**Data Description**

* DataFrame **df1**: It contains “sales data” filtered based on certain criteria such as billing date, brand, sold-to party code, **bill quantity**, and sales type. It includes columns such as sales region, sales zone, sold-to party district name, billing date, sold-to party code, sales type, and bill quantity.

billing\_date > '2022-01-01', brand = 'DALMIA', sold\_to\_party\_code > 7\*10^6, sold\_to\_party\_code < 71\*10^5, bill\_quantity >= 0, sales\_type = 'Trade'

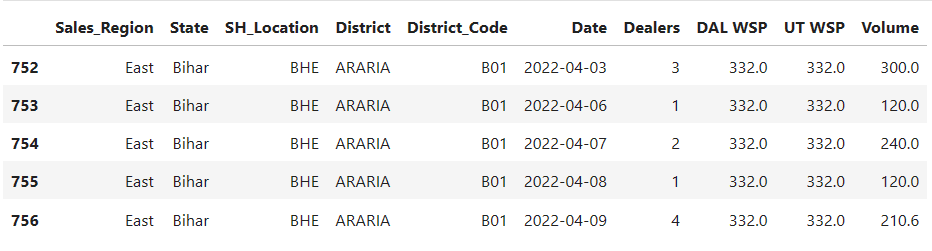
* DataFrame **df2**: It includes all columns from the "sales\_pd\_wsp" table in the "Dalmia sales" database. It contains dal\_wsp price and ut\_wsp price columns.

**Data Handling and Cleaning**

Convert both the data frame into day-level and done forward and backword fill for price columns for each district. Daily price is average value and volume is sum. Merged both the data frame, considering primary columns 'District', 'Date'.

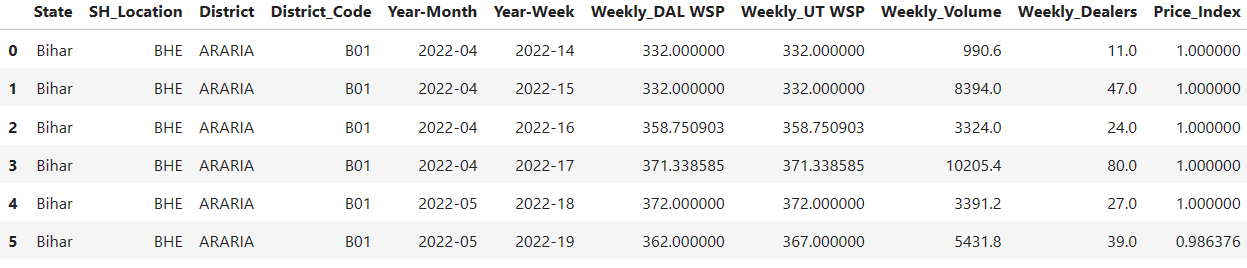
**Observation**: Some of districts are coming under two different sales\_region. Considered based on maximum number of occurrence and recent occurrence.

Considered east\_region sales data, because sales\_pd\_wsp were also of east region.



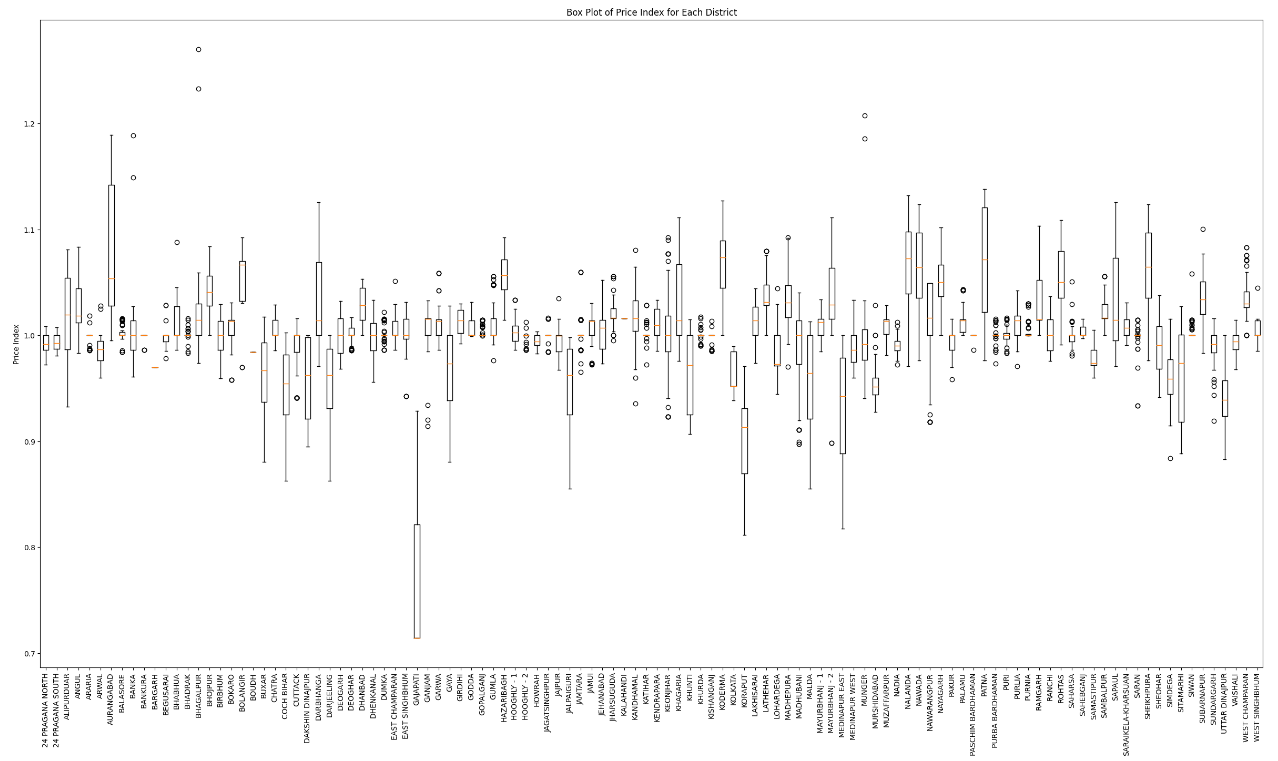
Converted into week level. Taken Weighted average of DAL WSP and UT WSP with Volume. Weekly Volume and Weekly Dealers are the sum of Dealers and Volume for each group of 'State', 'SH\_Location', 'District', 'District\_Code', 'Year-Month', 'Year-Week'.

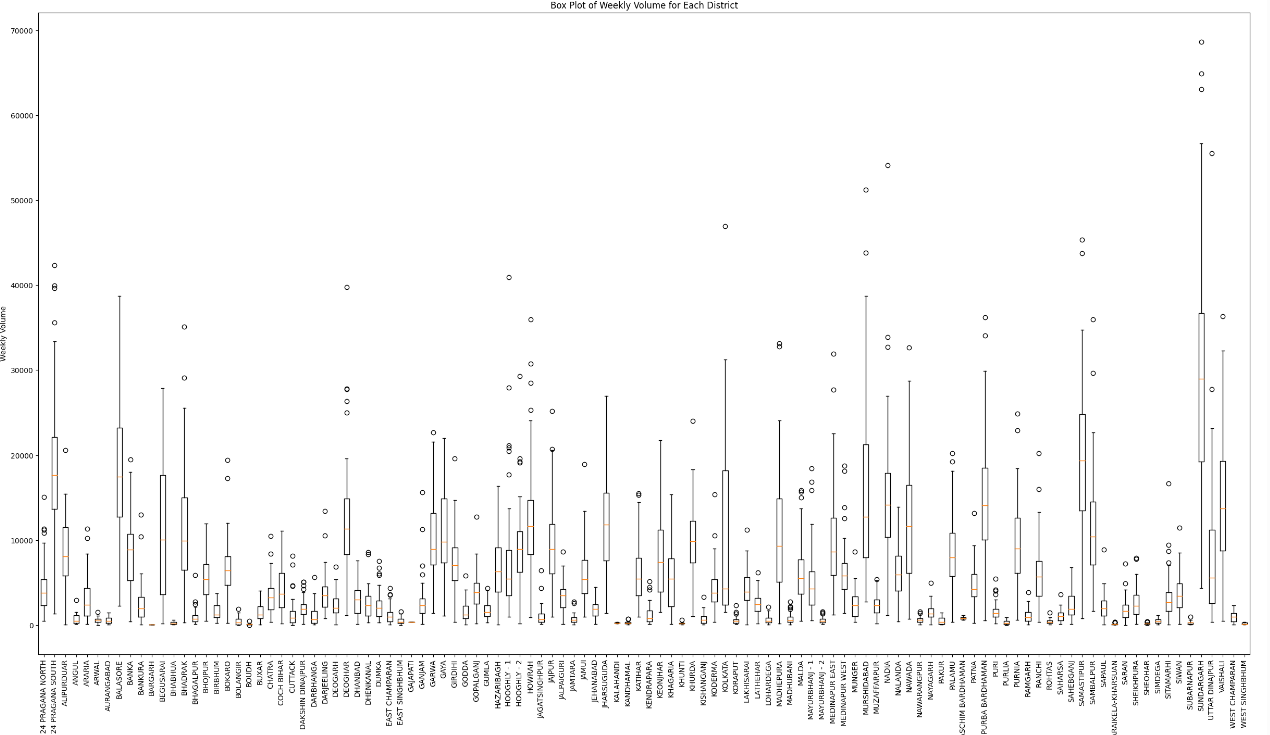
This is the final data after cleaning, which is used for volume optimization.



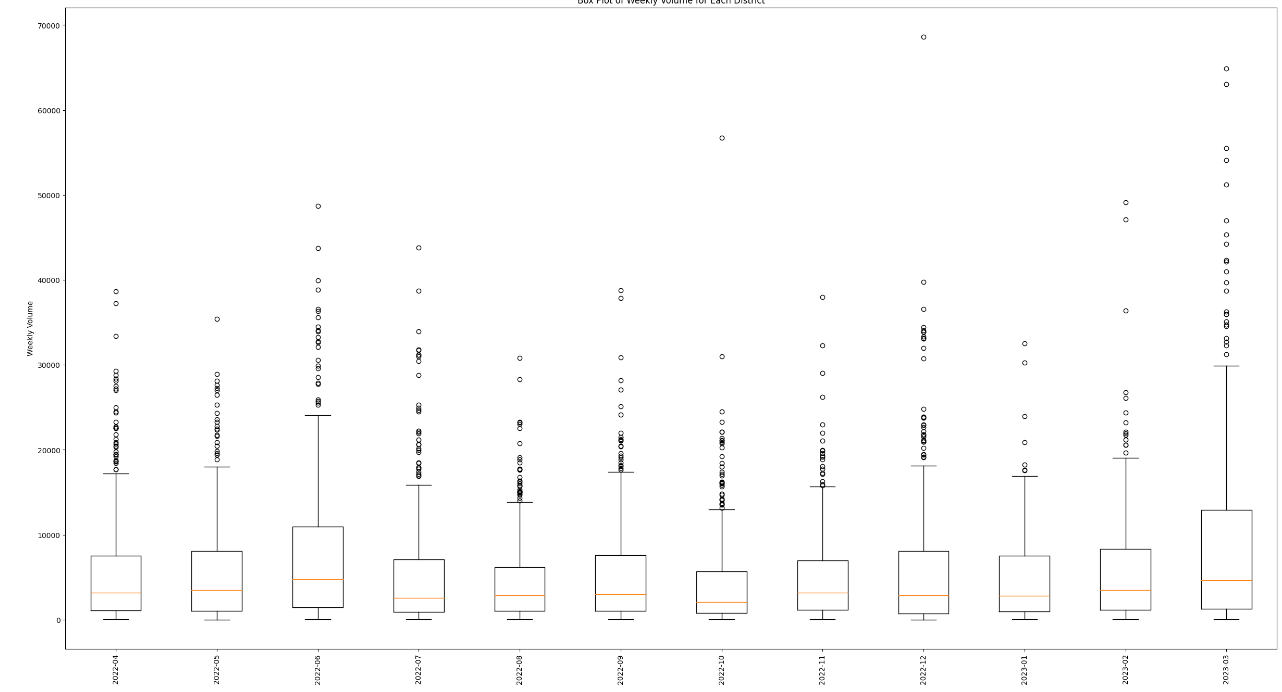
**Initial Analysis**

Checked the outliers of volume and PI for each district.

****

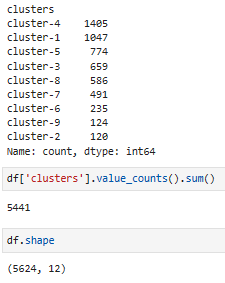
****

Checked the outliers of district for each month

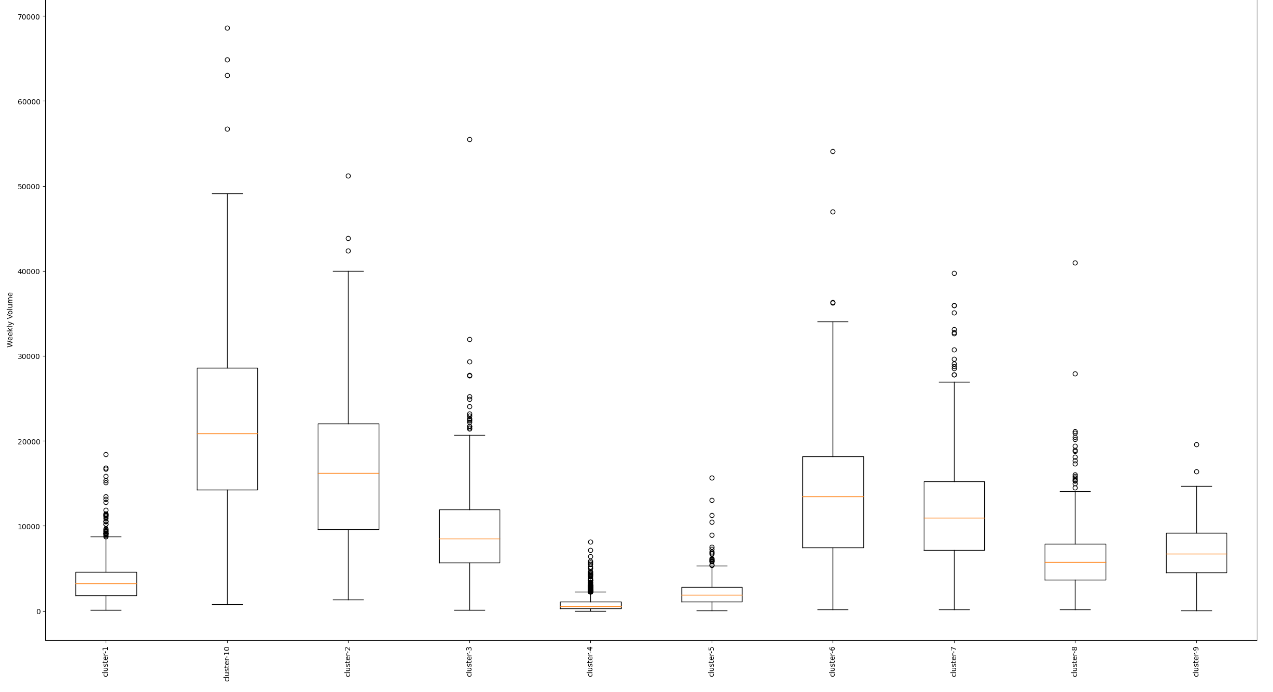
****

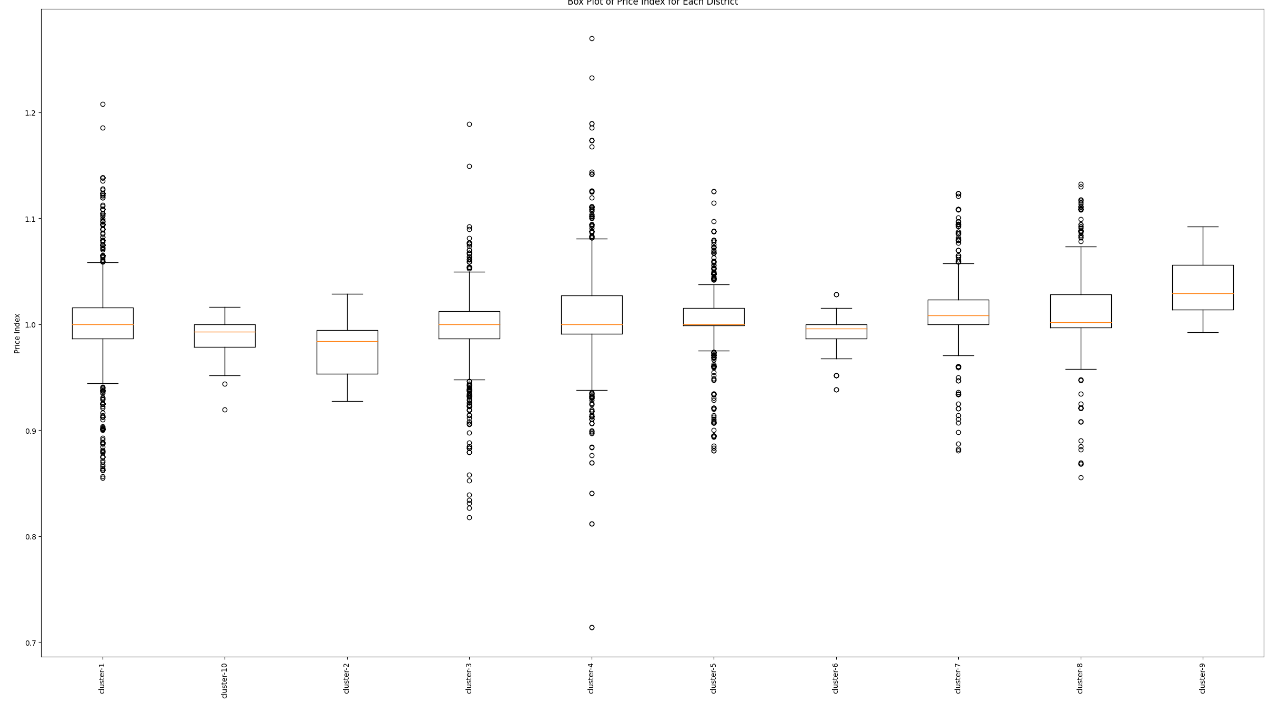
There are many outliers, if we remove them data points will get reduced. So, I didn’t remove them.

Tried to create clusters of districts to check the threshold of volume. Cluster should be like that if the difference between 75 percentile Weekly volume of any two district is less than 1000 bags then these two districts will consider in a same cluster. From this, out of 5624 datapoints, created cluster remaining datapoints has assigned with new cluster.



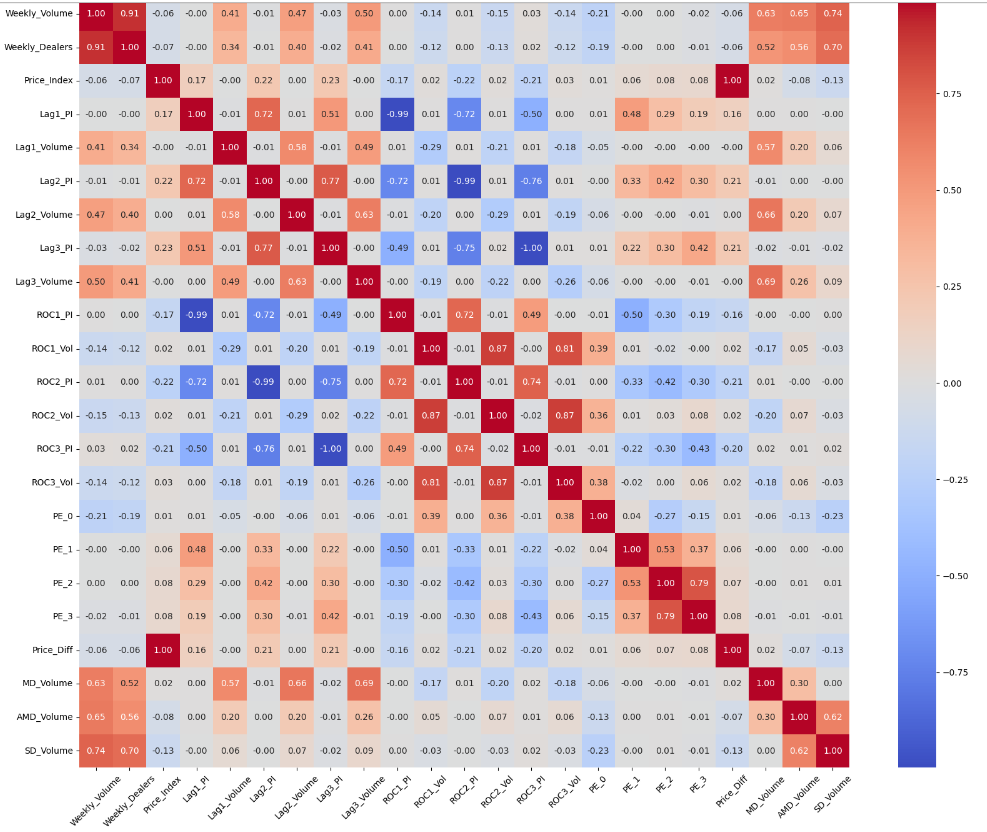
Based on the cluster checked the outliers for price index and weekly volume.

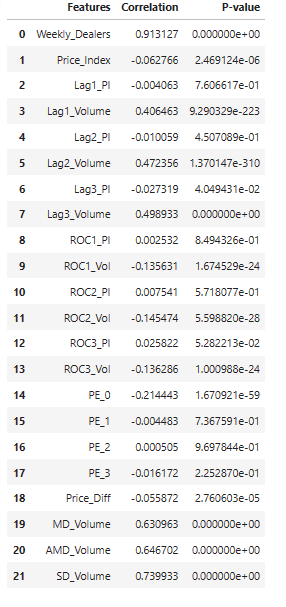




**Further Analysis**

To analysis further, created many lag variables and new variables. Analyse correlation matric heatmap, p-values and correlations with Weekly Volume (Output Features) for each feature.





Select features based on correlation analysis and p-values. And checked the accuracy of line of fit with linear regression (OLS) and Quantile regression in different group like (District-Level, SH-Location-Level, Cluster-Level, State-Level). Accuracy was coming well for some of the districts/ SH-Location…etc, around 0.93, 0.87 but in same cases it was not good.

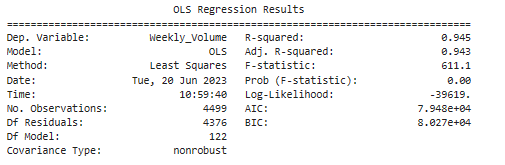
To improve the accuracy and also the objective function should be one at national level, I have selected features based on Recursive Feature Elimination and also created dummies of State, SH\_Location, District. Run the OLS model for whole dataset.

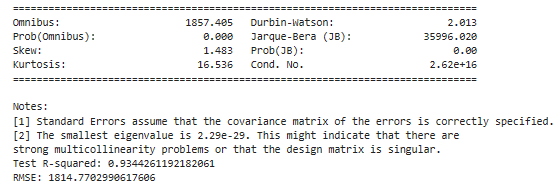
Selected Features:

Index(['Weekly\_Dealers', 'Price\_Index', 'Lag1\_PI', 'Lag2\_PI', 'Lag3\_PI',

'ROC3\_PI', 'PE\_0', 'PE\_1', 'PE\_2', 'PE\_3'],

dtype='object')

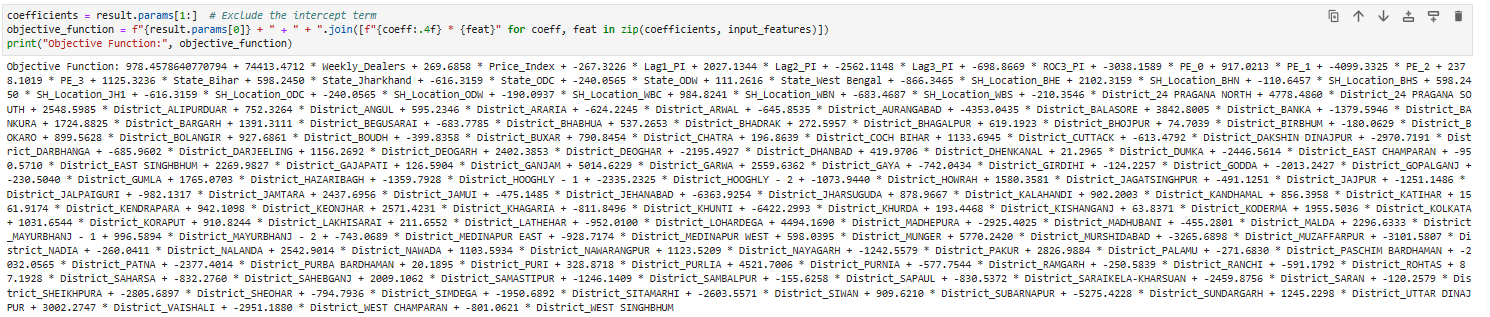




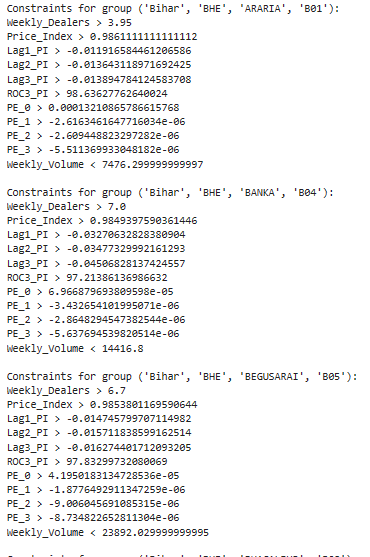
**Objective Function and Constraints**

Based on that formed equation has considered as objective function.

**Objective Function:** 978.4578640770794 + 74413.4712 \* Weekly\_Dealers + 269.6858 \* Price\_Index + -267.3226 \* Lag1\_PI + 2027.1344 \* Lag2\_PI + -2562.1148 \* Lag3\_PI + -698.8669 \* ROC3\_PI + -3038.1589 \* PE\_0 + 917.0213 \* PE\_1 + -4099.3325 \* PE\_2 + 2378.1019 \* PE\_3 + 1125.3236 \* State\_Bihar + 598.2450 \* State\_Jharkhand + -616.3159 \* State\_ODC + -240.0565 \* State\_ODW + 111.2616 \* State\_West Bengal + -866.3465 \* SH\_Location\_BHE + 2102.3159 \* SH\_Location\_BHN + -110.6457 \* SH\_Location\_BHS + 598.2450 \* SH\_Location\_JH1 + -616.3159 \* SH\_Location\_ODC + -240.0565 \* SH\_Location\_ODW + -190.0937 \* SH\_Location\_WBC + 984.8241 \* SH\_Location\_WBN + -683.4687 \* SH\_Location\_WBS + -210.3546 \* District\_24 PRAGANA NORTH + 4778.4860 \* District\_24 PRAGANA SOUTH + 2548.5985 \* District\_ALIPURDUAR + 752.3264 \* District\_ANGUL + 595.2346 \* District\_ARARIA + -624.2245 \* District\_ARWAL + -645.8535 \* District\_AURANGABAD + -4353.0435 \* District\_BALASORE + 3842.8005 \* District\_BANKA + -1379.5946 \* District\_BANKURA + 1724.8825 \* District\_BARGARH + 1391.3111 \* District\_BEGUSARAI + -683.7785 \* District\_BHABHUA + 537.2653 \* District\_BHADRAK + 272.5957 \* District\_BHAGALPUR + 619.1923 \* District\_BHOJPUR + 74.7039 \* District\_BIRBHUM + -180.0629 \* District\_BOKARO + 899.5628 \* District\_BOLANGIR + 927.6861 \* District\_BOUDH + -399.8358 \* District\_BUXAR + 790.8454 \* District\_CHATRA + 196.8639 \* District\_COCH BIHAR + 1133.6945 \* District\_CUTTACK + -613.4792 \* District\_DAKSHIN DINAJPUR + -2970.7191 \* District\_DARBHANGA + -685.9602 \* District\_DARJEELING + 1156.2692 \* District\_DEOGARH + 2402.3853 \* District\_DEOGHAR + -2195.4927 \* District\_DHANBAD + 419.9706 \* District\_DHENKANAL + 21.2965 \* District\_DUMKA + -2446.5614 \* District\_EAST CHAMPARAN + -950.5710 \* District\_EAST SINGHBHUM + 2269.9827 \* District\_GAJAPATI + 126.5904 \* District\_GANJAM + 5014.6229 \* District\_GARWA + 2559.6362 \* District\_GAYA + -742.0434 \* District\_GIRDIHI + -124.2257 \* District\_GODDA + -2013.2427 \* District\_GOPALGANJ + -230.5040 \* District\_GUMLA + 1765.0703 \* District\_HAZARIBAGH + -1359.7928 \* District\_HOOGHLY - 1 + -2335.2325 \* District\_HOOGHLY - 2 + -1073.9440 \* District\_HOWRAH + 1580.3581 \* District\_JAGATSINGHPUR + -491.1251 \* District\_JAJPUR + -1251.1486 \* District\_JALPAIGURI + -982.1317 \* District\_JAMTARA + 2437.6956 \* District\_JAMUI + -475.1485 \* District\_JEHANABAD + -6363.9254 \* District\_JHARSUGUDA + 878.9667 \* District\_KALAHANDI + 902.2003 \* District\_KANDHAMAL + 856.3958 \* District\_KATIHAR + 1561.9174 \* District\_KENDRAPARA + 942.1098 \* District\_KEONJHAR + 2571.4231 \* District\_KHAGARIA + -811.8496 \* District\_KHUNTI + -6422.2993 \* District\_KHURDA + 193.4468 \* District\_KISHANGANJ + 63.8371 \* District\_KODERMA + 1955.5036 \* District\_KOLKATA + 1031.6544 \* District\_KORAPUT + 910.8244 \* District\_LAKHISARAI + 211.6552 \* District\_LATHEHAR + -952.0100 \* District\_LOHARDEGA + 4494.1690 \* District\_MADHEPURA + -2925.4025 \* District\_MADHUBANI + -455.2801 \* District\_MALDA + 2296.6333 \* District\_MAYURBHANJ - 1 + 996.5894 \* District\_MAYURBHANJ - 2 + -743.0689 \* District\_MEDINAPUR EAST + -928.7174 \* District\_MEDINAPUR WEST + 598.0395 \* District\_MUNGER + 5770.2420 \* District\_MURSHIDABAD + -3265.6898 \* District\_MUZAFFARPUR + -3101.5807 \* District\_NADIA + -260.0411 \* District\_NALANDA + 2542.9014 \* District\_NAWADA + 1103.5934 \* District\_NAWARANGPUR + 1123.5209 \* District\_NAYAGARH + -1242.5579 \* District\_PAKUR + 2826.9884 \* District\_PALAMU + -271.6830 \* District\_PASCHIM BARDHAMAN + -2032.0565 \* District\_PATNA + -2377.4014 \* District\_PURBA BARDHAMAN + 20.1895 \* District\_PURI + 328.8718 \* District\_PURLIA + 4521.7006 \* District\_PURNIA + -577.7544 \* District\_RAMGARH + -250.5839 \* District\_RANCHI + -591.1792 \* District\_ROHTAS + 87.1928 \* District\_SAHARSA + -832.2760 \* District\_SAHEBGANJ + 2009.1062 \* District\_SAMASTIPUR + -1246.1409 \* District\_SAMBALPUR + -155.6258 \* District\_SAPAUL + -830.5372 \* District\_SARAIKELA-KHARSUAN + -2459.8756 \* District\_SARAN + -120.2579 \* District\_SHEIKHPURA + -2805.6897 \* District\_SHEOHAR + -794.7936 \* District\_SIMDEGA + -1950.6892 \* District\_SITAMARHI + -2603.5571 \* District\_SIWAN + 909.6210 \* District\_SUBARNAPUR + -5275.4228 \* District\_SUNDARGARH + 1245.2298 \* District\_UTTAR DINAJPUR + 3002.2747 \* District\_VAISHALI + -2951.1880 \* District\_WEST CHAMPARAN + -801.0621 \* District\_WEST SINGHBHUM



Constraints is defined for each district. All selected input feature should be greater than 5 percentile value and Weekly Volume should be less than 95 percentiles.



….