SUPPLY CHAIN CAPSTHONE PROJECT

FINAL REPORT

Business Problem:

A FMCG company has entered 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

By: Yogender Singh

Post-Graduation in Data Sci & Business Analysis Mar`22

Contents

1. Introduction

1.1 Brief introduction about the problem statement and the need of solving it.

ANS:

Business Problem:

A FMCG company has entered 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 vice versa. 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 warehouse in entire country.

Goal & Objective:

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.

Also try to analysis the demand pattern in different pockets of the country so management can drive the advertisement campaign particular in those pockets.¶

Benefits:

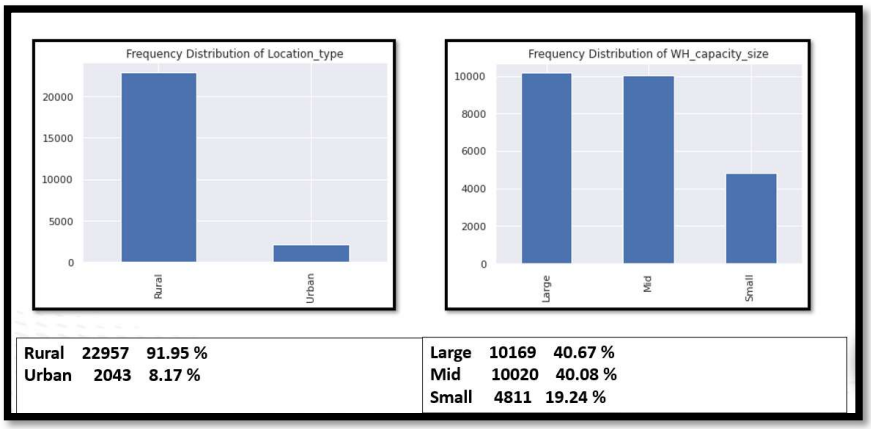
1. Demands of customer can be fulfilled at right time.
2. Noodles wastage can be reduced.
3. Unwanted manpower can be removed.
4. Better monitoring can be done.
5. Business expansion can be done.

2. EDA and Business Implication

2.1Uni-variate / Bi-variate / multi-variate analysis to understand relationship b/w variables. How is your analysis impacting the business?

ANS:

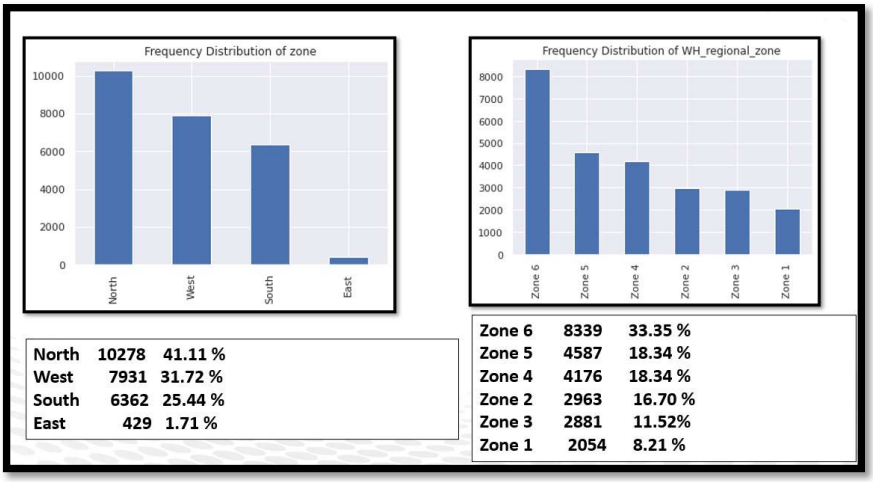
Uni-variate Analysis:



Rural	22957	91.95 %
Urban	2043	8.17 %

Large	10169	40.67 %
Mid	10020	40.08 %
Small	4811	19.24 %

Figure 1:Univariate Analysis: Location Type & Warehouse Capacity Size



North	10278	41.11 %
West	7931	31.72 %
South	6362	25.44 %
East	429	1.71 %

Zone 6	8339	33.35 %
Zone 5	4587	18.34 %
Zone 4	4176	18.34 %
Zone 2	2963	16.70 %
Zone 3	2881	11.52%
Zone 1	2054	8.21 %

Figure 2:Univariate Analysis: Zone & WH_Regional_Zone

Bi variate Analysis:

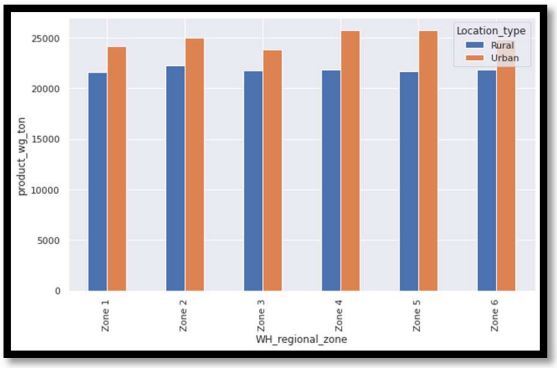


Figure 3 Product Wt Tone VS WH Reginal Zone

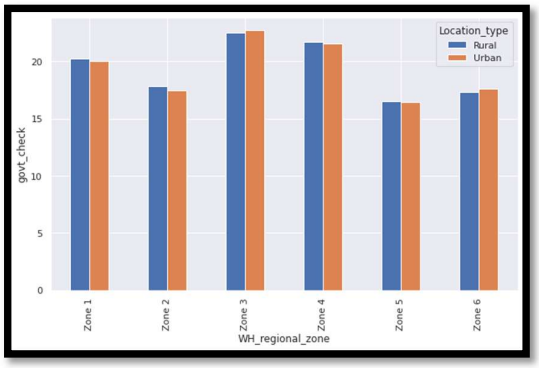


Figure 4 Govt Check VS WH Reginal Zone

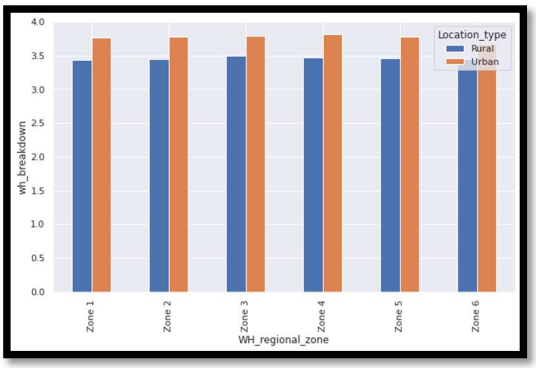


Figure 5 wh breakdown VS WH Reginal Zone

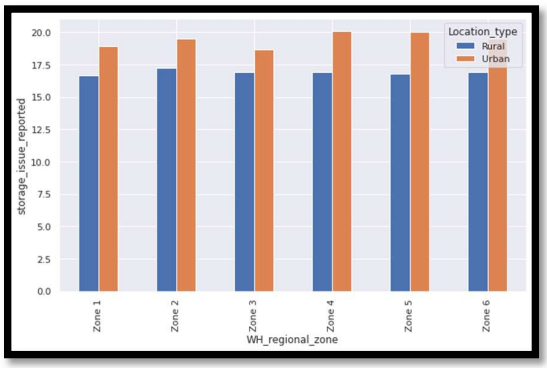


Figure 6:Storage Issue Reported VS WH Reginal Zone

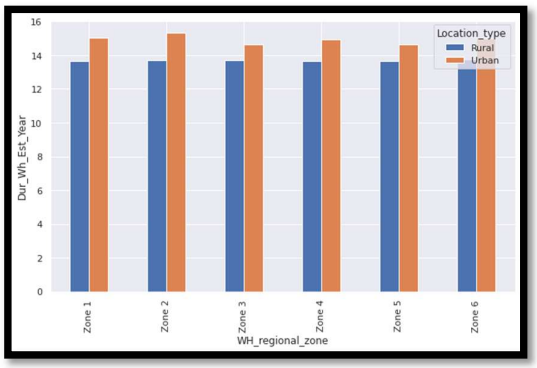


Figure 7 Dur Wh Est Year VS WH Reginal Zone

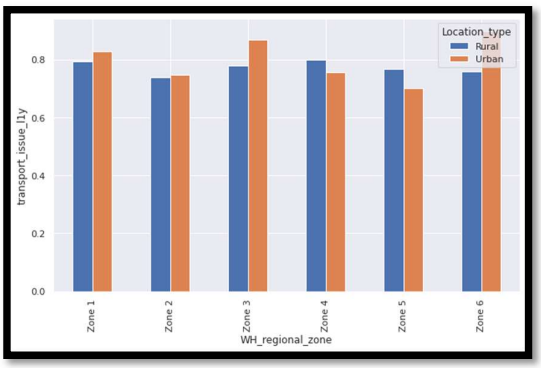


Figure 8 Transport Issues VS WH Reginal Zone

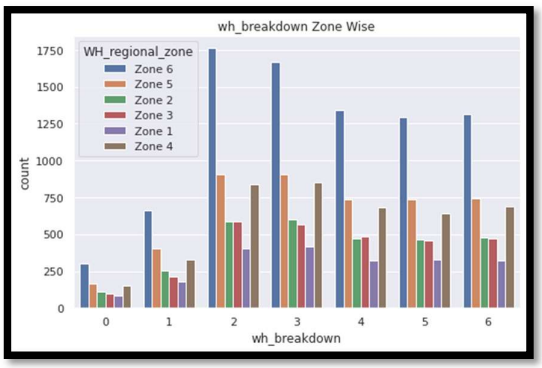


Figure 9 wh_breakdown with WH Reginal Zone

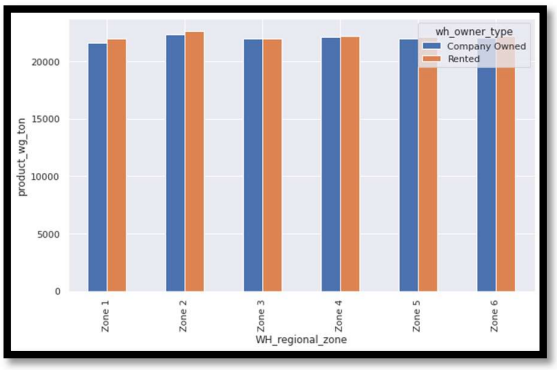


Figure 10 : Product Wt. Tone VS WH Reginal Zone

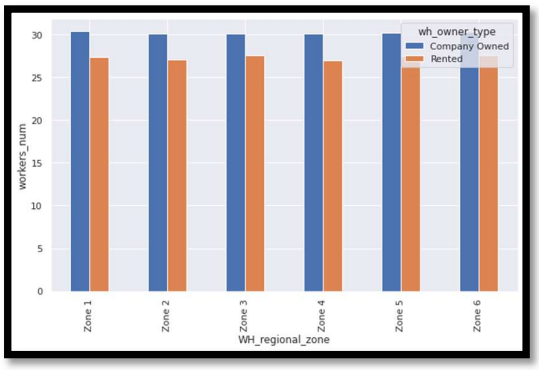
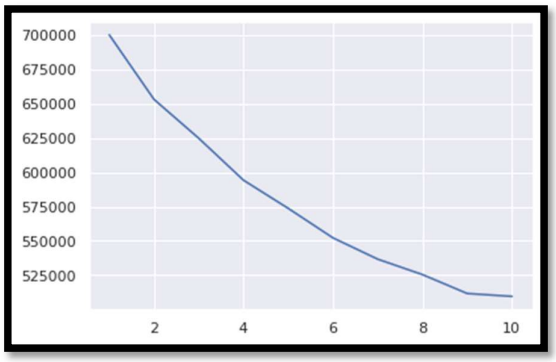


Figure 11 Worker Num. VS WH Reginal Zone

After K_Means Clustering

Two Cluster formed.

1. **Cluster_0 –High Supply - Mean (Target) Product Wt. Tone 30651**
2. **Cluster_1 –Low Supply - Mean (Target) Product Wt. Tone 12459**



Silhouette_score = 0.065 With WSS Plot

Bi variate Analysis:

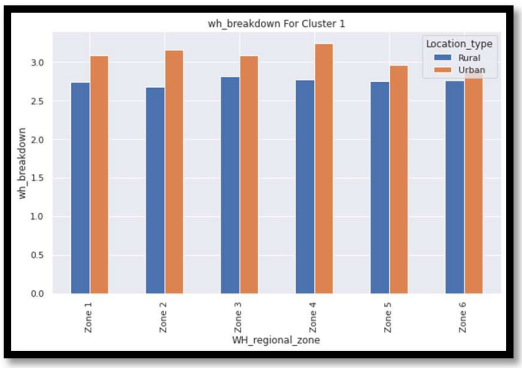


Figure 12:wh breakdown VS WH Reginal Zone

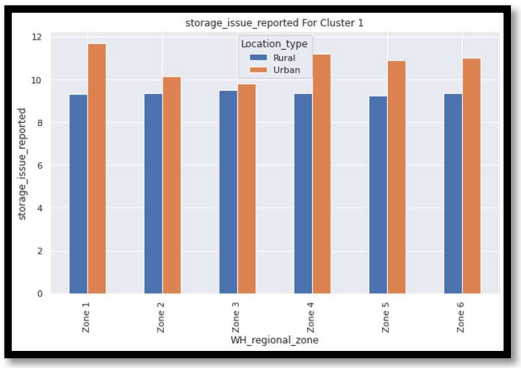
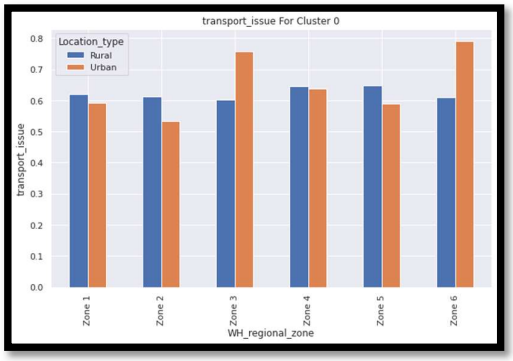


Figure 13 Storage Issue Reported VS WH Reginal Zone



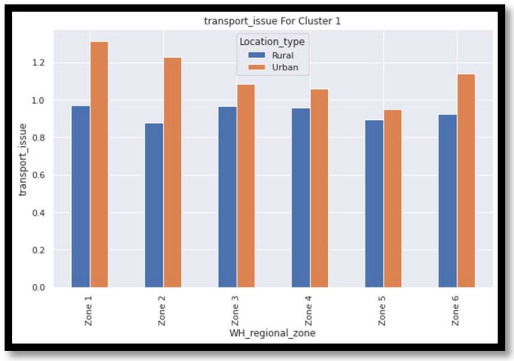


Figure 14 Transport issue VS WH Reginal Zone

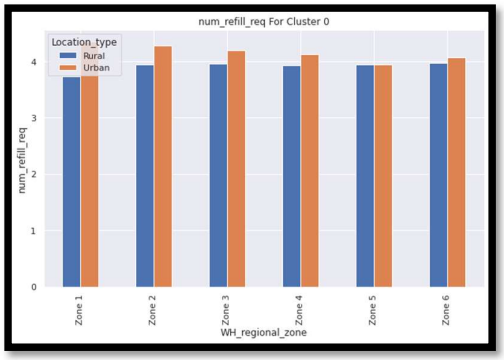


Figure 15 num refill req. VS WH Reginal Zone

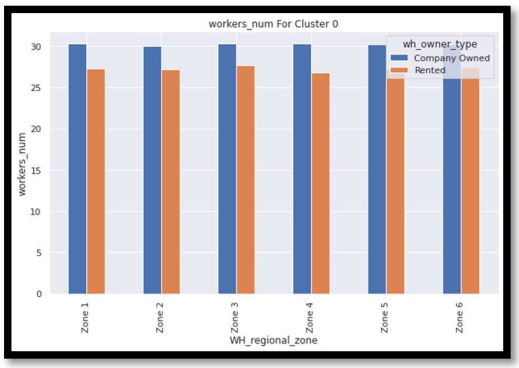


Figure 16workers_num VS WH Reginal Zone

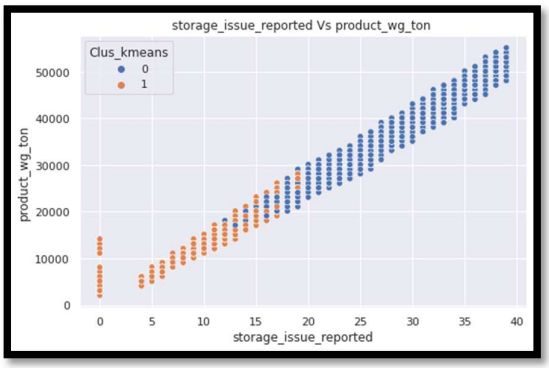


Figure 17:Product wg ton vs Storage Issue Reported

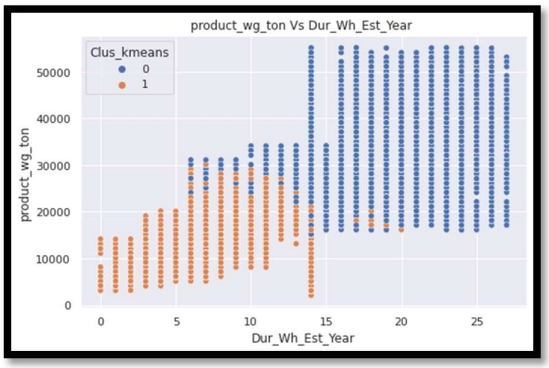


Figure 18 Product wg ton vs dur est year

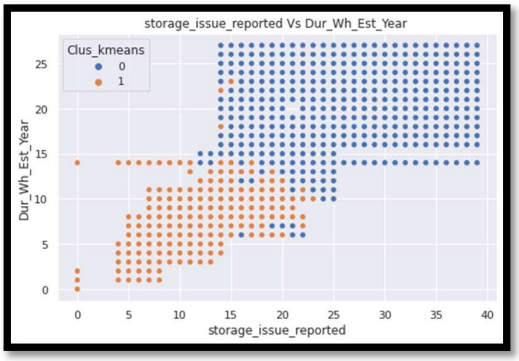


Figure 19 Storage Issue Reported vs Dur est year

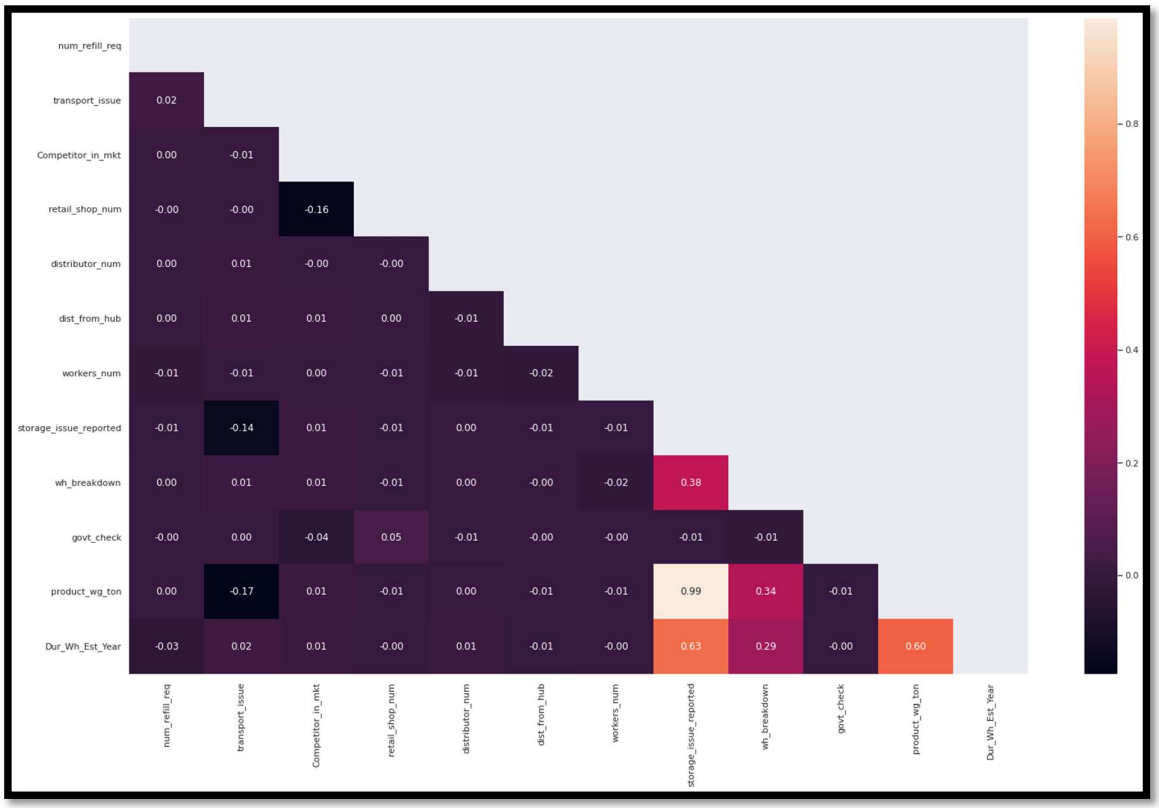
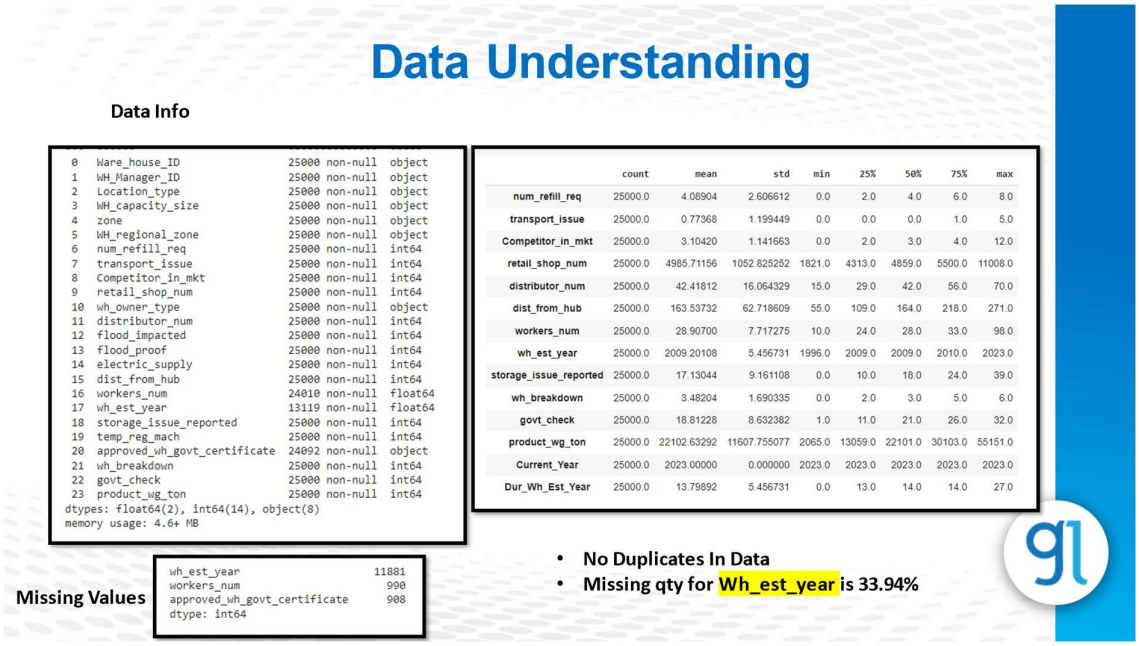


Figure 20 Heat Map

2.2Both visual and non-visual understanding of the data.

ANS:



3. Data Cleaning and Pre-processing

3.1Approach used for identifying and treating missing values and outlier treatment (and why)

ANS:

1. Missing values are identified by info. Of data frame.
2. Missing Numerical Values are replaced by Median.
3. Missing categorical Values are replaced by Mode.
4. Outliers are treated otherwise it will impact our model performance.
5. If the outliers are not treated in the first step while doing the exploratory data analysis, it can lead to biases in the results.

3.2Need for variable transformation (if any)

ANS:

Yes, There is need of variable transformation whose skewness lies outside range [-0.5 to 0.5] to bring data under bell shape

index	Skewness
transport_issue	1.610
workers_num	1.090
Competitor_in_mkt	0.970
retail_shop_num	0.908
product_wg_ton	0.330
storage_issue_reported	0.110
distributor_num	0.015
dist_from_hub	-0.006
num_refill_req	-0.075
Dur_Wh_Est_Year	-0.117
govt_check	-0.363

3.3Variables removed or added and why (if any)

ANS:

Removed: Wh_Est_Year

Added: Dur_Wh_Est_Year

Until business says to remove feature then a data analysist should not remove feature otherwise treat missing values & treat outliers…..& perform EDA.

4. Model building

4.1Clear on why a particular model(s) was chosen.

ANS:

Modelling Approach Used & Why

A). Model building and interpretation

a. Build various models The models are built to establish the relationship between the Target variable and predictor variables.
Target Variable: product _ wg _ton is the target variable in these models

Data spliced in train & test in 70 & 30 Ratio
Four Models are built:

1. Linear Regression Model
2. Decision Tree –Simple Model
3. Random Forest -(Ensemble Bagging Technique)
4. Artificial Neural Network –Supplied scaled data

Modelling Approach Used & Why

A). Linear Regression Model
Assumption : Target variable should be normal distributed

- If skewness is in range of -0.5 to 0.5 then Feature can be assumed well normal distributed.
- For linear reg Target variable should have linear relationship with independent feature
- Using ANOVA : Three categorical variable has no relationship with target variable.

Skewness

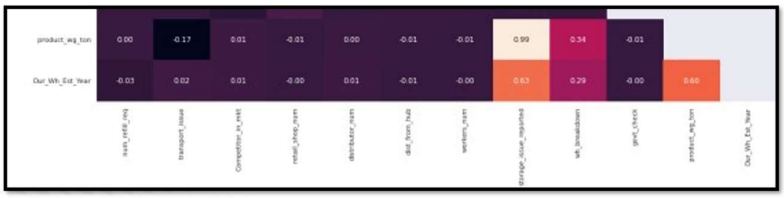
	Skewness
product_wg_ton	0.331611
Dur_Wh_Est_Year	-0.117099
govt_check	-0.363240
storage_issue_reported	0.113338
workers_num	1.095016
dist_from_hub	-0.005998
distributor_num	0.015212
retail_shop_num	0.908247
Competitor_in_mkt	0.978397
transport_issue	1.610810
num_refill_req	-0.075212

1. Temp reg Mach
2. App wh govt certificate
3. Location type

Modelling Approach Used & Why

A). For continuous variable

- Target variable has more dependence on storage issues reported & wh_brokedown



4.2Effort to improve model performance.

ANS:

Hyperparameter Tunning

A).Hyperparameter tuning is not possible for linear regression other model can use hyperparameter tuning.

B)GridsearchCV is used in knowing best parameter

Decision Tree Best Parameter:

{'max_depth': 10, 'min_samples_leaf ': 15, 'min_samples_split ': 40}

Random Forest Tree Best Parameter:

Decision Tree Best Parameter:

{'max_depth': 10, 'max_features': 6, 'min_samples_leaf ': 3, 'min_samples_split ': 30, 'n_estimators': 300}

Artificial Neural Network Best Parameter:

{'activation': 'relu', 'hidden_layer_sizes ': 100, 'solver': 'adam'}

Best Model: Random Forest Regressor

Before Hyperparameter Tuning

	Train RMSE	Test RMSE	Training Score	Test Score
Linear Regression	1744.605000	1760.647905	0.977398	0.977021
Decision Tree Regressor	0.000000	1250.725933	1.000000	0.988404
Random Forest Regressor	342.542504	931.400962	0.999129	0.993569
ANN Regressor	1072.283522	1157.639978	0.991462	0.990066

	Taining MAPE	Test MAPE
Linear Regression	8.906356	9.149702
Decision Tree Regressor	0.000000	5.093023
Random Forest Regressor	1.542697	4.270517
ANN Regressor	5.167553	5.654945

After Hyperparameter Tuning

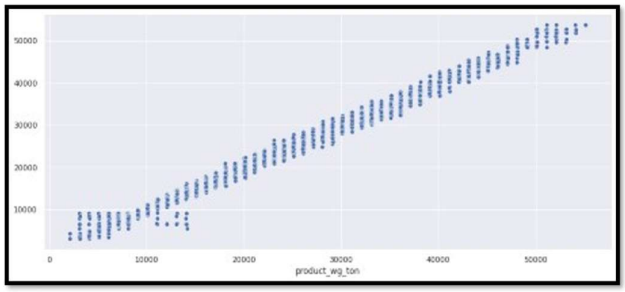
	Train RMSE	Test RMSE	Training Score	Test Score
Linear Regression	1744.605000	1760.647905	0.977398	0.977021
Decision Tree Regressor	847.965507	954.163756	0.994660	0.993251
Random Forest Regressor	2110.658138	2282.903578	0.966918	0.961367
ANN Regressor	1228.992631	1289.817158	0.988784	0.987668

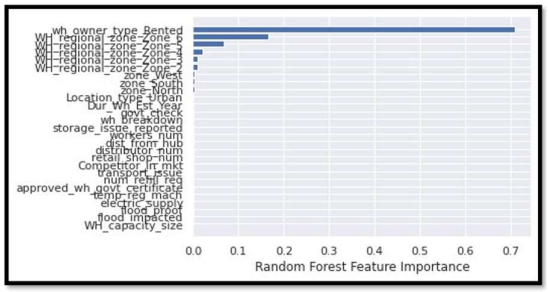
	Taining MAPE	Test MAPE
Linear Regression	8.906356	9.149702
Decision Tree Regressor	3.781706	4.338371
Random Forest Regressor	10.927756	11.831595
ANN Regressor	5.829211	6.198369

Test Parameter : RFR	Before Hyperparameter Tuning : RFR	After Hyperparameter Tuning : RFR
MAPE	4.27	11.83
Accuracy	99.35	96.13
RMSE	931.4	2282.90

Note: Because of less computation power of laptop, unable to perform proper hyperparameter tuning.

Pred V/s Actual For Best Model :RFR





5. Model validation

5.1How was the model validated? Just accuracy, or anything else too?

ANS:

Models are validated by checking three parameters:

1.MAPE

2.R2 or Score

3.RMSE

Thumb Rule: All above parameters should not have more/less than 10% difference in train & test.

Overfit: Train above parameters > 10% of Test above parameters

Underfit: Train above parameters < 10% of Test above parameters

6. Final interpretation / recommendation

6.1Detailed recommendations for the management/client based on the analysis done.

ANS:

Recommendations based on best model important features:

1. Wh_owner_type_rented.
2. Wh_regional_zone_6
3. Wh_regional_zone_5
4. Wh_regional_zone_4

Insight 1. 91.25 % warehouses are in rural areas in comparison of rural area has 8.17%.

Insight 2. 82 % warehouses are of large & mid type.

Insight 3: North zone & Zone 6 has majority of warehouses.

Insight 4:Avg Product wt tone is supplied more to urban area in comparison of rural area

Recommendation 1: Product wt tone should be supplied to be increased for rural area. Those warehouses are in urban areas should managed properly to cater more customer demand.

Insight 5: Govt checks are more in zone 3 & zone 4.

Recommendation 2:In zone 3 & zone 4,all documents should be maintained properly for audit purposes to avoid unnecessary breakdown.

Insight 6: Warehouse breakdown are more in urban zone than rural zone.

Recommendation 3: Strike from worker, flood, or electrical failure should be taken care properly by respective managers to reduce warehouse breakdowns.

Insight 7: Older warehouses are more in urban area comparison of rural area.

Recommendation 4: Renovation of roofs, pallets and others broken racks etc should be done to avoid unnecessary loss of product.

Insight 8: Transport issues are more in urban areas in comparison of rural area.

Recommendation 5: Latest technology could be used like drone to avoid delays.

Insight 9:Wh_Regional _Zone 6 has most number of breakdowns than 5 followed by zone 4.

Recommendation 6: In zone 6 specially ,electricity, worker strike & flood failure should be checked & planned in advance to avoid these failure.

Insight 10:Company owned warehouses has more numbers of workers than rented whereas supply is same.

Recommendation 7:HR manager should refer rented warehouse system as benchmark for company owned warehouse.

Insight 11:No of refills are more in urban areas than rural areas.

Recommendation 8: Transportation system of urban areas must be improved.

Insight 12: Storage issues are more where product is stored more.

Recommendation 9: Large & medium size specially older warehouses warehouses renovation must be done properly.