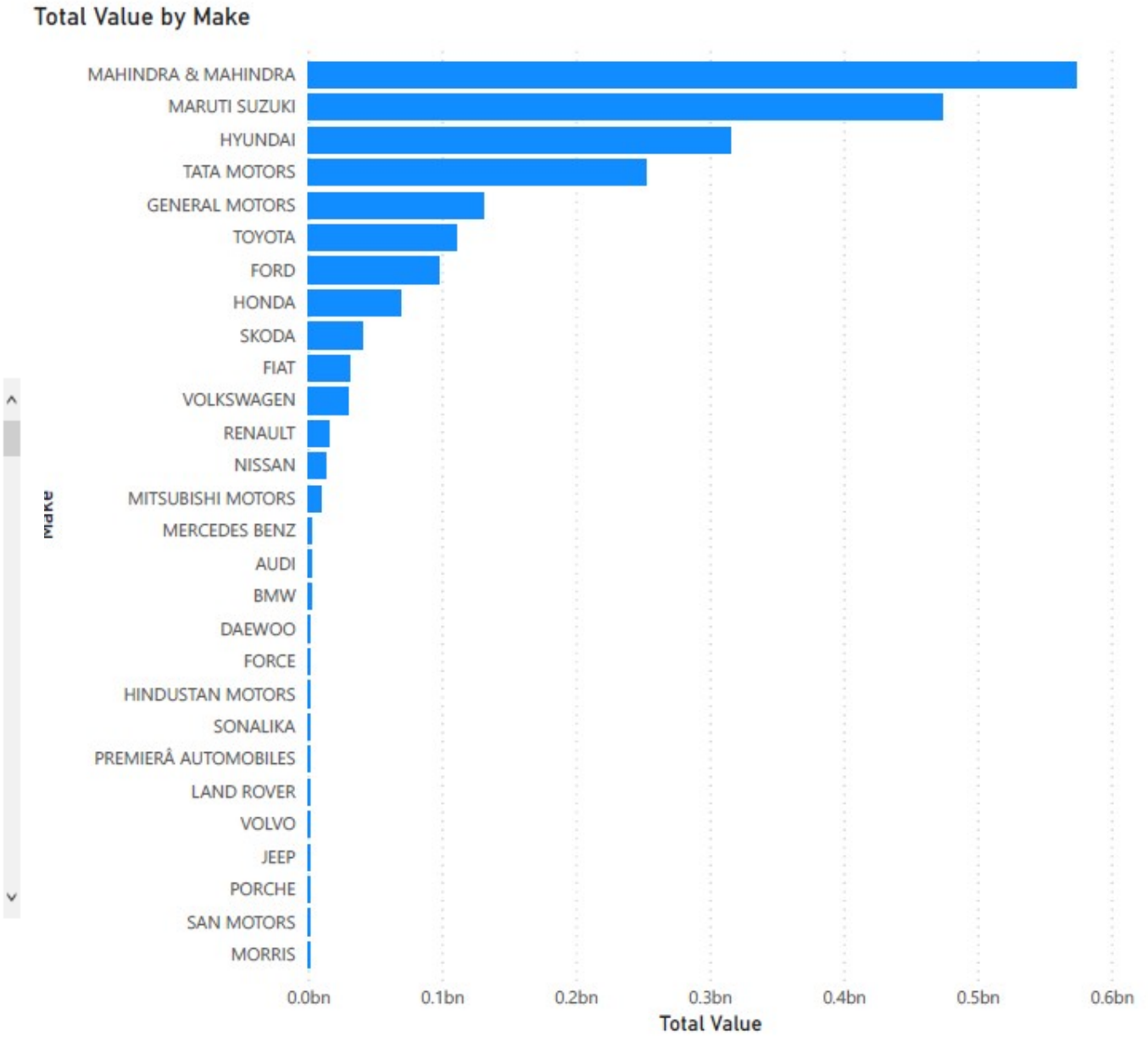
Revenue Based on Make and Model

Total Revenue

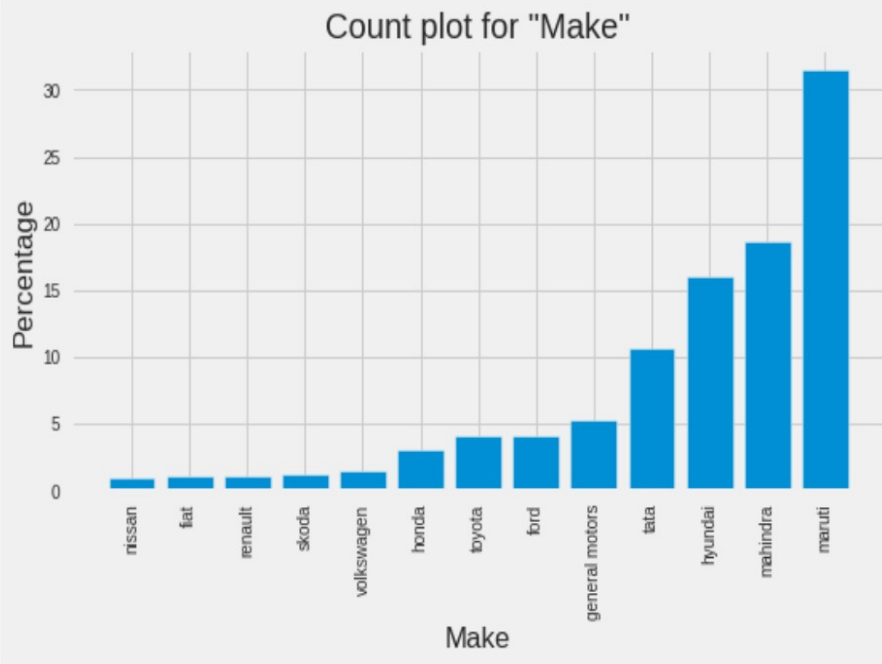
2.16bn

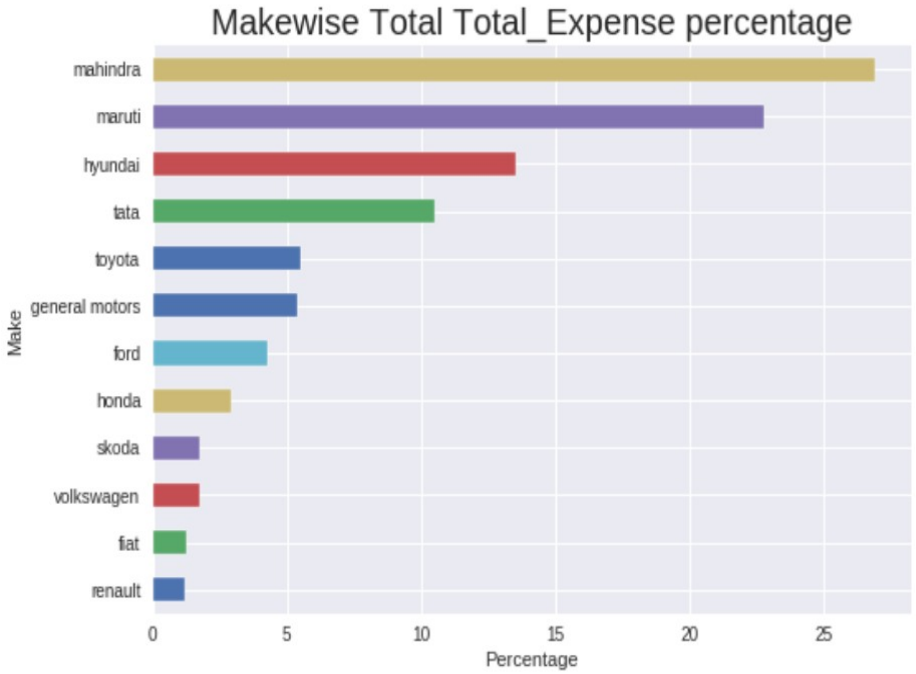
Sum of Total Value

Model	Total Amt Wtd Tax.
SCORPIO CRDE	161,632,919.86
SWIFT	124,054,357.05
BOLERO	102,505,371.55
XYLO	88,977,942.27
ALTO	81,797,043.43
SANTRO	81,488,349.43
INDICA	78,771,737.58
I10	72,758,381.14
INNOVA	68,878,551.50
WAGONR	66,817,665.81
SWIFT DZIRE	63,799,355.70
I20	52,640,250.98
LOGAN D	49,261,643.67
CITY	48,495,443.61
INDIGO	47,676,206.74
VISTA	44,349,416.23



Data Analysis

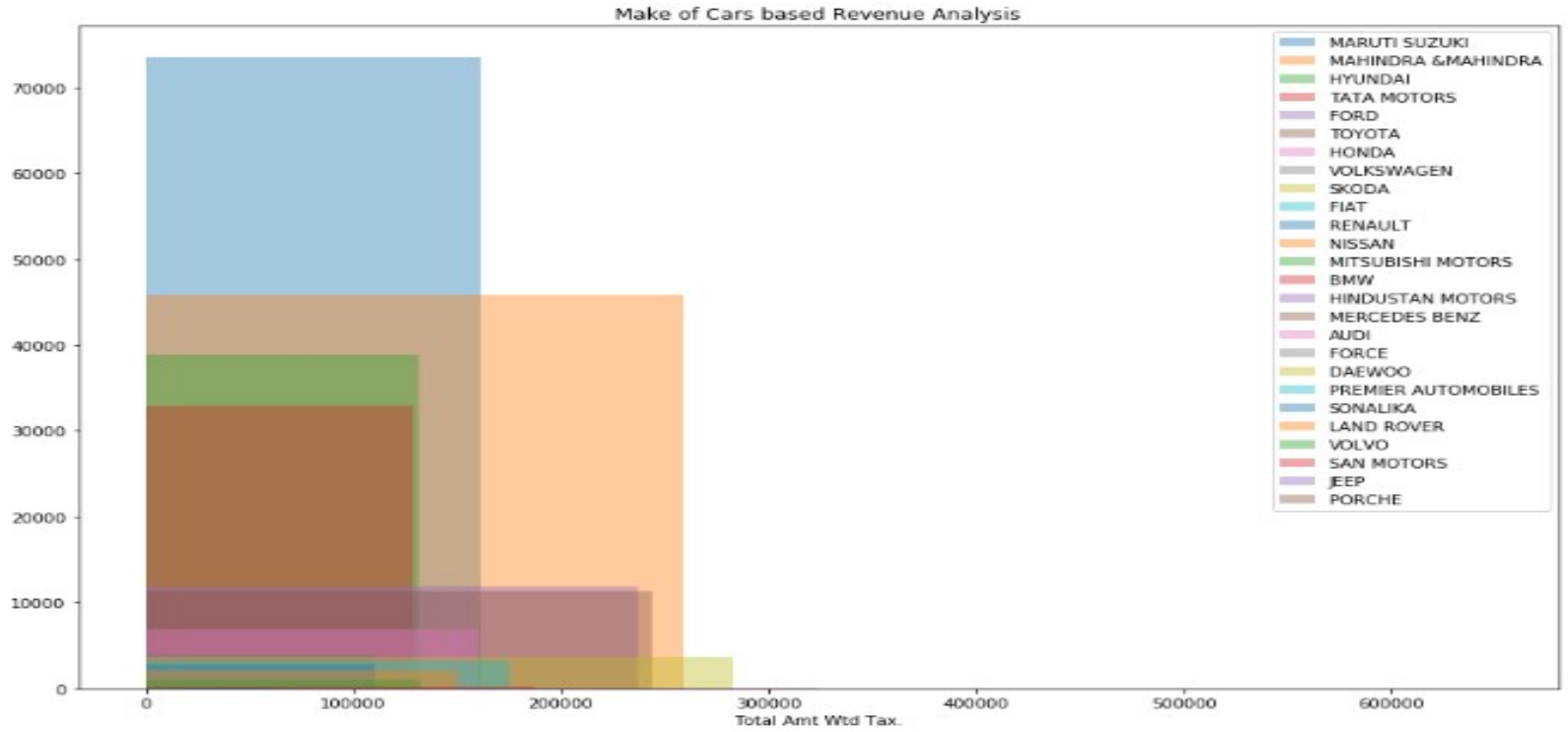




Observation:

Revenue from ‘Mahindra’ is more though the number of repairs for Maruti is largest

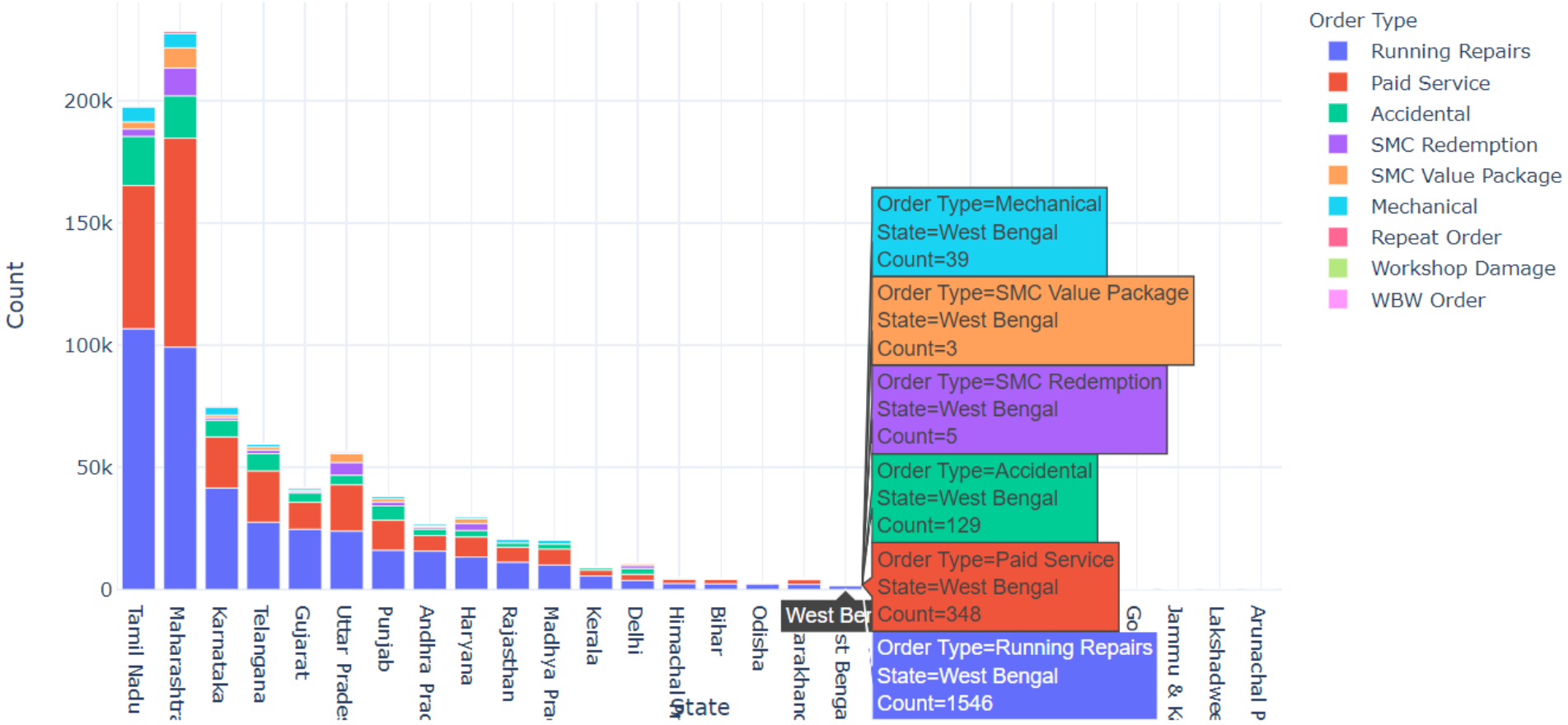
Car make wise revenue trend



Observations ➔

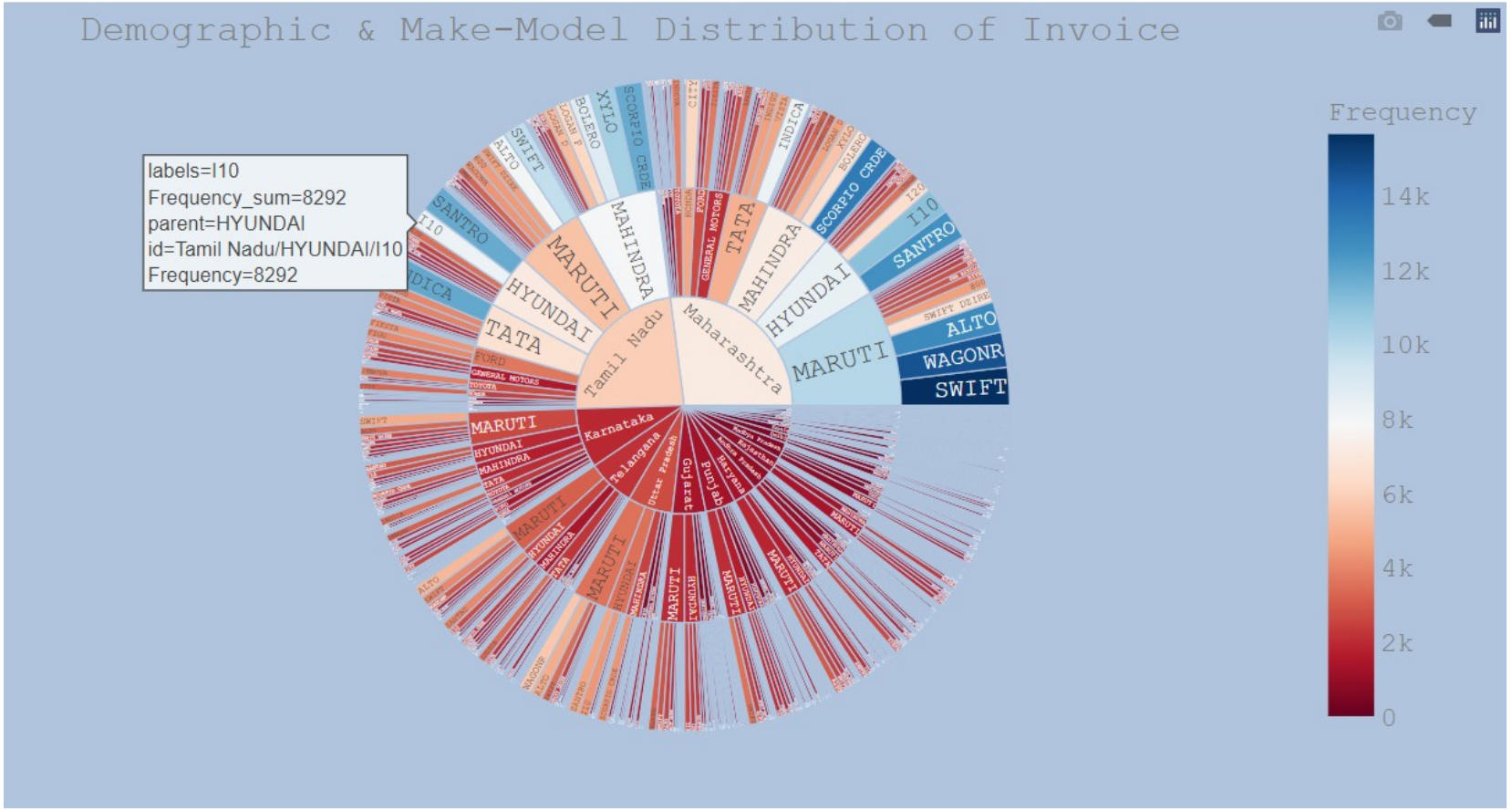
- Maruti Cars have highest frequency of visit
- Skoda cars generates highest revenue

Problem Statement 2 : Identifying Order Type Pattern



- Running Repairs & Paid Services are most common among top Manufacturers

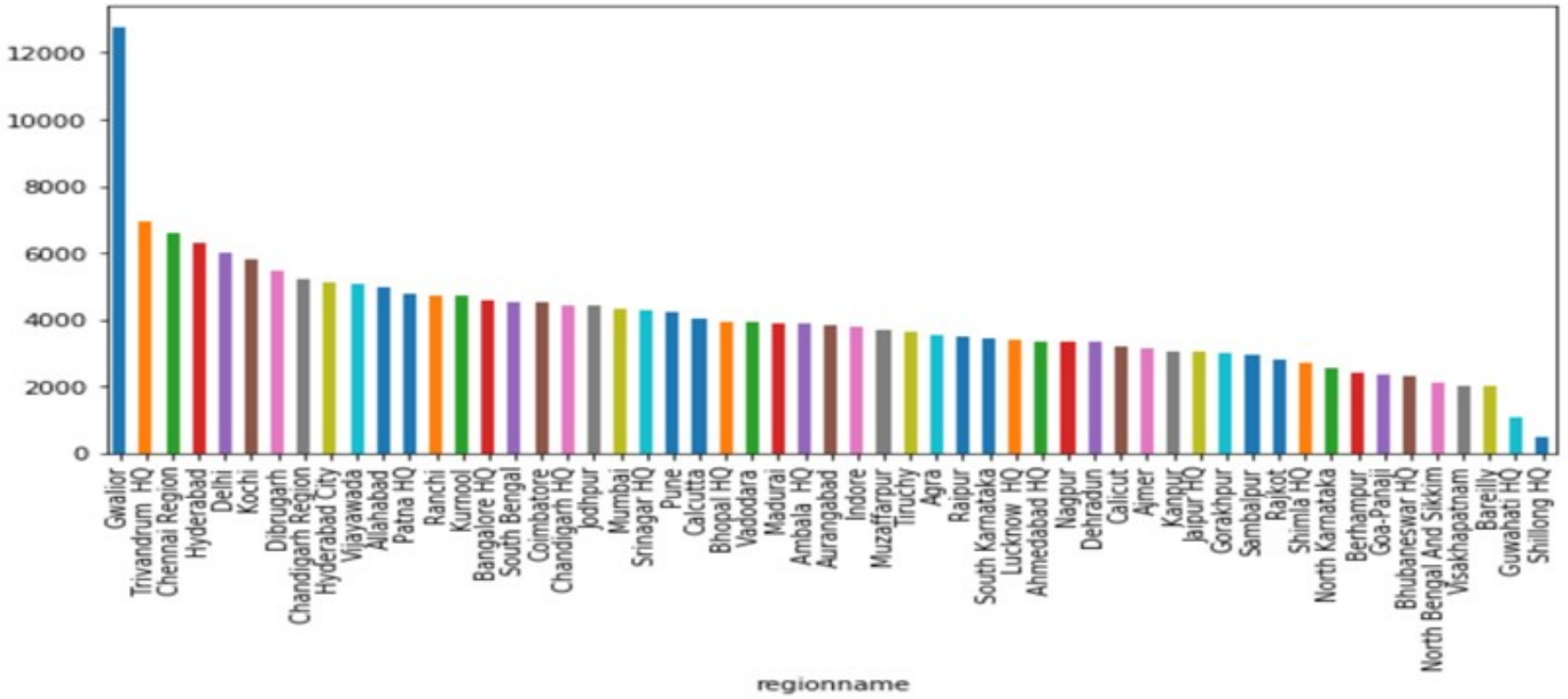
Make-Model Distribution



- Higher demand for ‘Maruti Suzuki’ & ‘Mahindra and Mahindra’ in ‘Tamil Nadu’ and ‘Maharashtra’.

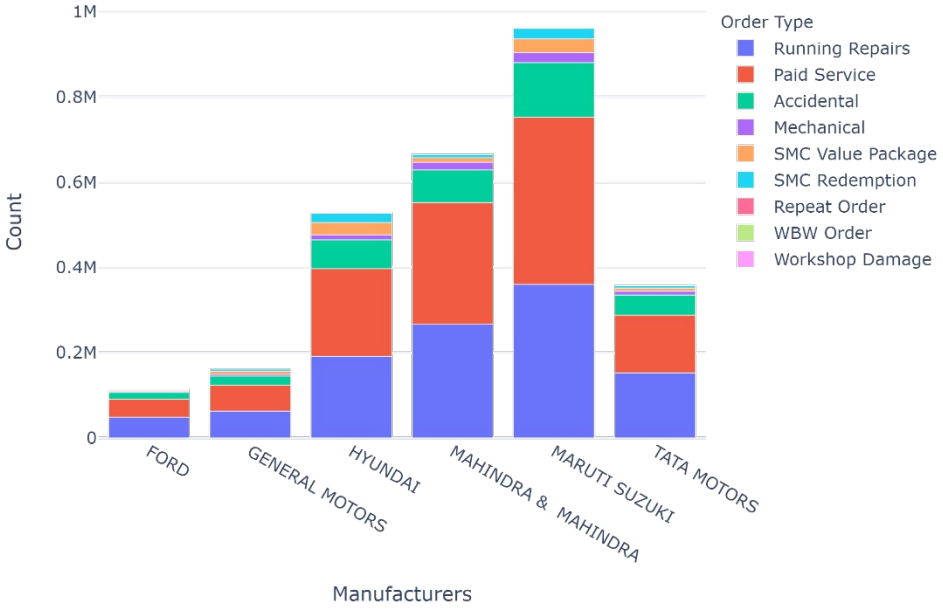
Data Analysis

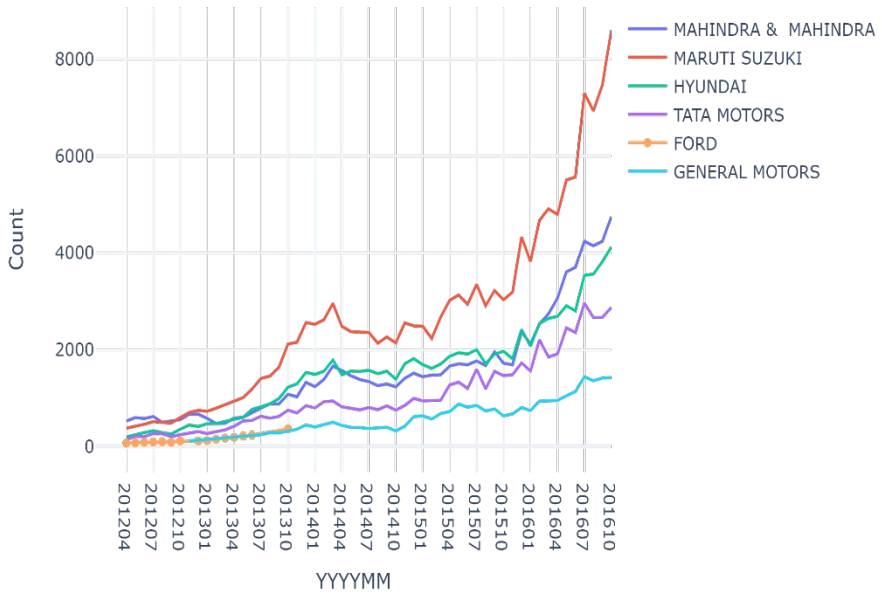
City by Average Revenue:



- 32% services are accidental cases and revenue generated from Gwalior is 94.83% from accidental cases

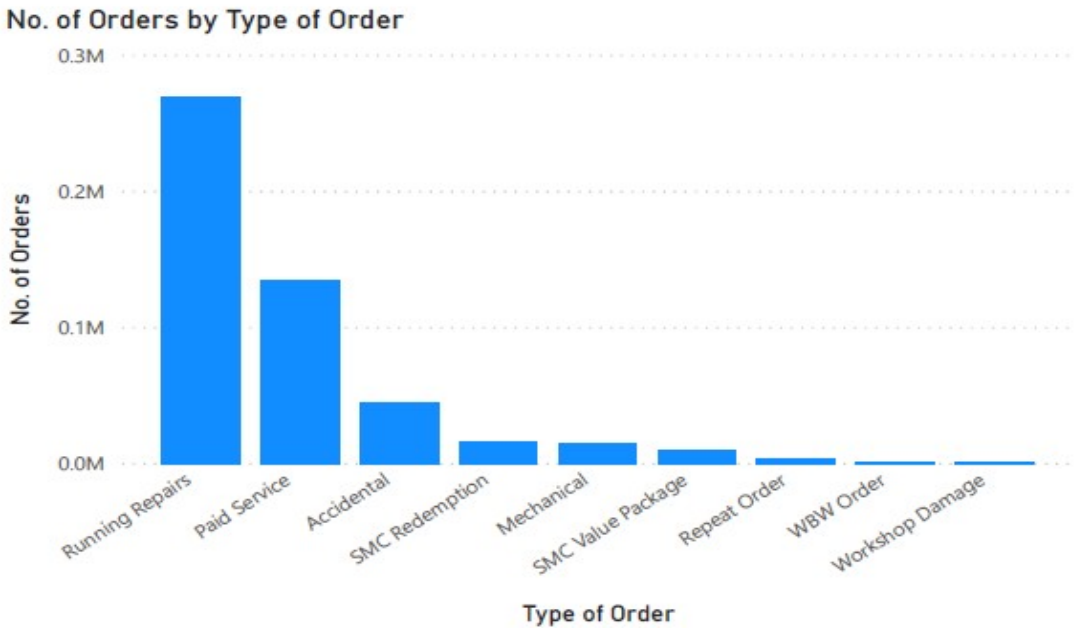
Top Manufacturers & their order type





- Running Repairs & Paid Services are most common among top Manufacturers
- Ford was 5th ranked car in Most popular car ranking during 2012 and 2013
- Since General Motors has taken its place

Revenue Based On Order Type



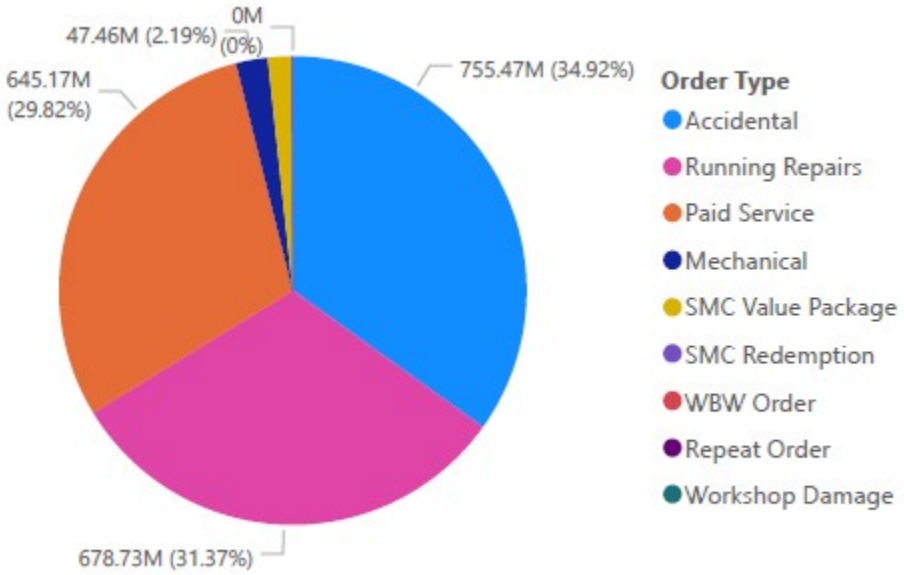
492K

No. of Orders

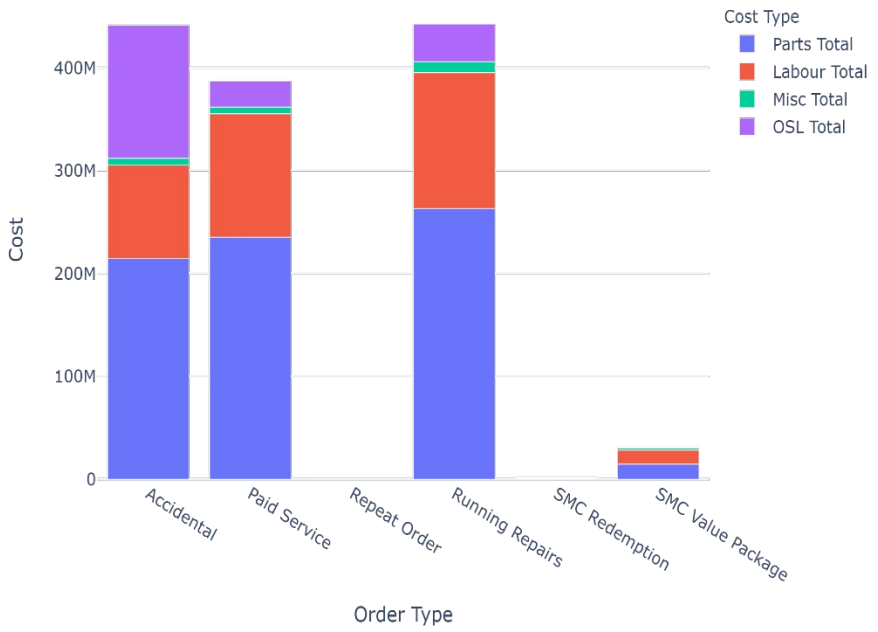
Total Revenue

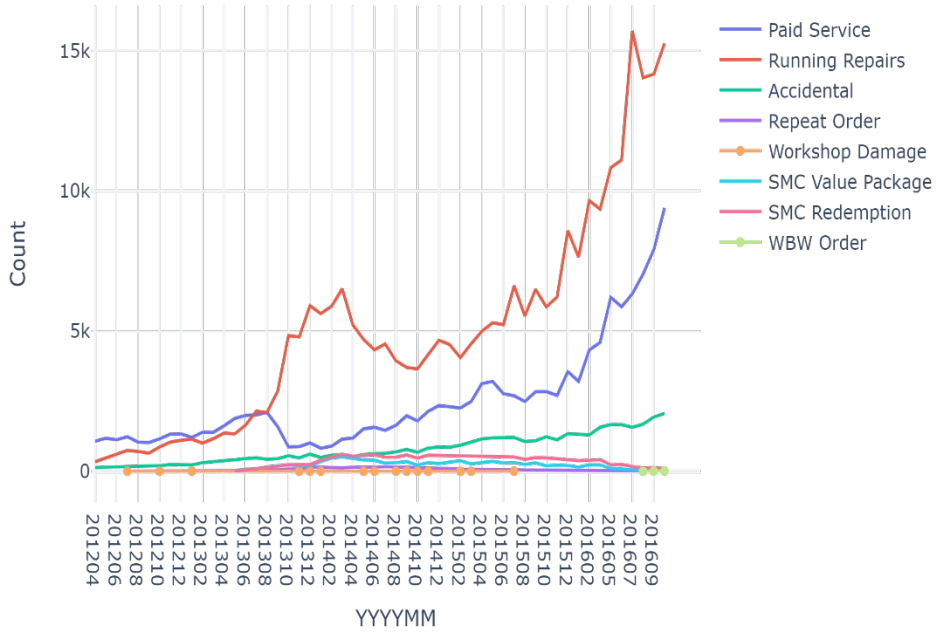
2.16bn

Sum of Total Value



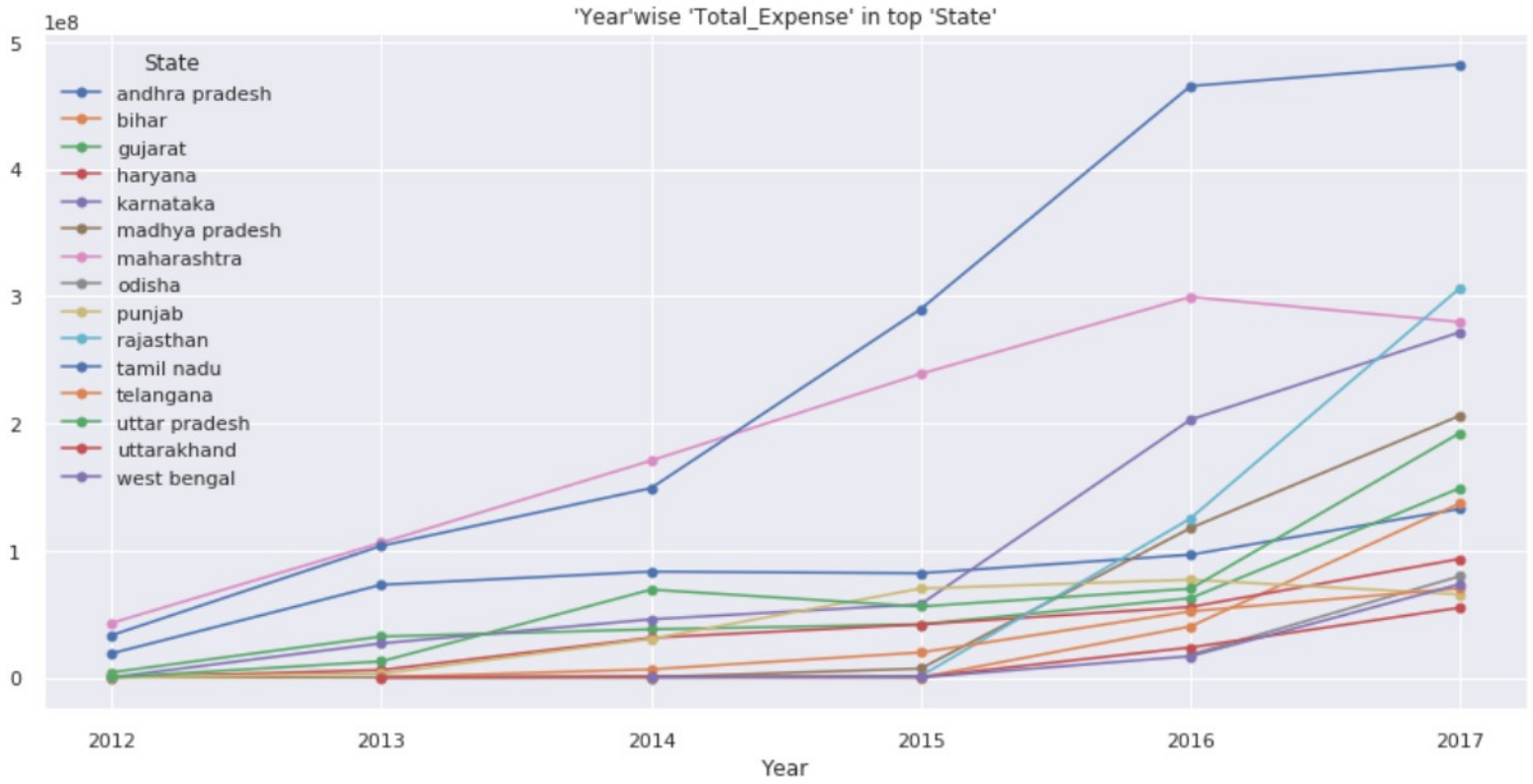
Cost for each order type





- For all type of orders parts & Labor cost are maximum
- For Accidental order type 25% of total amount goes towards outsourced work
- For SMC and Repeated order type outsourced amount is 0
- SMC package was introduced in May 2013

Data Analysis



Observation:

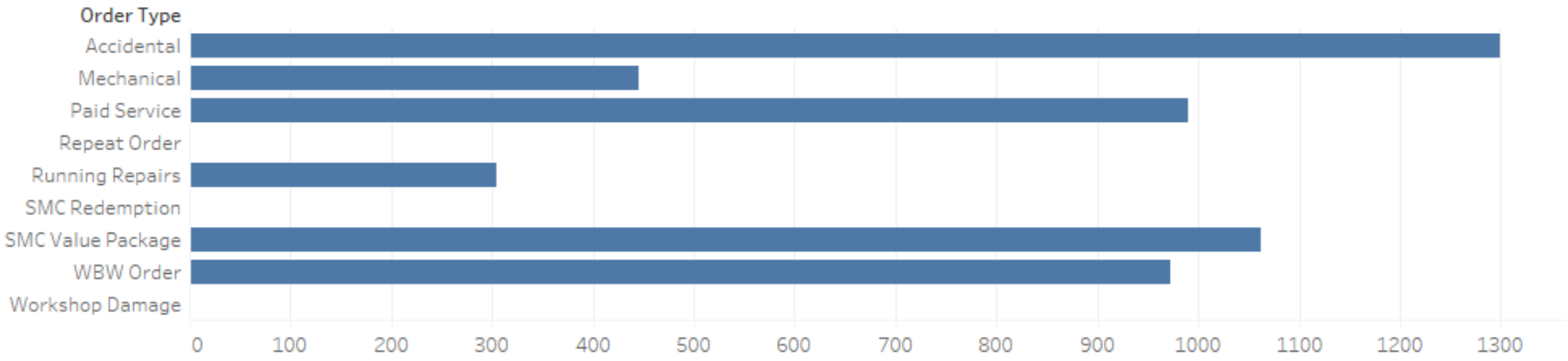
Uttarakhand has the lowest ‘Total_Expense’ but the ‘Tamilnadu’ has increase a lot from 2014 to 2016

Difference in each service/labour costing

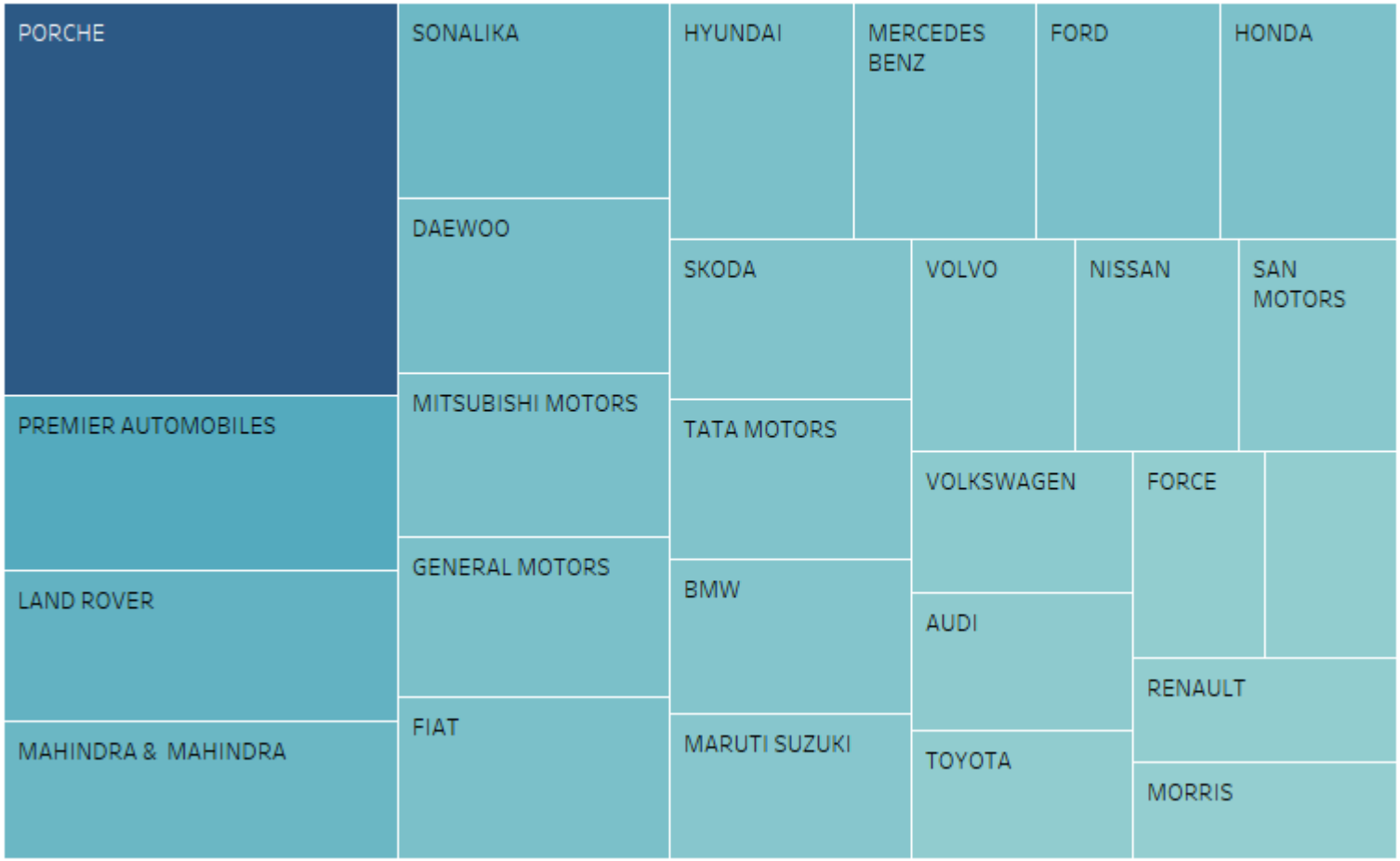
- **It depends on following factors:**

1) **Car Model (Make) :** type of vehicle, size of vehicle & petrol type
2) **Locality :** The basic labor charge varies based on the location as the land cost varies.
3) **Service type:** cleaning, general checkup, Car washing, Alignment, Tyre rotation

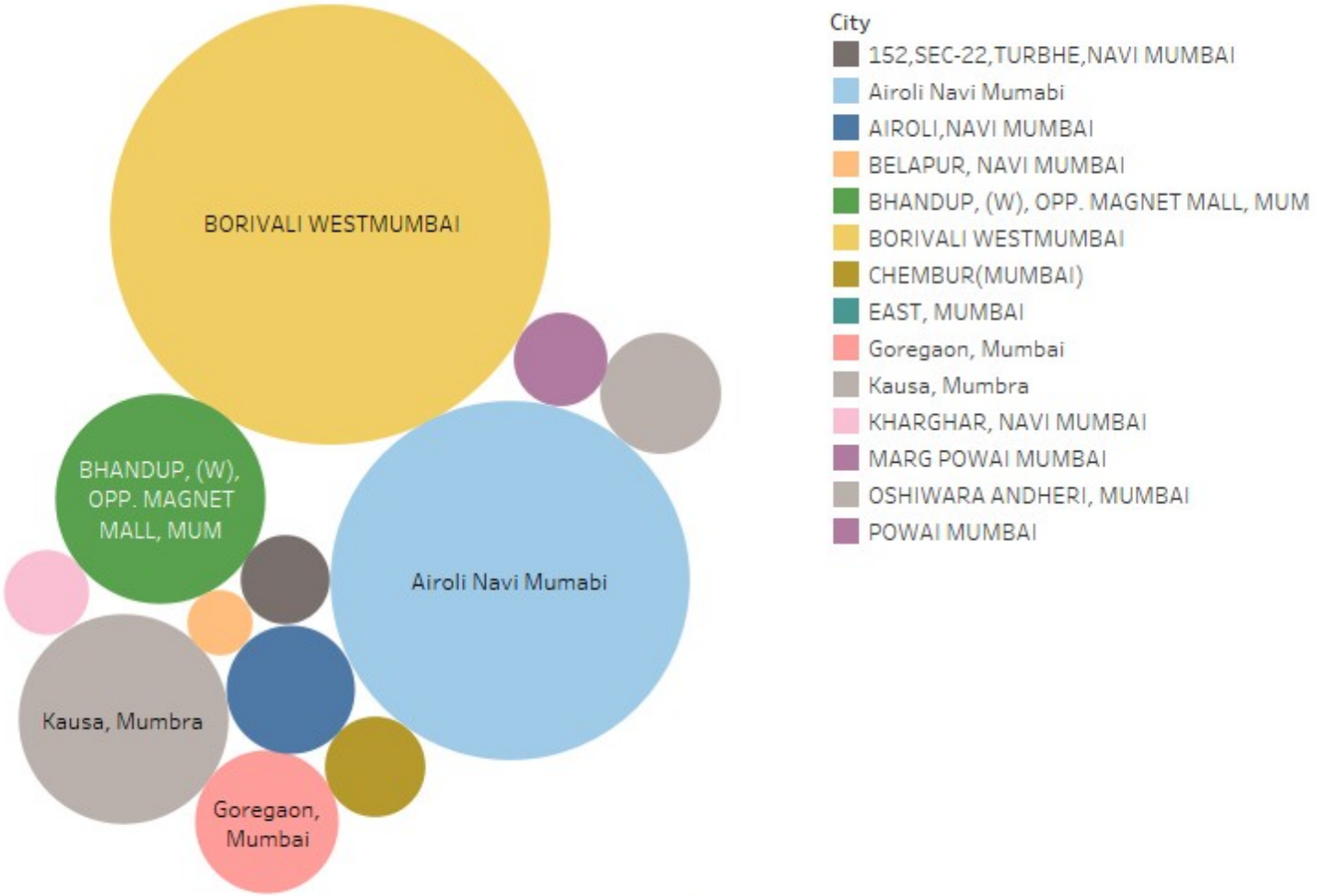
Labour costing for the factor service type



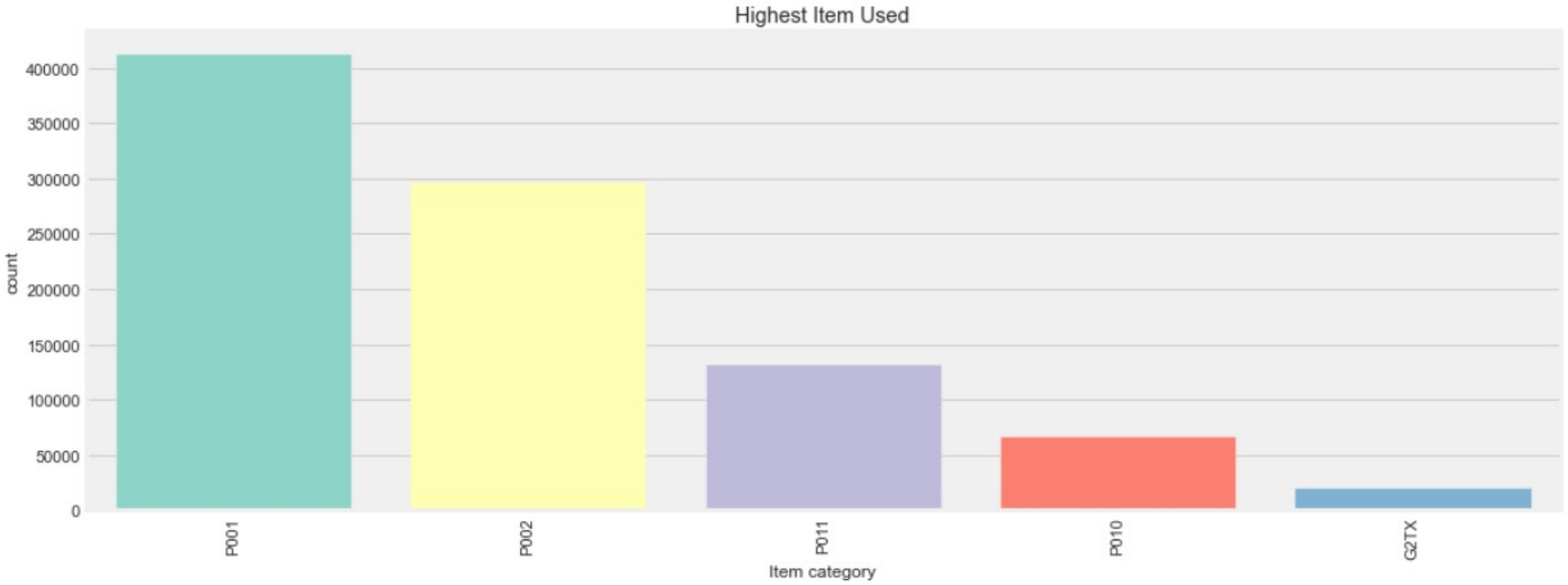
Labour Costing for the factor Make



Labour costing for the factor Locality (areas in Mumbai)

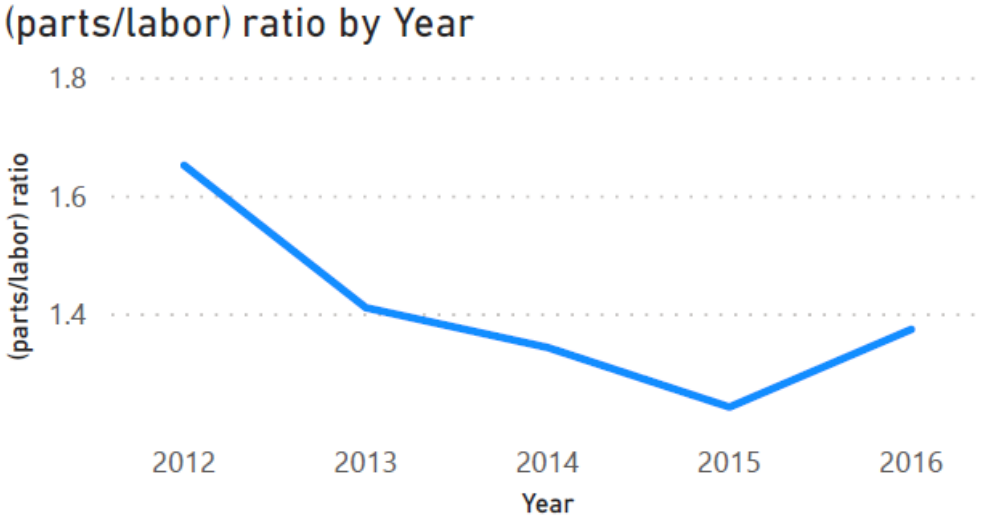


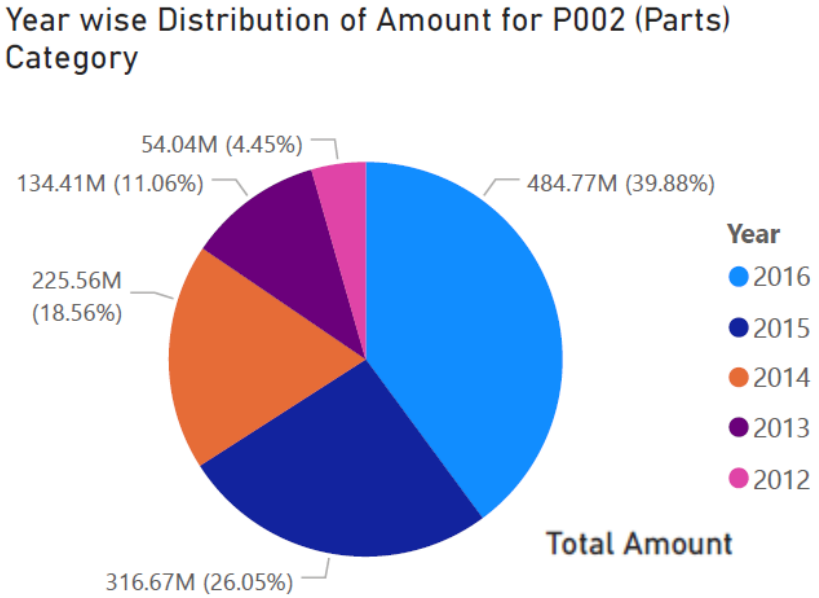
Popular items for Service orders

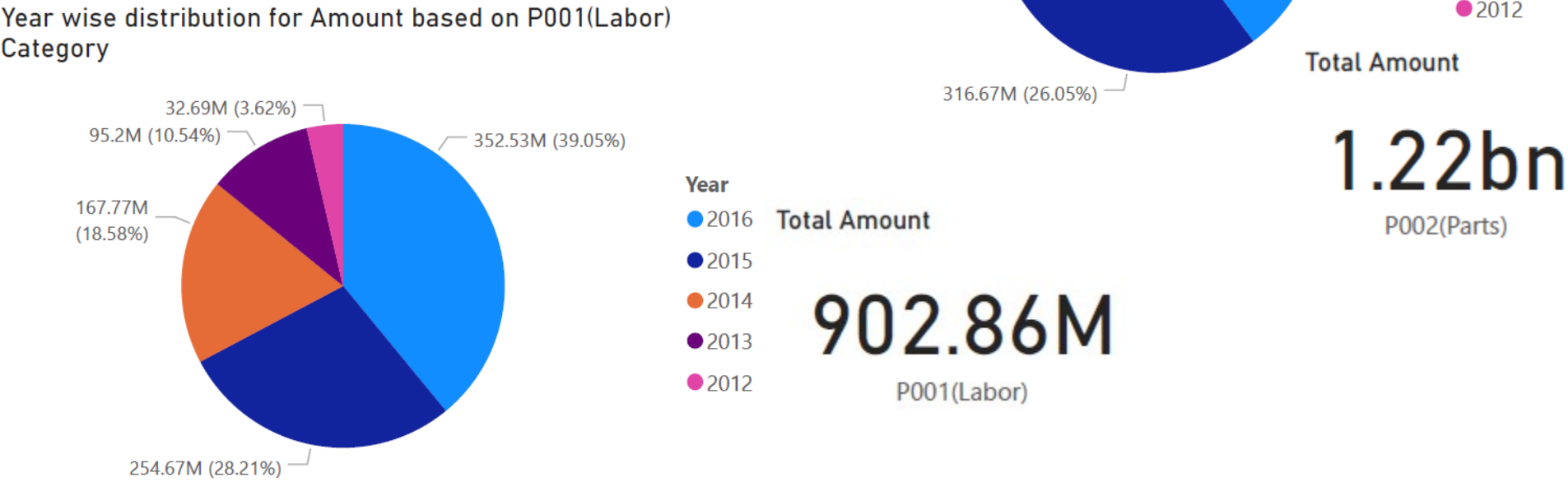


- P001, P002 are the most popular items utilized for fulfilment of service orders.

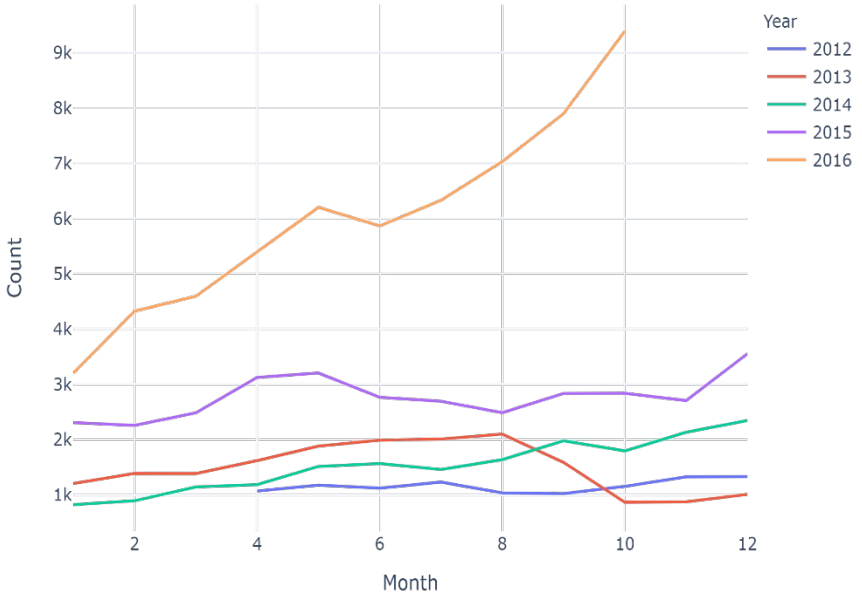
Parts To Labor Ratio

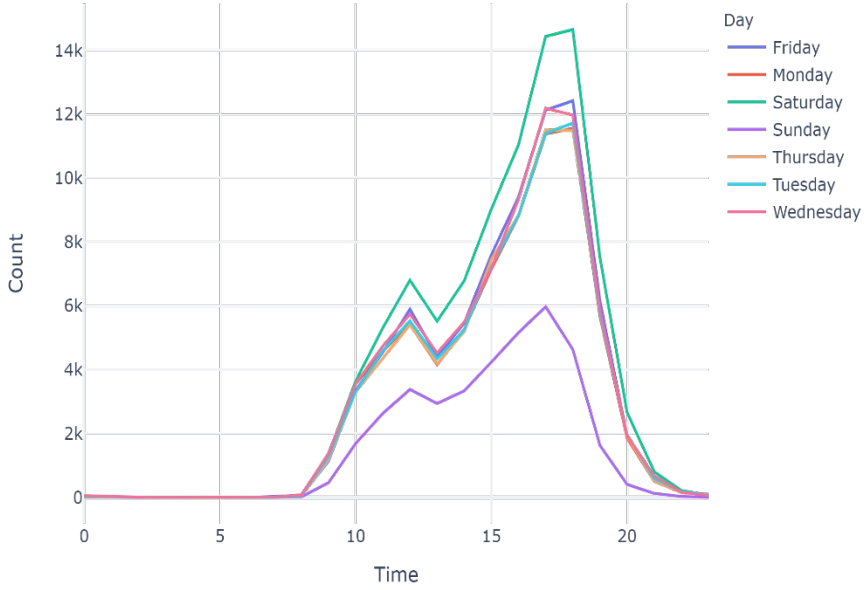






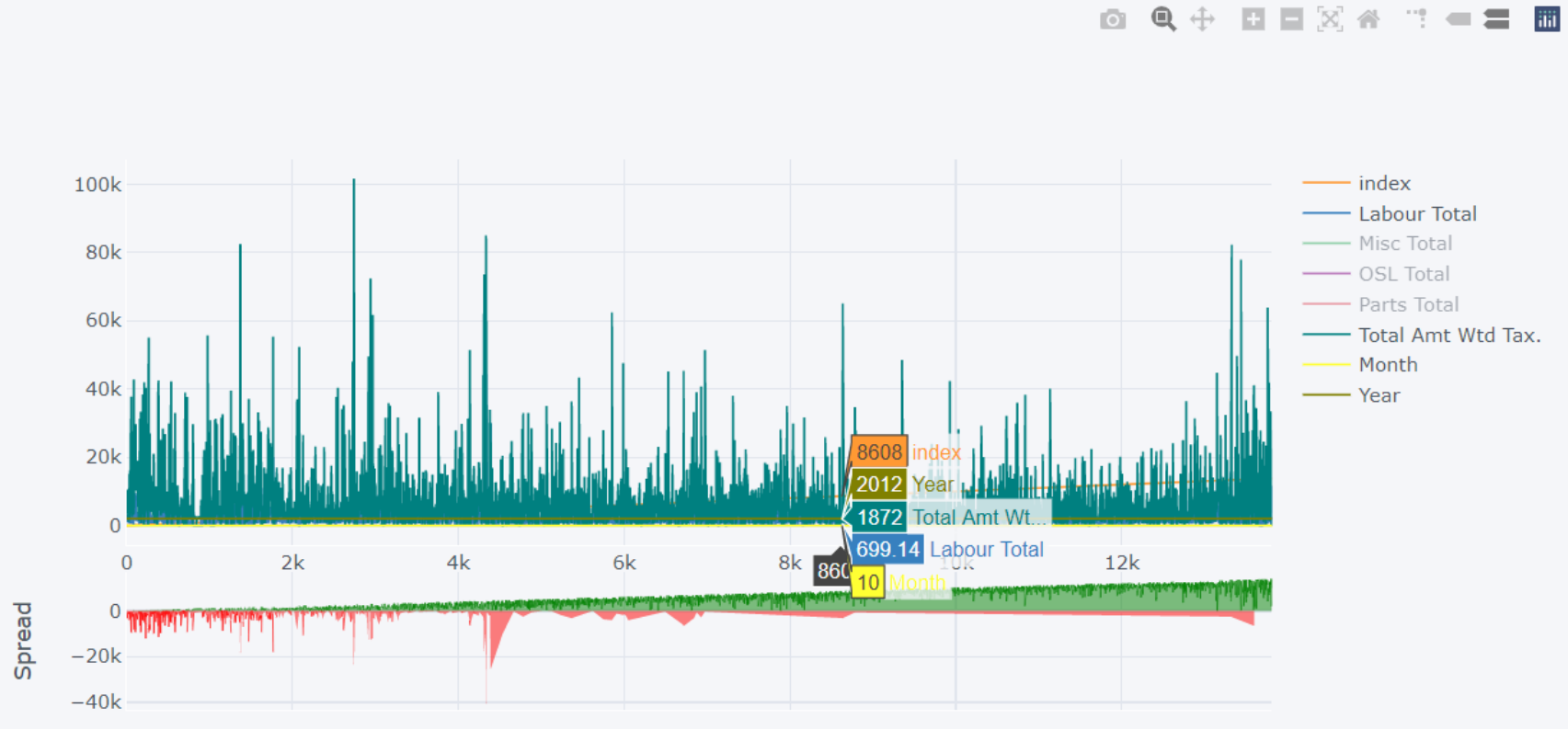
Is the service seasonal?



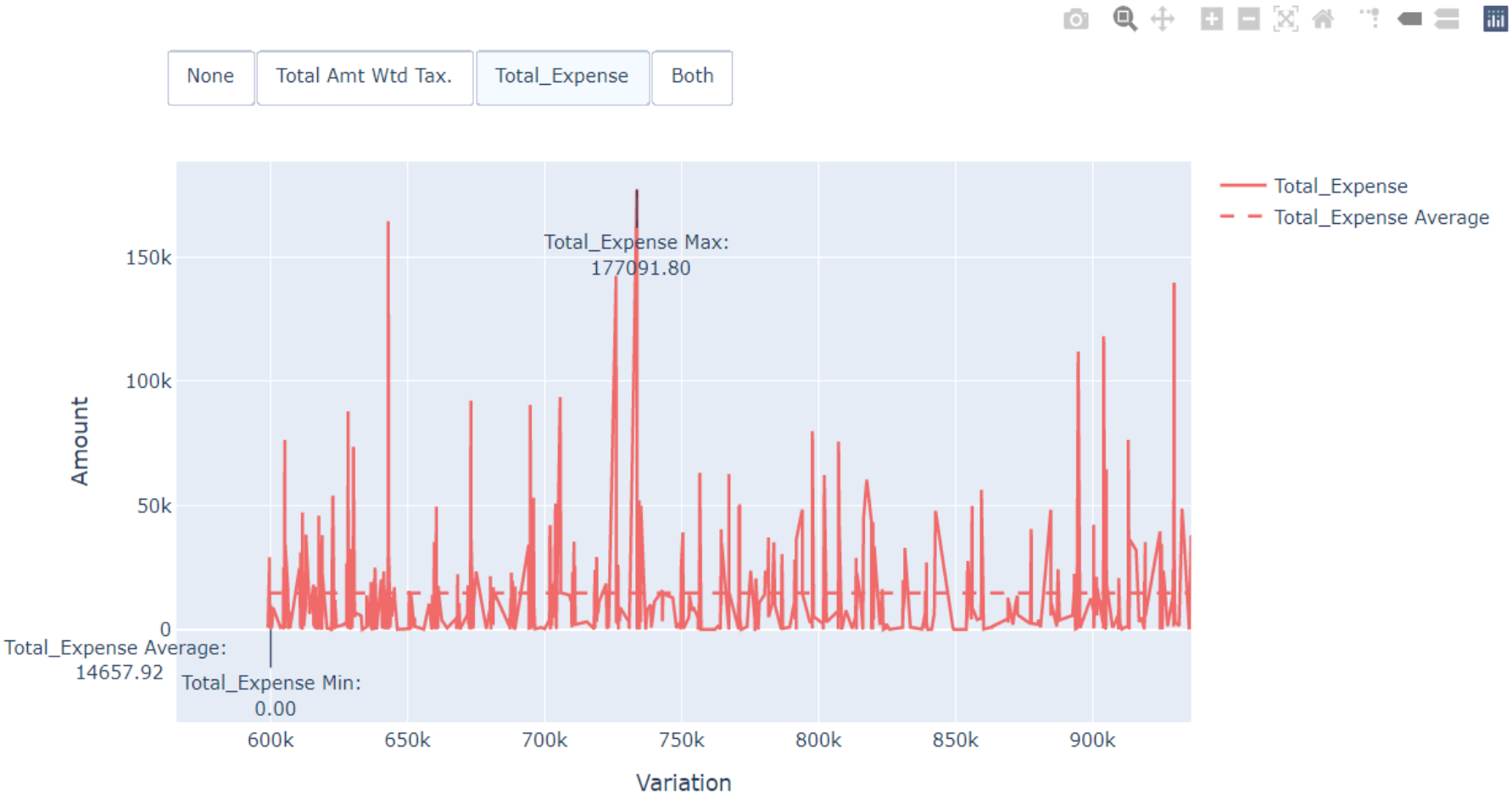


- There is no pattern of customer increment in Monthly customer count
- However there is 5PM-6PM is rush hour on daily basis
- And on Sundays least numbers of customers comes

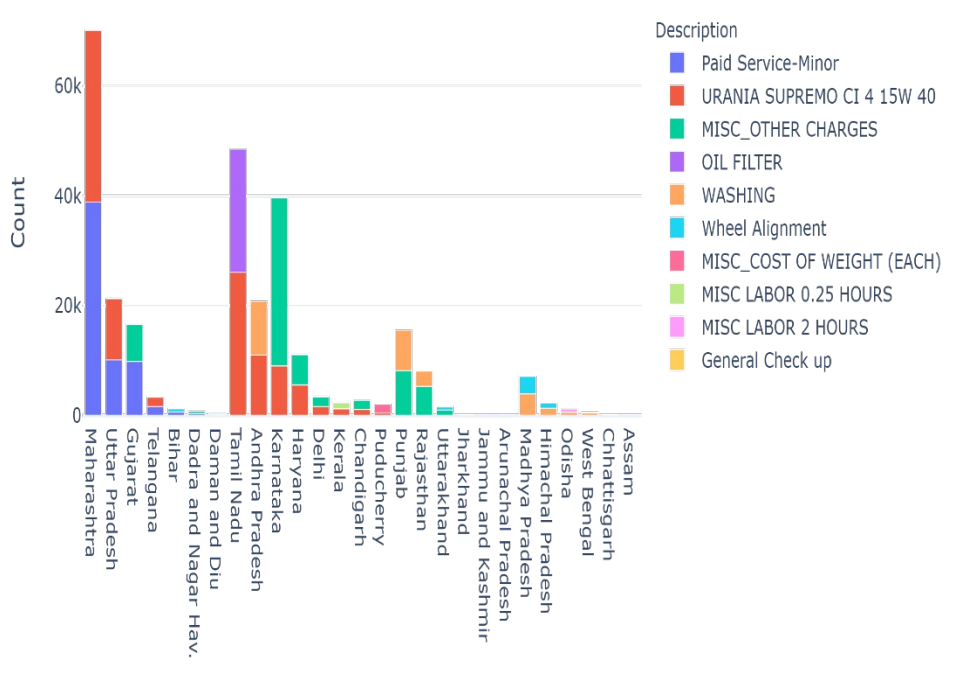
Spread of various amounts

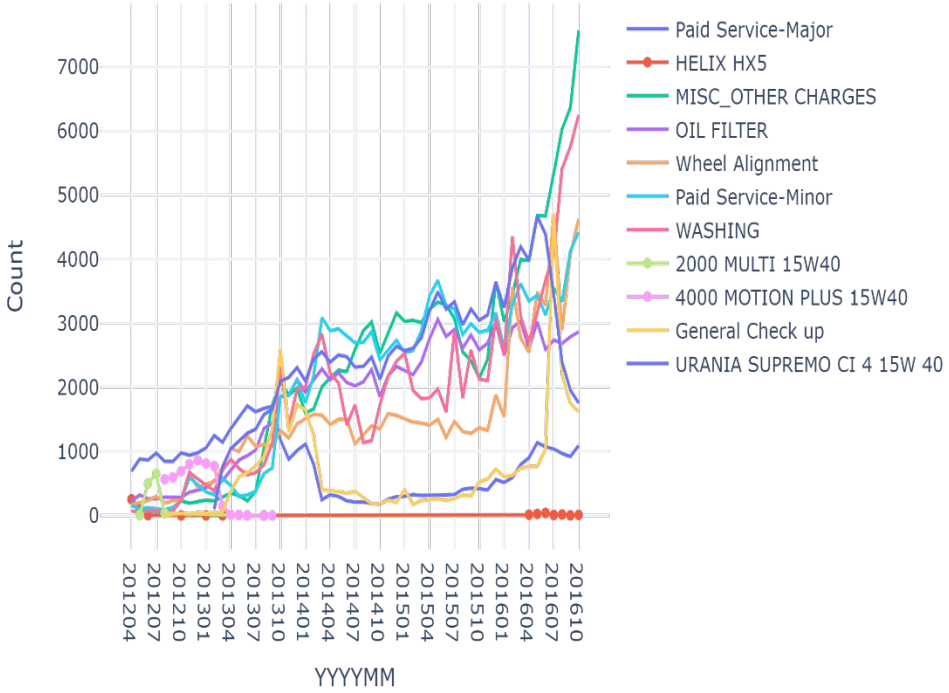


Data Analysis



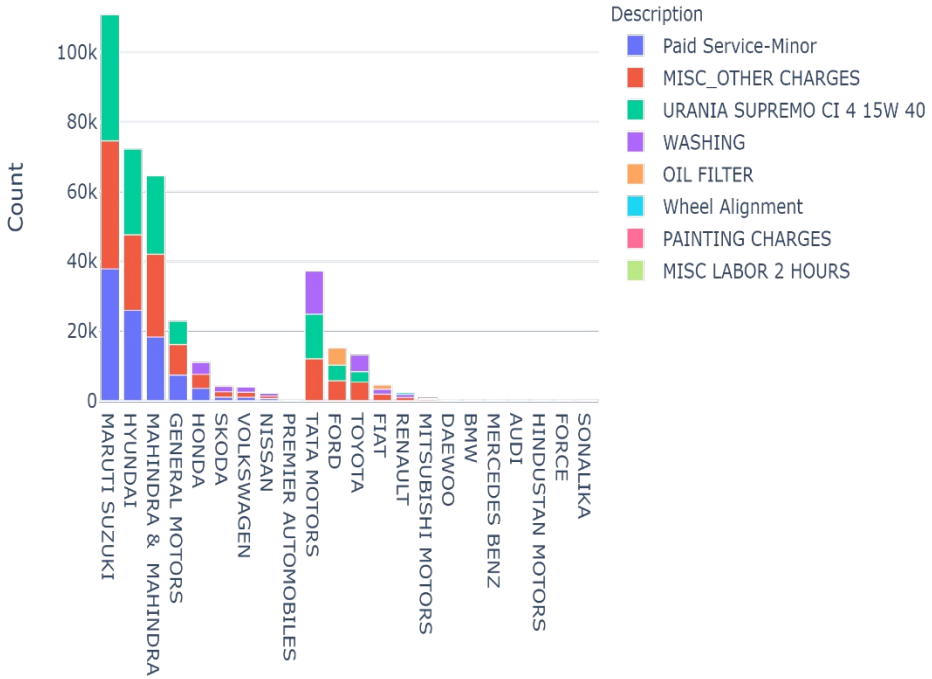
Inventory analysis : Top Inventory item

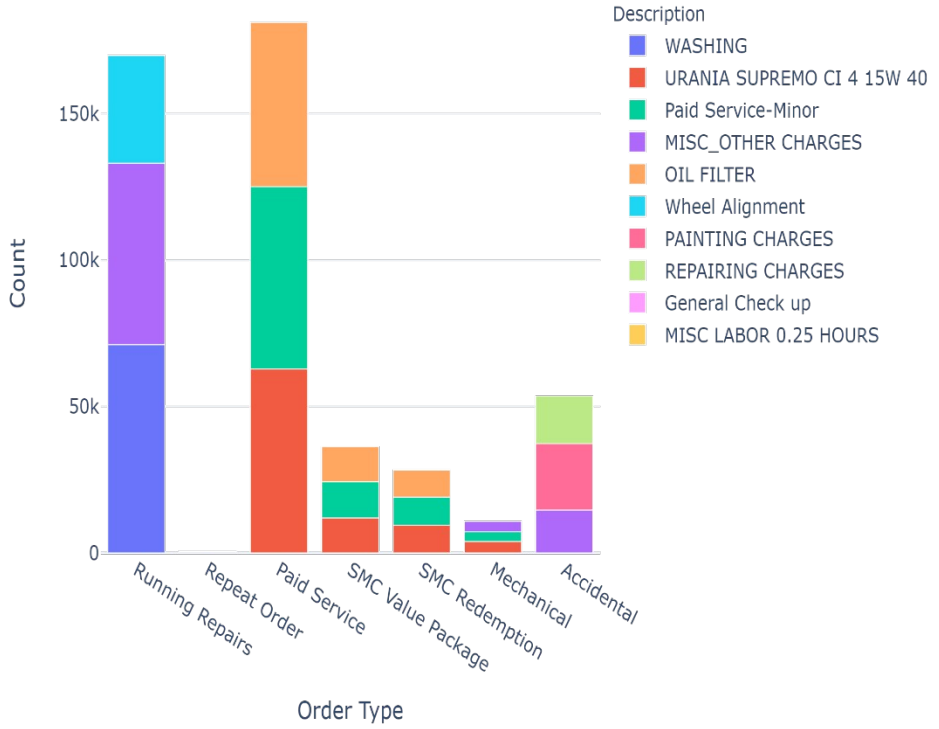




- Mirror Service, Engine Oil and Oil Filter most used item for most of the states
- Historically Oil Filter and Engine Oil are most used Inventory Items

Top Inventory item by Manufacturers & Order Type





- Mirror Service, Engine Oil and Oil Filter most used item for most of the states

Efficiency Of A Plant

- Efficiency was calculated Plant wise for each year & each month
- Efficiency = (Total Revenue in a year / Total servicing time in hrs * Count of invoices) *100

Plants	Total Revenue	Total Servicing time in hrs	Count of Invoices	Efficiency
BC01				
2012	5236.463258	27.77583333	2953	6.384215
2013	4878.202362	24.18916667	3794	5.315469
2014	5661.56783	24.59527778	3800	6.057611
2015	6282.654196	25.06611111	2778	9.022439
2016	4875.482893	25.69722222	1158	16.38411
BC02				
2012	4245.71192	24.39805556	1847	9.421681
2013	4457.701241	29.49986111	3195	4.729553
2014	5511.875385	29.57861111	2648	7.037261
2015	5139.167115	28.80555556	519	34.37551

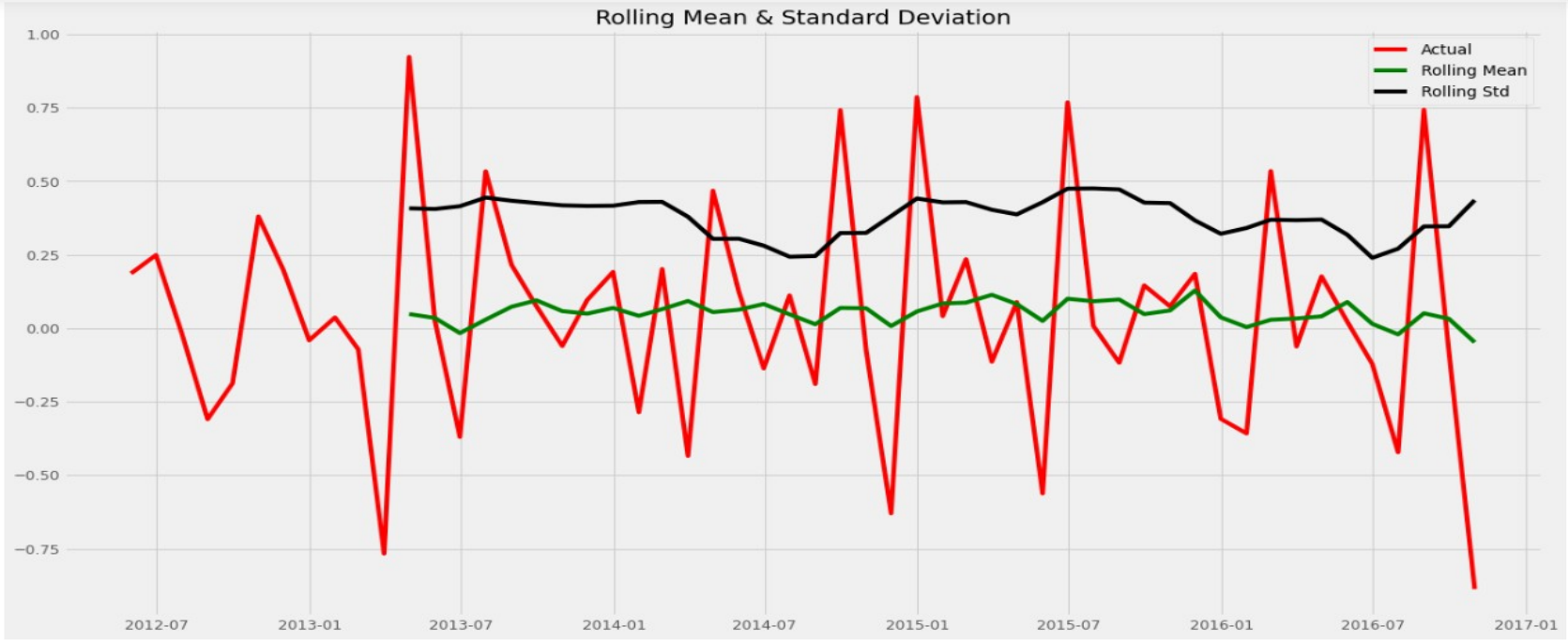
Arima model – Revenue Forecast

Stationary??
It is non-stationary because mean and std is not constant

Transformations

- Using differencing to make this time series stationary

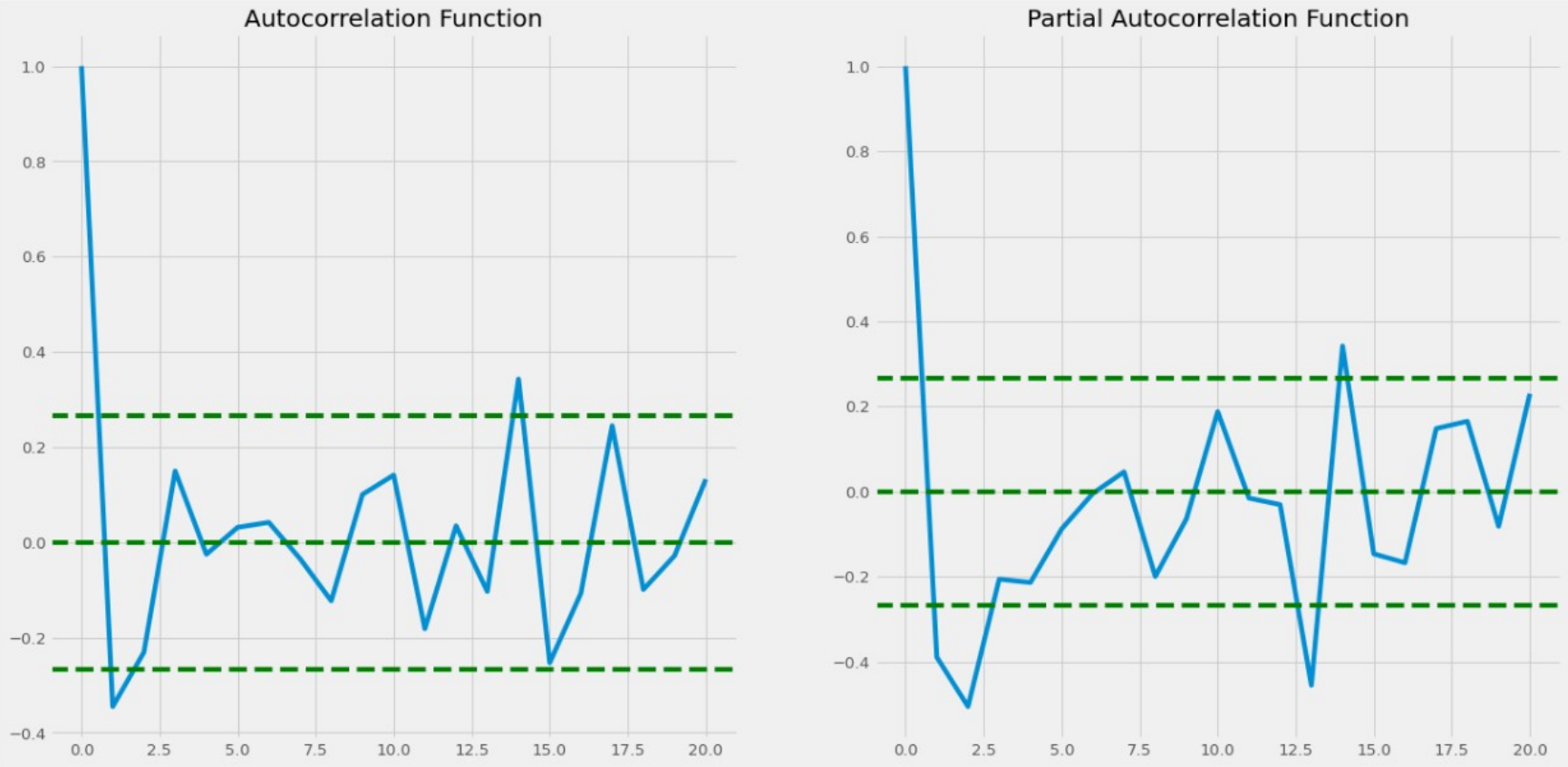
Dickey-Fuller Test AND Rolling Test




```
Dickey-Fuller Test:
Test Statistic          -9.24
p-value                  0.00
Lags Used                1.00
No. of Obs              52.00
Critical Value (1%)     -3.56
Critical Value (5%)     -2.92
Critical Value (10%)    -2.60
```

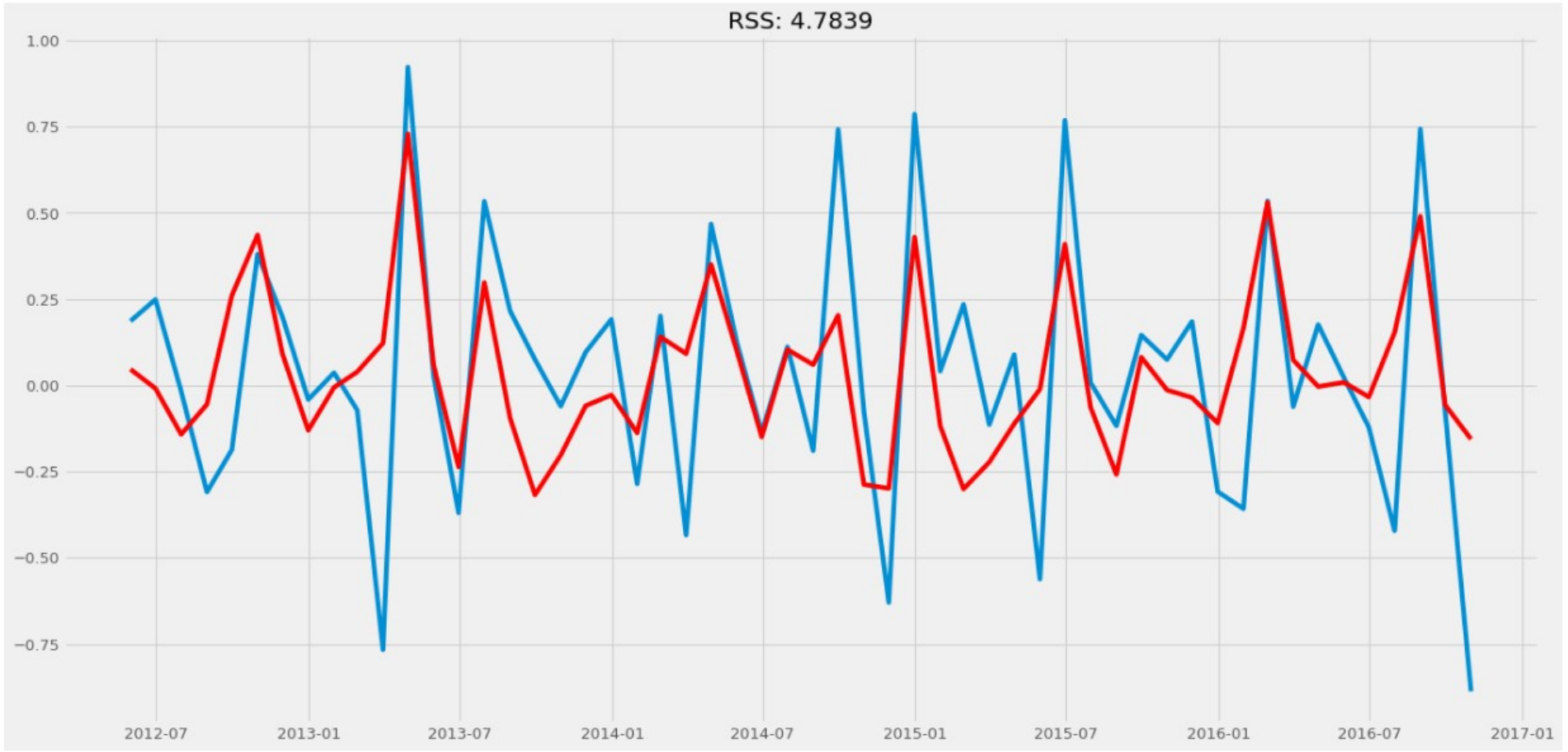
* The results show that the test statistic is significantly less than the 1% critical value, as its stationary

Plotting ACF & PACF

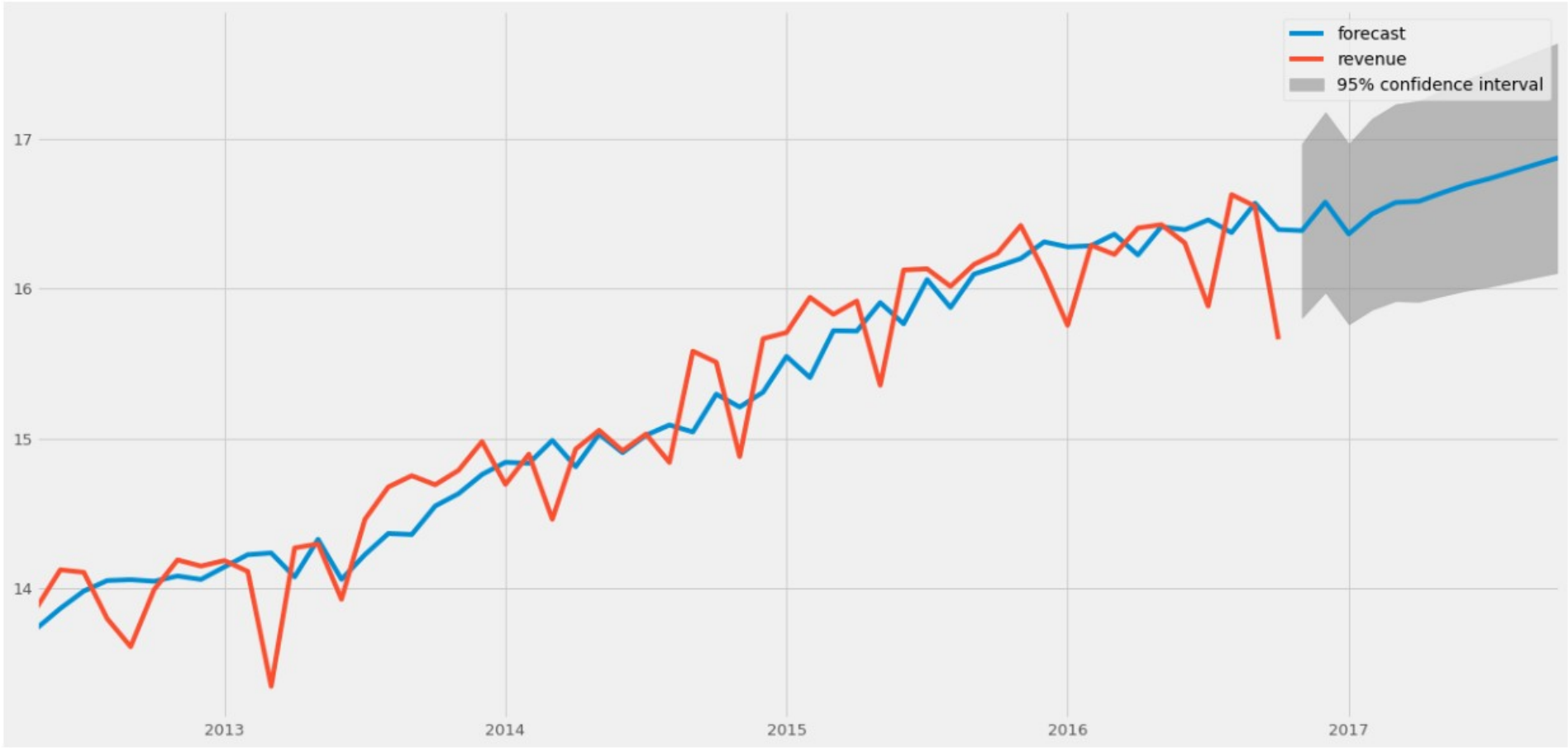


- In this p and q values can be determined as : p=1,q=1. This gives the optimal values for ARIMA model (p,d,q) which are (1,1,1)

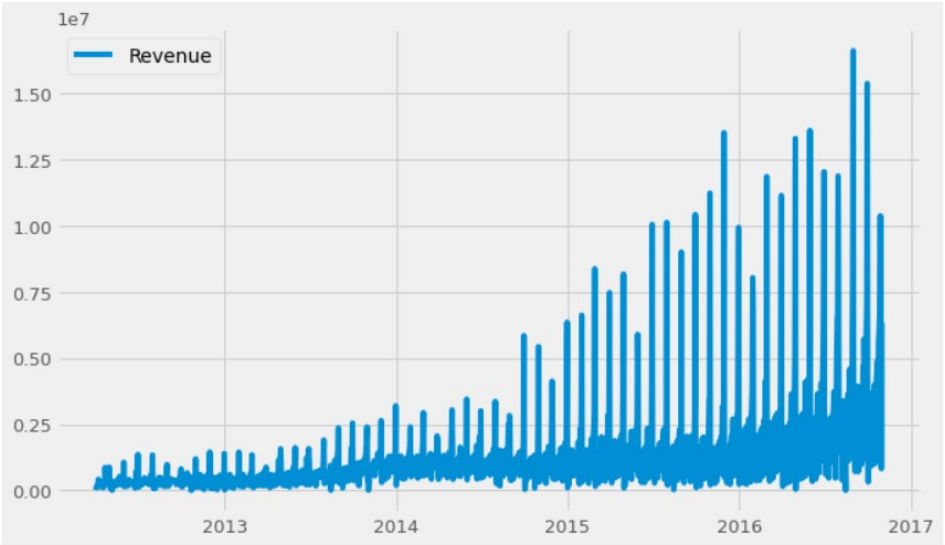
Plotting Arima model with order(1,1,1)



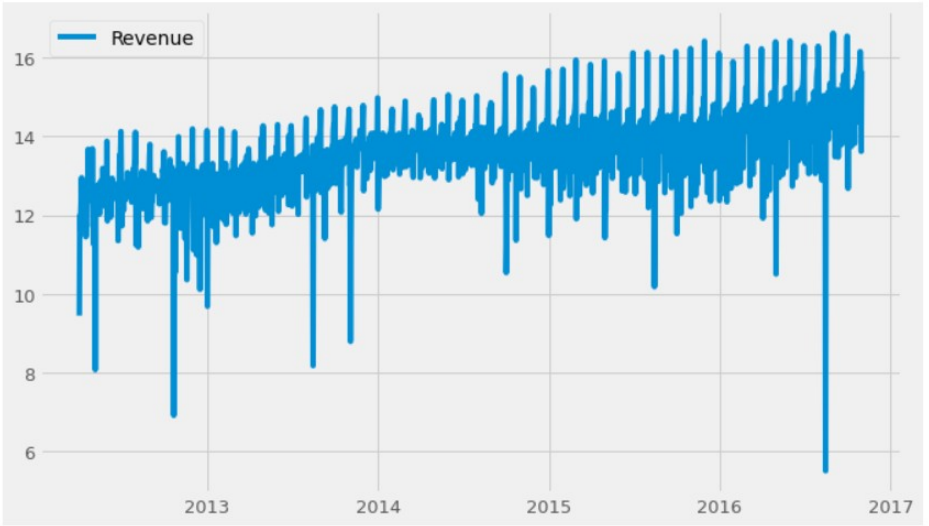
Forecasting with Arima model (period = 12 months)



Prophet Model

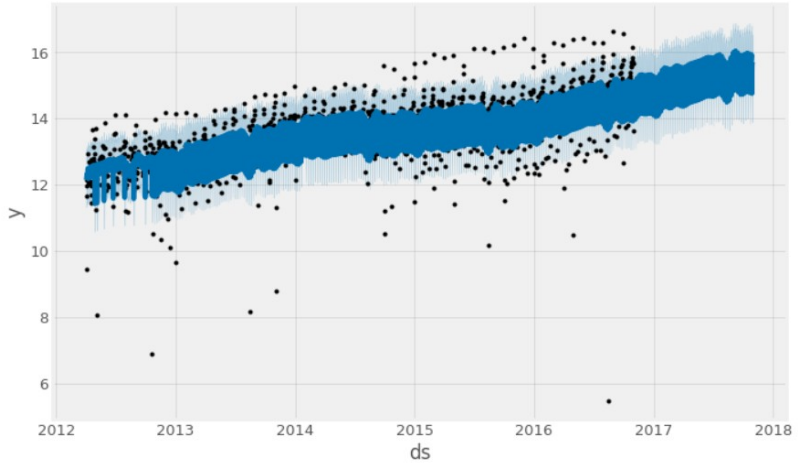


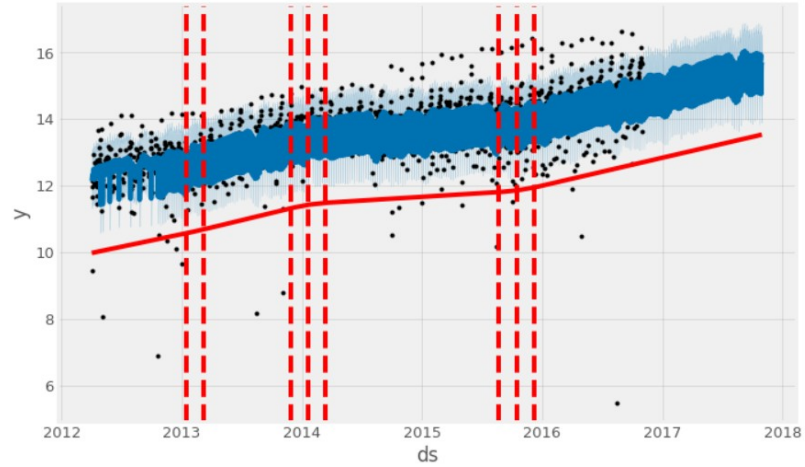
Plot for Revenue

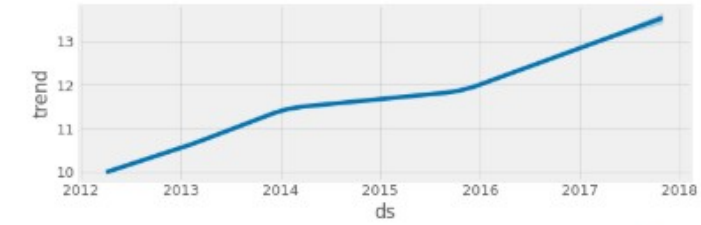


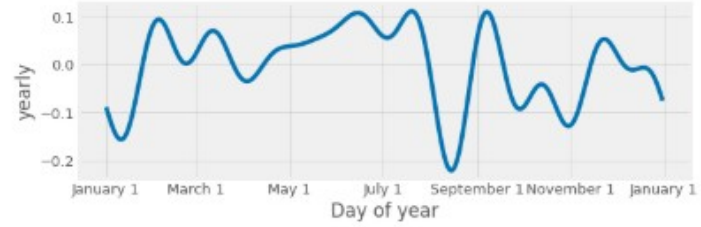
Plot for Log of Revenue

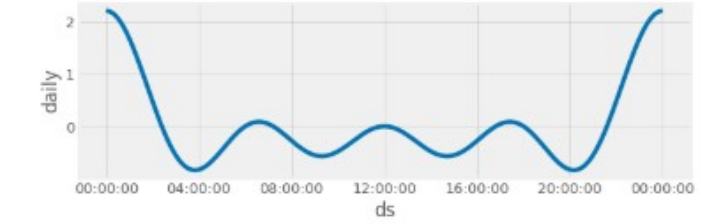
Basic & components Plots





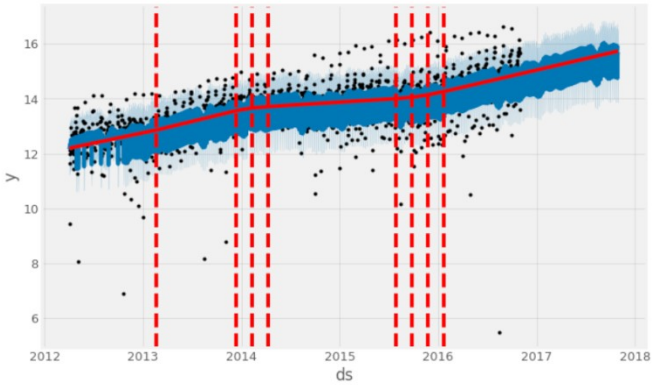




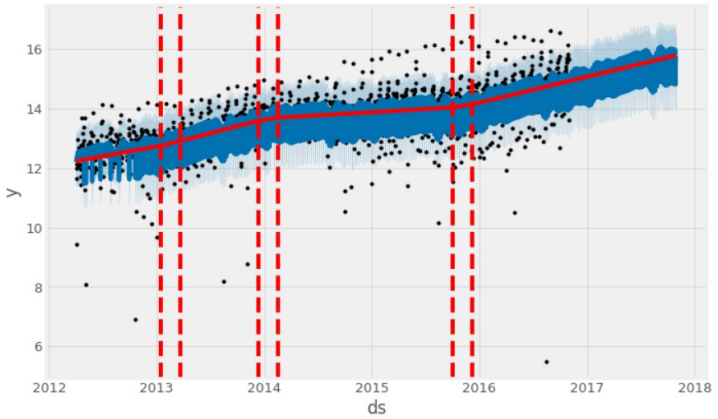


Model Tuning

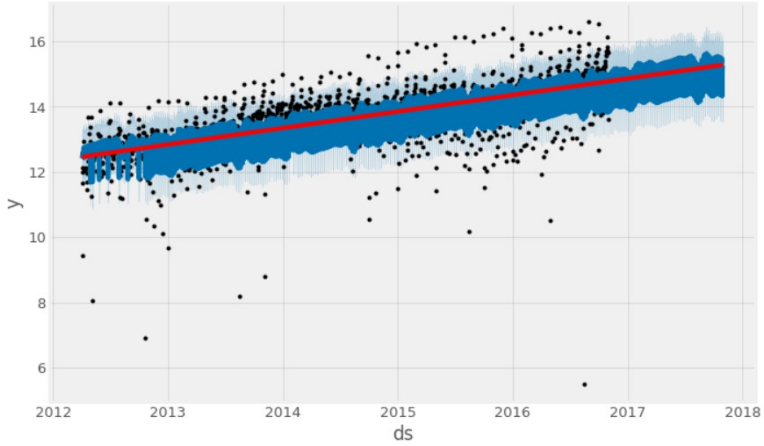
```
pro_change= Prophet(changepoint_range=0.9)
```



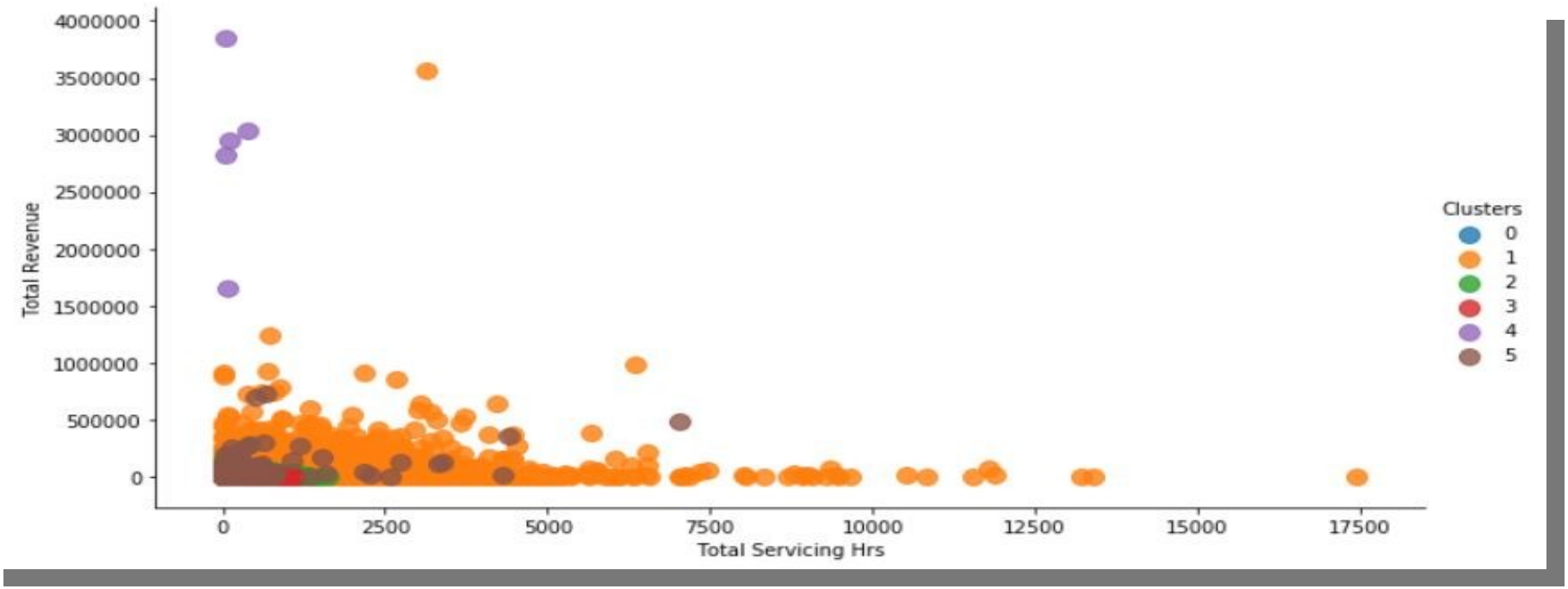

```
pro_change= Prophet(n_changepoints=20, yearly_seasonality=True, changepoint_prior_scale=0.08)
```



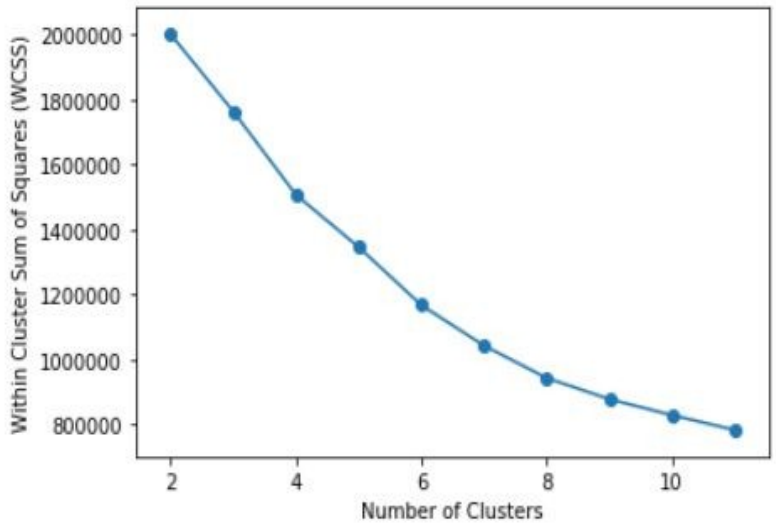

```
pro_change= Prophet(n_changepoints=20, yearly_seasonality=True, changepoint_prior_scale=0.001
```



Market Segmentation



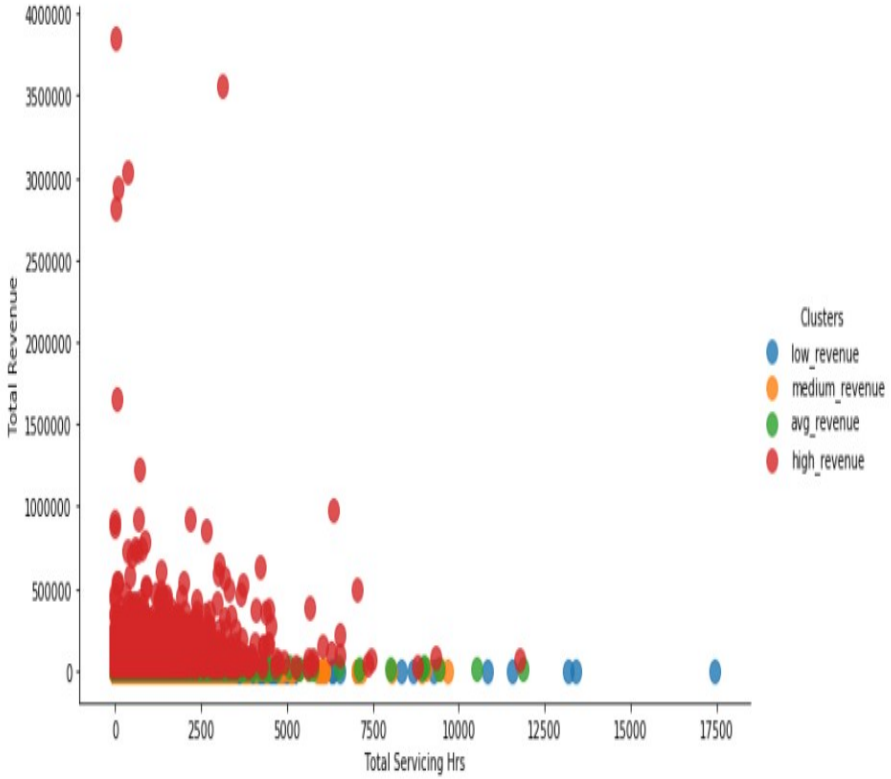
Clusters	Avg revenue	Avg Servicing Hrs	Customer Count
0	4069.758366	47.059950	90718
1	23231.093351	394.598749	34529
2	5037.272011	62.422289	52299
3	4738.955693	44.360695	106671
4	2862298.858000	120.571577	5
5	30004.909585	179.498966	434



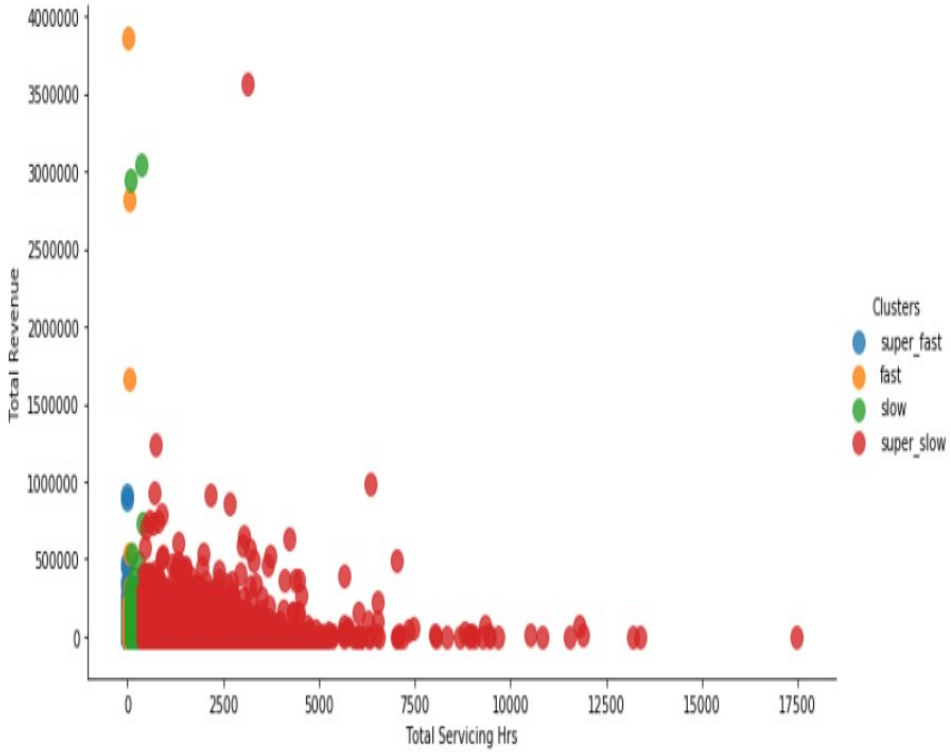
Cust Type	Average revenue	Average Servicing Hrs	Customers Count
0	**5033**	**54**	**90718**
Insurance	4235	49	1
MFCWL	6796	66	41
Retail	4069	47	90676
1	**15160**	**263**	**34529**
MFCWL	7085	131	10
Retail	23236	395	34519
2	**6212**	**74**	**52299**
MFCWL	7388	85	22
Retail	5036	62	52277
3	**6635**	**56**	**106671**
MFCWL	8533	68	40
Retail	4738	44	106631
4	**2941186**	**107**	**5**
Fleets	3335623	38	2
Retail	2546749	176	3
5	**25445**	**184**	**434**
Corporate- M&M	39127	207	73
Corporate others	25047	149	173
Fleets	32319	197	179
Franchise	5285	184	9
Grand Total	**403684**	**127**	**284656**

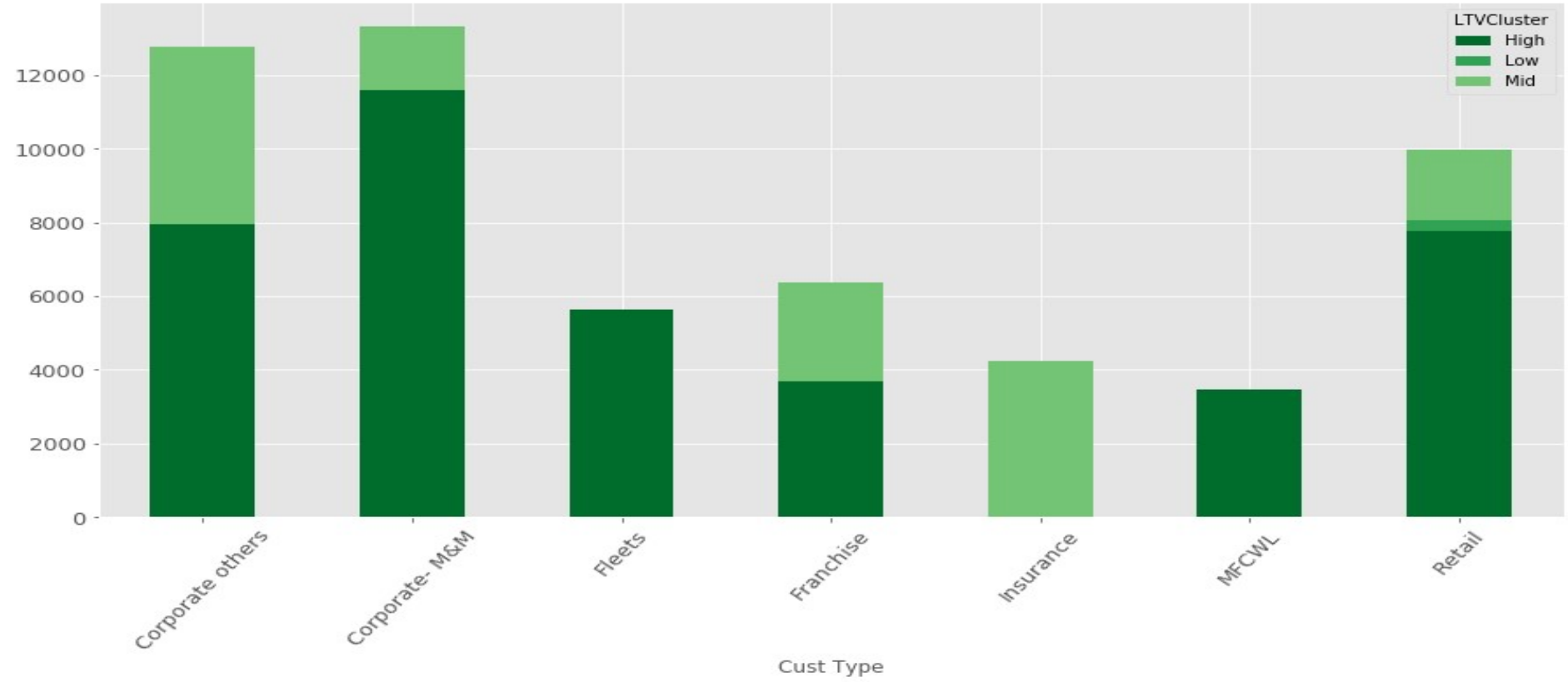
Rule Based Clustering

Clusters	Total Revenue	Total Servicing Hrs	Customer No.
high_revenue	56448.374909	349.862933	14233
avg_revenue	12768.362427	133.725734	56931
medium_revenue	4312.884887	76.767076	71163
low_revenue	915.945283	55.597661	142326



Clusters	Total Revenue	Total Servicing Hrs	Customer No.
super_slow	26637.683708	922.303598	14233
slow	12746.639368	164.518031	56931
fast	6793.334705	39.476048	71164
super_fast	2665.532217	4.685069	142328





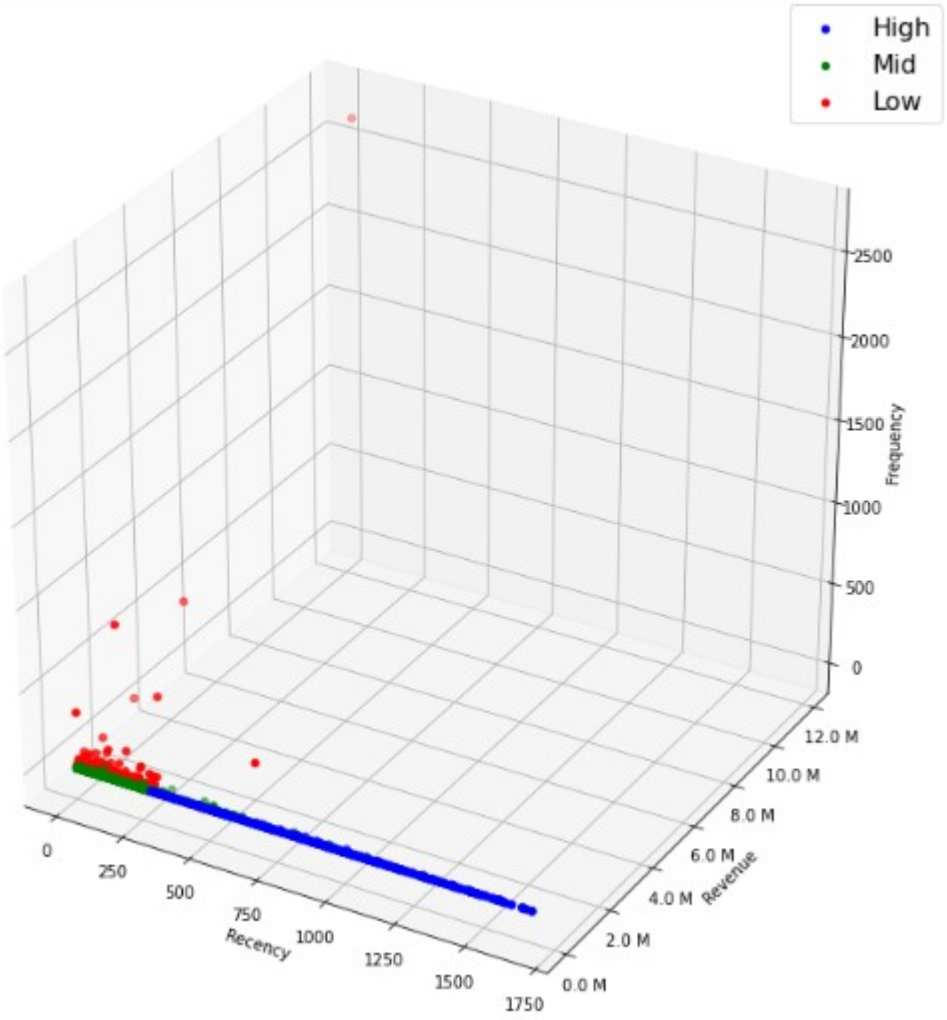
LTVCluster	count	mean	std	min	25%	50%	75%	max
High	235549.00	112.08	312.28	0.01	5.36	28.52	98.83	14138.89
Low	77512.00	48.03	247.59	0.01	0.82	2.81	17.40	5211.70
Mid	164028.00	60.47	236.41	-0.01	2.74	6.78	40.24	17769.26

RFM Clustering

- Customer are clustered based on there Recency, Frequency and Revenue
- KMean clustering algorithm is used for clustering the users
- We have clustered the customers in High, Mid and Low value clusters

High Level Approach

- Using Elbow Method 4 clusters were selected for Recency, Frequency and Monetary
- Overall Score = Recency Cluster + Frequency Cluster + Monetary Cluster
- Customer clustering is done based on Overall Score

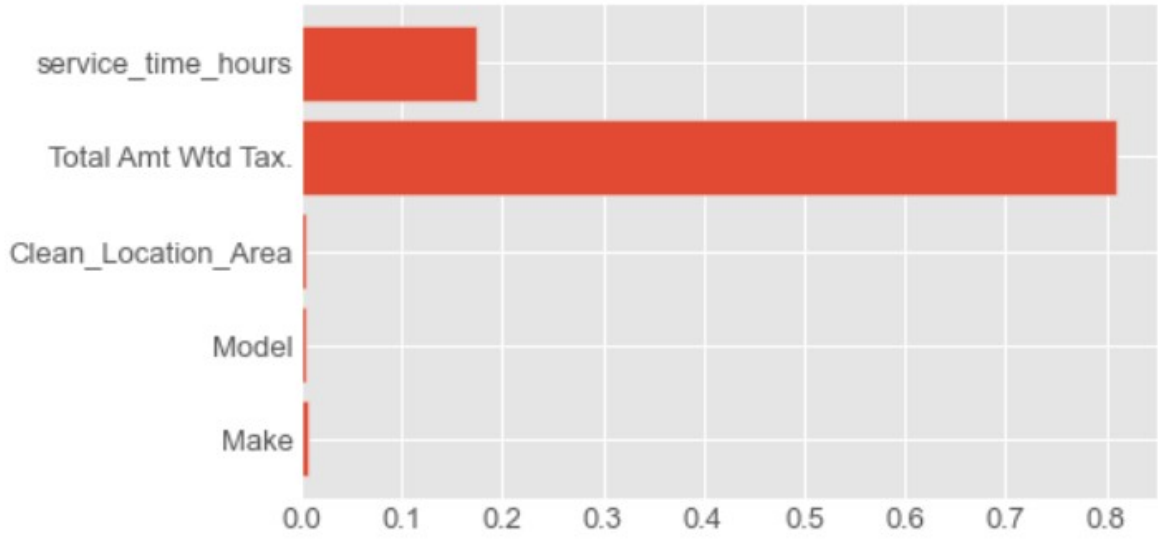


Customer Life Time Value Prediction

- Customer lifetime value is a metric that indicates the total revenue a business can reasonably expect from a single customer account.
- It is calculated as Average purchase frequency multiplied by Average spend per visit.
- Considered 2015 year data.

- Business can use this metric to increase ROI.

Feature Importance



- We can see that Total Amount feature is of highest value followed by service time

LTV Classification

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
XGBClassifier	0.86	0.82	None	0.86	10.43
LGBMClassifier	0.86	0.81	None	0.85	1.87
BaggingClassifier	0.84	0.80	None	0.84	3.76
RandomForestClassifier	0.83	0.79	None	0.83	1.50
DecisionTreeClassifier	0.80	0.77	None	0.80	0.60
ExtraTreesClassifier	0.80	0.76	None	0.80	1.11
SVC	0.80	0.75	None	0.80	75.68
GaussianNB	0.79	0.75	None	0.79	0.14
QuadraticDiscriminantAnalysis	0.79	0.75	None	0.79	0.13
NuSVC	0.79	0.73	None	0.78	199.18
LogisticRegression	0.79	0.72	None	0.77	0.53
ExtraTreeClassifier	0.76	0.72	None	0.76	0.21
CalibratedClassifierCV	0.79	0.72	None	0.77	41.05
KNeighborsClassifier	0.77	0.72	None	0.77	2.41
LinearSVC	0.78	0.70	None	0.74	23.00
PassiveAggressiveClassifier	0.72	0.70	None	0.73	0.46
AdaBoostClassifier	0.75	0.67	None	0.69	5.11
SGDClassifier	0.74	0.65	None	0.66	0.48
Perceptron	0.73	0.65	None	0.69	0.30
BernoulliNB	0.70	0.59	None	0.60	0.33
NearestCentroid	0.57	0.57	None	0.59	0.08
RidgeClassifier	0.60	0.45	None	0.49	0.43
RidgeClassifierCV	0.60	0.45	None	0.49	0.21
LinearDiscriminantAnalysis	0.59	0.45	None	0.49	0.58
CheckingClassifier	0.25	0.33	None	0.10	0.08
DummyClassifier	0.38	0.33	None	0.38	0.21

- *Classification Output*
  - *Low Ltv Group (<3036)*
  - *Medium Ltv Group(3036 & 8801)*
  - *High Ltv Group (>8801)*

LTV Regression

Model	R-Squared	RMSE	Time Taken
ExtraTreesRegressor	0.98	3180.36	0.93
ExtraTreeRegressor	0.97	3679.84	0.14
BaggingRegressor	0.97	3761.94	1.29
RandomForestRegressor	0.96	3836.54	1.39
GradientBoostingRegressor	0.96	4125.34	1.85
KNeighborsRegressor	0.94	5114.60	1.82
LGBMRegressor	0.88	6936.91	0.36
HistGradientBoostingRegressor	0.88	6991.13	1.09
XGBRegressor	0.81	8786.86	2.78
MLPRegressor	0.77	9569.73	54.49
LassoLars	0.77	9742.66	0.03
LassoCV	0.76	9786.43	0.69
RidgeCV	0.76	9789.57	0.05
Lasso	0.76	9789.99	0.03
Ridge	0.76	9790.17	0.03
BayesianRidge	0.76	9790.18	0.03
OrthogonalMatchingPursuitCV	0.76	9790.24	0.16
LassoLarsIC	0.76	9790.24	0.04
LassoLarsCV	0.76	9790.24	0.06
LarsCV	0.76	9790.24	0.07
Lars	0.76	9790.24	0.20
TransformedTargetRegressor	0.76	9790.24	0.03
LinearRegression	0.76	9790.24	0.04
SGDRegressor	0.76	9966.79	0.22
PassiveAggressiveRegressor	0.72	10748.85	0.15
HuberRegressor	0.72	10766.35	0.70
ElasticNet	0.71	10811.02	0.06
DecisionTreeRegressor	0.68	11377.35	0.26
LinearSVR	0.68	11463.95	0.08
RANSACRegressor	0.65	11882.59	0.05
OrthogonalMatchingPursuit	0.64	12124.81	0.02
AdaBoostRegressor	0.59	12897.56	0.94
ElasticNetCV	0.10	19167.66	0.70
SVR	0.09	19233.64	201.11
NuSVR	0.09	19268.06	215.87
DummyRegressor	-0.00	20170.81	0.02

- *Regression Output - Continuous LTV value*

Value delivered to the business

- For Accidental order type 25% of total amount goes towards outsourced work, repairing the some part can reduce outsourced work
- There is no seasonally rush, but on daily basis 5PM-6PM are rush hours
- Mid-Value customers are most revenue generating cluster
- Recently we have been losing lots of High-Value customers
- We Should Stock Engine Oil, Oil Filter since they are high demand items
- Highest revenue & repairs during mid year (around July) -> Hire contract staff
- Use less load day (i.e. Sunday) -> for training on soft skills & technical aspects
- Major customer types are retail -> Work on tie ups with corporate & fleet (ola, uber, etc). Use corporate discounts & offers.
- Investigate what are the reasons for decrease in revenue in the Plants present in state of Telangana , Uttar Pradesh
- Please take feedback from Customers after service is given in for example (rating out of 5). Since it can be used to find the quality of service