|  |  |
| --- | --- |
| A picture of a winding road and trees  **[Capstone Project Notes -1**]  [**Supply Chain Management**  **(2022-2023)**  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**

6

|  |  |  |
| --- | --- | --- |
| **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 |

**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.

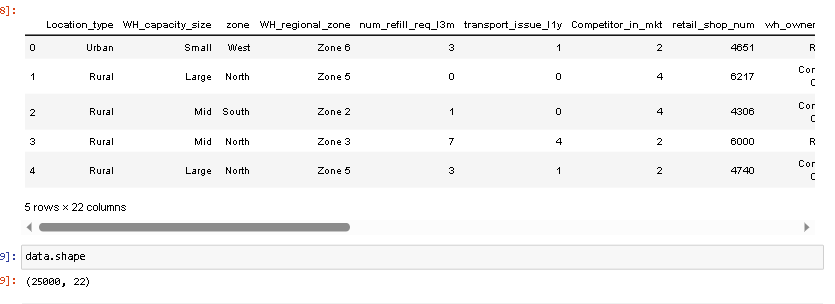
Noodles. The goal of the management is to maximise supply in each and every warehouse. across the entire nation. This project's aim is to create a model utilising historical data that will Identify the ideal product weight that should be sent each time to the warehouse. PORDUCT\_WG\_TON is the target 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.

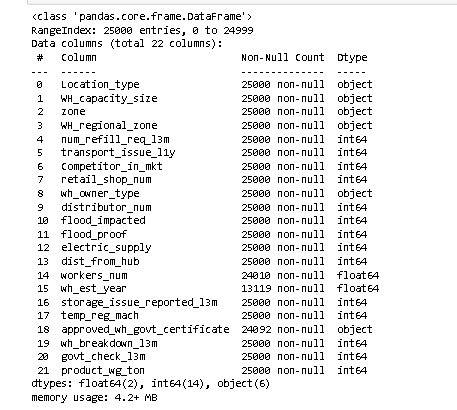
This research compares various well-known machine learning classifiers and evaluates their effectiveness to determine

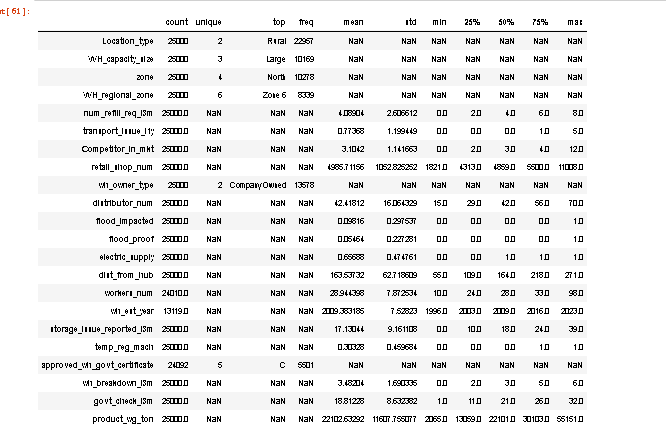
**B. Need of the study/project**

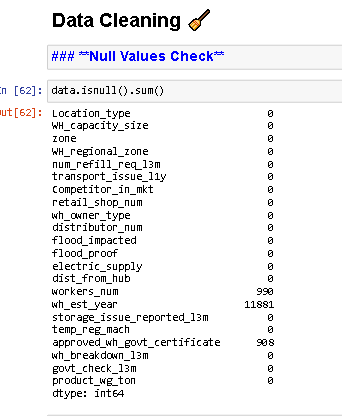
* **Objective** - To predict the weight of products of a FMCG company for various warehouses with different conditions, size & locality. To determine the Ideal Quantity of Product Weight Shipped to the various Ware Houses of FMCG Instant Noodles Company in order to reduce wastage of the Product, Bridge 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.

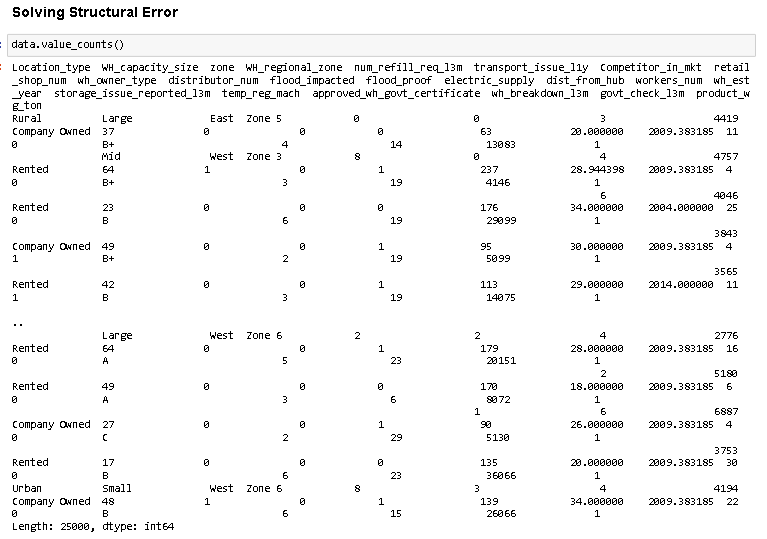
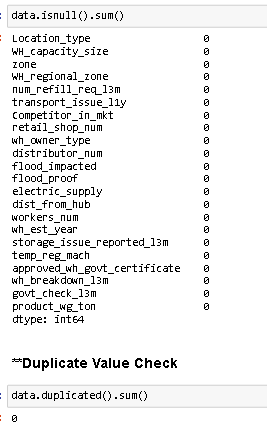
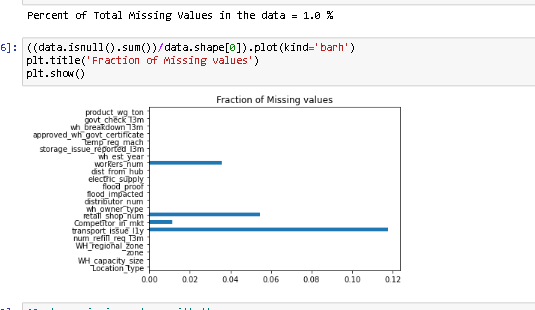
**2. Data Report**

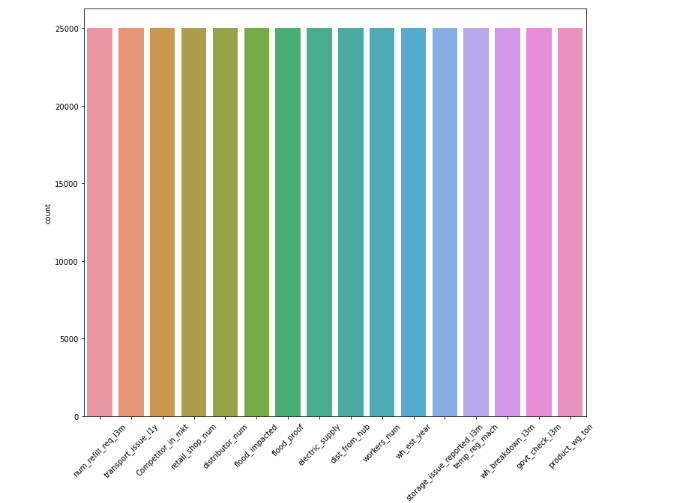


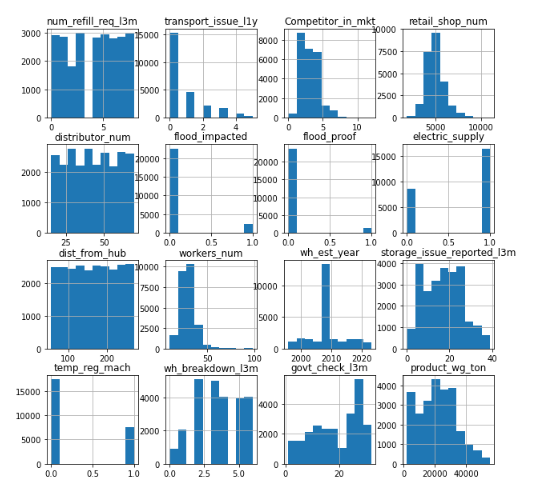




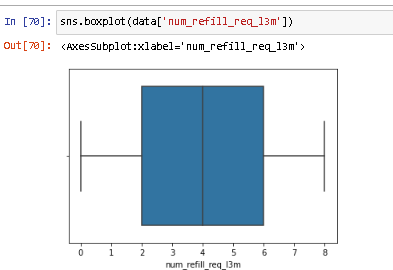


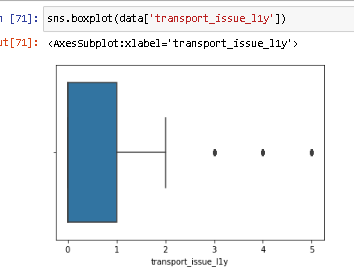
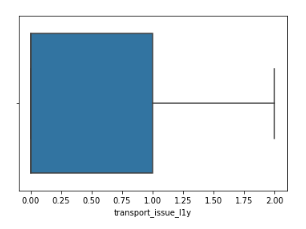




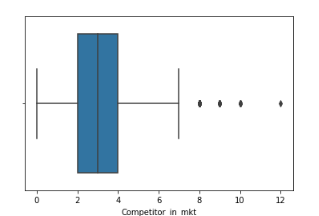
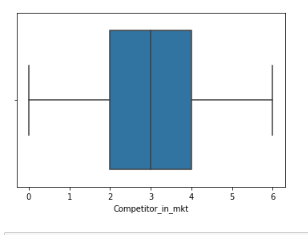
**Checking data distribution****n**

### **Outlier Management**

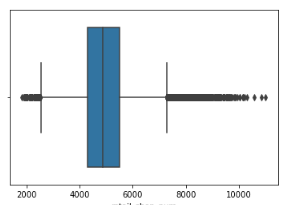
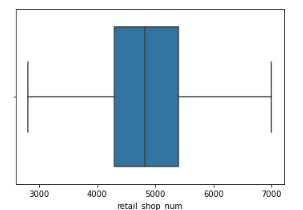


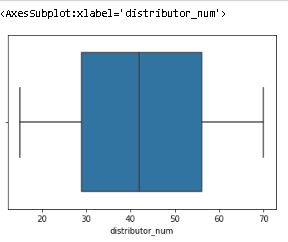
 

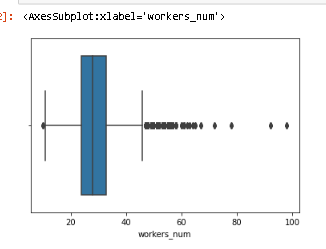
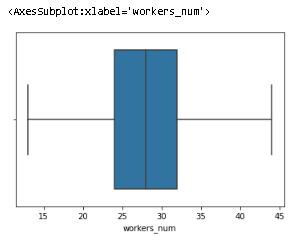
**'Competitor\_in\_mkt**

'retail\_shop\_num'

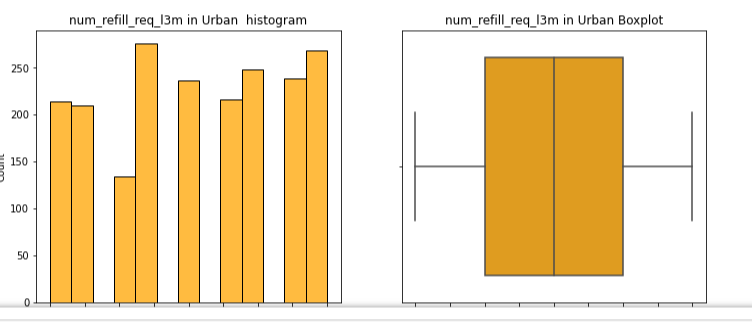
 

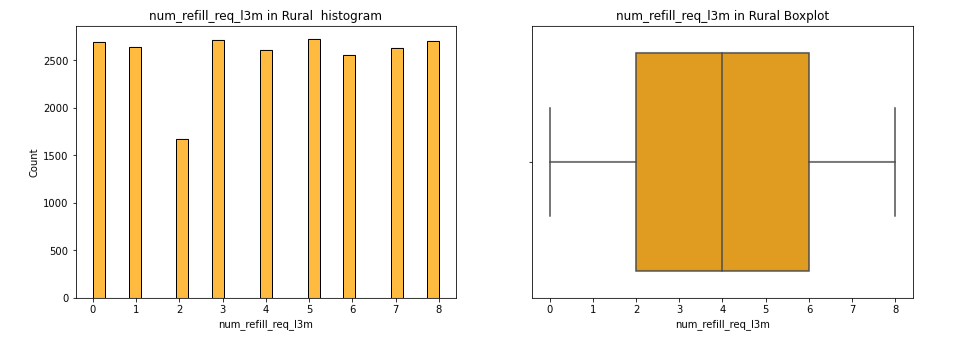


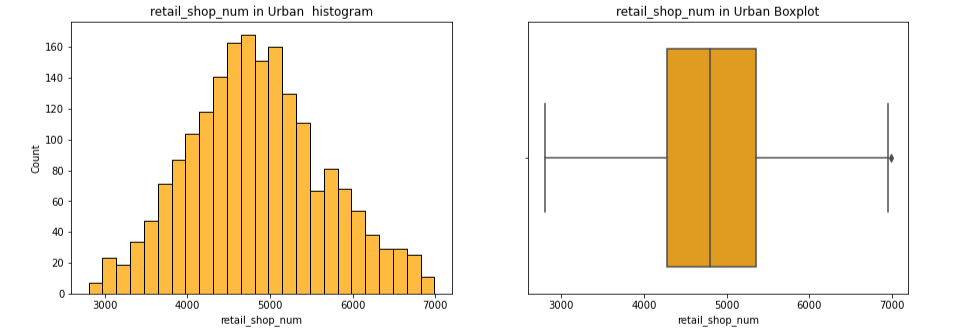
 

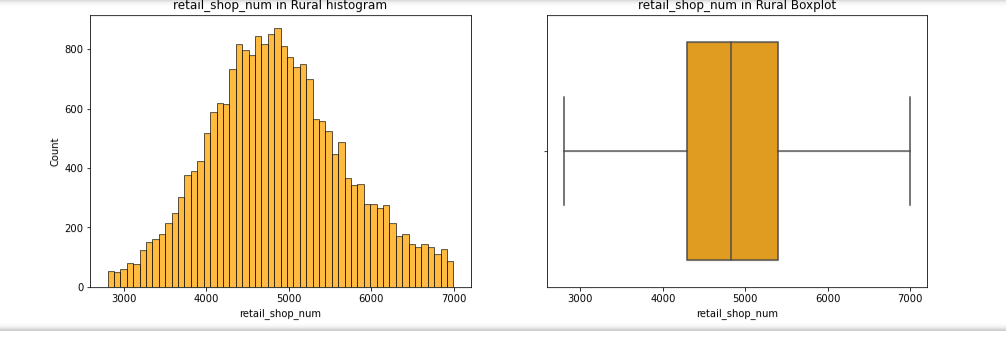
**3) Exploratory Data Analysis**

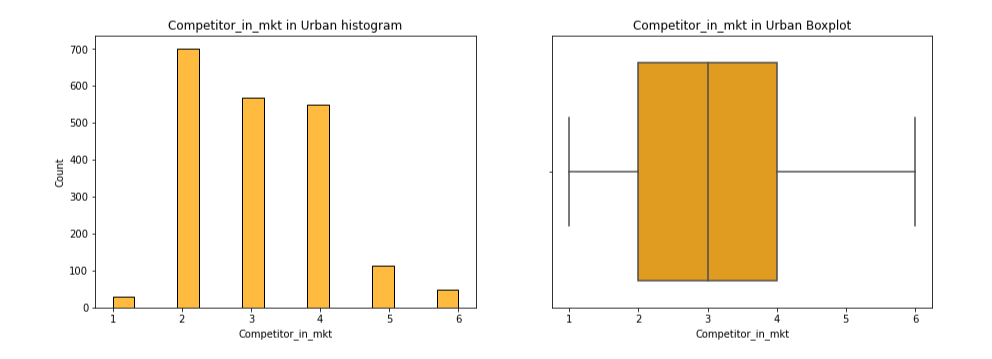
# Univariate Analysis

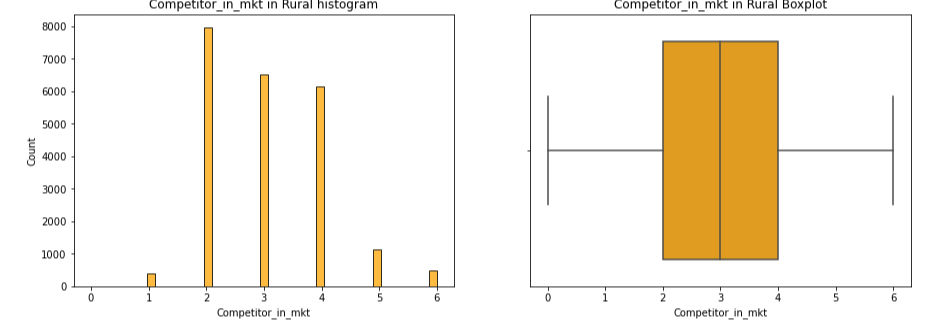


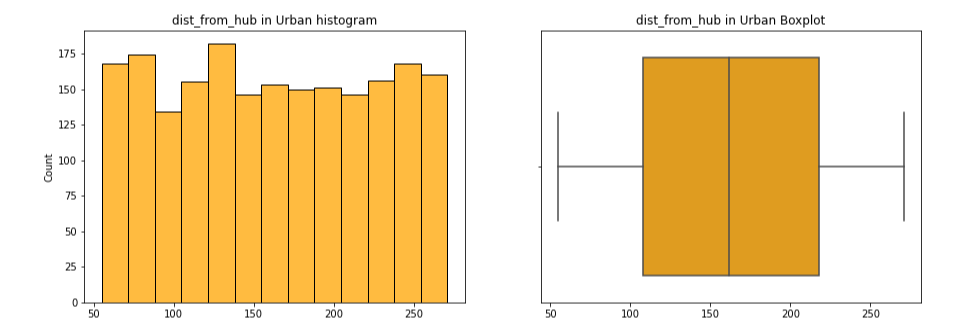


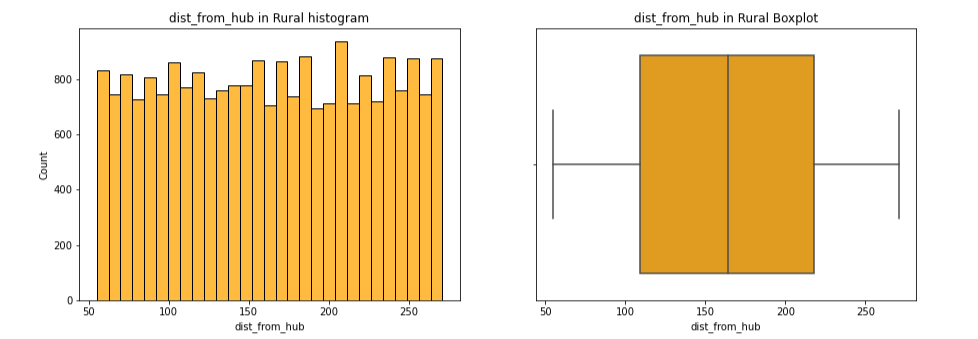


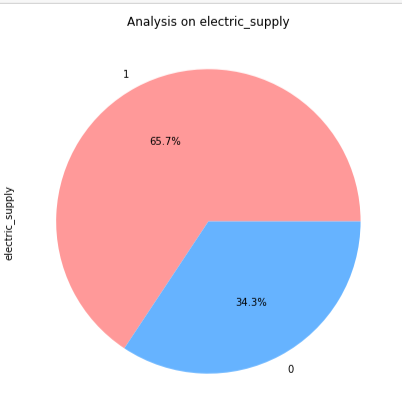
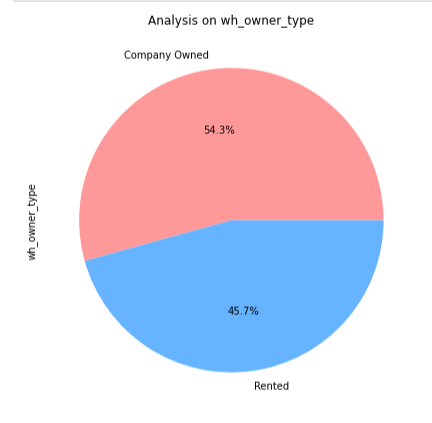




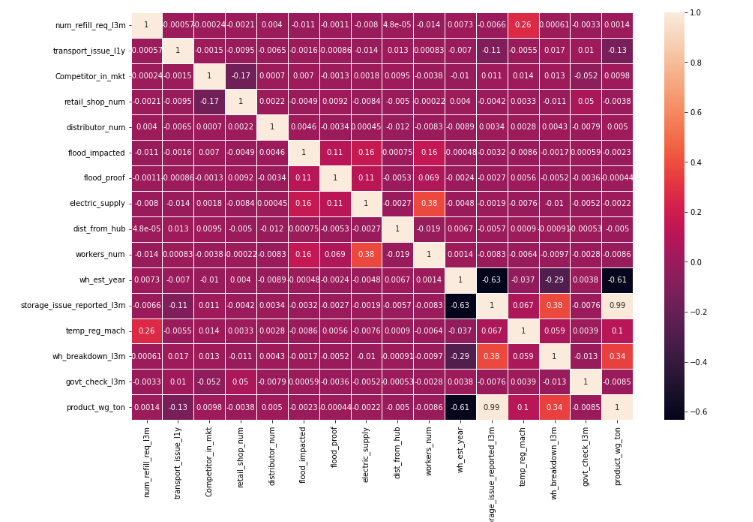


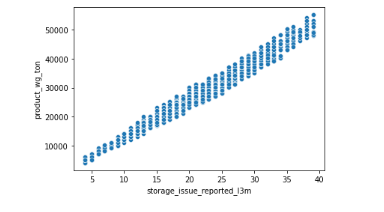


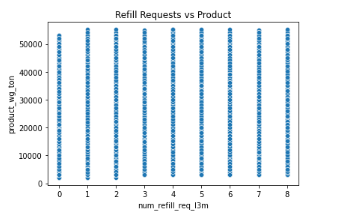


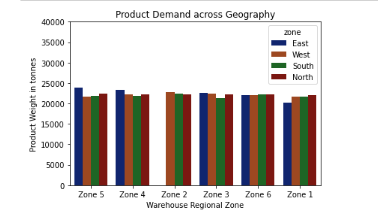
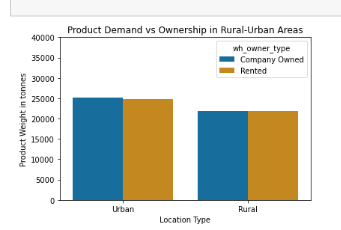
 

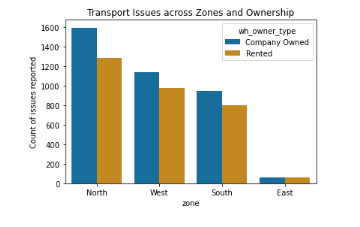
**Bivariate Analysis**

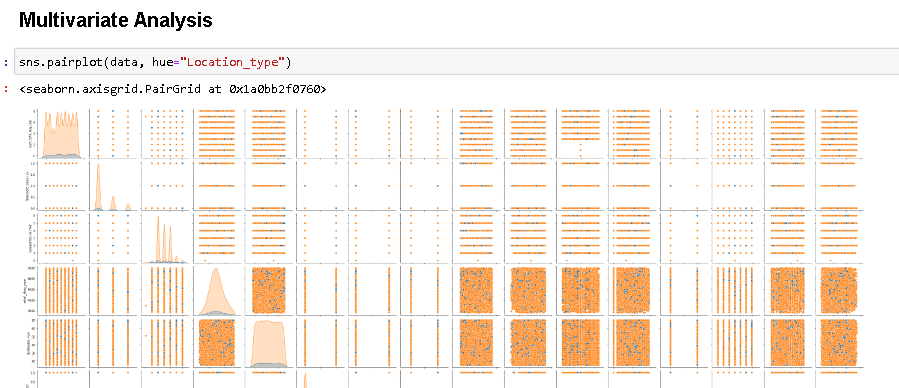


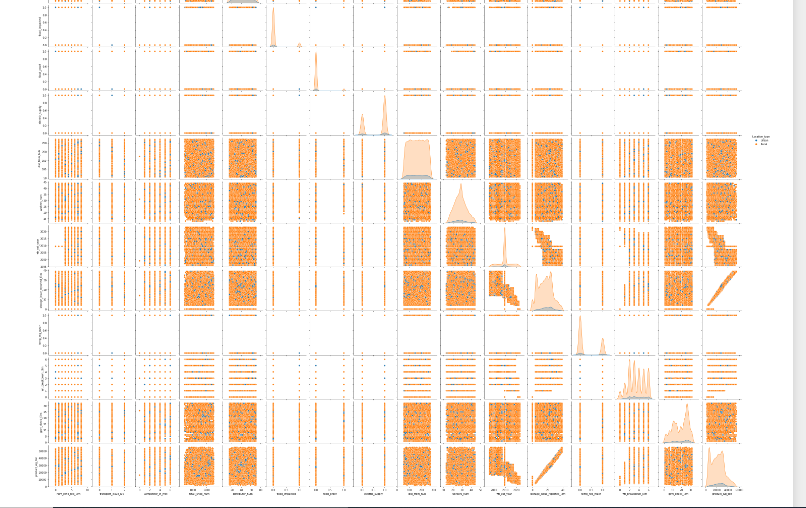










**Business Insights**

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.