Capstone Report
Supply Chain Analysis

Submitted by:

N. Aishwarya
PGP-DSBA

October 2022

Business Report Outline

List of Figures

List of Tables

List of Formulas

1. Introduction to the Business Problem

Problem Statement:

A FMCG company has entered into the instant noodles business two years back. Their higher management has notices that there is a mismatch 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.

The objective of this exercise is to build a model, using historical data that will determine an optimum weight of the product to be shipped each time to the warehouse.

Purpose of Study:

The company is incurring inventory loss due to inadequate demand-supply management of its product, instant noodles. The management needs to optimize the supply quantity across all warehouses in the country.

Hence the objective of this project is to build a model using historical data, to determine an optimum product weight to be shipped each time to the warehouse. With multiple options available to analyse data, it is challenging to decide the appropriate machine learning model to use since the performance of the model vary on the parameters available in the data. This project aims to compare different popular machine learning classifiers, and measure their performance to find out which machine learning model performs better.

Approach:

Data exploration will be carried out using the Python, through univariate and bivariate analysis. The presence of outliers interferes with the correct modelling and interpretation of the data. Thus, the outliers (extreme values) are identified and replaced.

Since the dataset used is related to supply chain, important parameters are identified and the machine learning models are trained with the dataset for detection of optimum weight. The study helps to analyse/optimize the supply quantity at each warehouse in the country & thereby determining the advertising strategies & campaigns for specific pockets.

In this problem, “PRODUCT_WG_TON is the target variable.

Data dictionary as described below:

S. No	Field Name	Description
1	Ware_house_ID	Product warehouse ID
2	WH_Manager_ID	Employee ID of warehouse manager
3	Location_type	Location of warehouse like in city or village
4	WH_capacity_size	Storage capacity size of the warehouse
5	zone	Zone of the warehouse
6	WH_regional_zone	Regional zone of the warehouse under each zone
7	num_refill_req_l3m	Number of times refilling has been done in last 3 months
8	transport_issue_l1y	Any transport issue like accident or goods stolen reported in last one year
9	Competitor_in_mkt	Number of instant noodles competitor in the market
10	retail_shop_num	Number of retails shop who sell the product under the warehouse area
11	wh_owner_type	Company is owning the warehouse or they have got the warehouse on rent
12	distributor_num	Number of distributers works in between warehouse and retail shops
13	flood_impacted	Warehouse is in the Flood impacted area indicator
14	flood_proof	Warehouse is flood proof indicators. Like storage is at some height not directly on the ground
15	electric_supply	Warehouse have electric back up like generator, so they can run the warehouse in load shedding
16	dist_from_hub	Distance between warehouse to the production hub in Kms
17	workers_num	Number of workers working in the warehouse
18	wh_est_year	Warehouse established year
19	storage_issue_reported_l3m	Warehouse reported storage issue to corporate office in last 3 months. Like rat, fungus because of moisture etc.
20	temp_reg_mach	Warehouse have temperature regulating machine indicator
21	approved_wh_govt_certificate	What kind of standard certificate has been issued to the warehouse from government regulatory body
22	wh_breakdown_l3m	Number of time warehouse face a breakdown in last 3 months. Like strike from worker, flood, or electrical failure
23	govt_check_l3m	Number of time government Officers have been visited the warehouse to check the quality and expire of stored food in last 3 months
24	product_wg_ton	Product has been shipped in last 3 months. Weight is in tons

Table 1: Data dictionary for the dataset

There is total 24 variables in this dataset. It contains various measures related to company business.

2. Data report:

2.1. Visual inspection of data

The dataset “Data.csv” is loaded using pandas and the dataset has 25,000 observations (rows) and 24 variables (columns). A quick glimpse of the data is shown below:

Ware_house_ID	WH_Manager_ID	Location_type	WH_capacity_size	zone	WH_regional_zone	num_refill_req_l3m	transport_issue_l1y	Competitor_in_mkt
WH_100000	EID_50000	Urban	Small	West	Zone 6	3	1	2
WH_100001	EID_50001	Rural	Large	North	Zone 5	0	0	4
WH_100002	EID_50002	Rural	Mid	South	Zone 2	1	0	4
WH_100003	EID_50003	Rural	Mid	North	Zone 3	7	4	2
WH_100004	EID_50004	Rural	Large	North	Zone 5	3	1	2
WH_100005	EID_50005	Rural	Small	West	Zone 1	8	0	2
WH_100006	EID_50006	Rural	Large	West	Zone 6	8	0	4
WH_100007	EID_50007	Rural	Large	North	Zone 5	1	0	4
WH_100008	EID_50008	Rural	Small	South	Zone 6	8	1	4
WH_100009	EID_50009	Rural	Small	South	Zone 6	4	3	3

Figure 1: Dataset overview

Checking columns of dataset:

```
Index(['Ware_house_ID', 'WH_Manager_ID', '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_wg_ton'],
      dtype='object')
```

Checking the shape of the data:

Number of rows: 25,000
Number of columns: 24

Observations:

- The data set contains 25,000 rows 24 columns.
- 6 Categorical Variables, 4 Nominal Variables & 14 Continuous Variables.
- The product_wg_ton is the target variable, that is the objective of the study is to study and model the product weight in order to estimate the future demands.
- For our analysis, Ware_house_ID and WH_Manager_ID are dropped.

2.2. Understanding of attributes:

Description of variables are as below, to understand the data better:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25000 entries, 0 to 24999
Data columns (total 24 columns):
 #   Column                        Non-Null Count  Dtype  
---  ------                        --------------  -----  
 0   Ware_house_ID                 25000 non-null  object 
 1   WH_Manager_ID                 25000 non-null  object 
 2   Location_type                 25000 non-null  object 
 3   WH_capacity_size              25000 non-null  object 
 4   zone                          25000 non-null  object 
 5   WH_regional_zone              25000 non-null  object 
 6   num_refill_req_l3m            25000 non-null  int64  
 7   transport_issue_l1y           25000 non-null  int64  
 8   Competitor_in_mkt             25000 non-null  int64  
 9   retail_shop_num               25000 non-null  int64  
 10  wh_owner_type                 25000 non-null  object 
 11  distributor_num               25000 non-null  int64  
 12  flood_impacted                25000 non-null  int64  
 13  flood_proof                   25000 non-null  int64  
 14  electric_supply               25000 non-null  int64  
 15  dist_from_hub                 25000 non-null  int64  
 16  workers_num                   24010 non-null  float64
 17  wh_est_year                   13119 non-null  float64
 18  storage_issue_reported_l3m    25000 non-null  int64  
 19  temp_reg_mach                 25000 non-null  int64  
 20  approved_wh_govt_certificate  24092 non-null  object 
 21  wh_breakdown_l3m              25000 non-null  int64  
 22  govt_check_l3m                25000 non-null  int64  
 23  product_wg_ton                25000 non-null  int64  
dtypes: float64(2), int64(14), object(8)
memory usage: 4.6+ MB
```

Figure 2: Description of variables

Statistical description of the dataset:

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
Ware_house_ID	25000	25000	WH_100000	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
WH_Manager_ID	25000	25000	EID_50000	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
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

Figure 3: Statistical description of the dataset

The values of mean, standard deviation, minimum and maximum, 25th, 50th and 75th percentile is mentioned in the above tables.

Checking for Duplicate values in dataset:

Number of duplicate rows = 0

Summary statistics of the object variable:

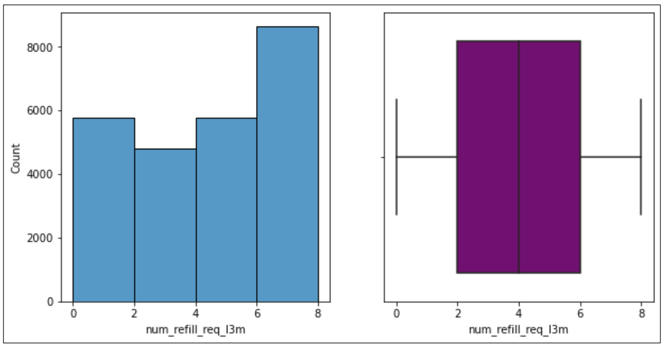
	count	unique	top	freq
Location_type	25000	2	Rural	22957
WH_capacity_size	25000	3	Large	10169
zone	25000	4	North	10278
WH_regional_zone	25000	6	Zone 6	8339
wh_owner_type	25000	2	Company Owned	13578
approved_wh_govt_certificate	24092	5	C	5501

3. Exploratory Data Analysis:

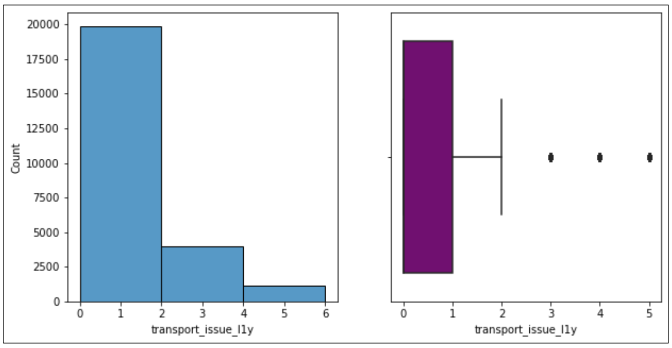
3.1. Univariate analysis:

Univariate analysis provides the distribution and spread for every continuous attribute, distribution of data in categories for categorical ones. Below charts provide the univariate analysis for the most critical variables:

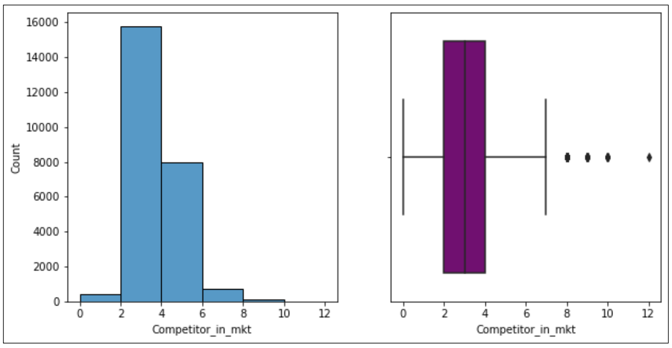
1. num_refill_req_l3m Variable:



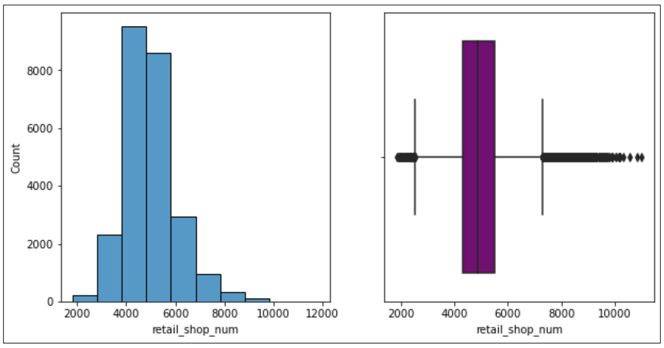
2. transport_issue_l1y Variable:



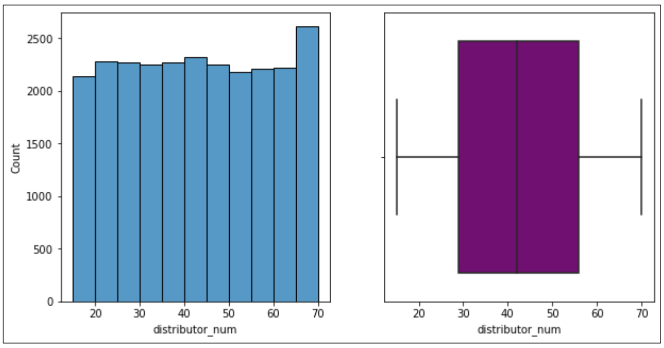
3. Competitor_in_mkt Variable:



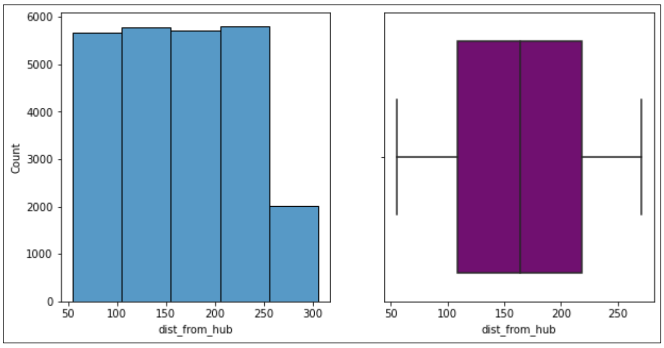
4. retail_shop_num Variable:



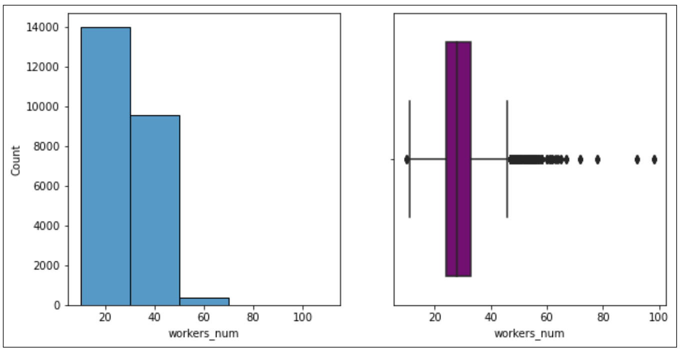
5. distributor_num Variable:



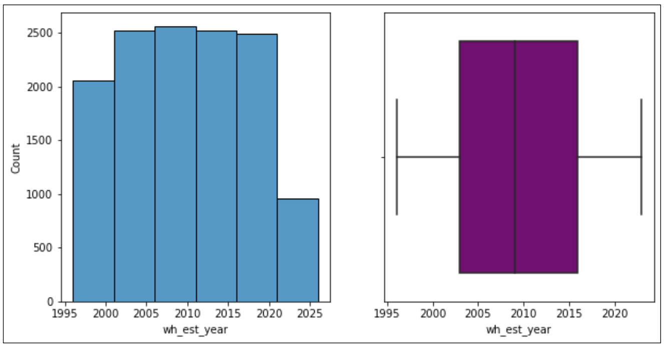
6. Dist_from_hub Variable:



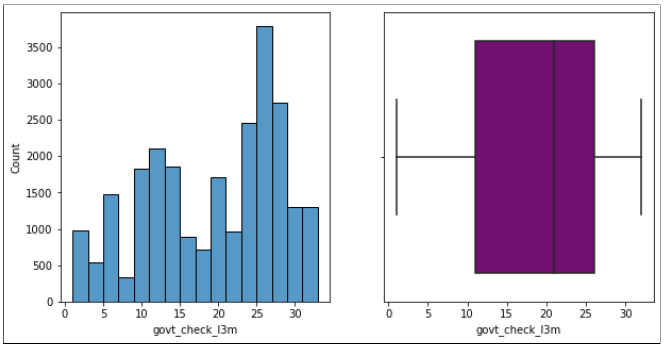
7. Workers_num Variable:



8. Wh_est_year Variable:



9. Govt_check_l3m Variable:



10. Product_wg_ton Variable:

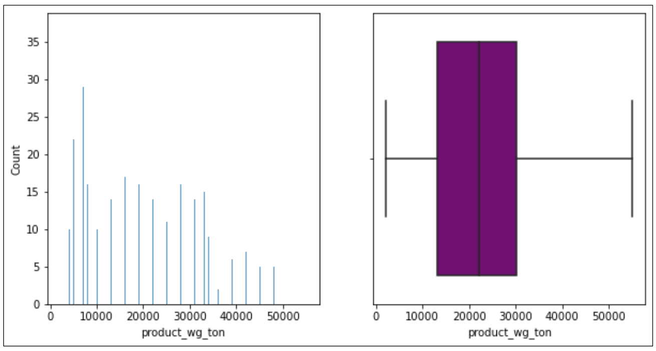


Figure 4: Univariate analysis

Observations:

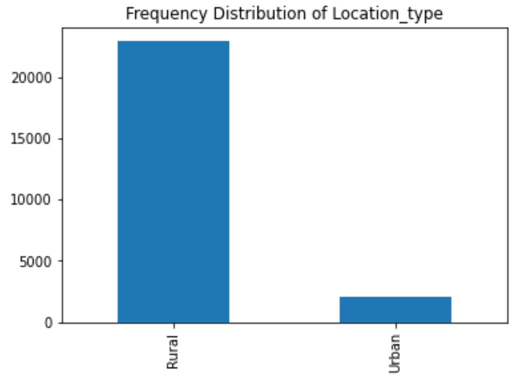
- Most of the variables are right skewed, left skewed variables are: wh_breakdown_13m, num_refill_req_13m, govt_check_13m, electric_supply.
- Retail shops are right skewed. Having 4000 to 5000 retail shops in normal.
- The number of distributors is observed to be fairly uniform across the respective ranges.
- The majority of warehouses have 20 to 40 workers.

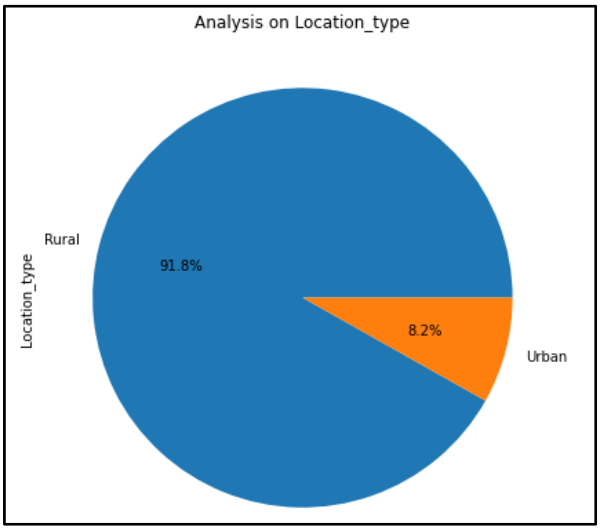
- The number of storage issues reported is right skewed, with the majority having less than 20 issues across the duration of 3 months. The bin width is set to 10.
- It is observed that the majority of warehouses have reported more than 4 breakdowns in past 3 months. The data is left skewed.
- The number of government checks have been 25 to 30 for most warehouses.
- Distance from hub is observed to be fairly uniform across the respective ranges

Getting unique counts of Categorical Variables:

1. Location_type Variable:

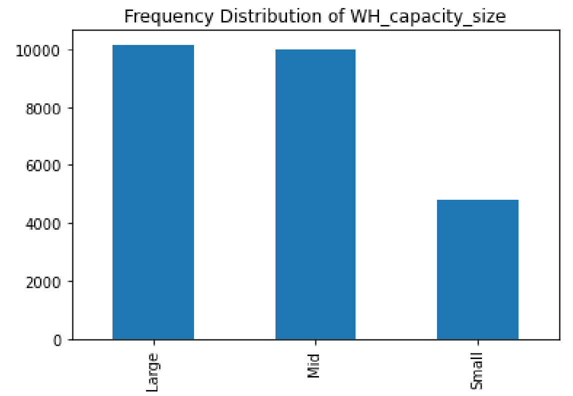
```
Details of Location_type
----------------------------------------------
Rural    22957
Urban     2043
Name: Location_type, dtype: int64
```

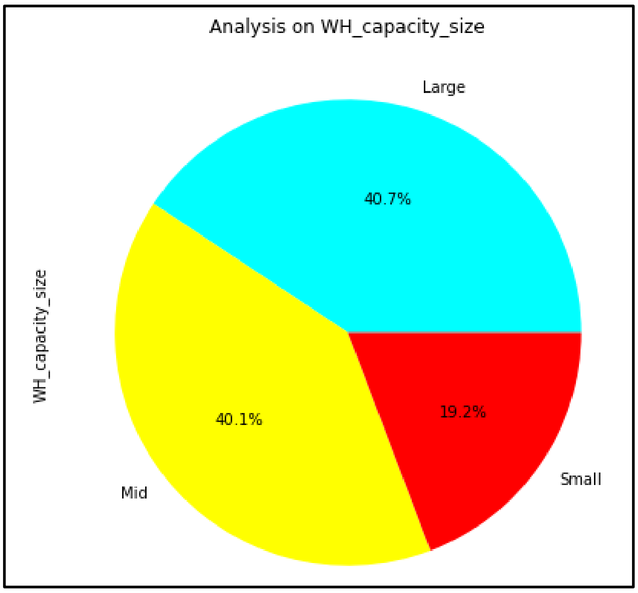




2. WH_capacity size Variable:

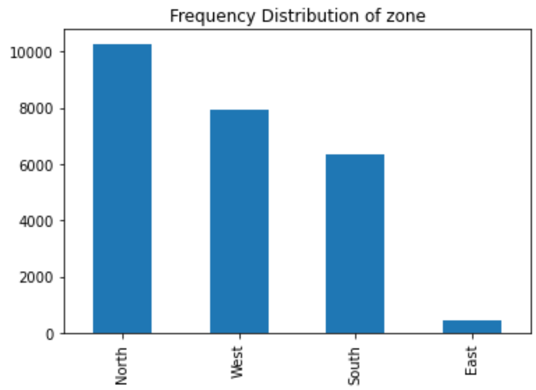
```
Details of WH_capacity_size
----------------------------------------------
Large    10169
Mid      10020
Small     4811
Name: WH_capacity_size, dtype: int64
```

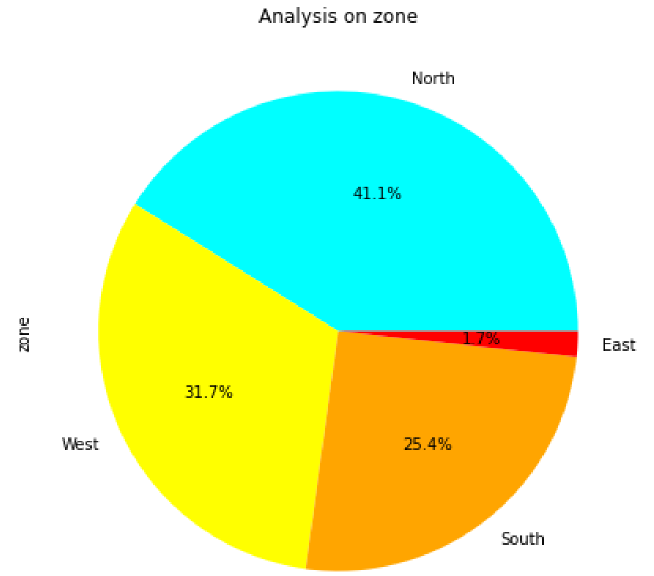




3. Zone Variable:

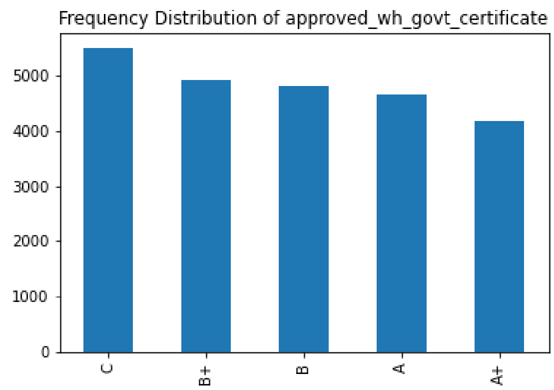
```
Details of zone
--------------------------------------------------
North    10278
West      7931
South     6362
East       429
Name: zone, dtype: int64
```

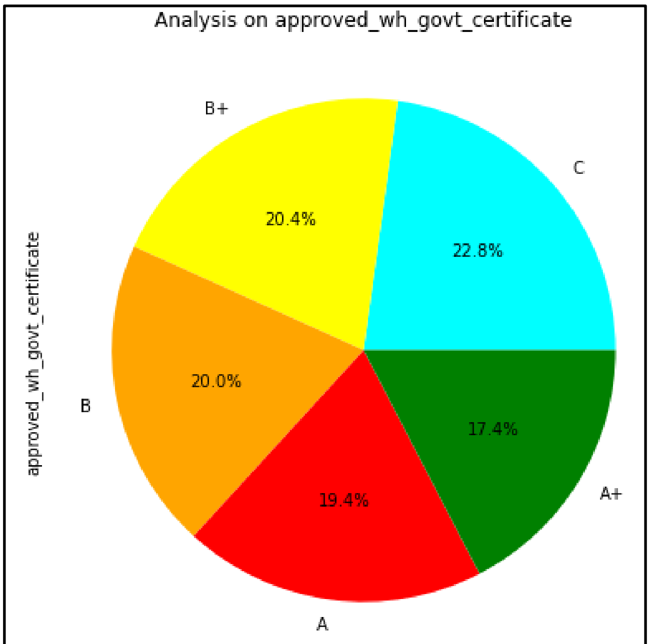




4. Approved_WH_Govt Variable:

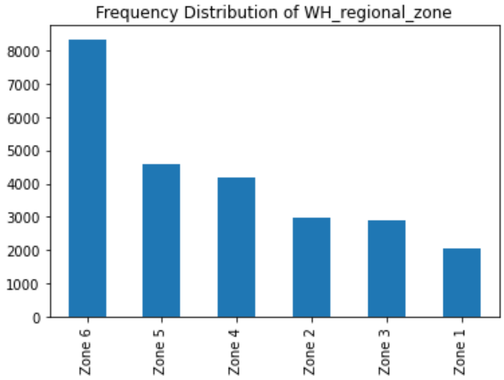
```
Details of approved_wh_govt_certificate
--------------------------------------------------
C      5501
B+     4917
B      4812
A      4671
A+     4191
Name: approved_wh_govt_certificate, dtype: int64
```

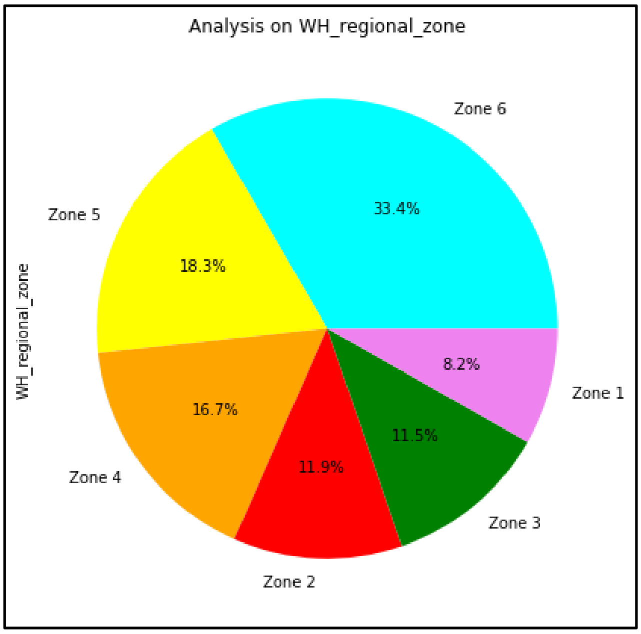




5. WH_regional Variable:

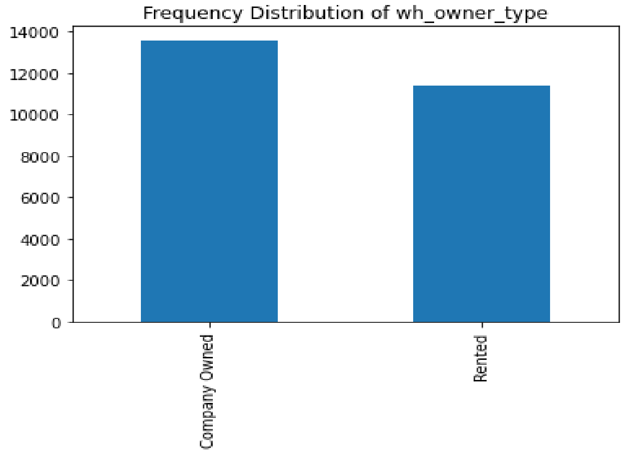
```
Details of WH_regional_zone
--------------------------------------------------
Zone 6    8339
Zone 5    4587
Zone 4    4176
Zone 2    2963
Zone 3    2881
Zone 1    2054
Name: WH_regional_zone, dtype: int64
```

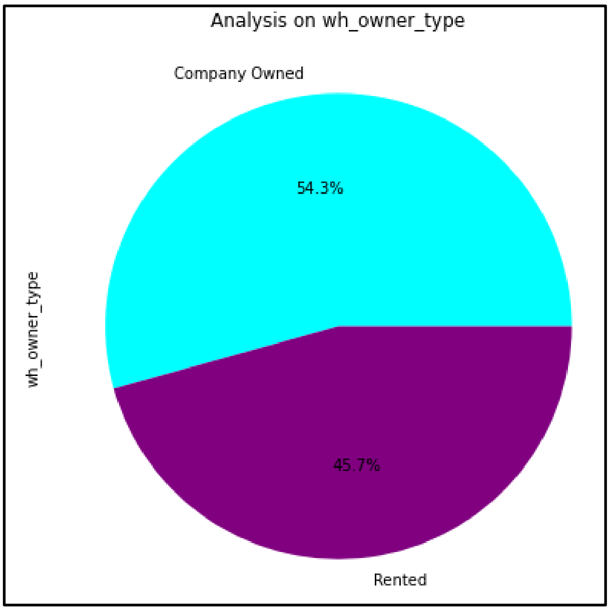




6. Owner type:

```
Details of wh_owner_type
--------------------------------------------------
Company Owned    13578
Rented           11422
Name: wh_owner_type, dtype: int64
```





Observations:

- Type of Location is Urban and Rural, with a higher count in rural type.
- Warehouse capacity are of 3 types: Small, Mid & large, it can be observed that small warehouses are less compared to that of Large and Mid.
- There are 2 Owner types: Rented & company owner. Company Owned warehouses are slightly more than the rented ones.
- The count of warehouses in the East zone are minimal compared to those in the North, West and South zones. majority of warehouses with electric supply have opted to go ahead without regulation of temperature facility. The warehouse also has a certificate (A+, A, B, B+ and C) based on the government standards.
- Regional zone numbered as Zone1 to Zone6, with Zone 1 has the lowest.
- It can be observed that the warehouses impacted by floods form a mere minority.

SKEWNESS VALUE:

Formula 1: Skewness = 3 * (Mean – Median) / Standard Deviation.

Variables	Skewness Value
flood_proof	3.92
flood_impacted	2.70
transport_issue_l1y	1.61
workers_num	1.06
Competitor_in_mkt	0.98
retail_shop_num	0.91
temp_reg_mach	0.86
product_wg_ton	0.33
storage_issue_reported_l3m	0.11
distributor_num	0.02
wh_est_year	0.01
dist_from_hub	-0.01
wh_breakdown_l3m	-0.07
num_refill_req_l3m	-0.08
govt_check_l3m	-0.36

electric_supply	-0.66

3.2. Bivariate analysis:

	WH_capacity_size	product_wg_ton	num_refill_req_l3m	wh_breakdown_l3m
0	Large	22100.487855	4.093815	3.475268
1	Mid	22202.298104	4.113473	3.496906
2	Small	21899.591561	4.028061	3.465392

	approved_wh_govt_certificate	product_wg_ton	workers_num
0	A+	26717.947984	28.879692
1	A	24122.532220	28.813673
2	B+	21456.008338	28.985403
3	B	21259.281588	28.967330
4	C	20938.889293	29.035566

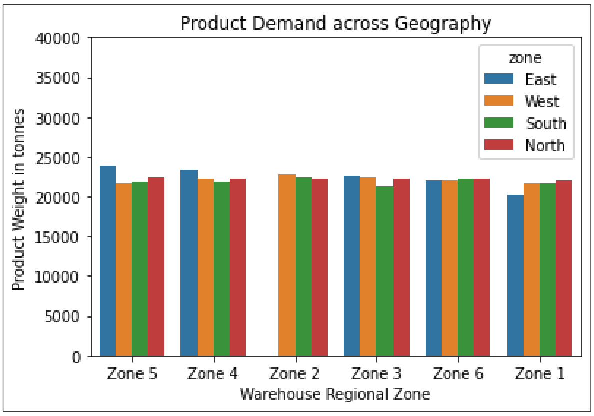


Figure 5: Bivariate analysis

It can be observed that East zone has no warehouses in Zone 2. In the North and the West, the Zone 6 has the highest number of warehouses. Further investigation shows that the East zone has fewer number of distributors and retail shops, but more competitors in the market, relative to the other zones.

In summary, this zone has fewer warehouses, retail outlets and distributors. It has much higher number of competitors yet has the same product demand as other zones. This might be a result of the popularity of the product in this region, encouraging the marketing department to pay greater attention to this region.

Correlation analysis:

Formula 2:

$$Correlation = \frac{Cov(x,y)}{\sigma_x * \sigma_y}$$

Where,

$Cov(x, y) =$ Covariance of x and y

$\sigma_x =$ Standard deviation of x

$\sigma_y =$ Standard deviation of y

	num_refill_req_l3m	transport_issue_l1y	Competitor_in_mkt	retail_shop_num	distributor_num	flood_impacted	flood_proof
num_refill_req_l3m	1.000	0.019	0.003	-0.001	0.004	-0.011	-0.001
transport_issue_l1y	0.019	1.000	-0.006	-0.002	0.009	-0.010	0.000
Competitor_in_mkt	0.003	-0.006	1.000	-0.157	-0.001	0.009	-0.003
retail_shop_num	-0.001	-0.002	-0.157	1.000	-0.000	-0.004	0.007
distributor_num	0.004	0.009	-0.001	-0.000	1.000	0.005	-0.003
flood_impacted	-0.011	-0.010	0.009	-0.004	0.005	1.000	0.107
flood_proof	-0.001	0.000	-0.003	0.007	-0.003	0.107	1.000
electric_supply	-0.008	-0.009	0.002	-0.009	0.000	0.165	0.115
dist_from_hub	0.000	0.014	0.008	0.000	-0.012	0.001	-0.005
workers_num	-0.014	-0.009	0.000	-0.005	-0.015	0.168	0.041
wh_est_year	0.015	-0.013	-0.011	0.006	-0.012	-0.001	-0.003
storage_issue_reported_l3m	-0.007	-0.144	0.010	-0.007	0.003	-0.003	-0.003
temp_reg_mach	0.261	0.018	0.010	-0.001	0.003	-0.009	0.006
wh_breakdown_l3m	0.001	0.013	0.013	-0.008	0.004	-0.002	-0.005
govt_check_l3m	-0.003	0.002	-0.043	0.046	-0.008	0.001	-0.004
product_wg_ton	0.001	-0.174	0.009	-0.007	0.005	-0.002	-0.000

Figure 6: Correlation analysis

Correlation Heatmap:

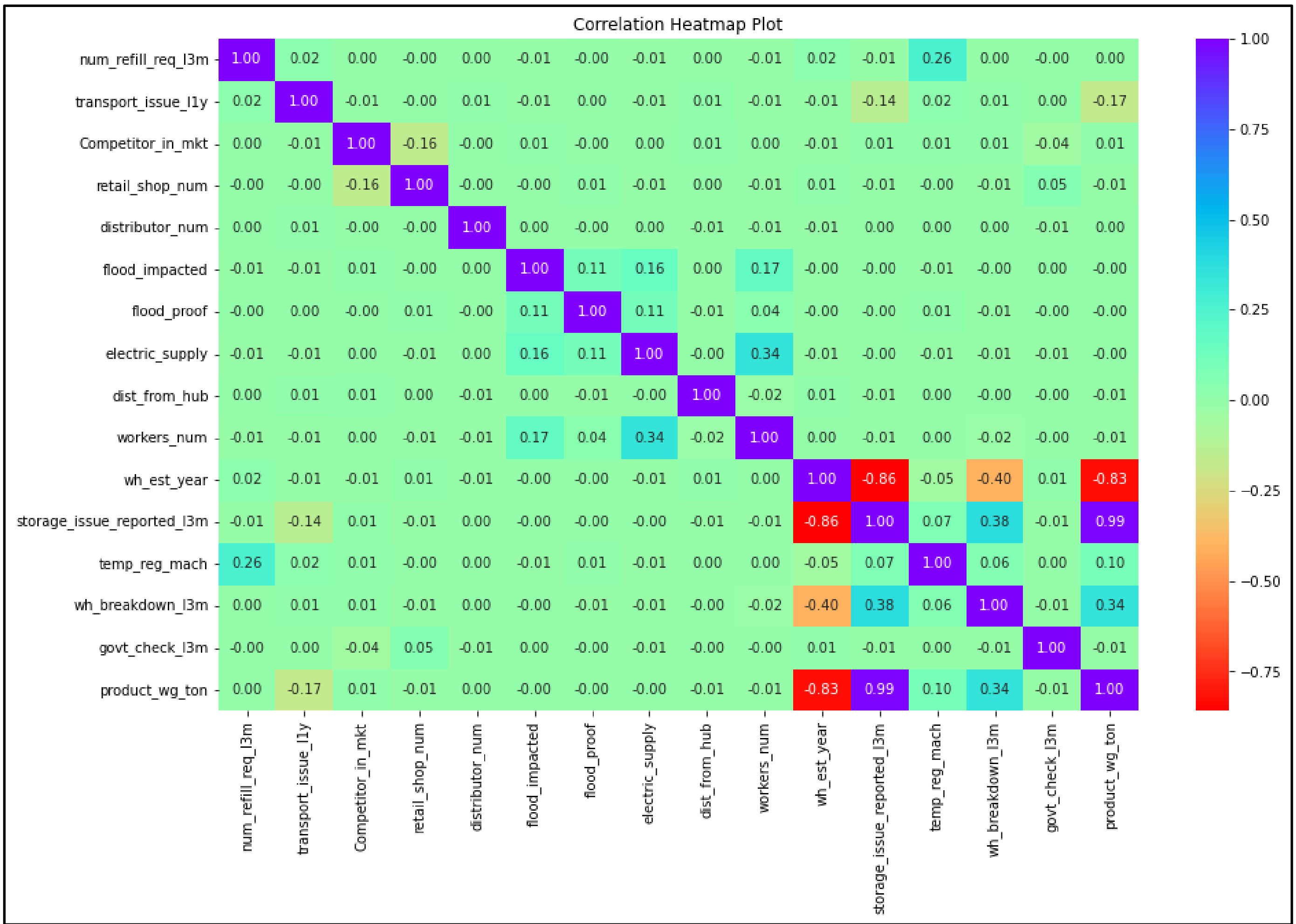


Figure 7: Correlation heatmap

- It is observed that the product demand has a strong positive correlation with the storage issues reported.
- The Pearson correlation coefficients calculated for all pairs of continuous variables obtained as a matrix, have been presented as a heatmap. The darker the colour, the lower is the magnitude of the correlation.
- As can be observed, "storage issues reported in 3 months" column has a high correlation with product weight in tons.

Pair plot

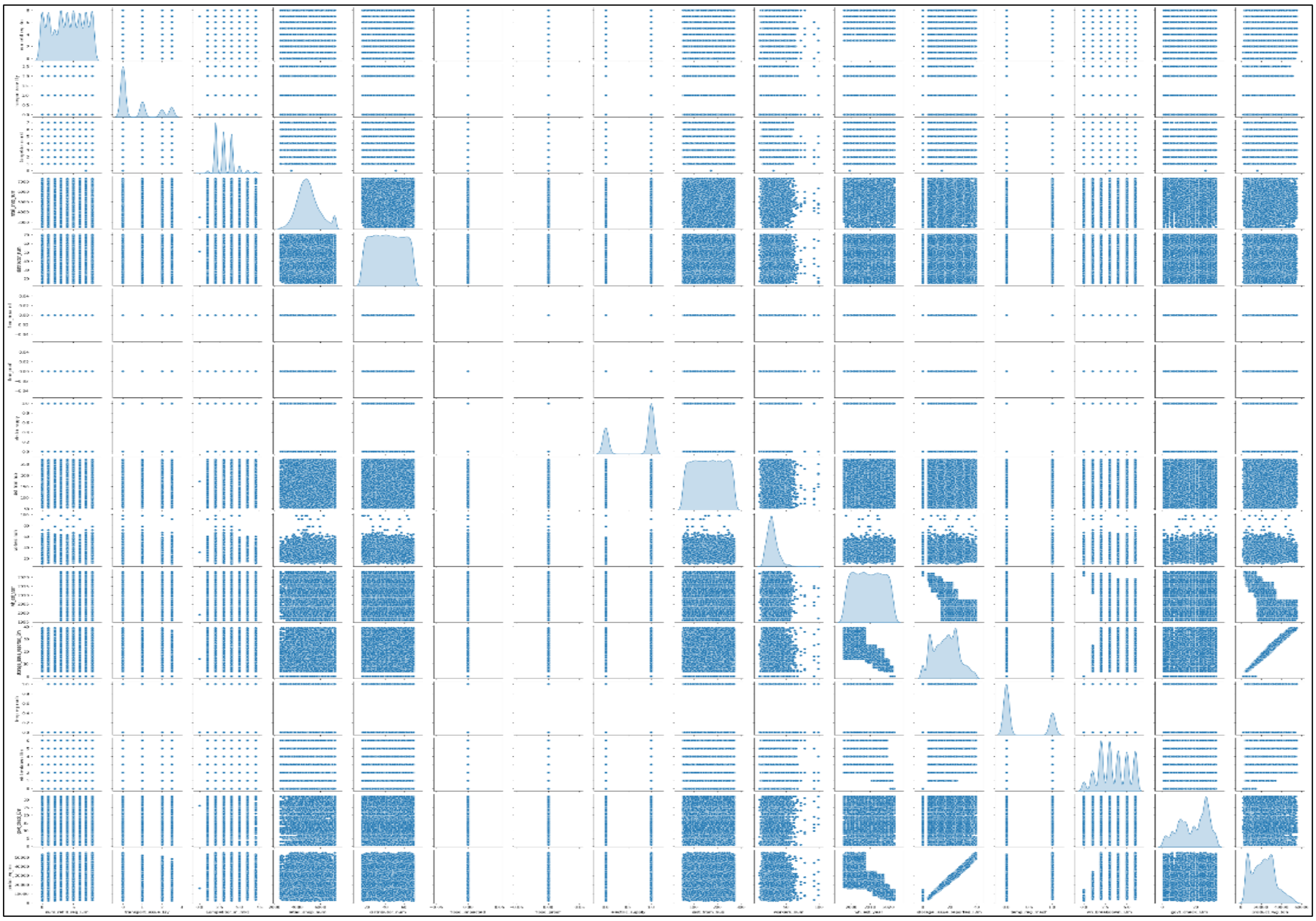


Figure 8: Pair plot analysis

3.3. Removal of unwanted variables:

There are two ID variables: Warehouse ID and Warehouse Manager ID. These columns have not been included in this analysis.

3.4. Missing Value treatment:

Checking the missing value among all variables	
Ware_house_ID	0
WH_Manager_ID	0
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

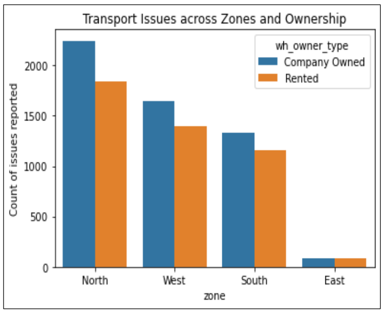
Applied the KNN imputer, to remove the missing values in the dataset	
num_refill_req_l3m	0
transport_issue_l1y	0
Competitor_in_mkt	0
retail_shop_num	0
distributor_num	0
flood_impacted	0
flood_proof	0
electric_supply	0
dist_from_hub	0
workers_num	0
wh_est_year	0
storage_issue_reported_l3m	0
temp_reg_mach	0
wh_breakdown_l3m	0
govt_check_l3m	0
product_wg_ton	0
Location_type_Urban	0
WH_capacity_size_Mid	0
WH_capacity_size_Small	0
zone_North	0
zone_South	0
zone_West	0
WH_regional_zone_Zone 2	0
WH_regional_zone_Zone 3	0
WH_regional_zone_Zone 4	0
WH_regional_zone_Zone 5	0
WH_regional_zone_Zone 6	0
wh_owner_type_Rented	0
approved_wh_govt_certificate_A+	0
approved_wh_govt_certificate_B	0
approved_wh_govt_certificate_B+	0
approved_wh_govt_certificate_C	0

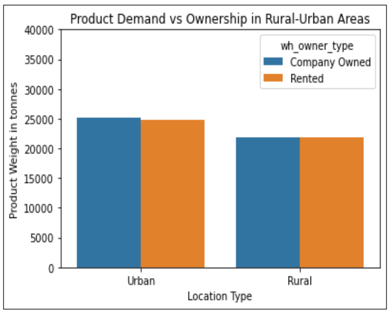
Table 2: Missing value treatment

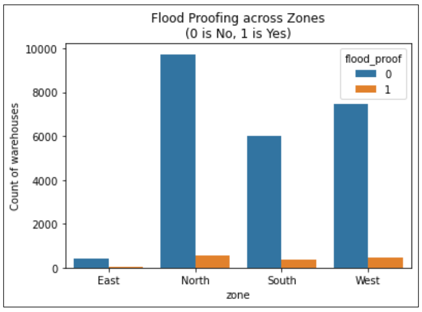
Observations:

- 'workers_num' variable has 990 missing values.
- 'wh_est_year' variable has 11881 missing values.
- 'approved_wh_govt_certificate' variable has 908 missing values.
- In total, 2.3% of the data has null values.
- Post KNN imputation method, all missing values have been removed.
- The number of nearest neighbours to be considered is kept at 1. However, as they are categorical in nature, these are converted to dummy variables eventually. The output after the KNN imputation is shown above (right side)

Additional observations:







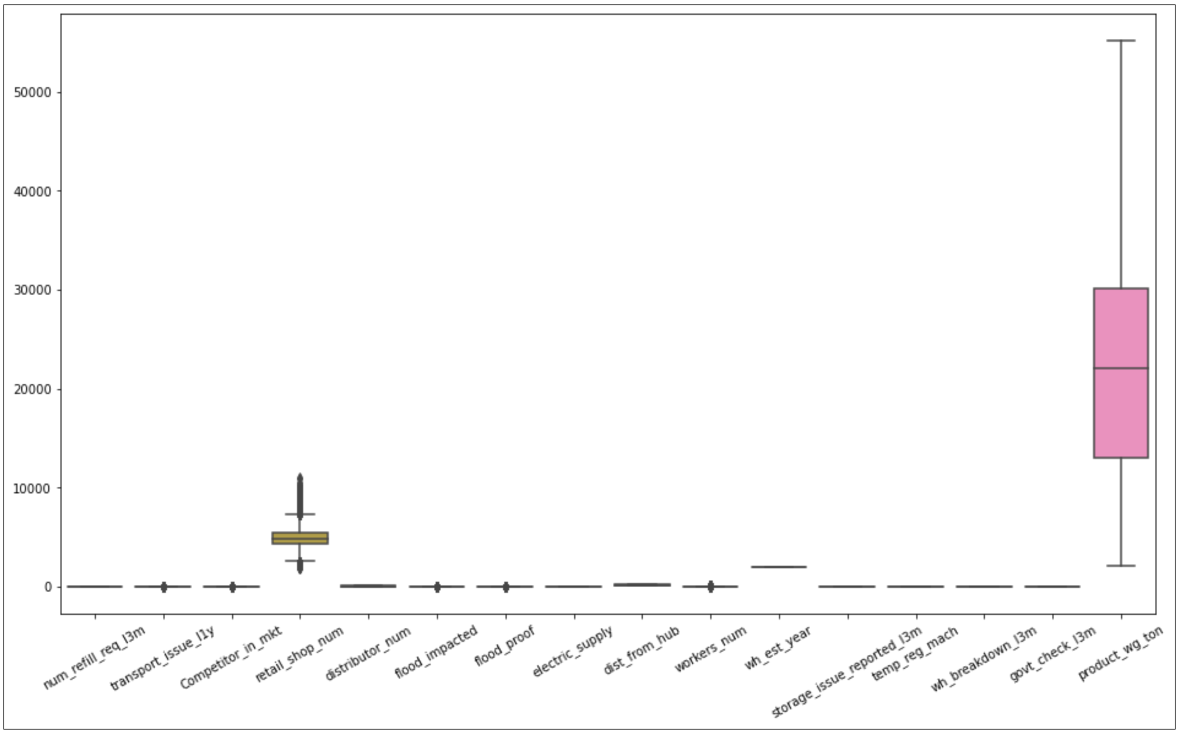
- ❖ The ownership of the warehouse seems to have an impact on the average product weight shipped, which also varies based on the location of the warehouse. The urban company-owned warehouses order more than the rented ones in the same location. In rural areas, this difference is negligible.
- ❖ A greater number of transport issues have been reported by company-owned warehouses, in comparison to the rented ones. This trend is consistent in all zones, except the East, were there is no difference.
- ❖ Flood proofing though a desirable facility at any storage facility, it can be observed that the warehouses impacted by floods form a mere minority. This justifies why warehouses have not opted for flood proofing.

3.5. Outlier treatment:

The descriptive analysis of the data has been followed by outlier detection and removal. This is limited to continuous variables. Below are the variables having outliers

- ➢ Competitor_in_mkt, retail_shop_num, transport_issue_l1y, workers_num, the outlier treatment is done in the further analysis.
- ➢ Flood_impacted & flood_proof outliers are not treated as they are nominal columns.

Before treating outliers



The IQR (Interquartile Range) method has been used to detect and remove outliers. IQR is the range between the first and the third quartiles namely Q1 and Q3: IQR = Q3 – Q1. The data points which fall below Q1 – 1.5 IQR or above Q3 + 1.5 IQR are outliers.

For the treatment of such outliers, the values lesser than lower bound are replaced with the value of lower bound, and the values greater than upper bound are replaced by the value of upper bound. Thus, the range of the values shrinks.

After treating outliers:

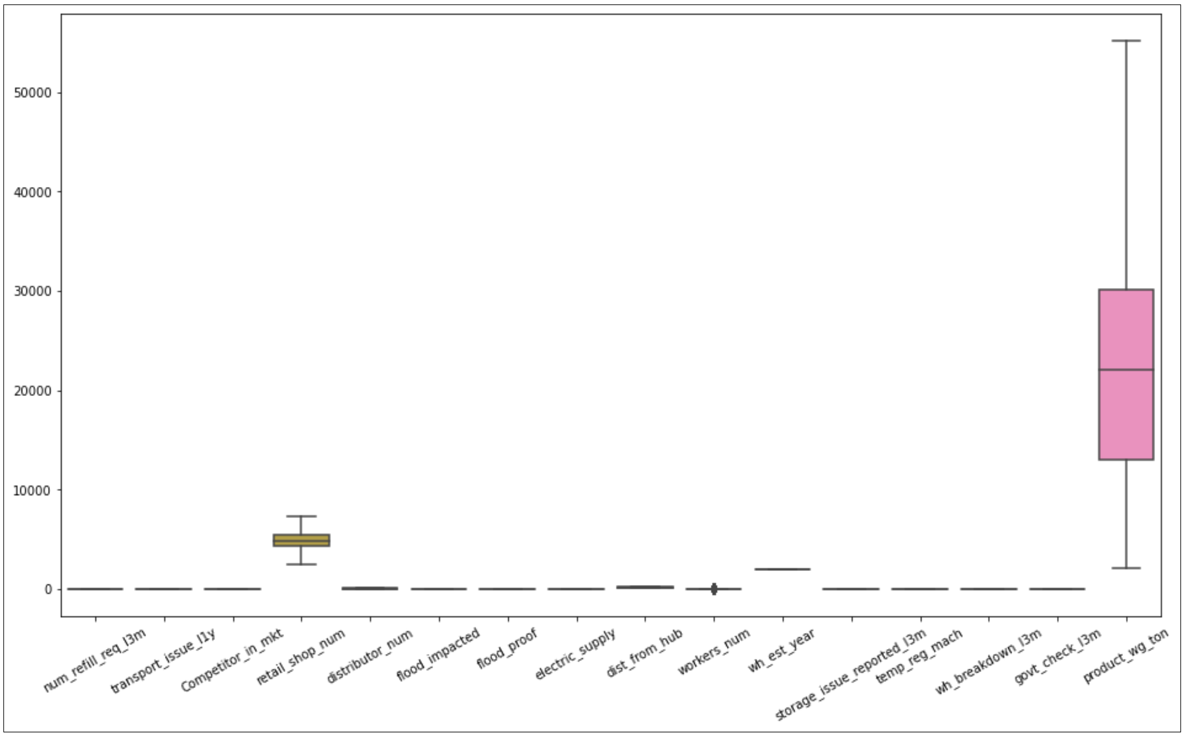


Figure 9: Outlier analysis

It can be observed that in the “product_wg_ton” variable, outliers could not be removed. However, there is no impact in the overall analysis.

Addition of new variables - Age group:

- Created a new variable, “Age group” by using imputation and lambda function
- This is used to create bins -> [0,5,10,15,20,25]
- Below is the distribution across the bins of the age group

0-5	5,155
5-10	4,720
10-15	5,789
15-20	4,411
20-25	4,263

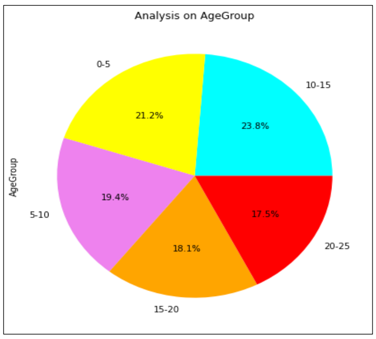


Figure 10: Additional of new variables

4. Business insights from EDA:

4.1. Data Imbalance and its Treatment:

Data imbalance is applicable to the classification problem. Since this is a Regression problem, we no need do any data Imbalance treatment here. Clustering is done for unsupervised data, when no historical data is given. Here in our study, we are given with the historical data and it is a supervised problem, hence clustering is not applicable.

4.2. Business insights:

A. Warehouse location:

- ❖ More warehouses are situated in the rural regions, than urban.
- ❖ There are unusually low number of warehouses in the East region. This region has more competition and more scope of growth. Thus, the marketing team can focus on this region. The legal team needs to figure out why this place gets more government checks than other regions.

B. Warehouse features:

- ❖ One third of the warehouses have no electricity and hence this needs close monitoring during product storage to avoid any perishability.
- ❖ A minority have opted for flood proofing and temperature regulation. The ones with a temperature regulator, report more breakdowns and hence the quality of these machines should be improved.
- ❖ The warehouses with A+ certifications have on average, greater demand for products.

C. Warehouse ownership:

- ❖ The ownership of the warehouse seems to have an impact on the average product weight shipped, which also varies based on the warehouse location. The urban company owned warehouses order more than the rented ones in the same location.
- ❖ A greater number of transportation issues have been reported by company-owned warehouses, in comparison to the rented ones. This trend is consistent across all zones, except the East, were there is no difference. This could be an additional area of improvement.