Capstone Project

Mahindra First Choice Services (MFCS)

Cohort – DSMP 28th July 2019

Team-13
Pankaj Kumar
Rinkesh Gaikwad
Piyusha Khot
Devangi Sheth

Geolocation Based Customer Analysis:

Problem Statement-1 :
Identifying the ownership pattern of cars through out the country.

Problem Statement-2:

Identify the type of order each state receives.

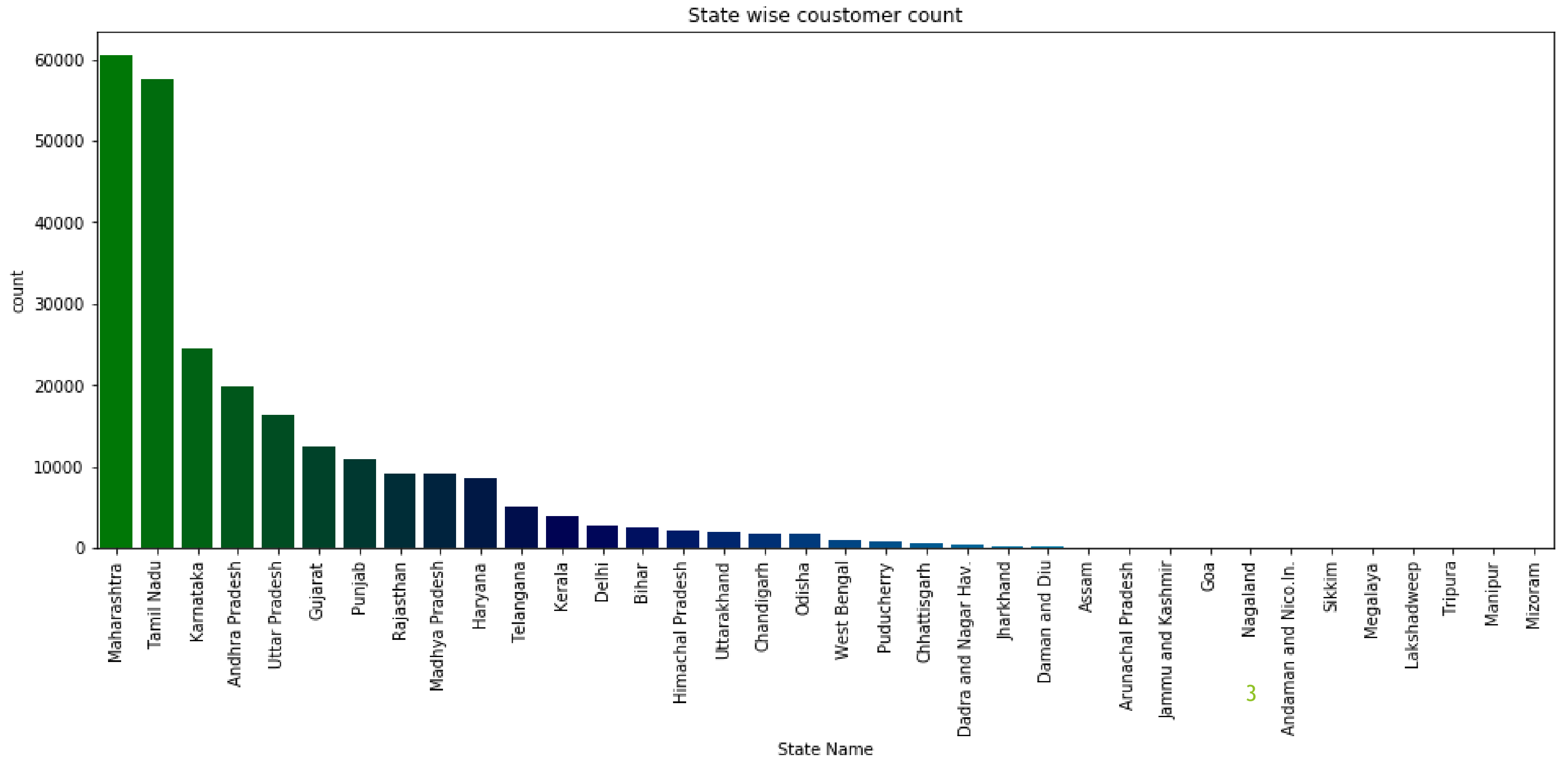
Problem Statement-3:

Customer Lifetime value prediction - Based on Customer segments.

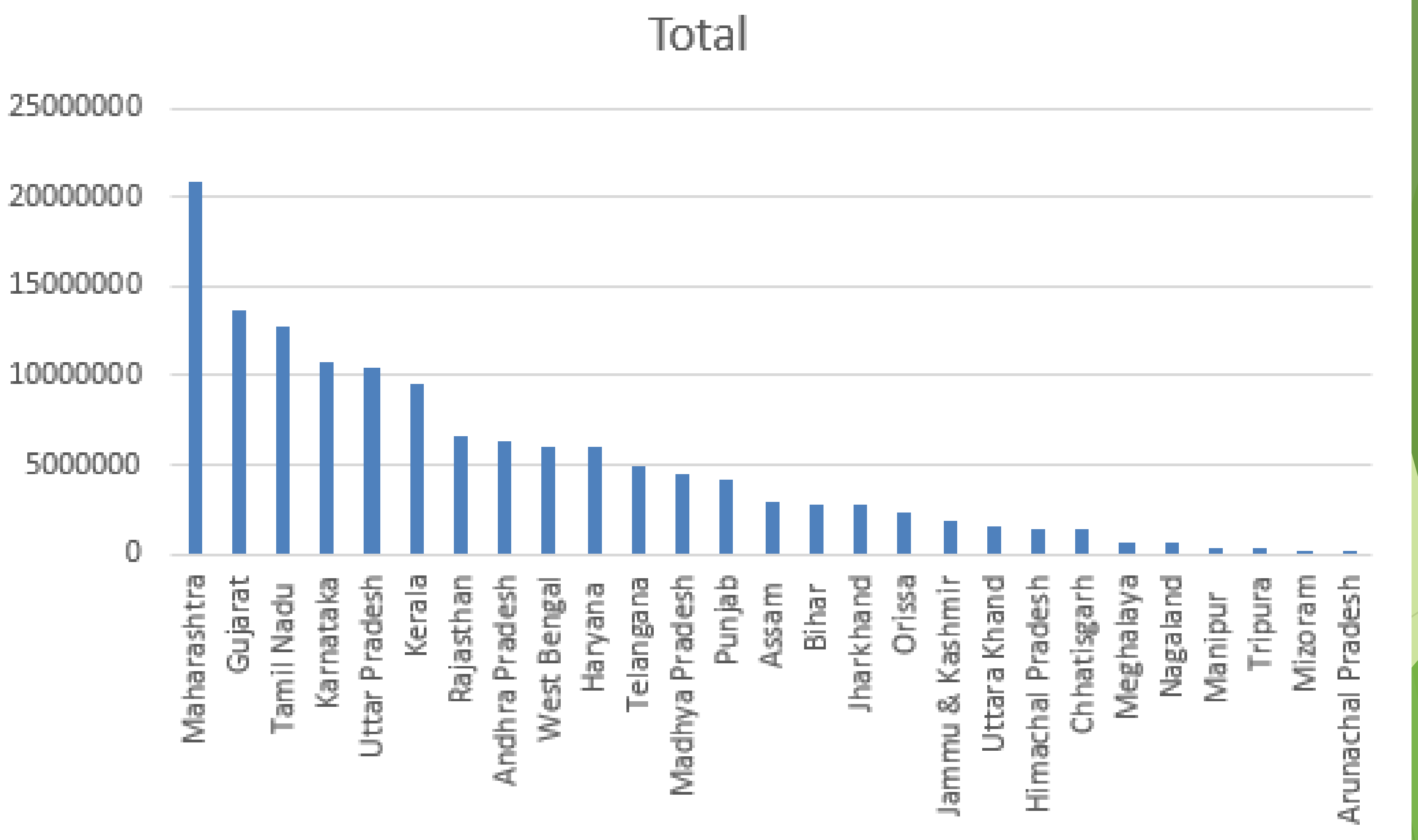
Problem Statement - 1
Ownership Patterns

STATE WISE CUSTOMER COUNT

- Demand for customers from various states
- Highest demand seen in Maharashtra and Tamil Nadu

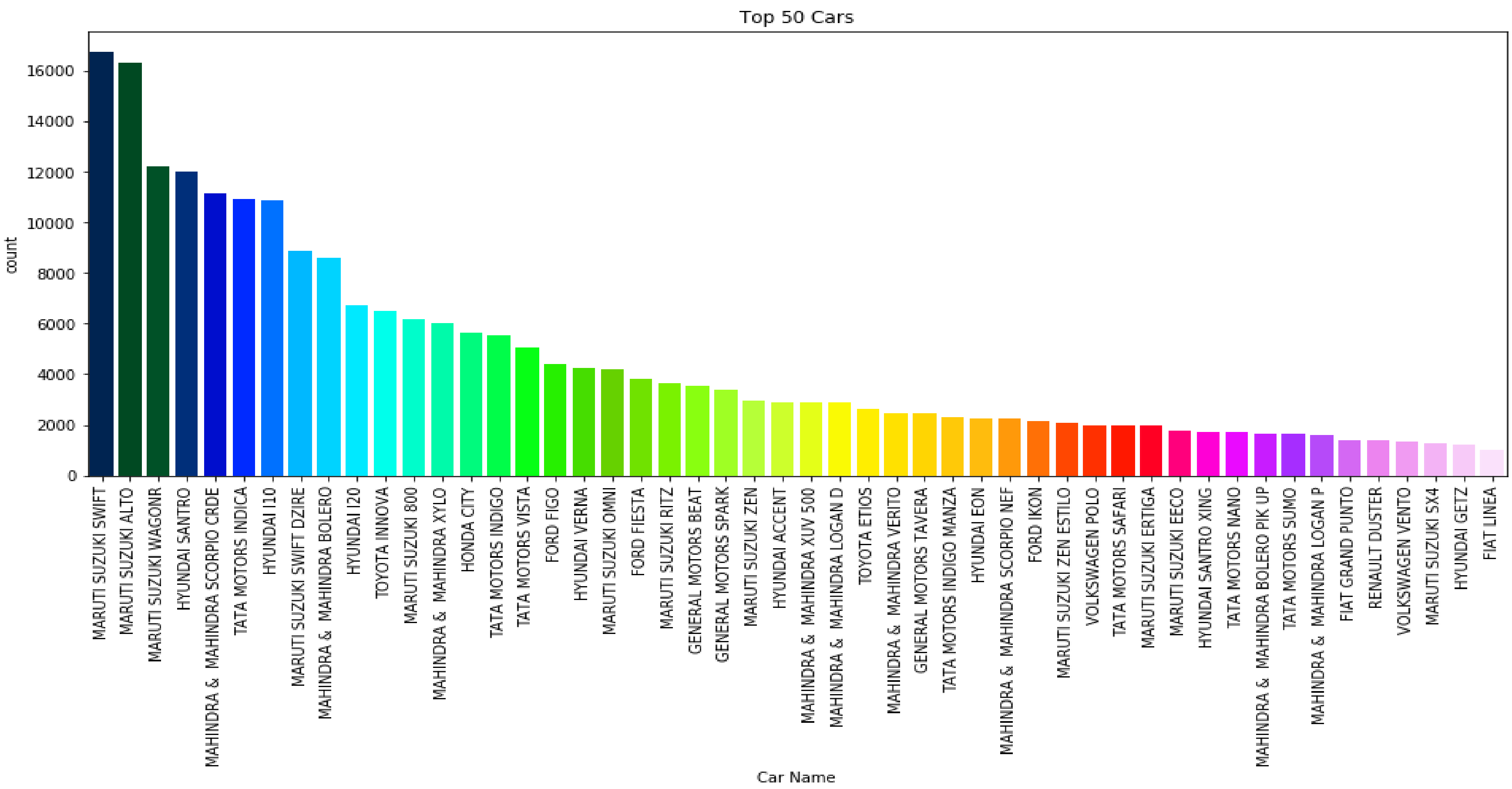


Govt Data – Count of Car+ Jeep+ LMV

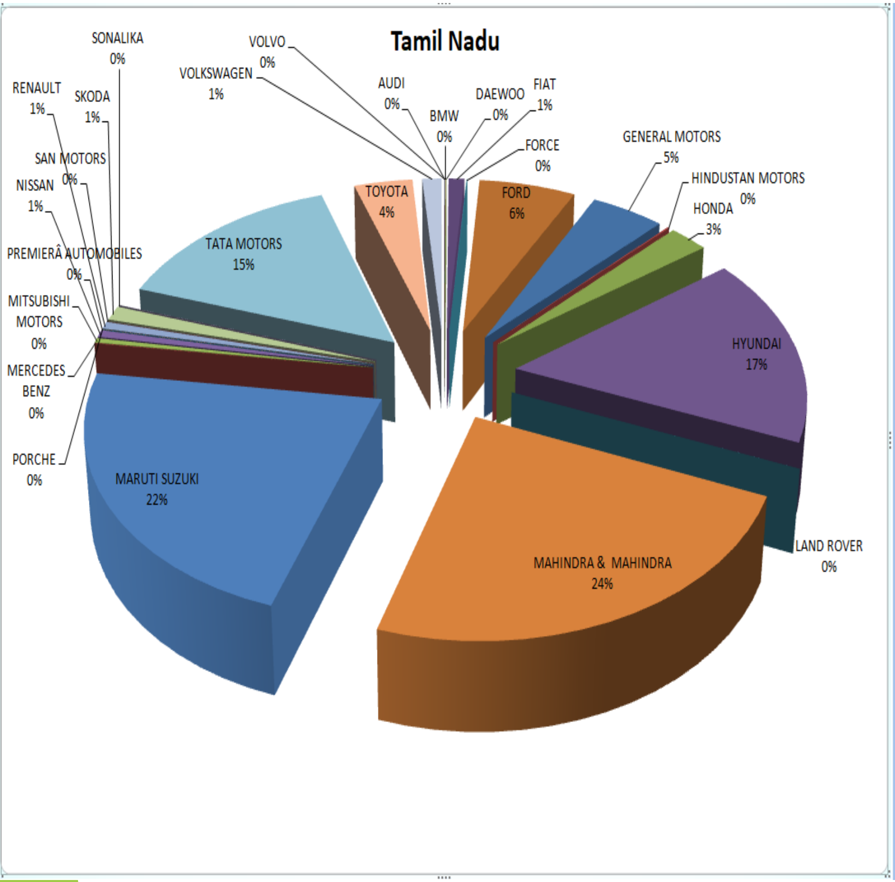


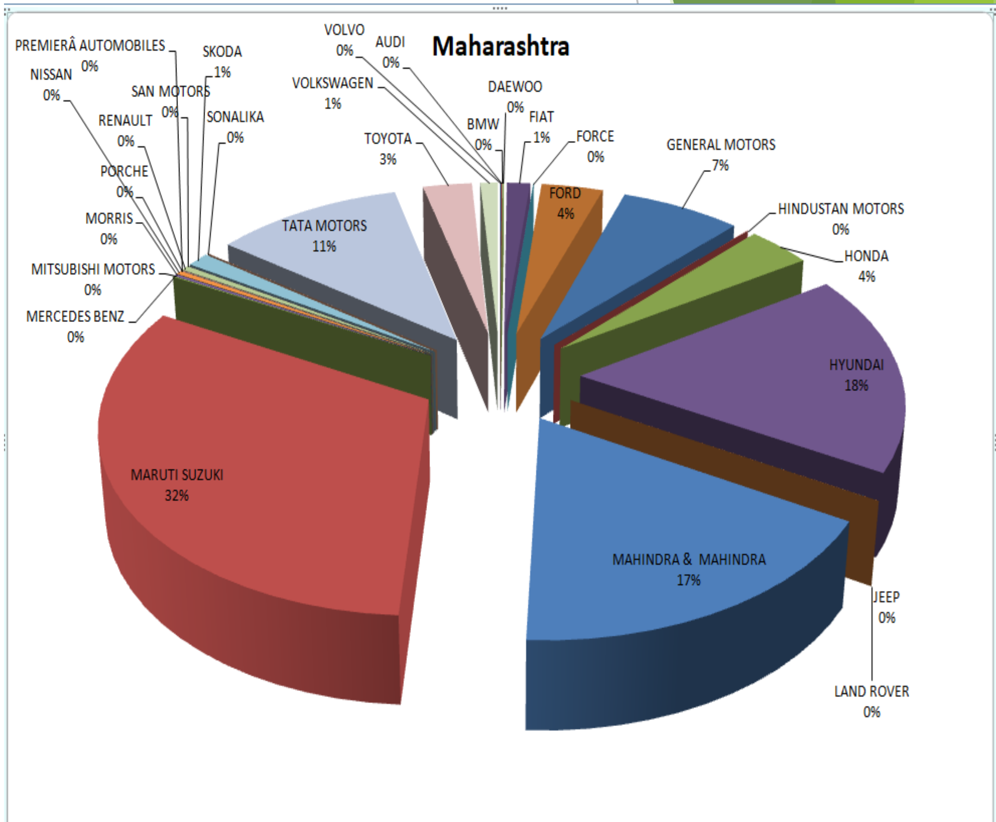
From Year 2012 to 2016

❑ Demand for Cars Make – Top 50

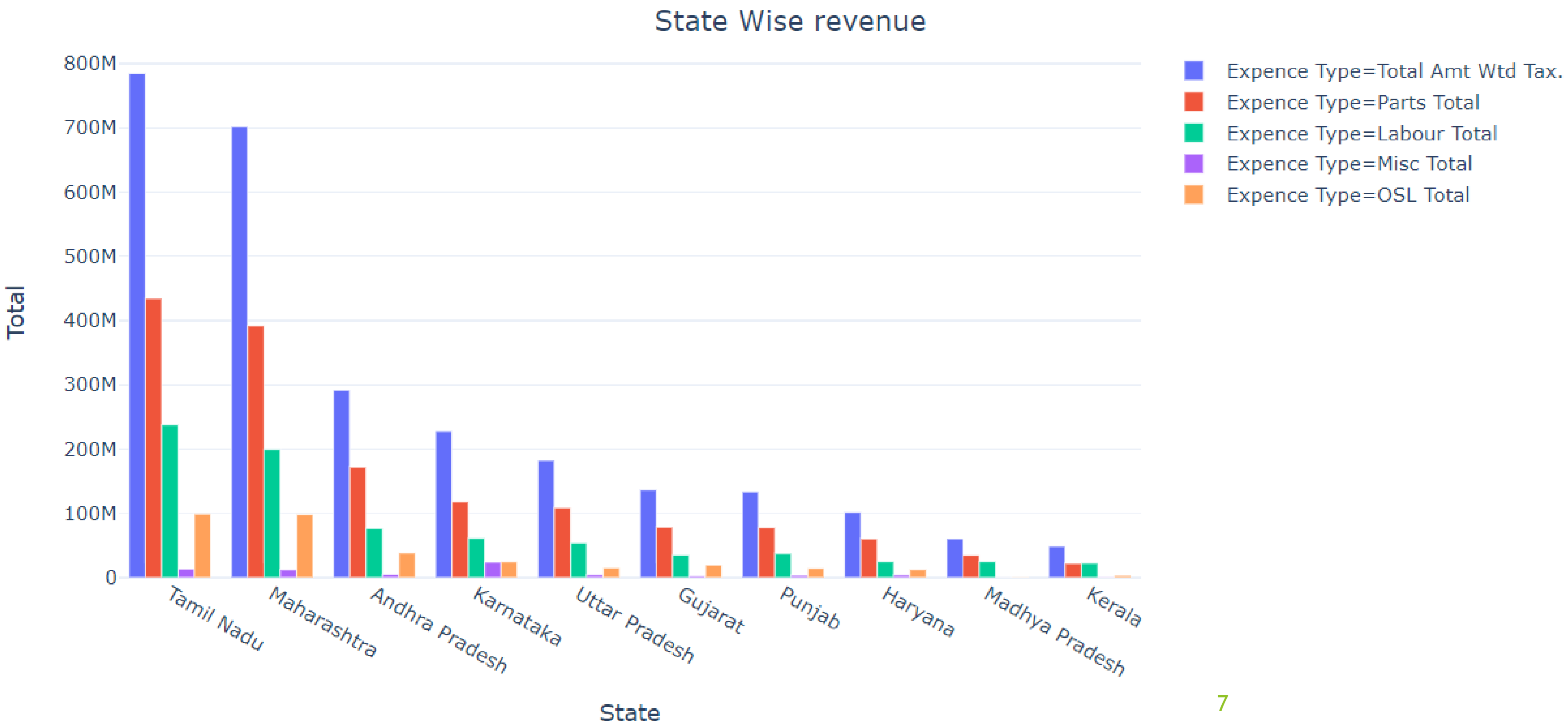


- Higher demand for ‘Maruti Suzuki’ & ‘Mahindra and Mahindra’ in ‘Tamil Nadu’ and ‘Maharashtra’.

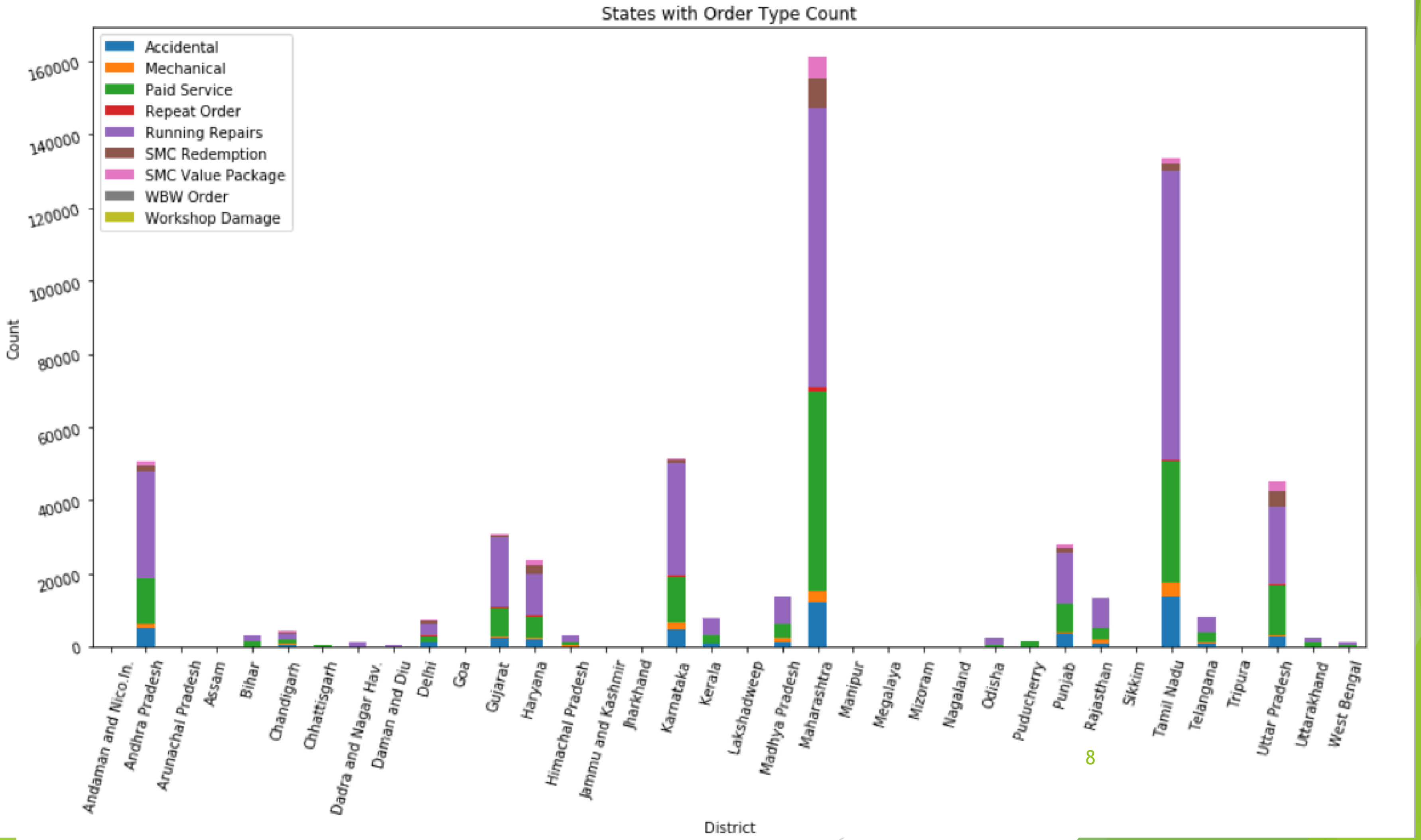




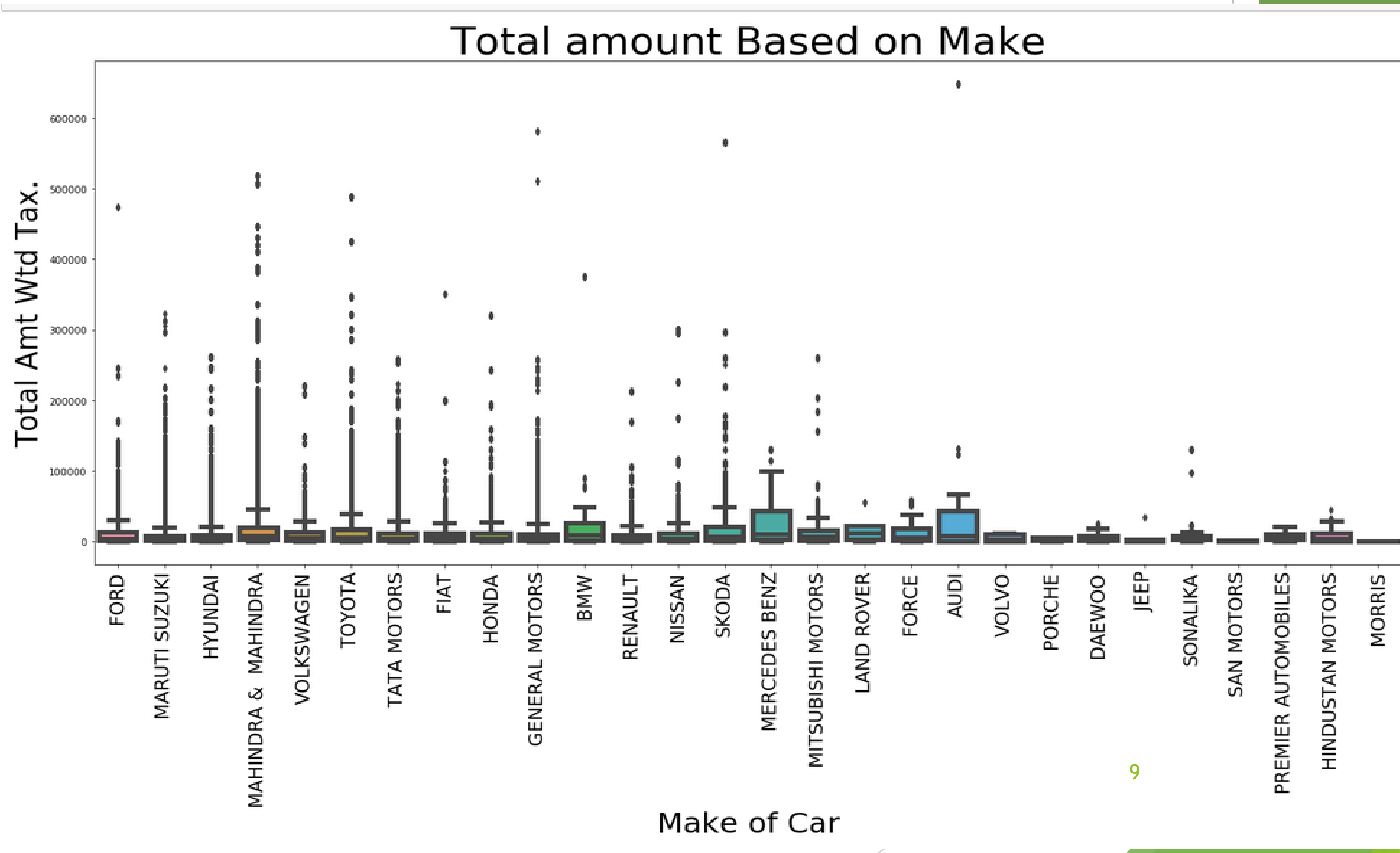
STATE WISE – Spending Patterns



TYPE OF ORDER EACH STATE RECEIVES



Revenue Outliers



Recommendations - Problem1

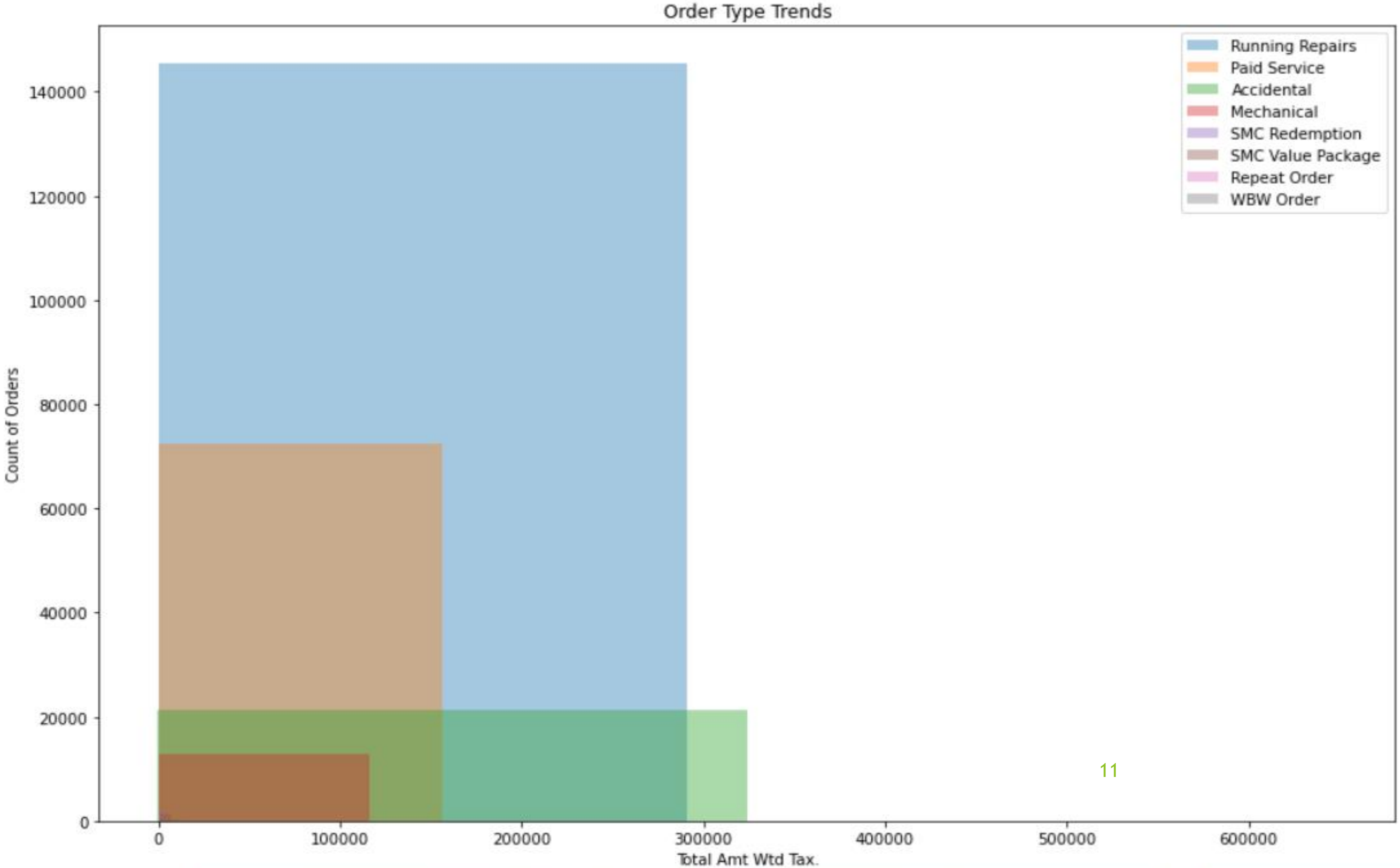
- States which are under performing while having great potential -> **Gujarat & Kerala**

- **Indian "Make"** are the biggest count makers **-> stock parts & hire experts**

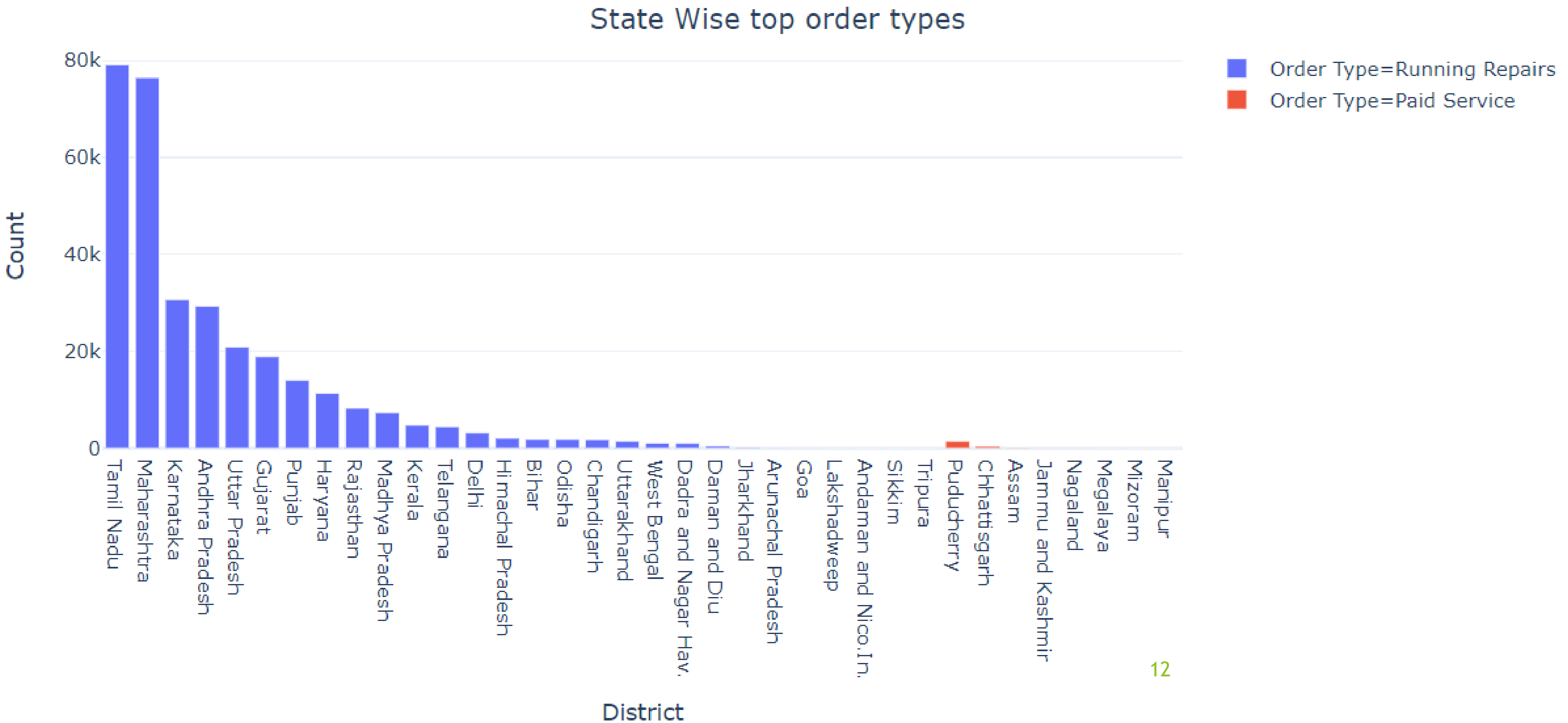
- Revenue Outliers are Audi, Mercedes & Skoda **-> Cash Cows**

ORDER TYPES TRENDS = Problem No 2

- Customers majorly come for running repairs in MFC

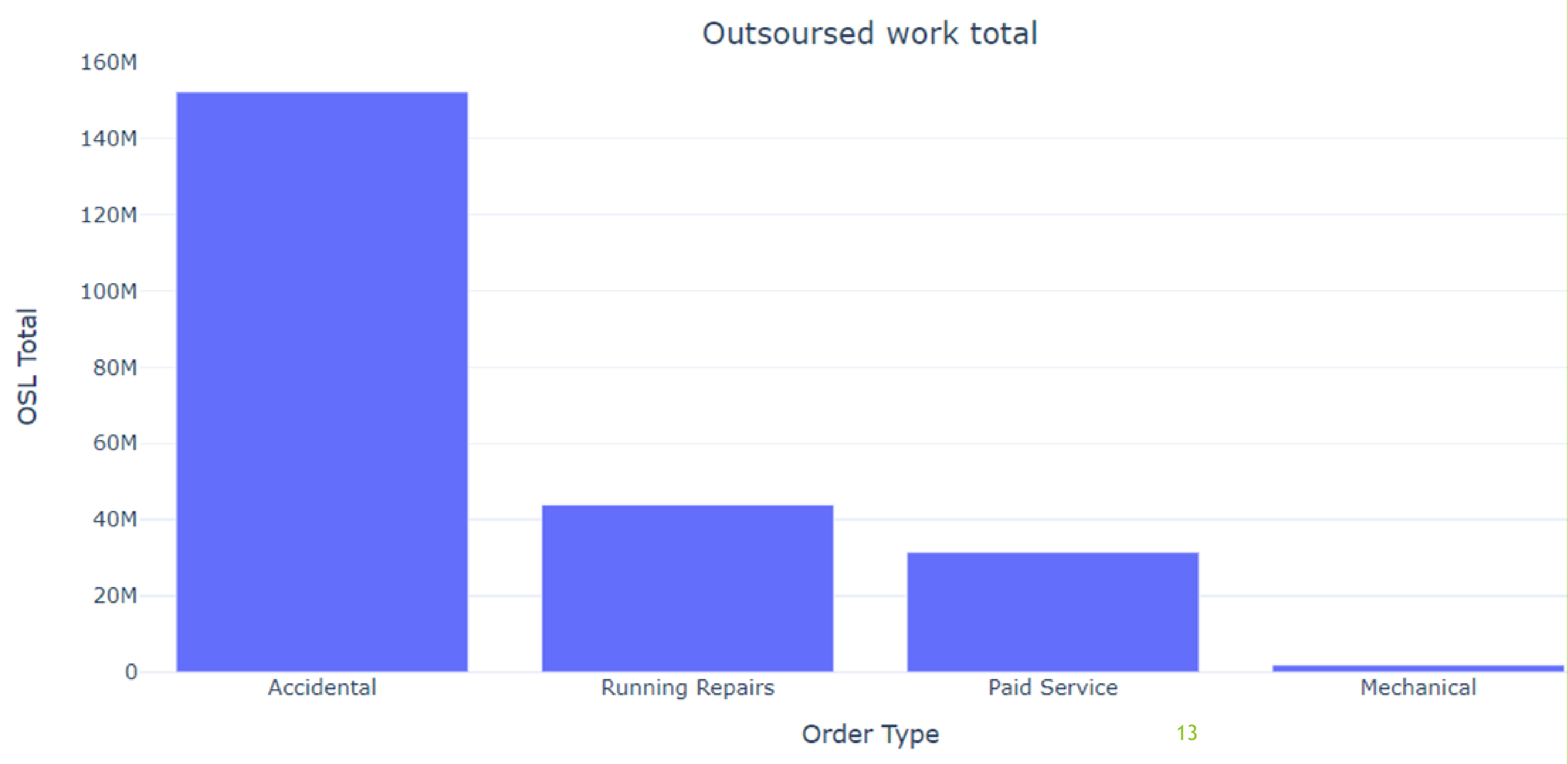


State Wise Top Order Types

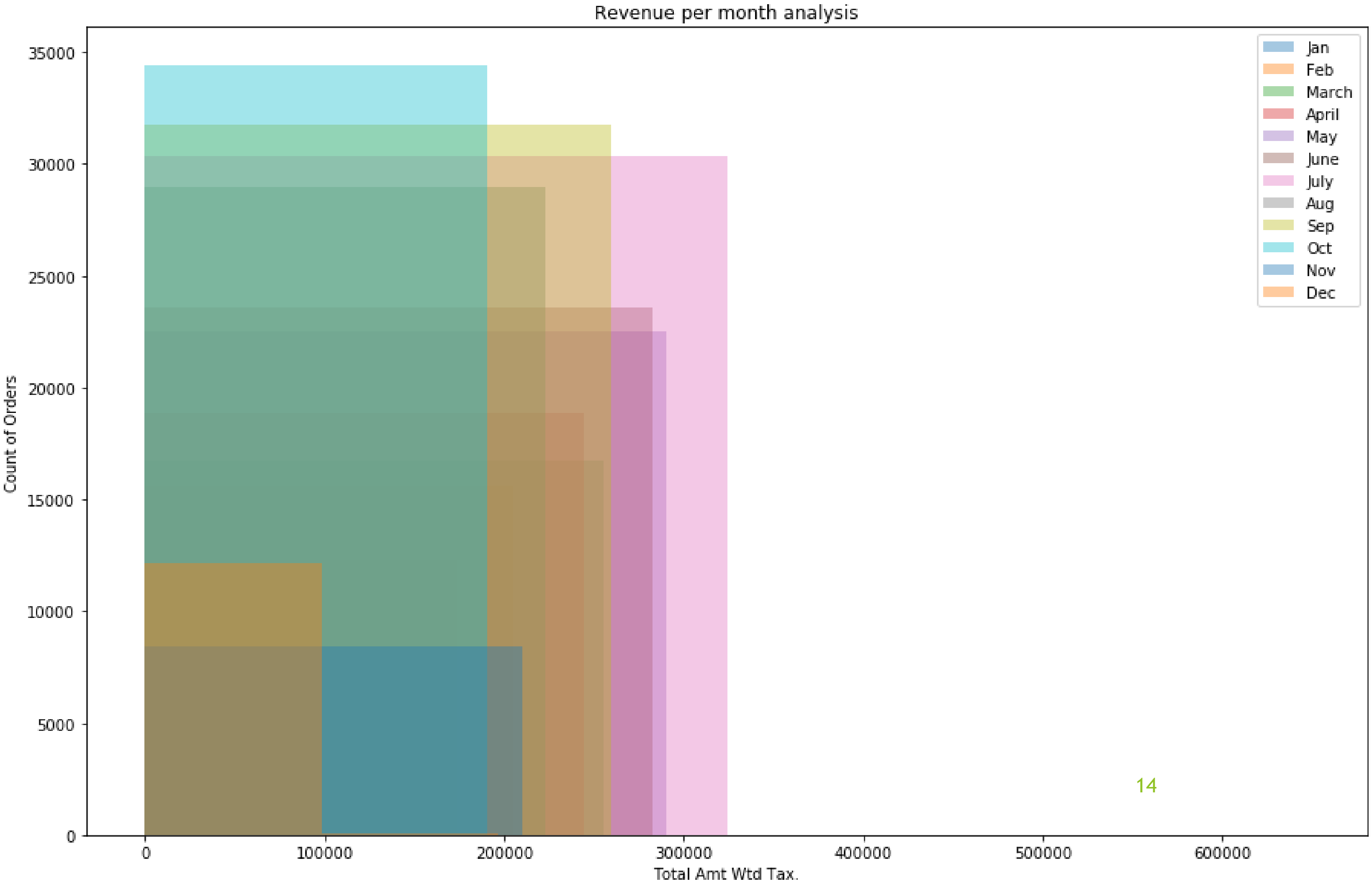


ORDER TYPE VS OSL TOTAL

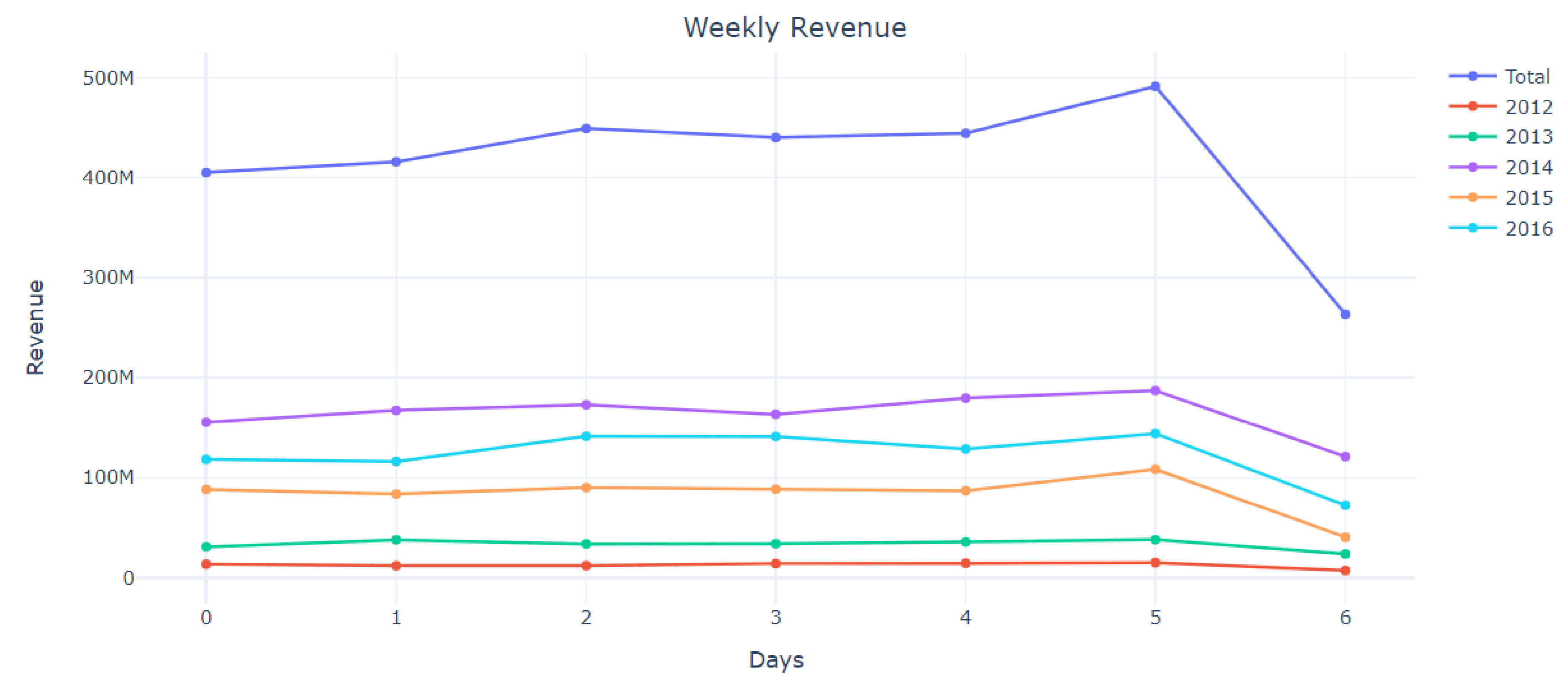
- Accidental orders are majorly outsourced



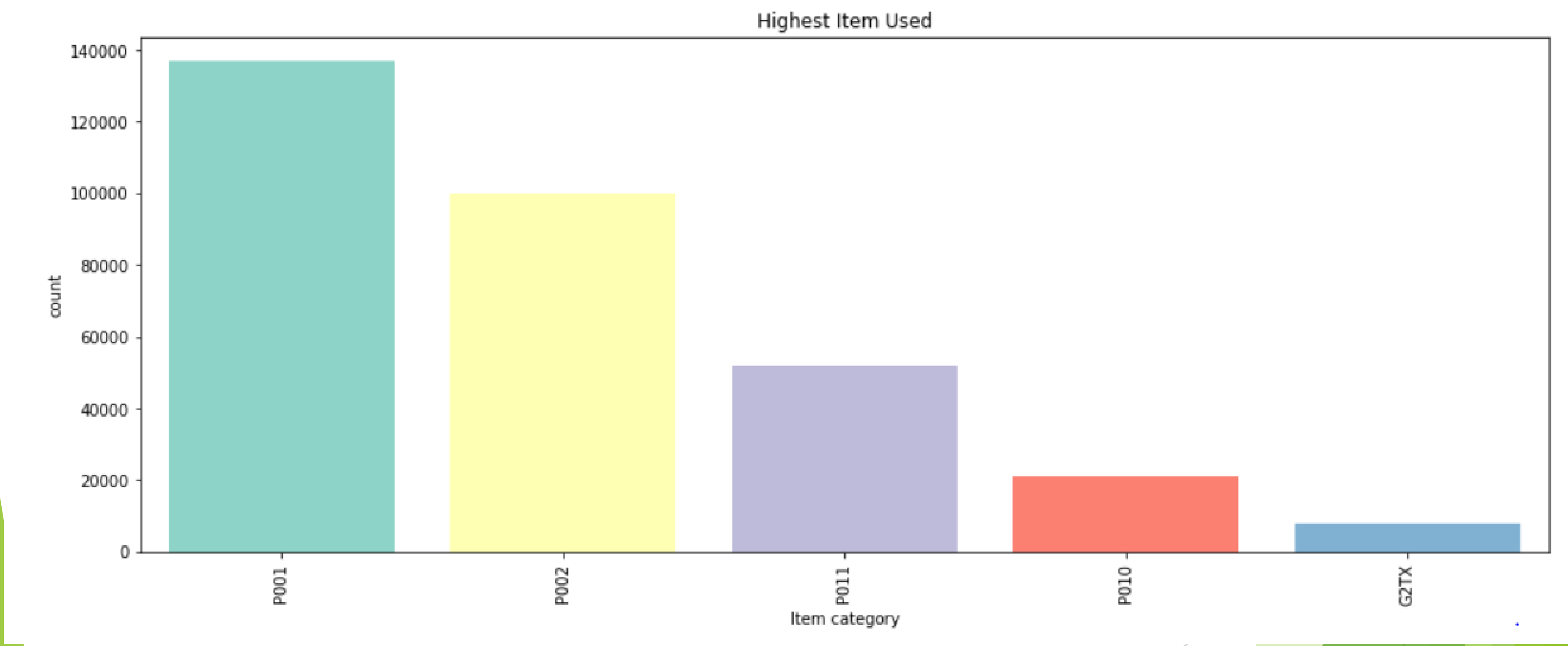
Revenue Per Month – Count Of Orders



Weekly Revenue



Highest Material Used



Recommendations – Problem2

- Highest revenue & repairs **during mid year** (around July) -> **Hire contract staff**

- Accidental orders are outsourced -> **Bit relax on TAT**

- Use less load day (i.e. Sunday) -> **for training on soft skills & technical aspects**

Problem Statement – 3

Cluster Analysis via KMeans

```
cluster_data.head()
```

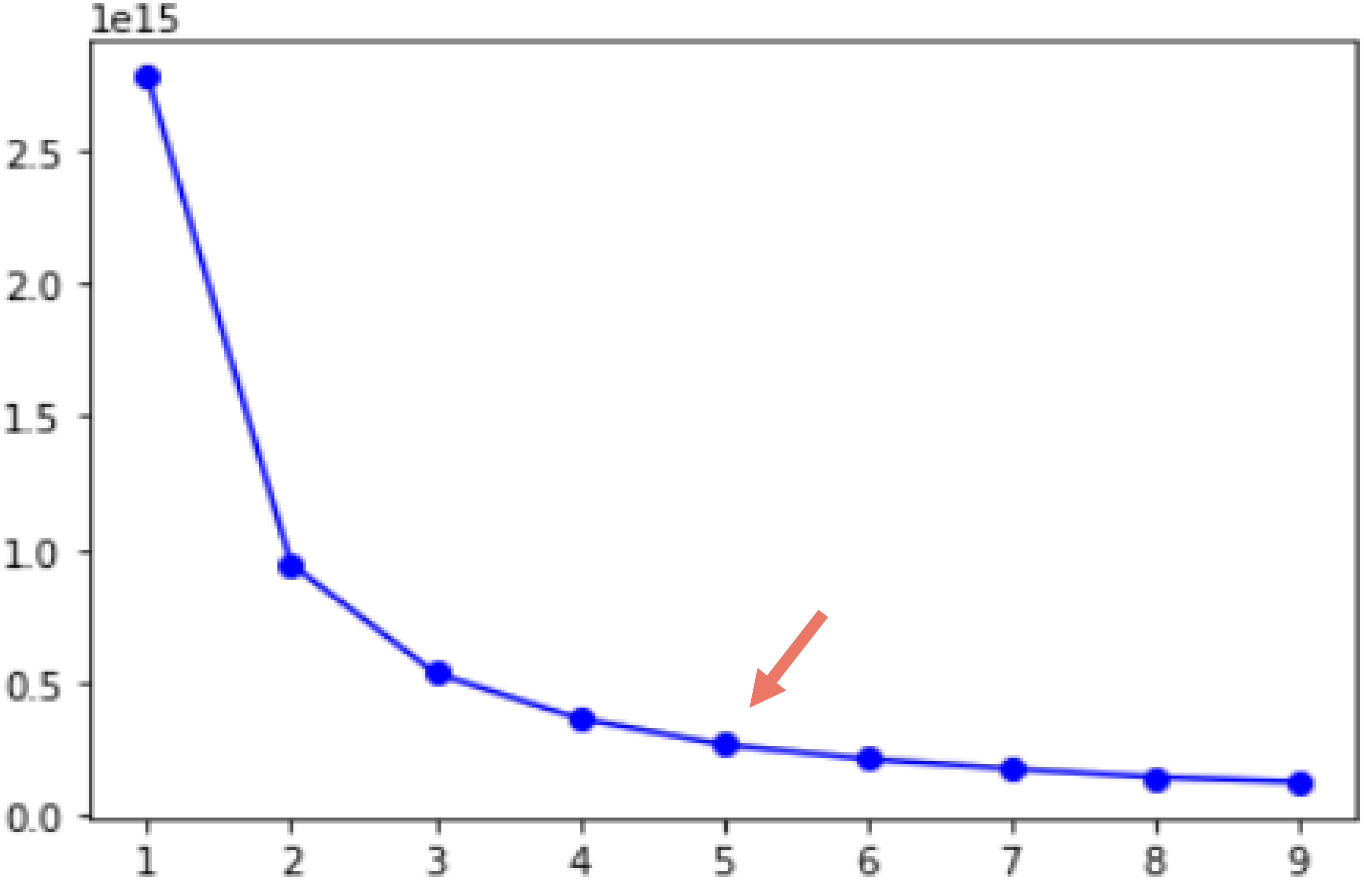
	Cust Type	Order Type	Make	State	ServiceTime_category	Data Origin	Job Card No	KMs Reading	Labour Total	Misc Total	OSL Total	Parts Total	Total Amt Wtd Tax.	Invoice month	Partner Type	Title	Order Quantity	Net value
0	5	2	5	20	3	9	227460	0	1802.05	0.0	0.0	399.73	2201.78	6	1.0	2.0	45.5	2201.78
1	5	4	13	20	1	0	413455	0	2874.80	0.0	0.0	3151.06	6025.86	7	1.0	2.0	202.5	6025.86
2	5	4	13	20	4	0	446448	0	2799.96	0.0	0.0	664.04	3464.00	9	1.0	2.0	189.0	3464.00
3	5	4	9	20	4	0	487708	0	554.95	0.0	0.0	1510.80	2065.75	10	1.0	2.0	7.0	2065.75
4	5	4	12	20	3	0	220109	1	75.02	0.0	0.0	578.43	653.45	5	1.0	2.0	13.0	653.45

Categorized based on Mileage & Service Efficiency

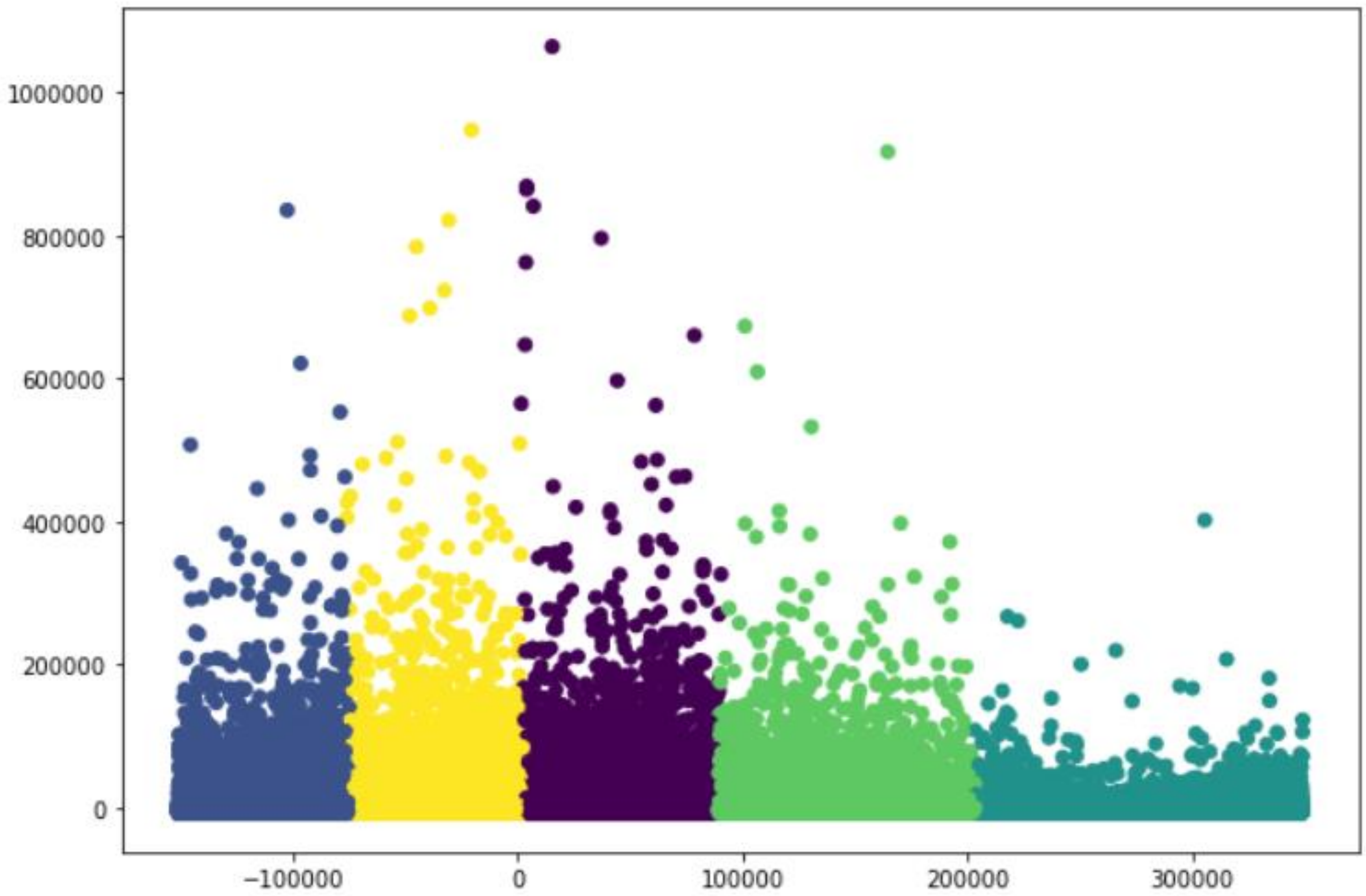
```
invoice_customer_plant["KMs Reading"] = invoice_customer_plant["KMs Reading"].astype(int)
invoice_customer_plant.loc[invoice_customer_plant["KMs Reading"] <= 100000, "KMs Reading"] = 0

invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 100000)
                           & (invoice_customer_plant["KMs Reading"] <= 200000), "KMs Reading"] = 1

invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 200000)
                           & (invoice_customer_plant["KMs Reading"] <= 300000), "KMs Reading"] = 2
invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 300000)
                           & (invoice_customer_plant["KMs Reading"] <= 400000), "KMs Reading"] = 3

invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 400000)
                           & (invoice_customer_plant["KMs Reading"] <= 500000), "KMs Reading"] = 4

invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 500000)
                           & (invoice_customer_plant["KMs Reading"] <= 600000), "KMs Reading"] = 5

invoice_customer_plant.loc[(invoice_customer_plant["KMs Reading"] > 600000)
                           & (invoice_customer_plant["KMs Reading"] <= 700000), "KMs Reading"] = 6

invoice_customer_plant.loc[invoice_customer_plant["KMs Reading"] > 700000, "KMs Reading"] = 7
```

```
<pd.Timedelta(hours=13),'ServiceTime_category']='super_fast'
]>=pd.Timedelta(hours=13))&(invoice_customer_plant['ServiceTime']<pd.Timedelta(hours=36)),'ServiceTime_category']='fast'
]>=pd.Timedelta(hours=36))&(invoice_customer_plant['ServiceTime']<pd.Timedelta(hours=61)),'ServiceTime_category']='mid'
]>=pd.Timedelta(hours=61))&(invoice_customer_plant['ServiceTime']<pd.Timedelta(hours=84)),'ServiceTime_category']='slow'
>=pd.Timedelta(hours=84),'ServiceTime_category']='super_slow'
```

Elbow Method - WCSS



Visualize clusters using PCA



LTV

- LTV = average purchase value **X** average purchase frequency **X** customer lifespan
  - invest in customer retention and acquisition

- Assuming cost on retention & acquisition as zero
- Considering 2015 data & lifespan of 1 year

TOP 20 special customer with highest LTV

:

	Customer No.	Make	Invoice No	Total Amt Wtd Tax.	LTV
32604	84810	TOYOTA	18	86761.728511	1.561711e+06
39442	E2003	MAHINDRA & MAHINDRA	49	30316.493458	1.485508e+06
31055	76500	MARUTI SUZUKI	9	138340.256423	1.245062e+06
30859	74406	HYUNDAI	5	188149.861583	9.407493e+05
30426	66248	MAHINDRA & MAHINDRA	5	163988.533882	8.199427e+05
30066	51811	GENERAL MOTORS	5	157266.226222	7.863311e+05
39497	MFCCM03	MAHINDRA & MAHINDRA	80	9397.866395	7.518293e+05
33571	87240	HYUNDAI	3	240759.304667	7.222779e+05
37773	96440	GENERAL MOTORS	5	140654.337400	7.032717e+05
30843	74073	MAHINDRA & MAHINDRA	15	39814.326508	5.972149e+05
20324	133669	MAHINDRA & MAHINDRA	2	294167.801684	5.883356e+05
30594	69319	GENERAL MOTORS	6	96673.902872	5.800434e+05
6176	111596	HONDA	4	144094.200189	5.763768e+05
35206	90834	SKODA	1	565487.200000	5.654872e+05
3536	106817	MAHINDRA & MAHINDRA	3	165008.090227	4.950243e+05
30025	4989	FORD	3	161254.616333	4.837638e+05
34627	89477	MAHINDRA & MAHINDRA	8	60044.711905	4.803577e+05
12372	121624	MAHINDRA & MAHINDRA	1	419840.660000	4.198407e+05
332	100620	NISSAN	3	139650.354848	4.189511e+05
4698	108907	MITSUBISHI MOTORS	5	82295.035484	4.114752e+05

LTV Modeling

```
df.head()
```

	Cust Type	State	KMs Reading	Make	Order Type	Invoice month	Partner Type	Title	ServiceTime_category	Item Category	Order Quantity	LTV
0	Retail	Maharashtra	0	FORD	Paid Service	6	1.0	2.0	super_fast	P001, P002	45.5	2201.78
4	Retail	Maharashtra	1	MAHINDRA & MAHINDRA	Running Repairs	5	1.0	2.0	super_fast	P001, P002	13.0	3827.00
5	Retail	Maharashtra	1	MAHINDRA & MAHINDRA	Paid Service	7	1.0	2.0	super_slow	P001, P002	50.0	3827.00
9	Retail	Maharashtra	0	VOLKSWAGEN	Running Repairs	1	1.0	2.0	mid	P001	85.0	35132.65
10	Retail	Maharashtra	0	VOLKSWAGEN	Running Repairs	1	1.0	2.0	fast	P001	300.6	35132.65

* Used scaling & one hot encoding

MODELS USED

Model	r2_Score	MAE	RMSE
K Neighbors Regressor(KNN)	0.31546	0.011233	0.027477
Decision Tree Regressor	0.49151	0.012431	0.0299624
Linear Regression	-1.3663	2257899	182081000
Ridge	0.27303	0.011618	0.0292
Lasso	-1.55650	0.0147682	0.0432848

Recommendations – Problem3

- LTV high for older cars-> **Old is Gold.** Appreciate customer for good maintenance & encourage for regular checkup

- Retention of high LTV customers

I. Discounts on accessories
II. Loyalty bonus
III. Referral Bonus
IV. Maintenance Packages

Further Recommendations

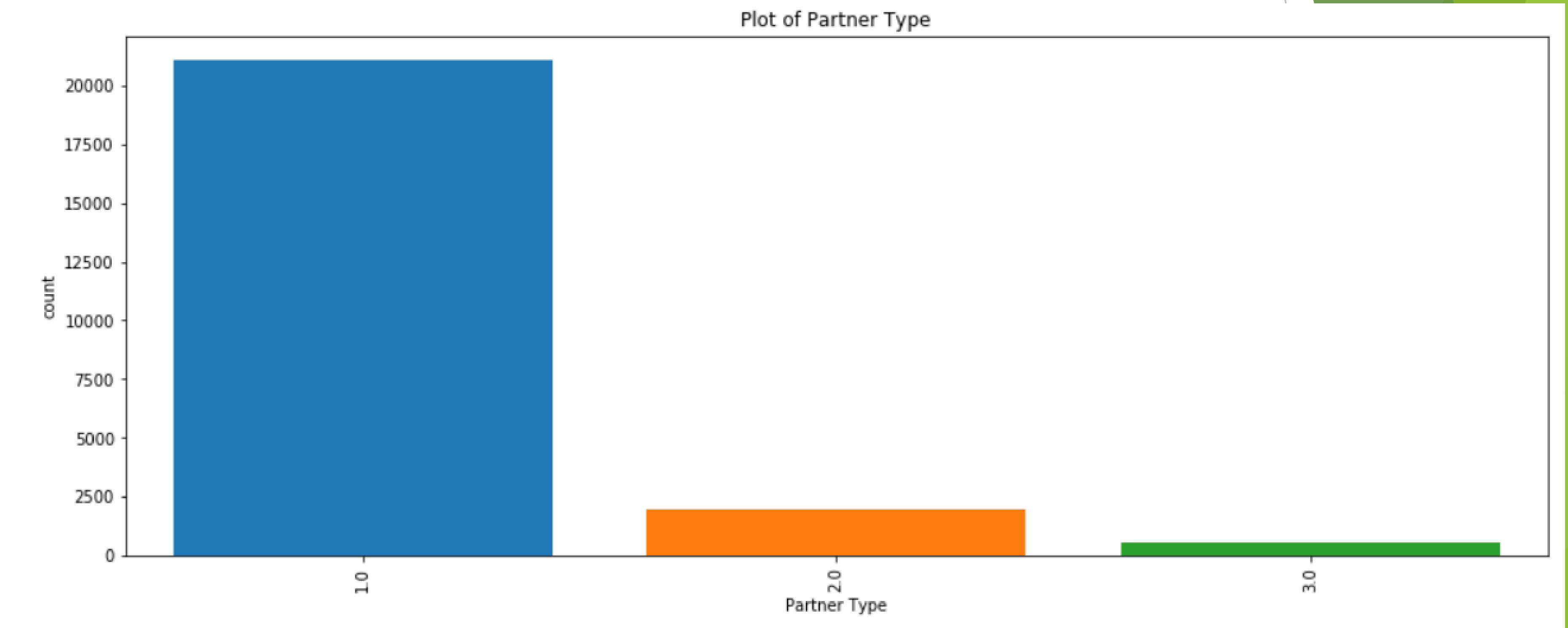
Marketing Recommendation

Customer & Revenue Prediction

Other Actions

Marketing Recommendation – Problem4

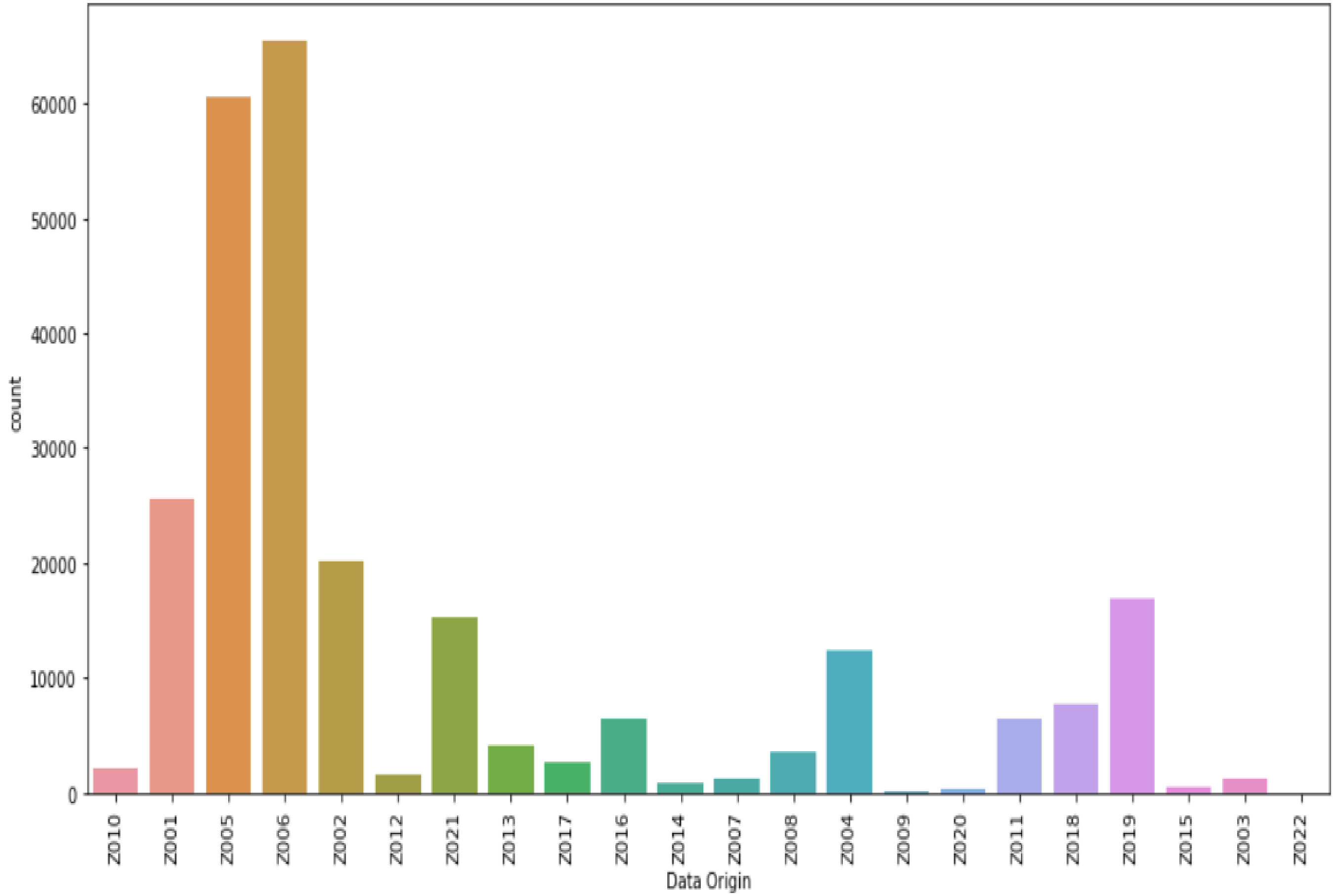
Partner Type Count Plot



* Retail, corporate & fleet

COUNT OF DATA ORIGIN

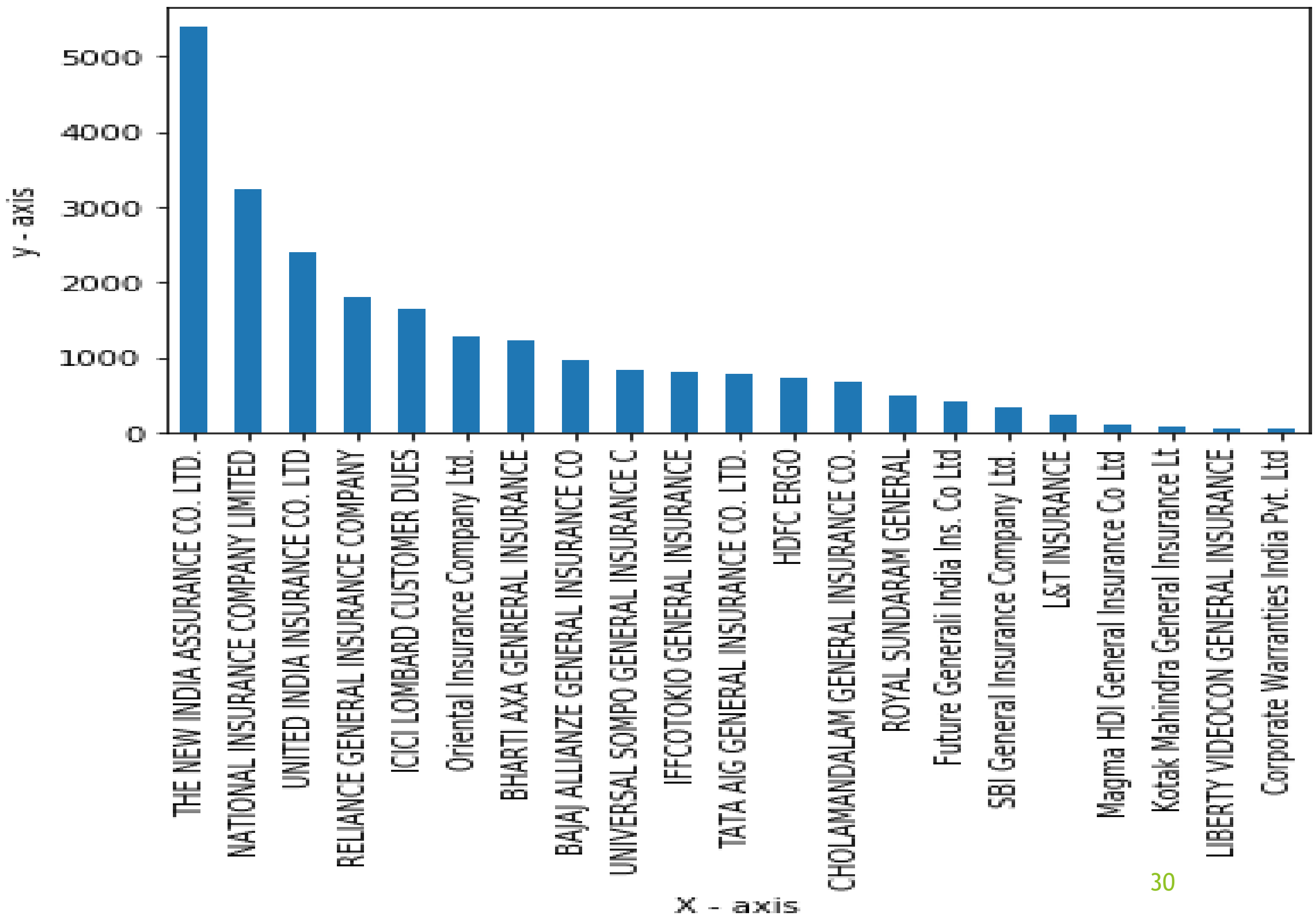
- According to the graph major pull of customers is seen from reference – employees and reference – customer i.e. **word-of-mouth**.



Z001	Camp - Outdoor
Z002	Camp - Workshop
Z003	Emailers
Z004	Fleet
Z005	Reference - Customer
Z006	Reference - Employee
Z007	Reference - Used car dealer
Z008	Just Dial/Other helpline
Z009	Snapdeal/Other websites
Z010	Company website
Z011	Float activity
Z012	Petrol pump activity
Z013	Hoardings/Outdoor Advertisements
Z014	Insurance Company
Z015	Television AD
Z016	Newspaper AD
Z017	Newspaper leaflet
Z018	Outdoor Sales Activity
Z019	Spotted the outlet
Z020	Mahindra Sister concern Employee
Z021	Other outdoor activity
Z022	Radio

INSURANCE TRENDS

- Most used Insurance Company
- ‘The New India Assurance Co. Ltd’



Recommendations – Problem4

- Major customer types are retail -> **Work on tie ups** with corporate & fleet (ola, uber, etc). Use corporate discounts & offers.

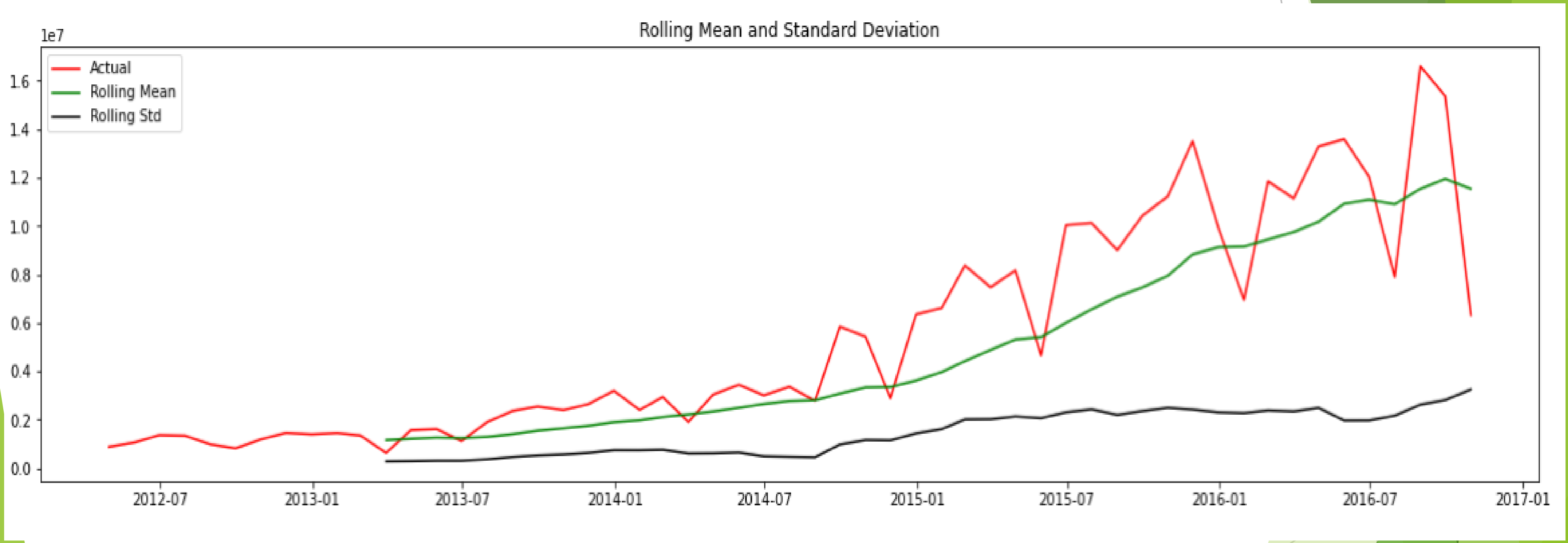
- Most used Insurance are Govt **-> Better tie ups with private vendors**

- Customers coming via word of mouth -> **Need better brand building to pull customers from other sources**

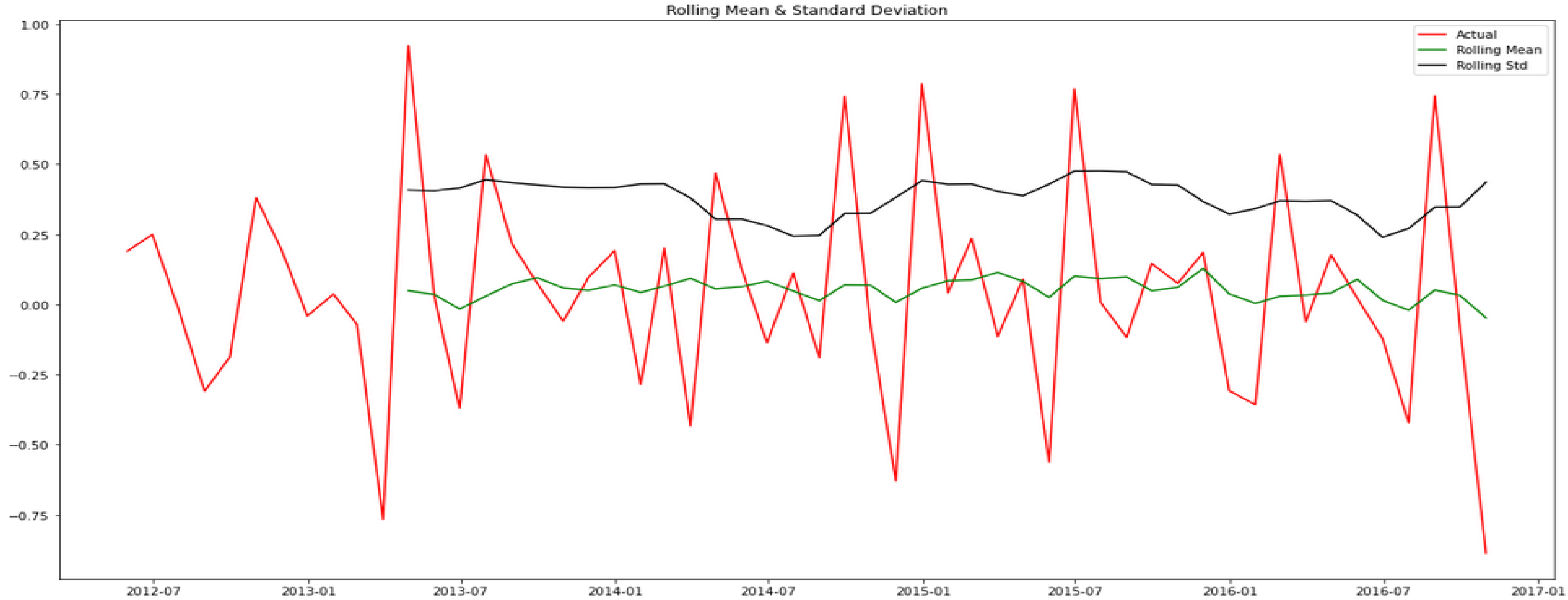
I. TV ads
II. Radio ads
III. Online Presence

Revenue Prediction – Problem5

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



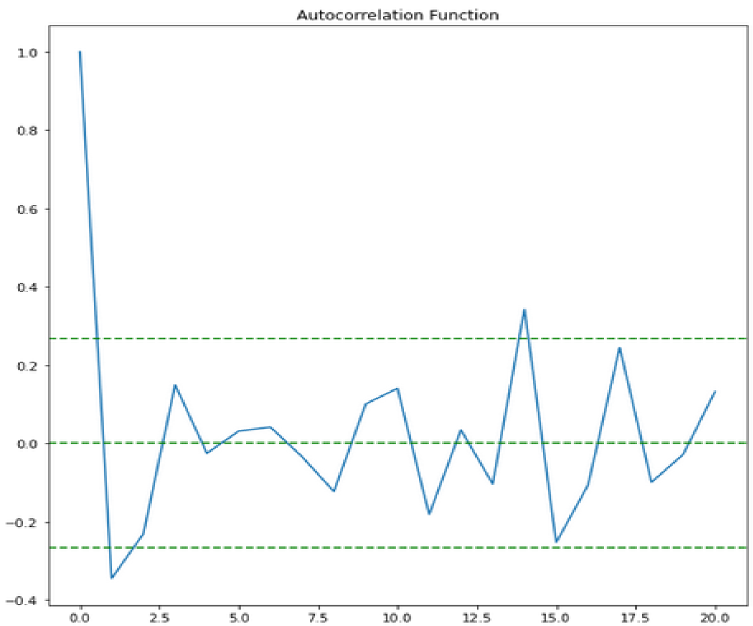
Using differencing to make this time series stationary

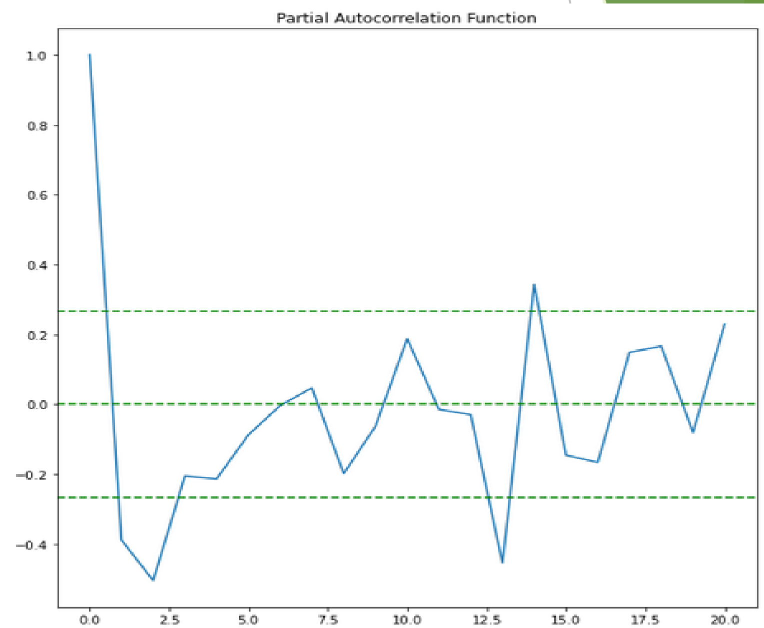



```
Dickey-Fuller Test:
Test Statistic           -9.240293e+00
p-value                   1.589295e-15
Lags Used                 1.000000e+00
No. of Obs                5.200000e+01
Critical Value (1%)      -3.562879e+00
Critical Value (5%)      -2.918973e+00
Critical Value (10%)     -2.597393e+00
dtype: float64
```


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

Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF)





In this plot, the 'p' and 'q' values can be determined as : p=2 , q=2.

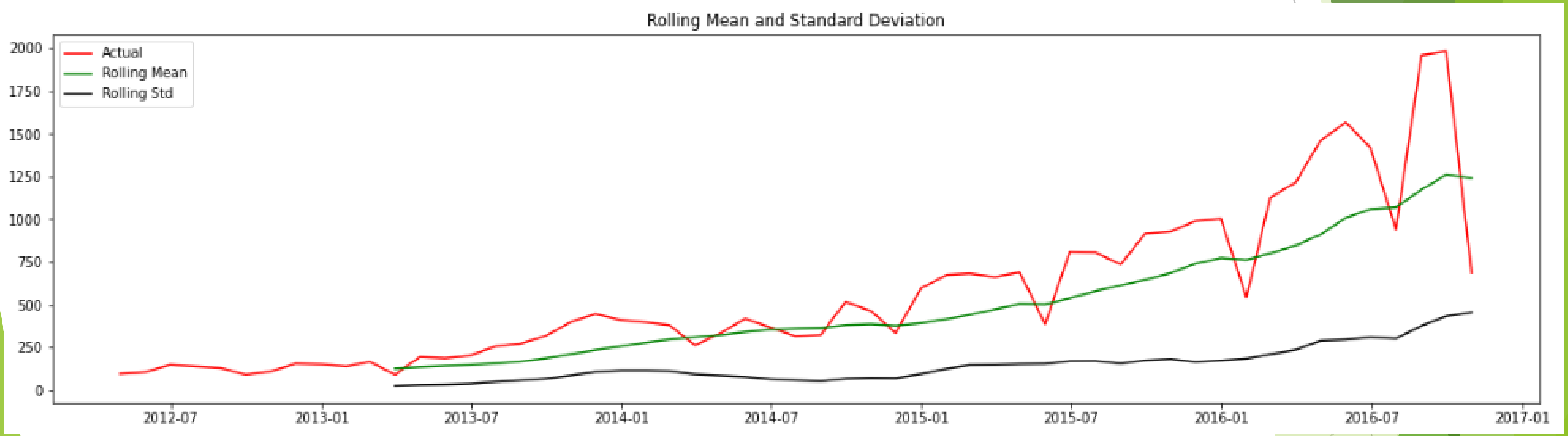
This means that the optimal values for the ARIMA(p,d,q) model are (2,1,2).

Plotting ARIMA model

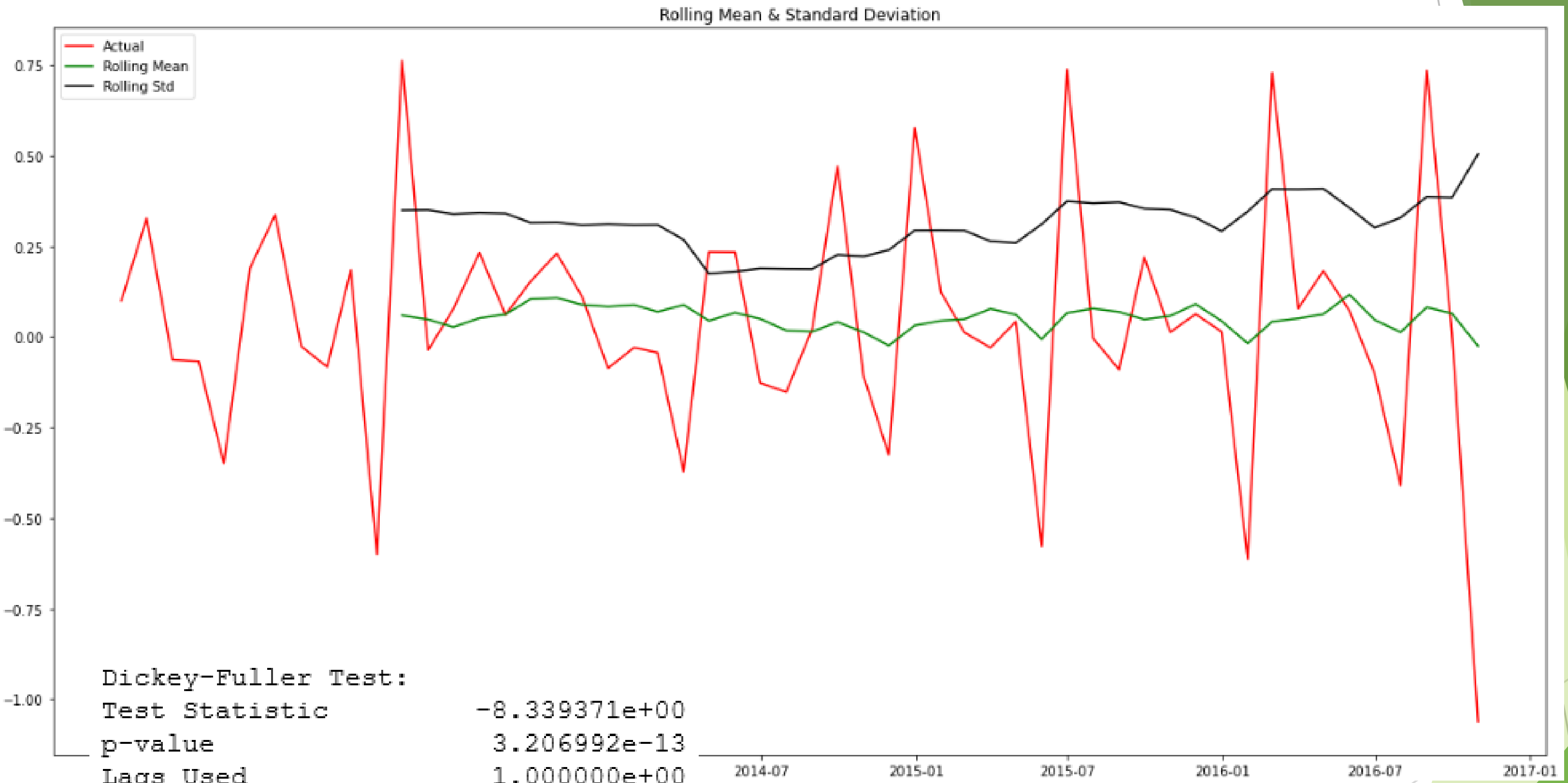
Month wise Revenue Forecasting

Customer count Prediction

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



Using differencing to make this time series stationary

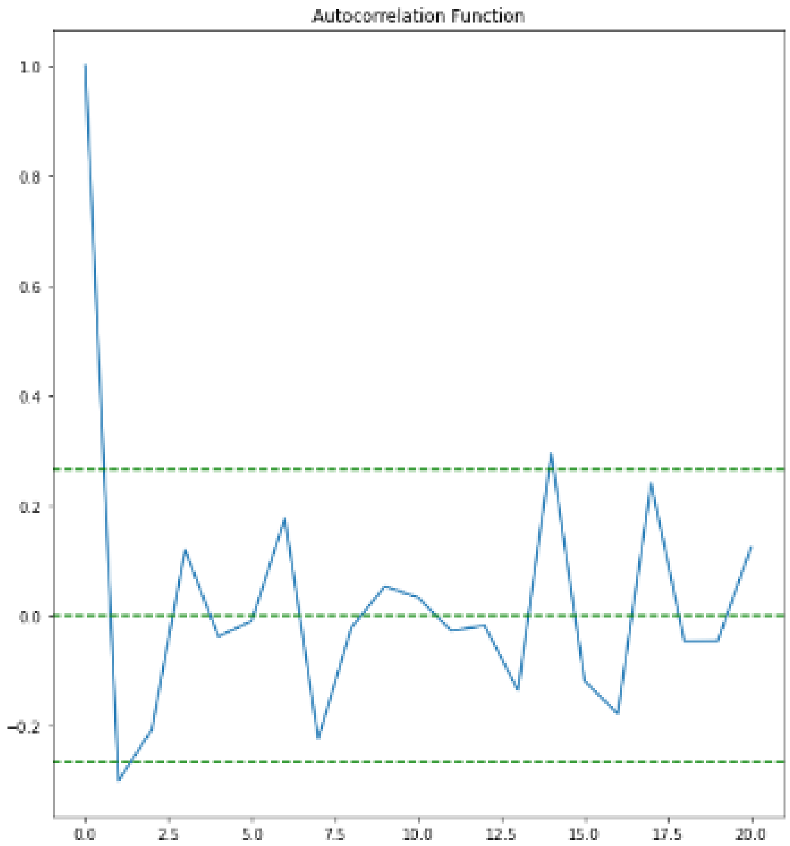


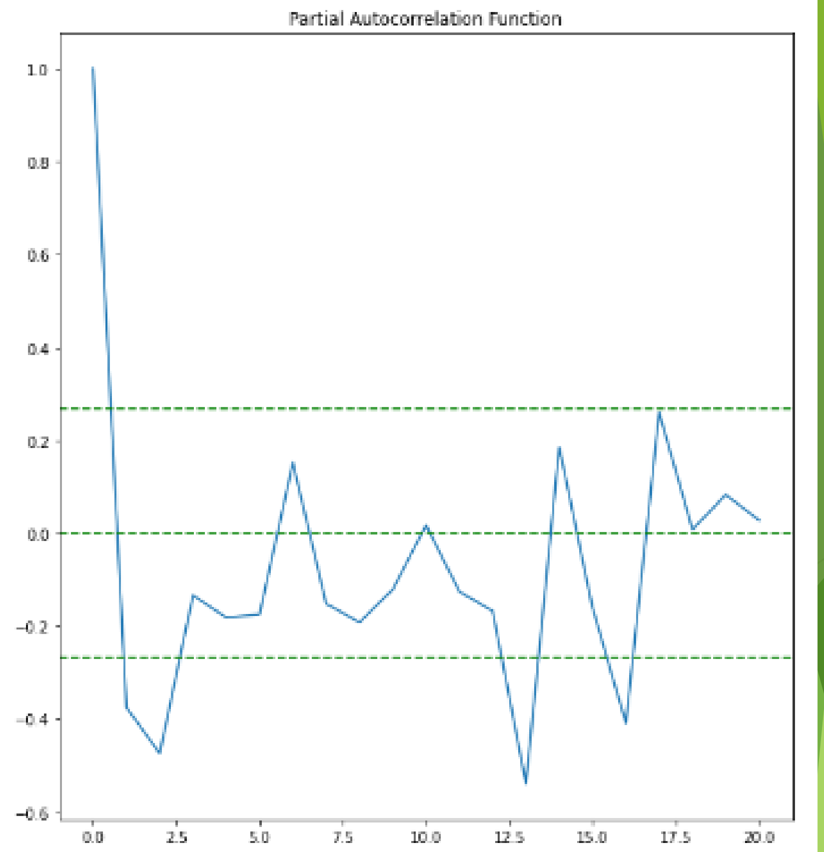

```
Dickey-Fuller Test:
Test Statistic            -8.339371e+00
p-value                    3.206992e-13
Lags Used                  1.000000e+00
No. of Obs                 5.200000e+01
Critical Value (1%)       -3.562879e+00
Critical Value (5%)       -2.918973e+00
Critical Value (10%)      -2.597393e+00
dtype: float64
```


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

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

Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF)

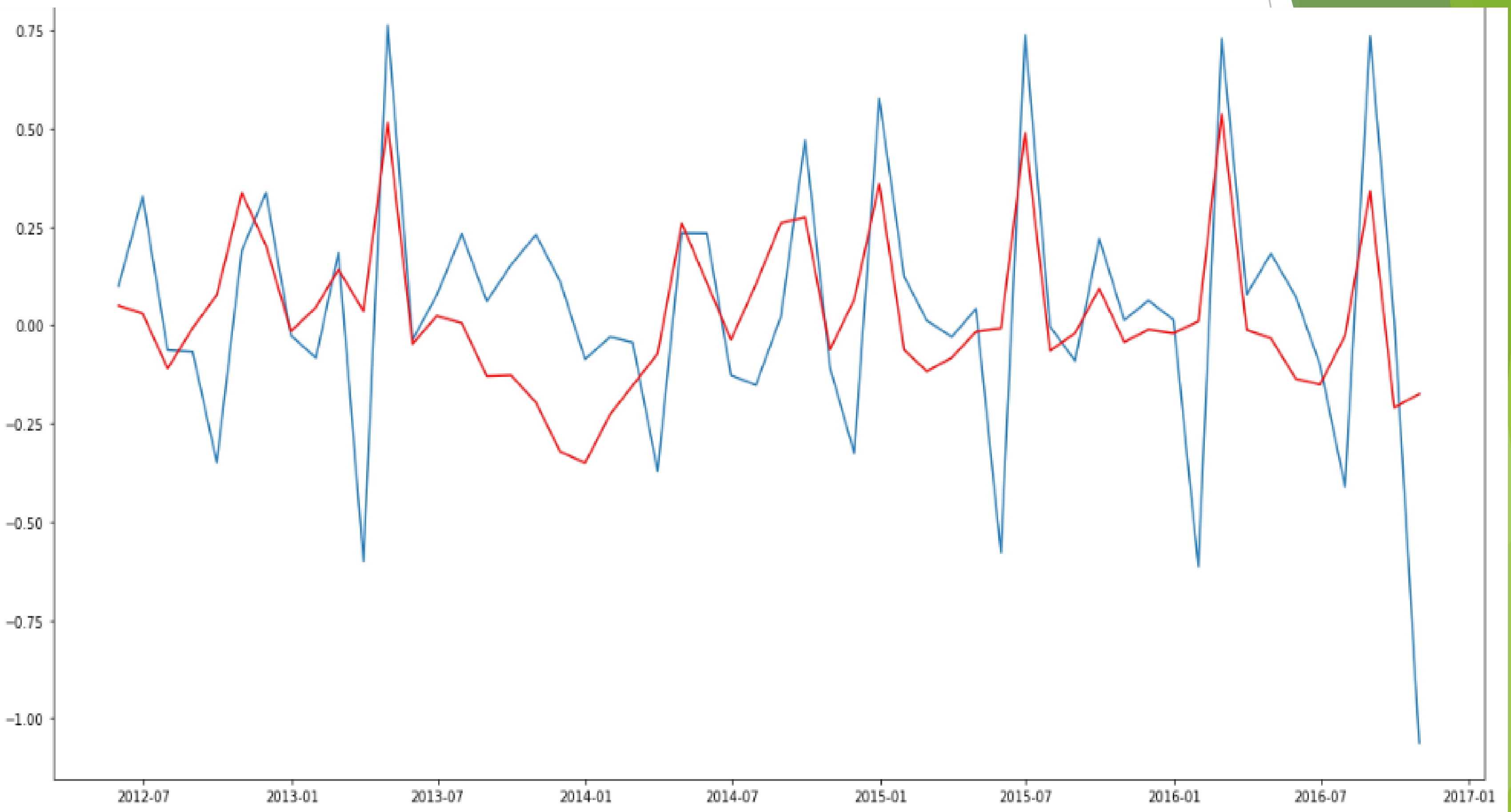




In this plot, the 'p' and 'q' values can be determined as : p=2 , q=2.

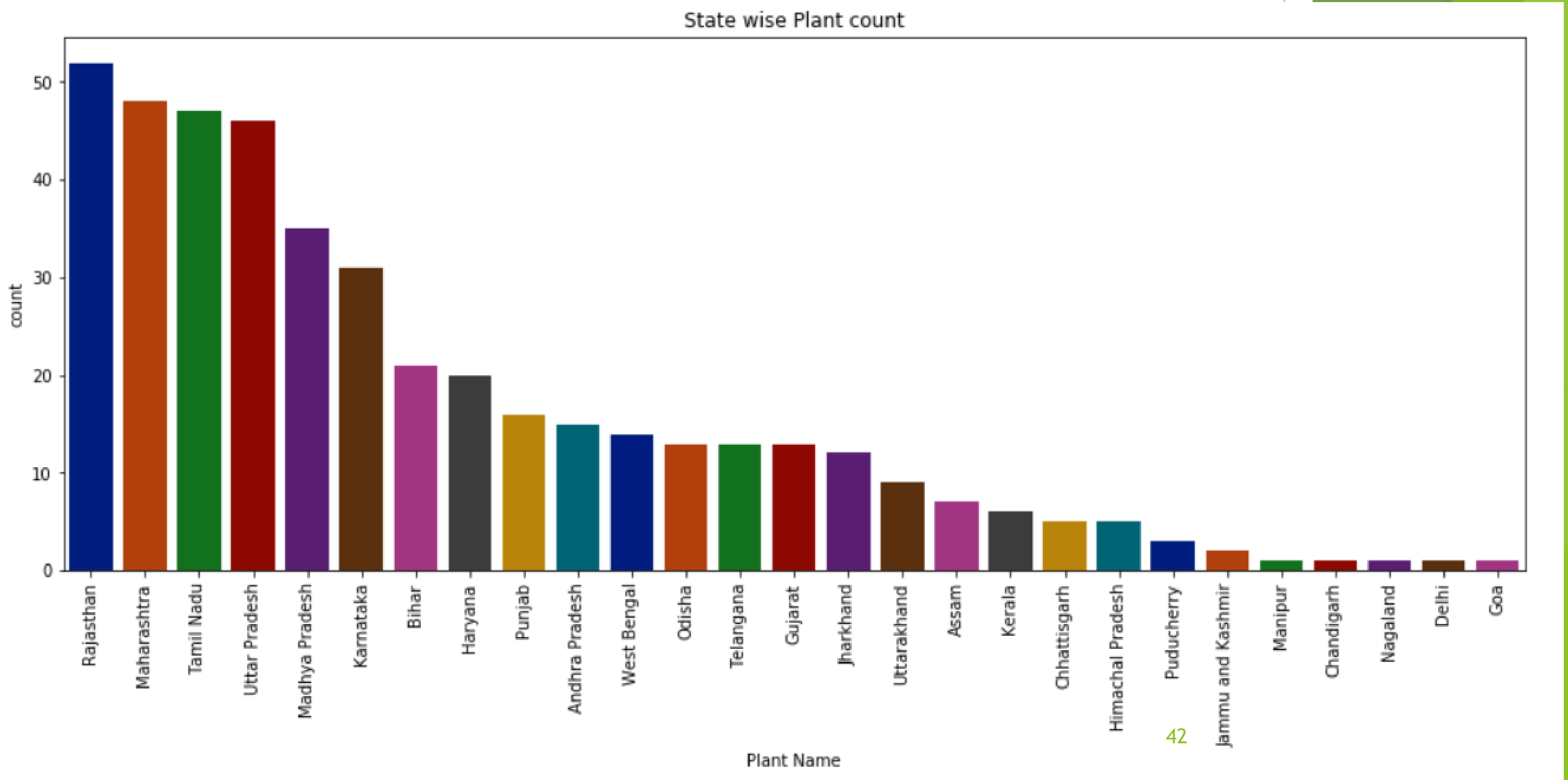
This means that the optimal values for the ARIMA(p,d,q) model are (2,1,2).

Plotting ARIMA model

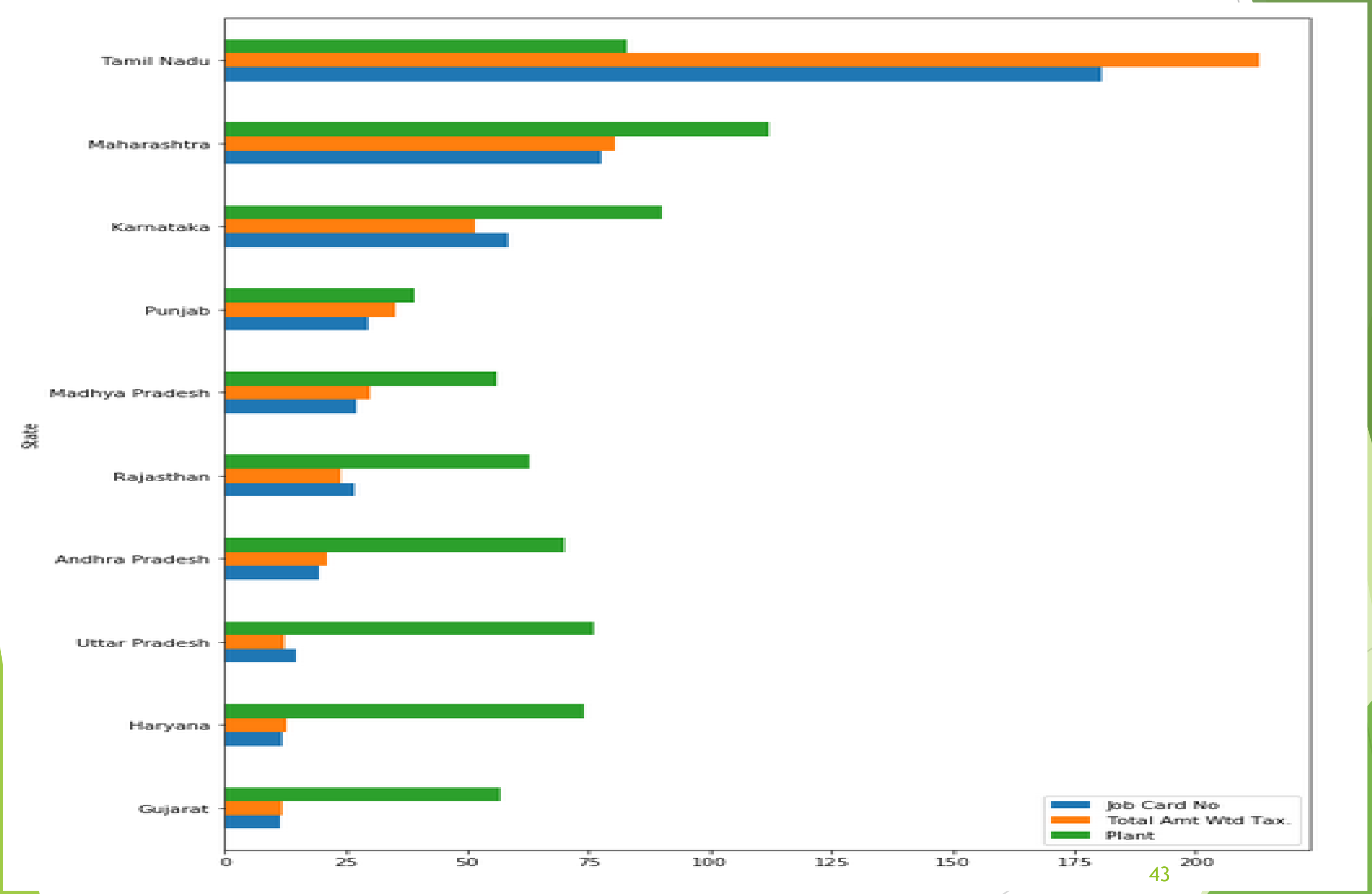


Month wise Customer Forecasting

MORE INSIGHTS - State wise plant Count



Finding New Plant Location

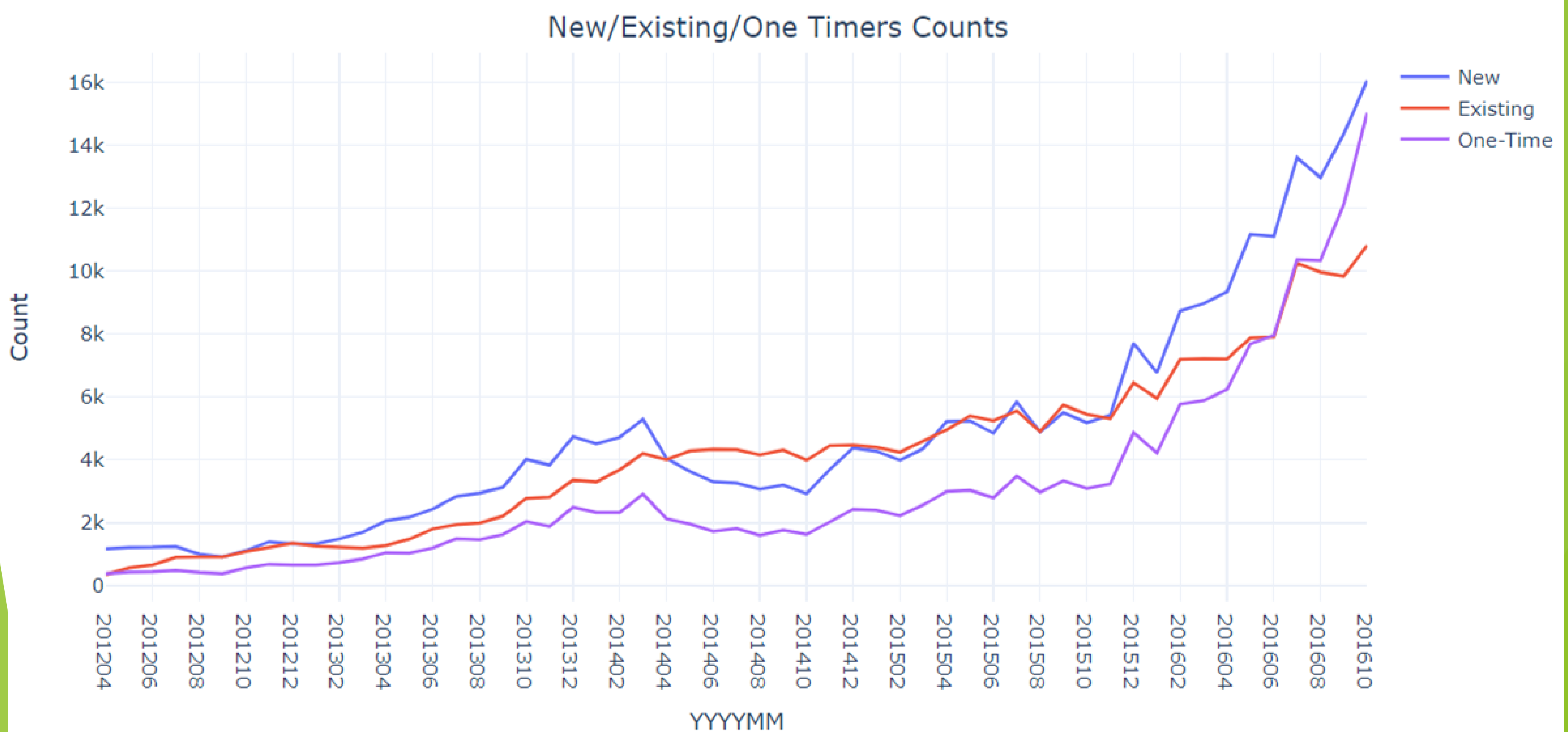


New Plant Location and Count

	State	Job Card No	Total Amt Wtd Tax.	Plant	New Garage Factor	New Garage Count
5	Chandigarh	5.118	7.688359	22	0.081300	0.0
1	Andhra Pradesh	19.602	21.164207	70	0.084665	0.0
25	Odisha	4.606	2.856602	11	0.108740	0.0
28	Rajasthan	26.876	24.129787	63	0.163394	1.0
19	Madhya Pradesh	26.968	29.983691	56	0.257844	1.0
16	Karnataka	58.286	51.638322	90	0.371579	2.0
31	Telangana	10.962	14.099799	20	0.386405	2.0
20	Maharashtra	77.748	80.453058	112	0.498650	2.0
27	Punjab	29.642	35.074158	39	0.683543	3.0
30	Tamil Nadu	180.828	212.942348	83	5.589482	23.0

```
df_grouped['Total Amt Wtd Tax.']=df_grouped['Total Amt Wtd Tax.']/2000000
df_grouped['Job Card No']= df_grouped['Job Card No']/500
df_grouped['New Garage Factor']= (df_grouped['Job Card No']/df_grouped['Plant'])* \
                                 (df_grouped['Total Amt Wtd Tax.']/df_grouped['Plant'])
df_grouped['New Garage Count']= (df_grouped['New Garage Factor']/df_grouped['New Garage Factor'].mean()).round()
```

CUSTOMER TYPE



Recommendations - extra

- New Plant locations -> **23 in Tamil Nadu**
- Invest on referral programs & ads -> **Gujarat, Haryana & Uttar Pradesh**
- Count of one time customers increasing **-> Loyalty programs needed**
- Efficient Garages **- > Efficient Garage Management Platform launched "dearo"**

All Action Points for MFCS

- States which are under performing while having great potential -> **Gujarat & Kerala**
- **Indian "Make"** are the biggest count makers -> **stock parts & hire experts**
- Revenue Outliers are Audi, Mercedes & Skoda -> **Cash Cows**
- LTV high for older cars-> **Old is Gold.** Appreciate customer for good maintenance & encourage for regular checkup
- Retention of high LTV customers

I. Discounts on accessories
II. Loyalty bonus
III. Referral Bonus
IV. Maintenance Packages

- New Plant locations -> **23 in Tamil Nadu**
- Invest on referral programs & ads -> **Gujarat, Haryana & Uttar Pradesh**
- Count of one time customers increasing -> **Loyalty programs needed**
- Efficient Garages - > **Efficient Garage Management Platform launched "dearo"**
- Highest revenue & repairs **during mid year** (around July) -> **Hire contract staff**
- Accidental orders are outsourced -> **Bit relax on TAT**
- Use less load day (i.e. Saturday) -> **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.
- Most used Insurance are Govt -> **Better tie ups with private vendors**
- Customers coming via word of mouth -> **Need better brand building to pull customers from other sources**

I. TV ads
II. Radio ads
III. Online Presence

Thank You