[CAPSTONE Project Notes -1]

[Supply Chain Management

(2022-2023)

TEXAS McCombs
The University of Texas at Austin
McCombs School of Business
]

Course Name

Post Graduate Program in
Data Science and Business Analytics

Batch Id

(PGP-DSBA-June22C)

Submitted by

Jayant Singh

Email Id: jayant101169@gmail.com

Table of Contents

S.No	Review Parameters	Page No.
1	Introduction of the business problem	2
	Defining problem statement	
	Need of the study/project	
2	Data Report	3 to 9
3	Exploratory Data Analysis	10 to 15
4	Business Insights	16

6

Review Parameters

1) Introduction of the business problem

A FMCG company has entered into the instant noodles business two years back. Their higher management has notices that there is a miss match in the demand and supply. Where the demand is high, supply is pretty low and where the demand is low, supply is pretty high. In both the ways it is an inventory cost loss to the company; hence, the higher management wants to optimize the supply quantity in each and every warehouse in entire country.

A. Defining problem statement

Due to the current supply management's inadequate practises, the organisation is experiencing an inventory cost loss.

Noodles. The goal of the management is to maximise supply in each and every warehouse. across the entire nation. This project's aim is to create a model utilising historical data that will Identify the ideal product weight that should be sent each time to the warehouse. PORDUCT_WG_TON is the target variable in this issue. With a variety of possibilities for analysis Considering that the performance of the model depends on the parameters included in the data, choosing the method and machine learning model to utilise can be highly challenging.

This research compares various well-known machine learning classifiers and evaluates their effectiveness to determine

B. Need of the study/project

- **Objective** - To predict the weight of products of a FMCG company for various warehouses with different conditions, size & locality. To determine the Ideal Quantity of Product Weight Shipped to the various Ware Houses of FMCG Instant Noodles Company in order to reduce wastage of the Product, Bridge the Demand – Supply Gap and avoid over-stacking of Products in the Ware Houses.
- **Scope** – To build various linear, non-linear & ensembled models to predict the weight of products of a FMCG company for various warehouses with different conditions, size & locality.
- **Significance of the project** – Demand forecasting also becomes very key as this is the driving force behind the entire process. Effective FSCM aims to create a value chain between the demand and supply, with optimum utilization of all resources.
- **Constraints (Out of Scope)** – No clear information about the distance between the production center & warehouses & sales in retail stores.

2. Data Report

8]:

	Location_type	WH_capacity_size	zone	WH_regional_zone	num_refill_req_l3m	transport_issue_l1y	Competitor_in_mkt	retail_shop_num	wh_owner
0	Urban	Small	West	Zone 6	3	1	2	4651	R
1	Rural	Large	North	Zone 5	0	0	4	6217	Cor C
2	Rural	Mid	South	Zone 2	1	0	4	4306	Cor C
3	Rural	Mid	North	Zone 3	7	4	2	6000	R
4	Rural	Large	North	Zone 5	3	1	2	4740	Cor C

5 rows × 22 columns

9]:
```
data.shape
```
9]:
```
(25000, 22)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25000 entries, 0 to 24999
Data columns (total 22 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Location_type                 25000 non-null  object 
 1   WH_capacity_size              25000 non-null  object 
 2   zone                          25000 non-null  object 
 3   WH_regional_zone              25000 non-null  object 
 4   num_refill_req_l3m            25000 non-null  int64  
 5   transport_issue_l1y           25000 non-null  int64  
 6   Competitor_in_mkt             25000 non-null  int64  
 7   retail_shop_num               25000 non-null  int64  
 8   wh_owner_type                 25000 non-null  object 
 9   distributor_num               25000 non-null  int64  
 10  flood_impacted                25000 non-null  int64  
 11  flood_proof                   25000 non-null  int64  
 12  electric_supply               25000 non-null  int64  
 13  dist_from_hub                 25000 non-null  int64  
 14  workers_num                   24010 non-null  float64
 15  wh_est_year                   13119 non-null  float64
 16  storage_issue_reported_l3m    25000 non-null  int64  
 17  temp_reg_mach                 25000 non-null  int64  
 18  approved_wh_govt_certificate  24092 non-null  object 
 19  wh_breakdown_l3m              25000 non-null  int64  
 20  govt_check_l3m                25000 non-null  int64  
 21  product_wg_ton                25000 non-null  int64  
dtypes: float64(2), int64(14), object(6)
memory usage: 4.2+ MB
```

Out[61]:

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Location_type	25000	2	Rural	22957	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WH_capacity_size	25000	3	Large	10169	NaN	NaN	NaN	NaN	NaN	NaN	NaN
zone	25000	4	North	10278	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WH_regional_zone	25000	6	Zone 6	8339	NaN	NaN	NaN	NaN	NaN	NaN	NaN
num_refill_req_l3m	25000.0	NaN	NaN	NaN	4.08904	2.606612	0.0	2.0	4.0	6.0	8.0
transport_issue_l1y	25000.0	NaN	NaN	NaN	0.77368	1.199449	0.0	0.0	0.0	1.0	5.0
Competitor_in_mkt	25000.0	NaN	NaN	NaN	3.1042	1.141663	0.0	2.0	3.0	4.0	12.0
retail_shop_num	25000.0	NaN	NaN	NaN	4985.71156	1052.825252	1821.0	4313.0	4859.0	5500.0	11008.0
wh_owner_type	25000	2	Company Owned	13578	NaN	NaN	NaN	NaN	NaN	NaN	NaN
distributor_num	25000.0	NaN	NaN	NaN	42.41812	16.064329	15.0	29.0	42.0	56.0	70.0
flood_impacted	25000.0	NaN	NaN	NaN	0.09816	0.297537	0.0	0.0	0.0	0.0	1.0
flood_proof	25000.0	NaN	NaN	NaN	0.05464	0.227281	0.0	0.0	0.0	0.0	1.0
electric_supply	25000.0	NaN	NaN	NaN	0.65688	0.474761	0.0	0.0	1.0	1.0	1.0
dist_from_hub	25000.0	NaN	NaN	NaN	163.53732	62.718609	55.0	109.0	164.0	218.0	271.0
workers_num	24010.0	NaN	NaN	NaN	28.944398	7.872534	10.0	24.0	28.0	33.0	98.0
wh_est_year	13119.0	NaN	NaN	NaN	2009.383185	7.52823	1996.0	2003.0	2009.0	2016.0	2023.0
storage_issue_reported_l3m	25000.0	NaN	NaN	NaN	17.13044	9.161108	0.0	10.0	18.0	24.0	39.0
temp_reg_mach	25000.0	NaN	NaN	NaN	0.30328	0.459684	0.0	0.0	0.0	1.0	1.0
approved_wh_govt_certificate	24092	5	C	5501	NaN	NaN	NaN	NaN	NaN	NaN	NaN
wh_breakdown_l3m	25000.0	NaN	NaN	NaN	3.48204	1.690335	0.0	2.0	3.0	5.0	6.0
govt_check_l3m	25000.0	NaN	NaN	NaN	18.81228	8.632382	1.0	11.0	21.0	26.0	32.0
product_wg_ton	25000.0	NaN	NaN	NaN	22102.63292	11607.755077	2065.0	13059.0	22101.0	30103.0	55151.0

Data Cleaning 🧹

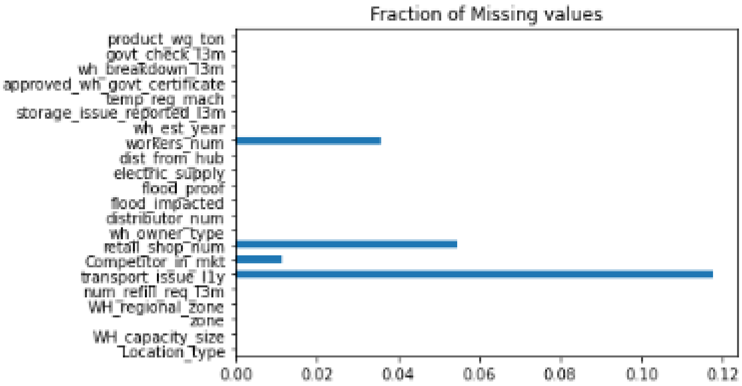
Null Values Check

In [62]:
```
data.isnull().sum()
```

Out[62]:
```
Location_type                      0
WH_capacity_size                   0
zone                               0
WH_regional_zone                   0
num_refill_req_l3m                 0
transport_issue_l1y                0
Competitor_in_mkt                  0
retail_shop_num                    0
wh_owner_type                      0
distributor_num                    0
flood_impacted                     0
flood_proof                        0
electric_supply                    0
dist_from_hub                      0
workers_num                      990
wh_est_year                    11881
storage_issue_reported_l3m         0
temp_reg_mach                      0
approved_wh_govt_certificate     908
wh_breakdown_l3m                   0
govt_check_l3m                     0
product_wg_ton                     0
dtype: int64
```

```
Percent of Total Missing Values in the data = 1.0 %
```

6]:
```
((data.isnull().sum())/data.shape[0]).plot(kind='barh')
plt.title('Fraction of Missing values')
plt.show()
```




```
data.isnull().sum()
```

```
Location_type                   0
WH_capacity_size                0
zone                            0
WH_regional_zone                0
num_refill_req_l3m              0
transport_issue_l1y             0
Competitor_in_mkt               0
retail_shop_num                 0
wh_owner_type                   0
distributor_num                 0
flood_impacted                  0
flood_proof                     0
electric_supply                 0
dist_from_hub                   0
workers_num                     0
wh_est_year                     0
storage_issue_reported_l3m      0
temp_reg_mach                   0
approved_wh_govt_certificate    0
wh_breakdown_l3m                0
govt_check_l3m                  0
product_wg_ton                  0
dtype: int64
```

**Duplicate Value Check

```
data.duplicated().sum()
```

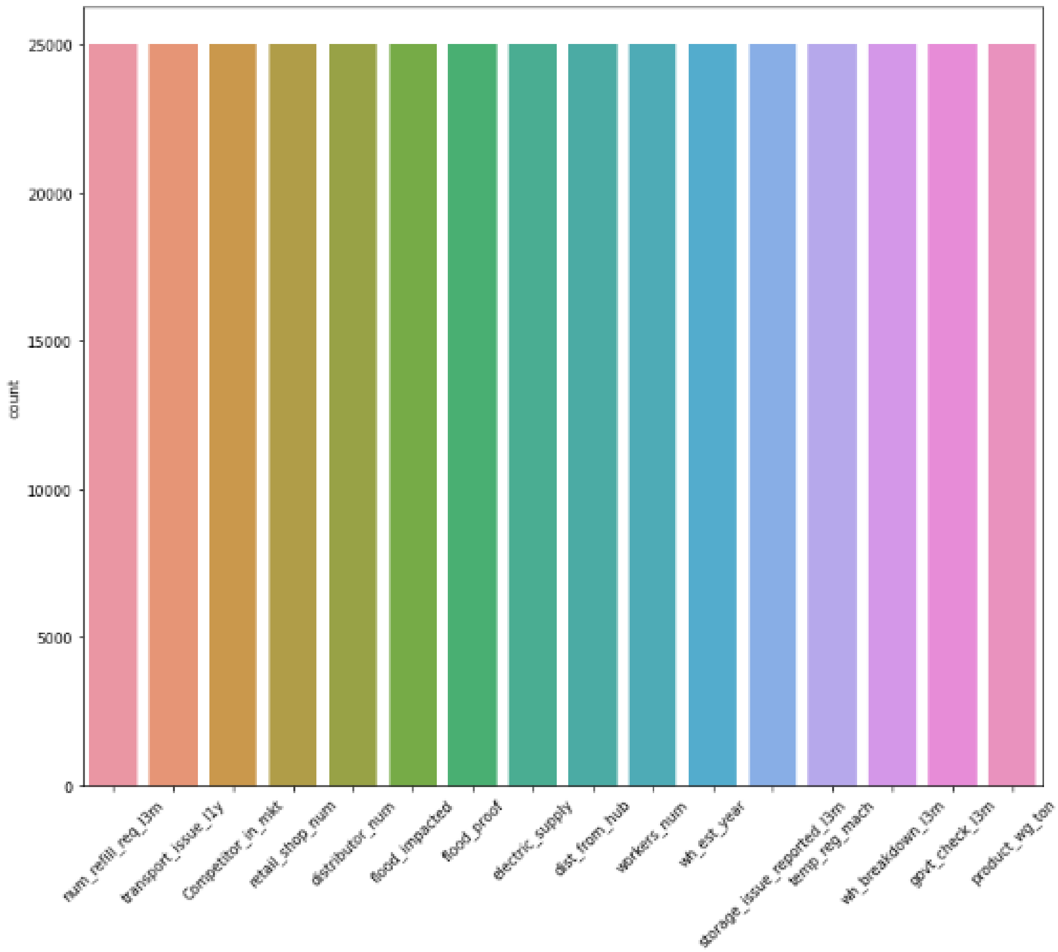
```
0
```

Solving Structural Error

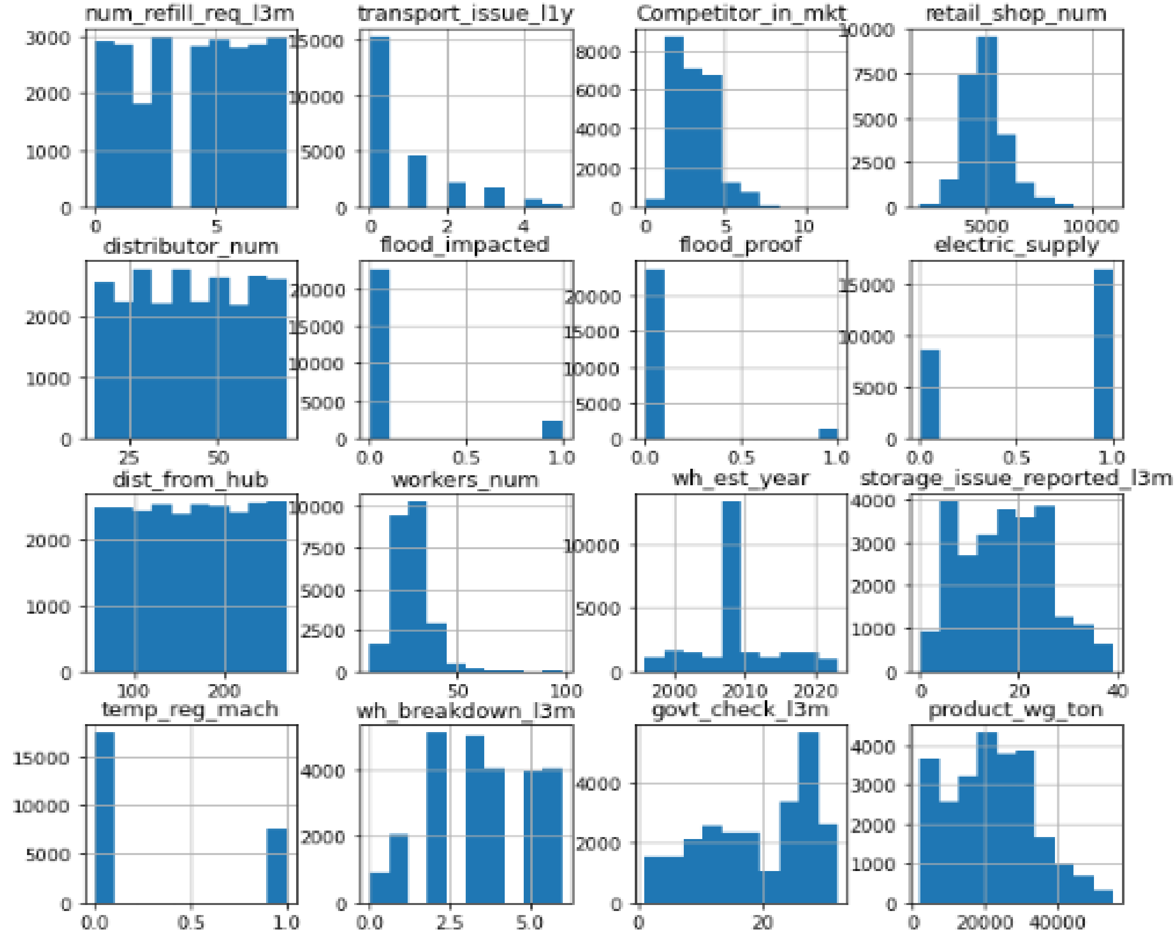
```
data.value_counts()
```

```
Location_type  WH_capacity_size  zone  WH_regional_zone  num_refill_req_l3m  transport_issue_l1y  Competitor_in_mkt  retail
_shop_num  wh_owner_type  distributor_num  flood_impacted  flood_proof  electric_supply  dist_from_hub  workers_num  wh_est
_year  storage_issue_reported_l3m  temp_reg_mach  approved_wh_govt_certificate  wh_breakdown_l3m  govt_check_l3m  product_w
g_ton
Rural          Large             East  Zone 5            0                   0                    3                  4419
Company Owned  37                0                0                0                63              20.000000     2009.383185  11
0              B+                                 4                14             13083           1
               Mid               West  Zone 3            8                   0                    4                  4757
Rented         64                1                0                1                237             28.944398     2009.383185  4
0              B+                                 3                19             4146            1
                                                                                                    6                  4046
Rented         23                0                0                0                176             34.000000     2004.000000  25
0              B                                  6                19             29099           1
                                                                                                                       3843
Company Owned  49                0                0                1                95              30.000000     2009.383185  4
1              B+                                 2                19             5099            1
                                                                                                                       3565
Rented         42                0                0                1                113             29.000000     2014.000000  11
1              B                                  3                19             14075           1

..
               Large             West  Zone 6            2                   2                    4                  2776
Rented         64                0                0                1                179             28.000000     2009.383185  16
0              A                                  5                23             20151           1
                                                                                                    2                  5180
Rented         49                0                0                0                170             18.000000     2009.383185  6
0              A                                  3                6              8072            1
                                                                     1                              6                  6887
Company Owned  27                0                0                1                90              26.000000     2009.383185  4
0              C                                  2                29             5130            1
                                                                                                                       3753
Rented         17                0                0                0                135             20.000000     2009.383185  30
0              B                                  6                23             36066           1
Urban          Small             West  Zone 6            8                   3                    4                  4194
Company Owned  48                1                0                1                139             34.000000     2009.383185  22
0              B                                  6                15             26066           1
Length: 25000, dtype: int64
```



Checking data distribution

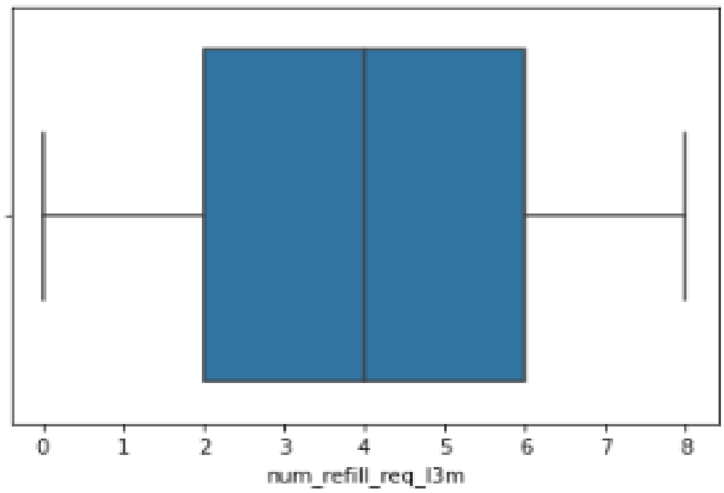


n

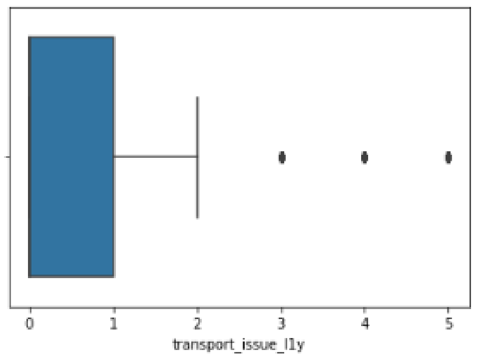
Outlier Management

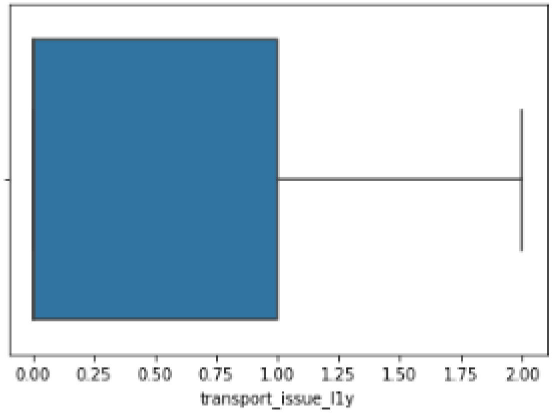
```
In [70]: sns.boxplot(data['num_refill_req_l3m'])
```

Out[70]: <AxesSubplot:xlabel='num_refill_req_l3m'>

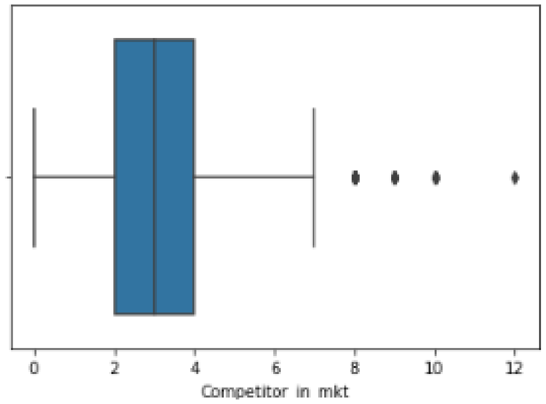


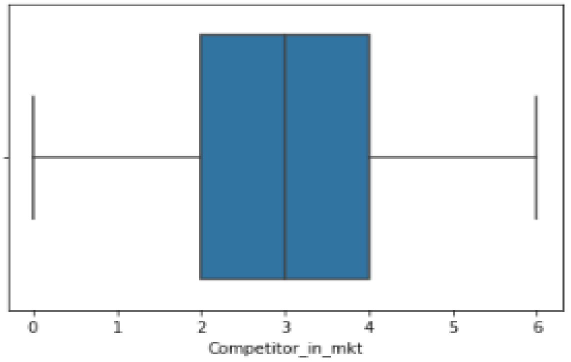
```
In [71]: sns.boxplot(data['transport_issue_l1y'])
Out[71]: <AxesSubplot:xlabel='transport_issue_l1y'>
```



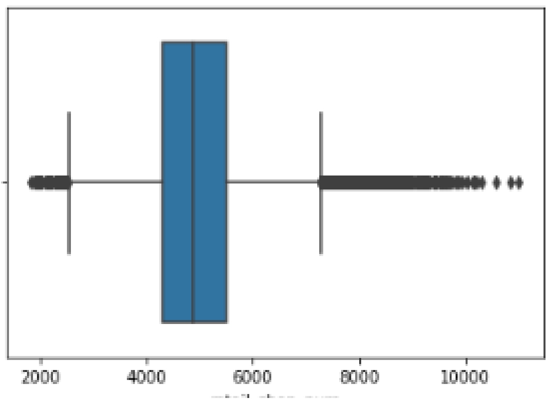


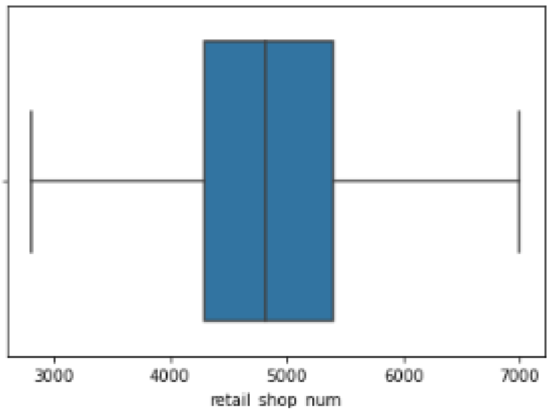
'Competitor_in_mkt



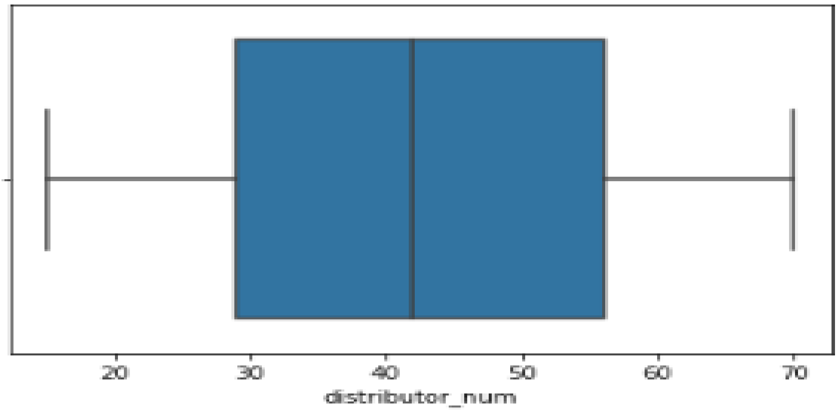


'retail_shop_num'

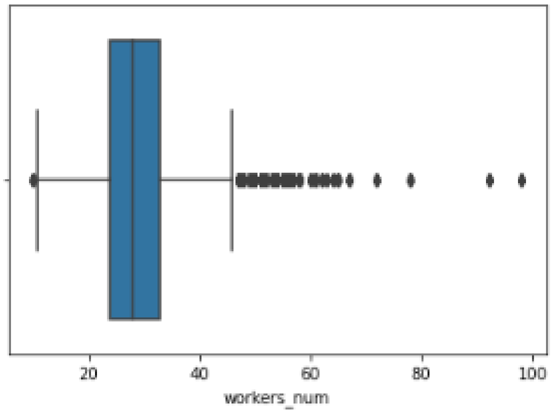




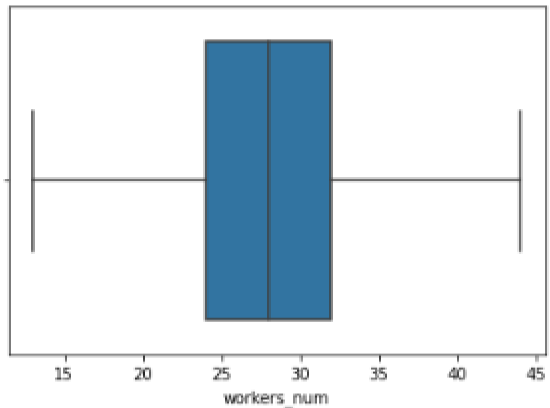
<AxesSubplot:xlabel='distributor_num'>



]: <AxesSubplot:xlabel='workers_num'>

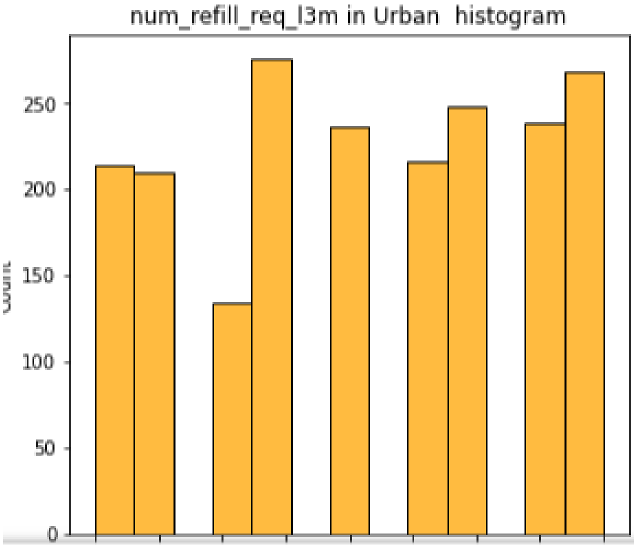


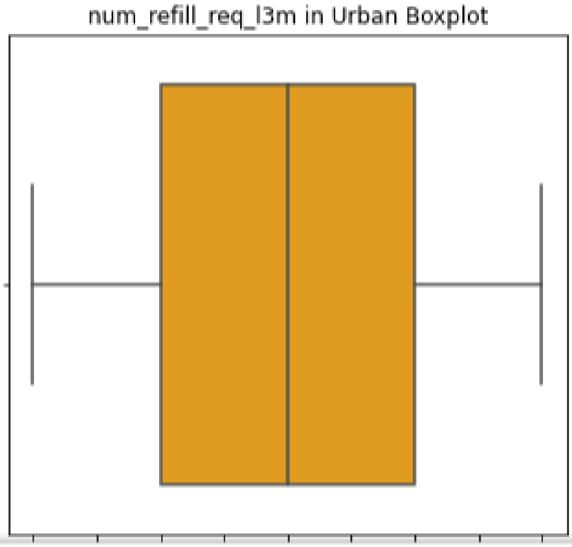
<AxesSubplot:xlabel='workers_num'>

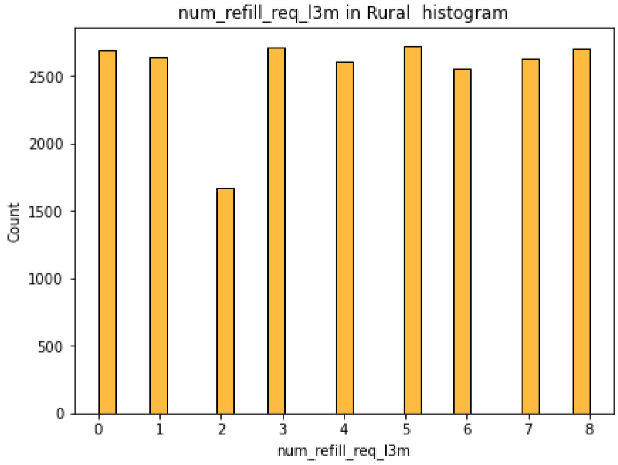


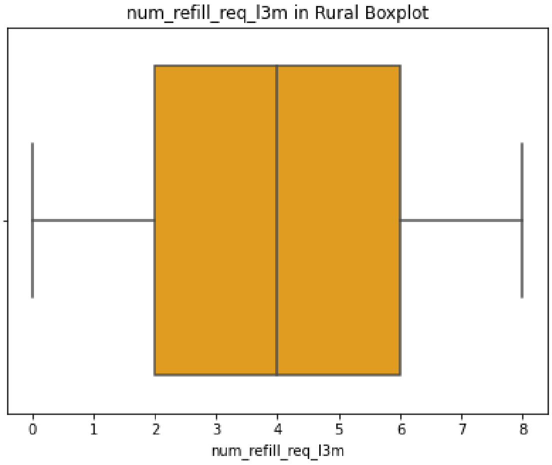
3) Exploratory Data Analysis

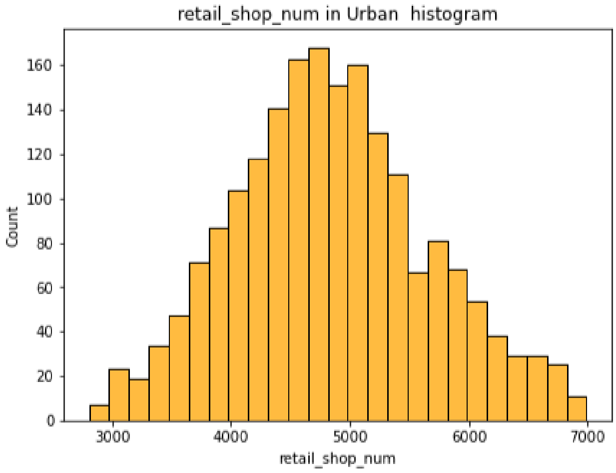
Univariate Analysis

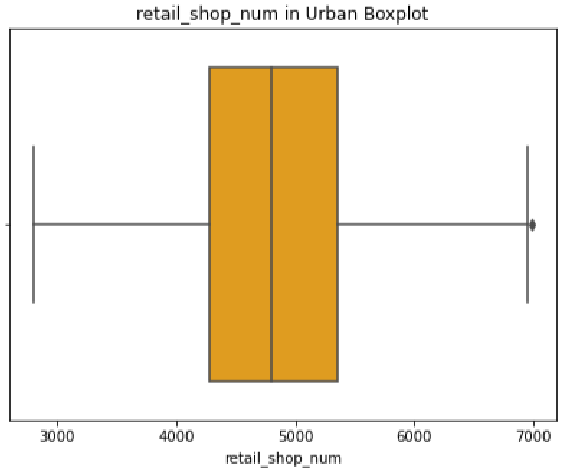


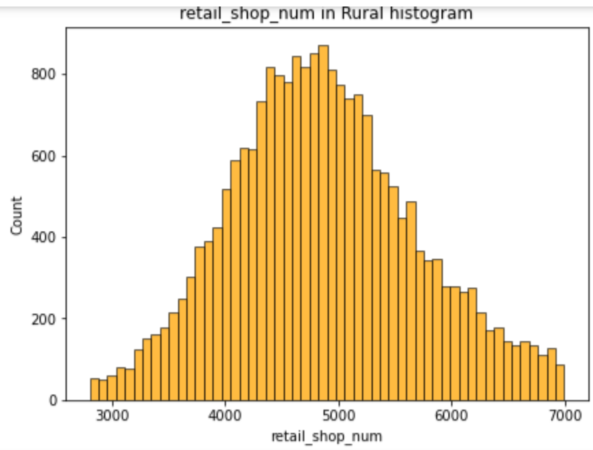


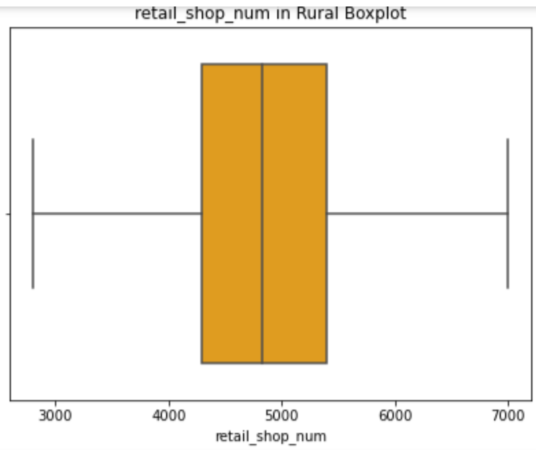


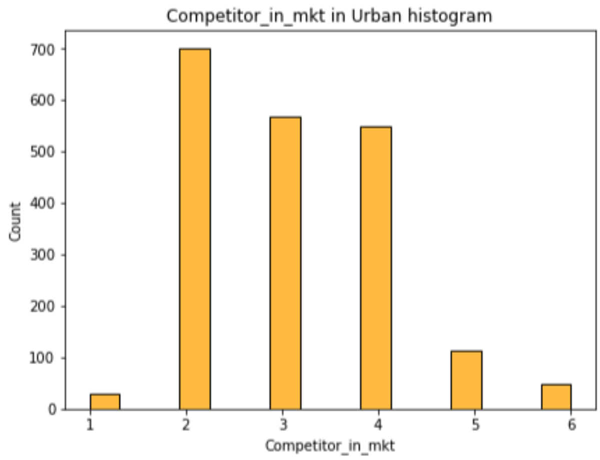


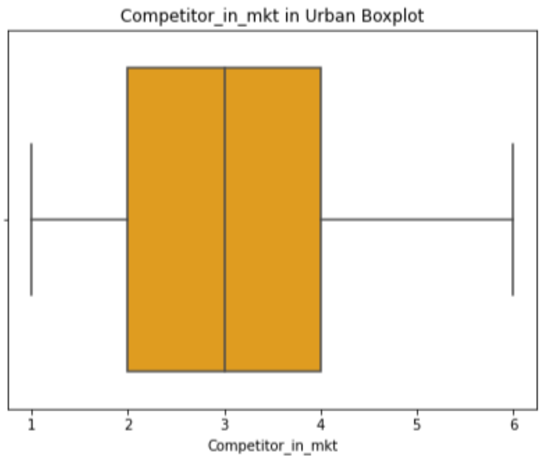

retail_shop_num in Urban histogram
Count
retail_shop_num

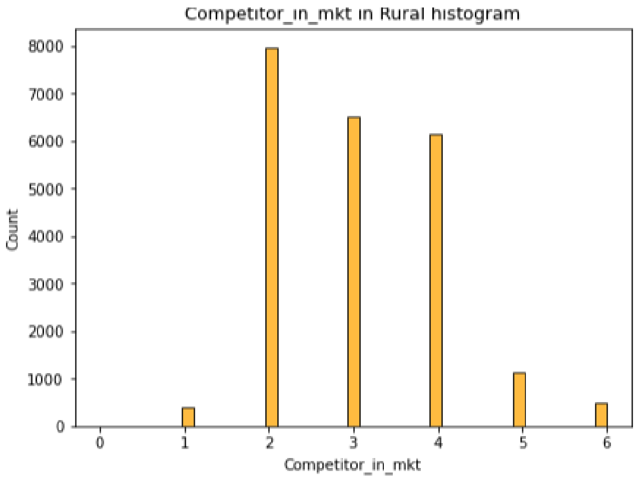

retail_shop_num in Urban Boxplot
retail_shop_num

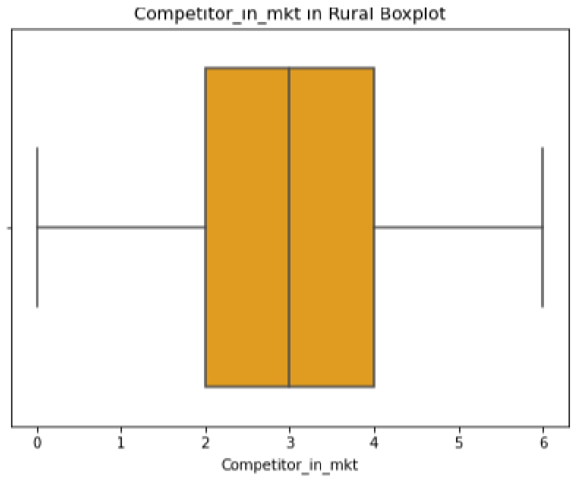

retail_shop_num in Rural histogram
Count
retail_shop_num

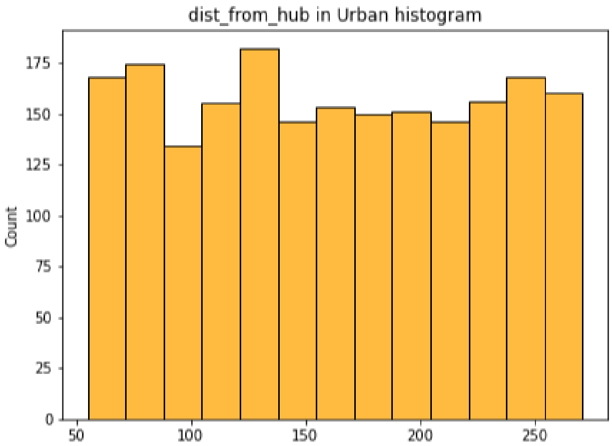

retail_shop_num in Rural Boxplot
retail_shop_num

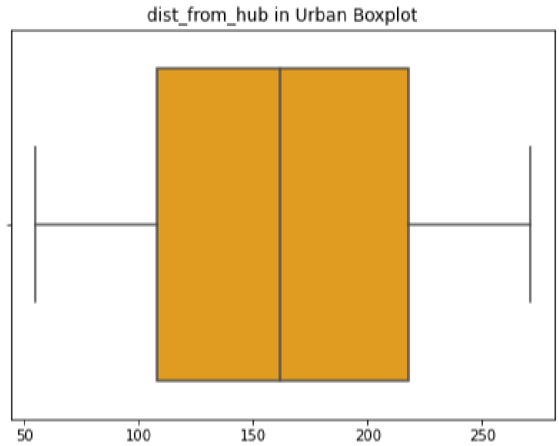

Competitor_in_mkt in Urban histogram
Count
Competitor_in_mkt

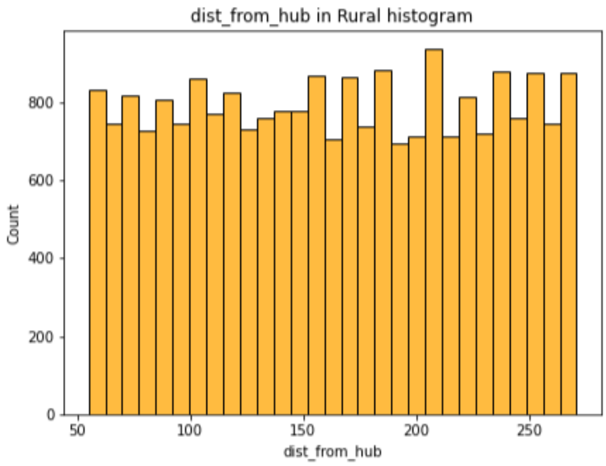

Competitor_in_mkt in Urban Boxplot
Competitor_in_mkt

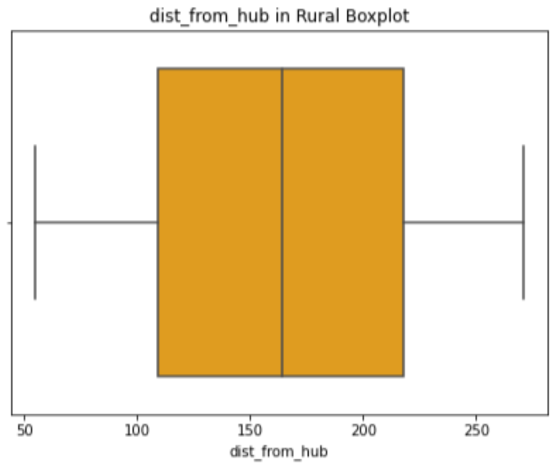

Competitor_in_mkt in Rural histogram
Count
Competitor_in_mkt
0
1000
2000
3000
4000
5000
6000
7000
8000
0
1
2
3
4
5
6

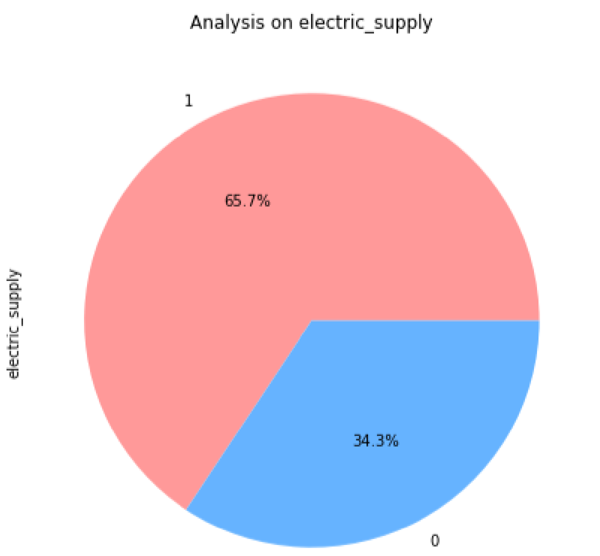

Competitor_in_mkt in Rural Boxplot
Competitor_in_mkt
0
1
2
3
4
5
6

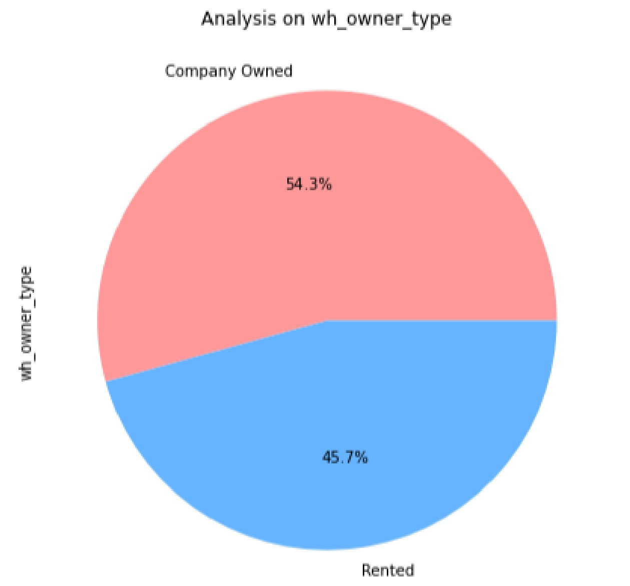

dist_from_hub in Urban histogram
Count
0
25
50
75
100
125
150
175
50
100
150
200
250


dist_from_hub in Urban Boxplot
50
100
150
200
250

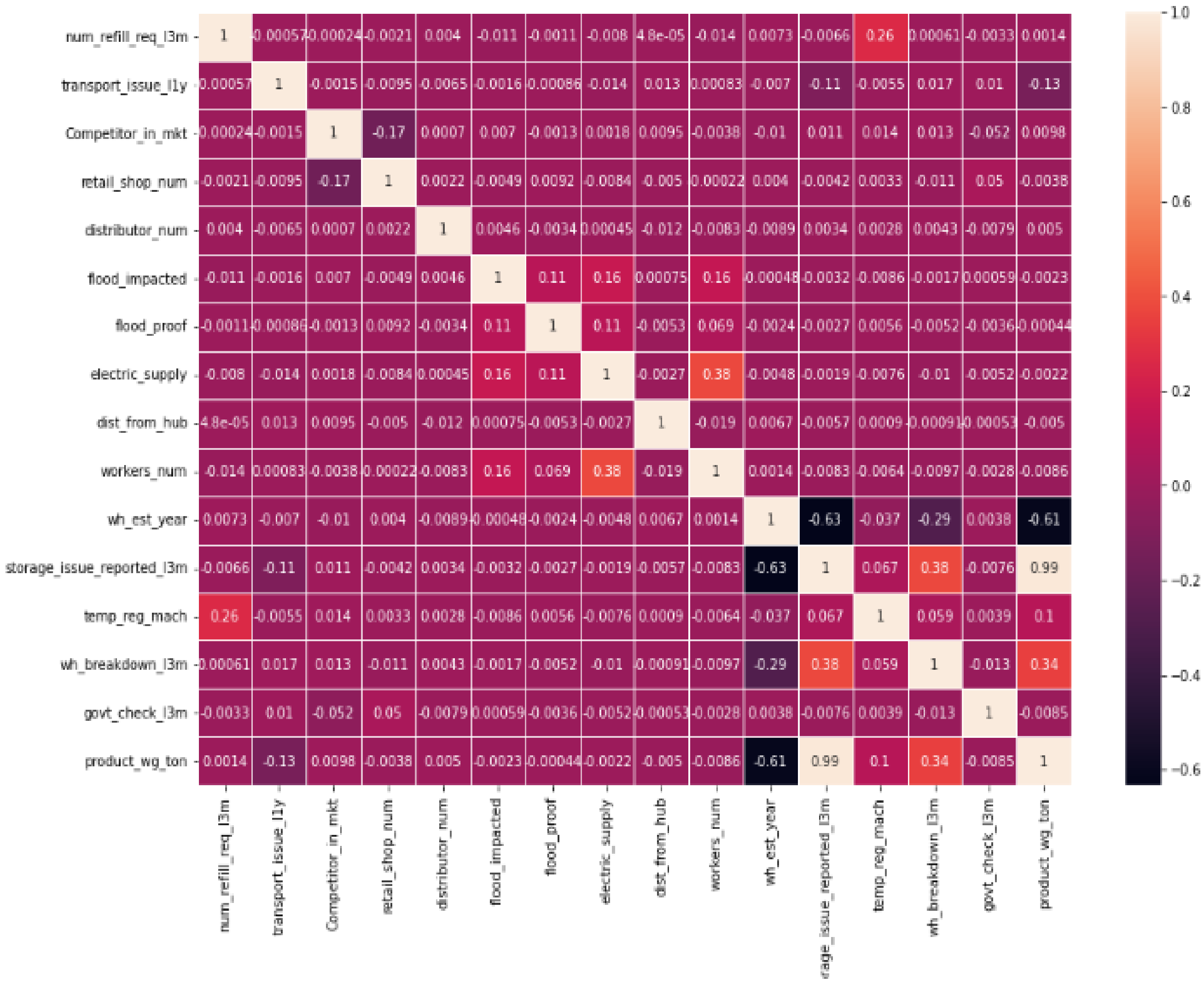

dist_from_hub in Rural histogram
Count
dist_from_hub
0
200
400
600
800
50
100
150
200
250

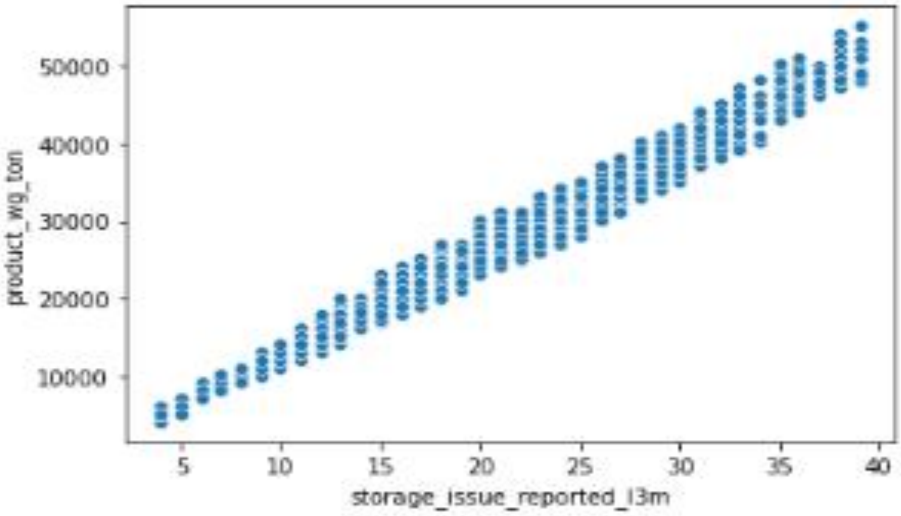

dist_from_hub in Rural Boxplot
dist_from_hub
50
100
150
200
250



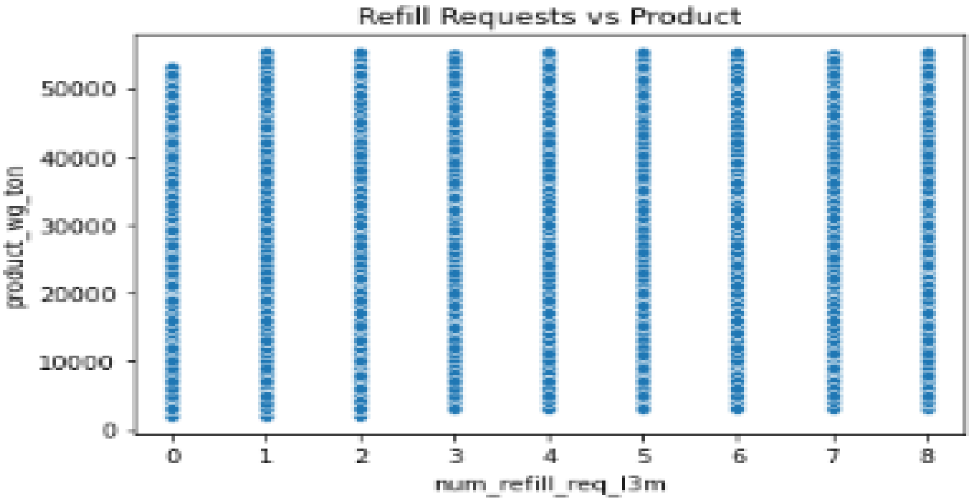


Bivariate Analysis

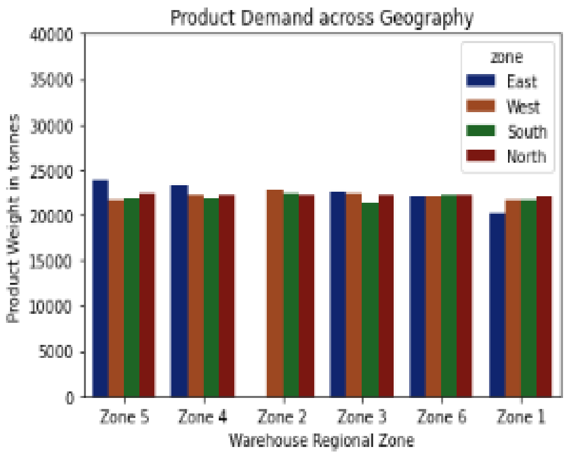




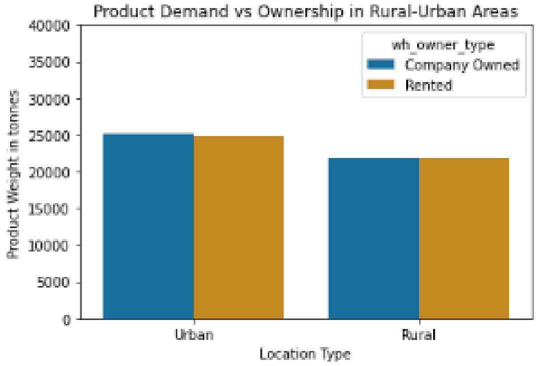
product_wg_ton
50000
40000
30000
20000
10000
5
10
15
20
25
30
35
40
storage_issue_reported_l3m



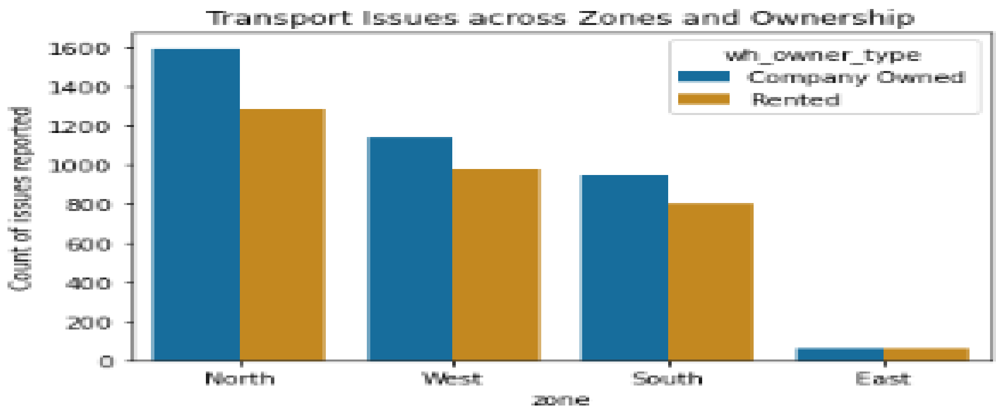
Refill Requests vs Product
product_wg_ton
50000
40000
30000
20000
10000
0
0
1
2
3
4
5
6
7
8
num_refill_req_l3m



Product Demand across Geography
Product Weight in tonnes
40000
35000
30000
25000
20000
15000
10000
5000
0
zone
East
West
South
North
Zone 5
Zone 4
Zone 2
Zone 3
Zone 6
Zone 1
Warehouse Regional Zone



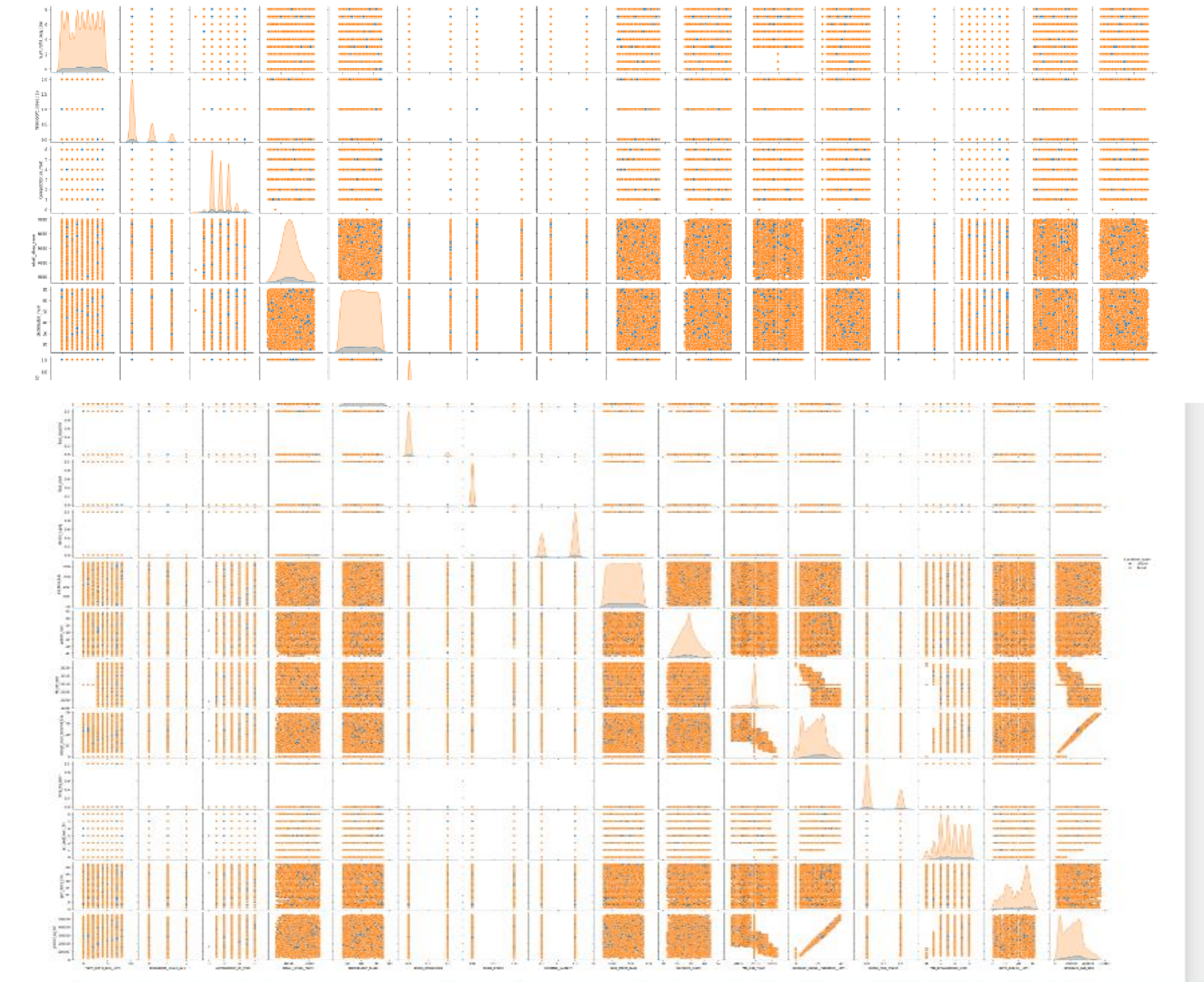
Product Demand vs Ownership in Rural-Urban Areas
Product Weight in tonnes
40000
35000
30000
25000
20000
15000
10000
5000
0
wh_owner_type
Company Owned
Rented
Urban
Rural
Location Type



Multivariate Analysis

```
: sns.pairplot(data, hue="Location_type")
```

```
: <seaborn.axisgrid.PairGrid at 0x1a0bb2f0760>
```



Business Insights

Q 4 a)

- Class Imbalance has been detected for features location_type , flood_impacted,flood_proof, and temp_reg_mach. We can generate more data using the oversampling technique to remove class imbalances or we can ask for more data regarding the location of type rural that are flood proofed and have Warehouse that has temperature regulating machine indicator.

Q 4 b)

- East side of Zone 5 has the highest product demand across the geography
- Company Owned Warehouses have more product demand in urban areas as compared to rural areas
- Company Owned Warehouses in the North zone have the highest transport issues than the other regions.

Q 4 c)

- storage_issue_reported_l3m and product_wg_ton are highly correlated features as they are 99% correlated.As we can say Intuitively due to storage issues like moisture,rat and fungus can lead to degraded product quality and weight.
- storage_issue_reported_l3m and wh_est_year are highly negatively correlated features as they are 63% negatively correlated. As we can intuitively say that warehouse standards have been improved over the years.
- product_wg_ton and wh_est_year are highly negatively correlated features as they are 61% negatively correlated. As we can intuitively say that product weight standards have been improved for warehousing over the years.
- From univariate analysis we can be sure that Number of times refilling has been done in the last 3 months were more for rural areas as compared to urban.
- From univariate analysis we can be sure that Number of instant noodles competitors in the market are more in rural areas as compared to urban locations as locals want to promote their local brands. Because there are more competitors in the market hence the retail shops.
- From univariate analysis we can be sure that Distance between the warehouse and the production hub is more for rural areas as compared to urban areas, which is contextually correct.