Big Mart Sales

Prediction

Big Mart Sales Prediction Analysis

Submitted by : Suresh Chaudhari

Background & Objectives

The Big Mart is an International Brand. Big Mart Brand has started its journey from America USA, 2007. Big Mart is a departmental and convenience store retail chain.

They stock an expansive range of daily need items including groceries, candies, personal care products, soft drinks, ready-to-eat food, ice-cream etc.

They started small with a single store in April 2007, today Big Mart has become a well-known name in the retail industry with more than 40 outlets in world.

Objective :-

The aim is to build a predictive model and predict the sales of each product at a particular outlet.

Using this model, BigMart will try to understand the properties of products and outlets which play a key role in increasing sales.

This report aims at identifying:

- The key factors affecting the sales at outlets.
- Identify the properties of products which can play key role in improving the sales.
- Identify the properties of outlets which can play key role in improving the sales.
- Machine Learning algorithm to predict sales of each product at a particular outlet.

Let's take a look at the available dataset & additional data requirement :

DATA OVERVIEW

- Training dataset contains the 12 features with 8523 observations.
- Testing dataset contains the 11 features with 5681 observations.

| Variable	Description
Item_Identifier	Unique product ID
Item_Weight	Weight of product
Item_Fat_Content	Whether the product is low fat or not
Item_Visibility	The % of total display area of all products in a store allocated to the particular product
Item_Type	The category to which the product belongs
Item_MRP	Maximum Retail Price (list price) of the product
Outlet_Identifier	Unique store ID
Outlet_Establishment_Year	The year in which store was established
Outlet_Size	The size of the store in terms of ground area covered
Outlet_Location_Type	The type of city in which the store is located
Outlet_Type	Whether the outlet is just a grocery store or some sort of supermarket
Item_Outlet_Sales	Sales of the product in the particular store. This is the outcome variable to be predicted.

DATA

From the BigMart perspective, the following additional data would prove to be important for analysis on sales of outlets:

- Price of each individual product.

- Outlet Timings in particular Area.

- Types of customers visited.

- Date of purchase.

- Offers on the floor with dates/season.

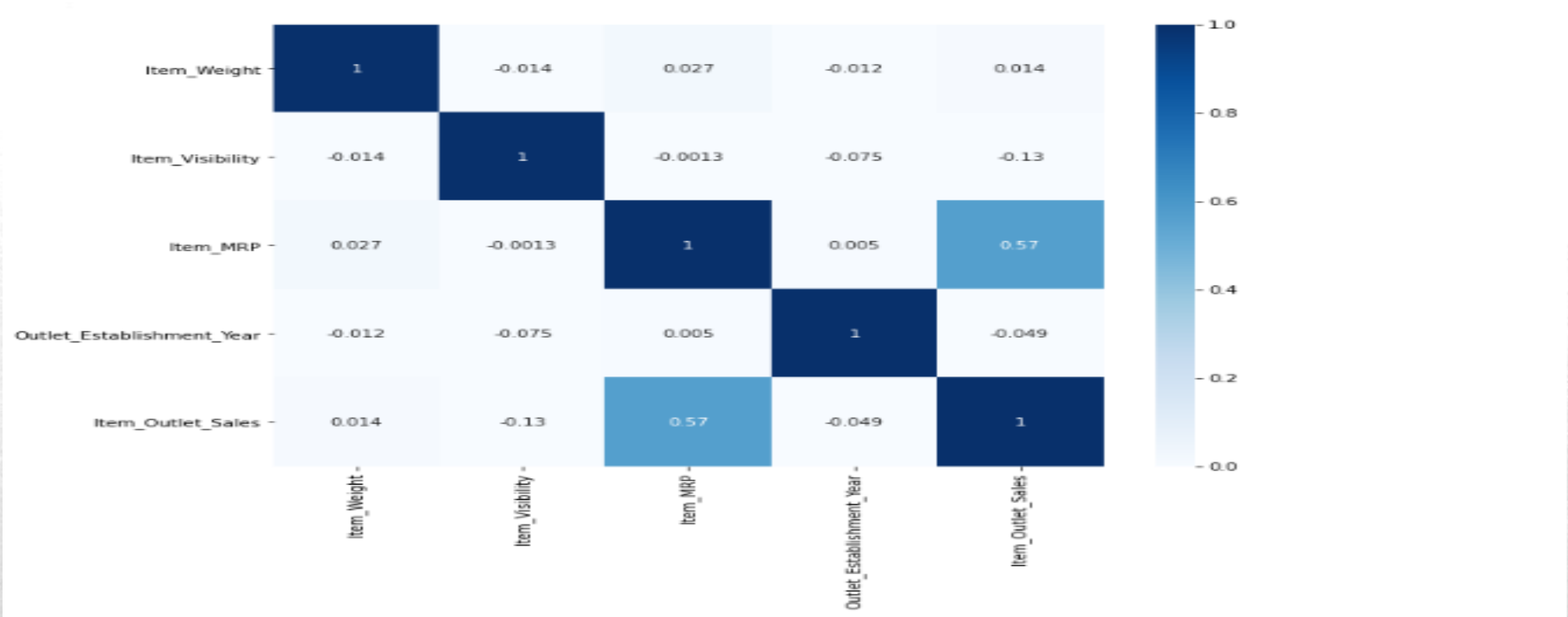# Let's take a look at EDA part :

EXPLORATORY DATA ANALYSIS
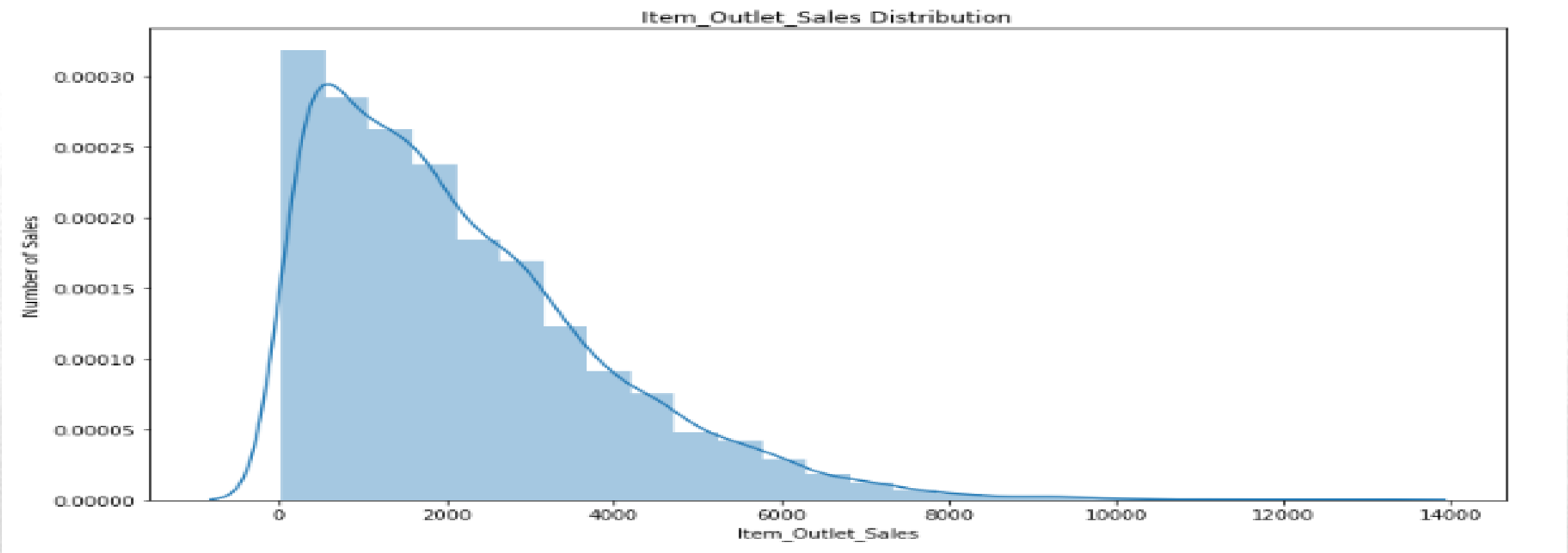
We can see that our target variable is skewed to the right, towards the higher sales, with higher concentration on lower sales.

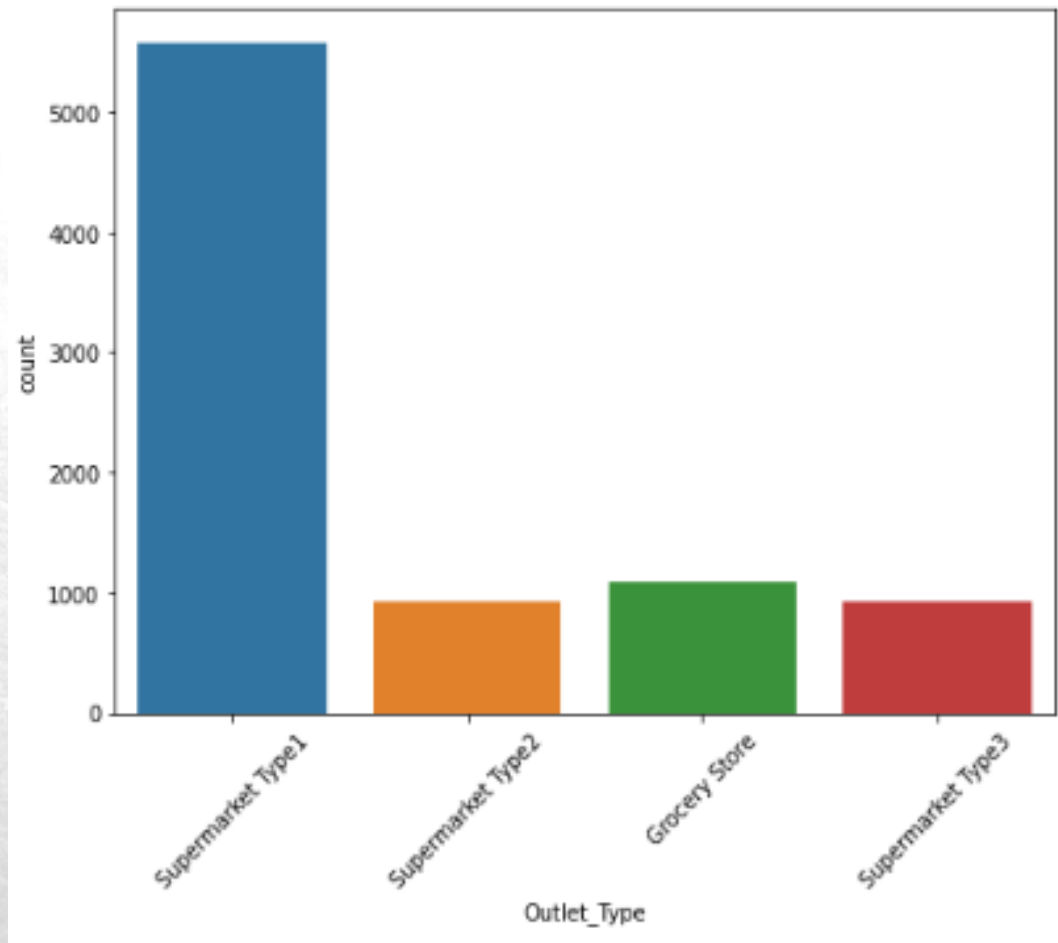**Target Variable Distribution**

EXPLORATORY DATA ANALYSIS

We can see feature Item MRP is highly correlated with target, While Outlets establishment year is least correlated with target.
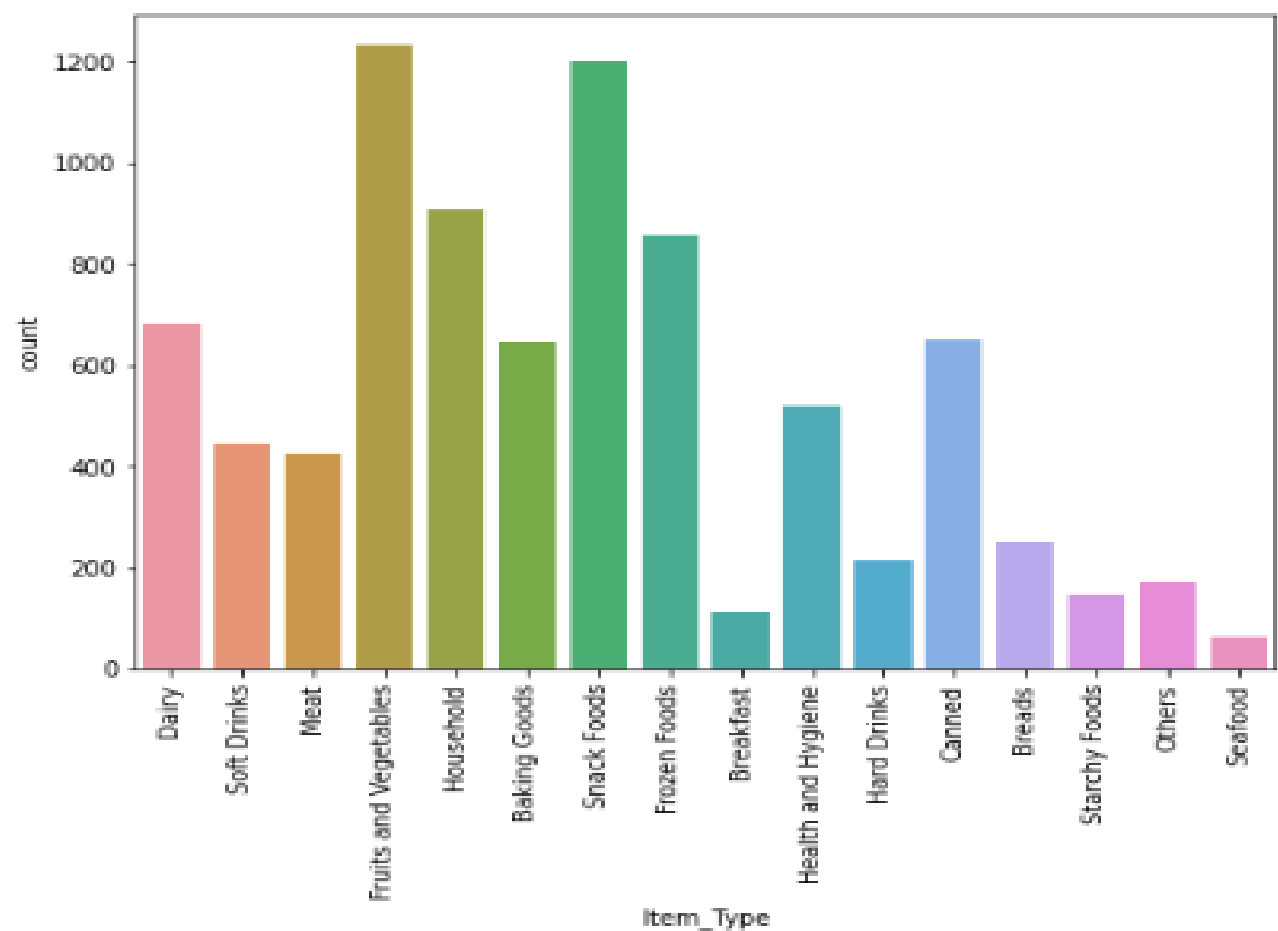
Correlation HeatMap

EXPLORATORY DATA ANALYSIS (Bivariate Analysis)
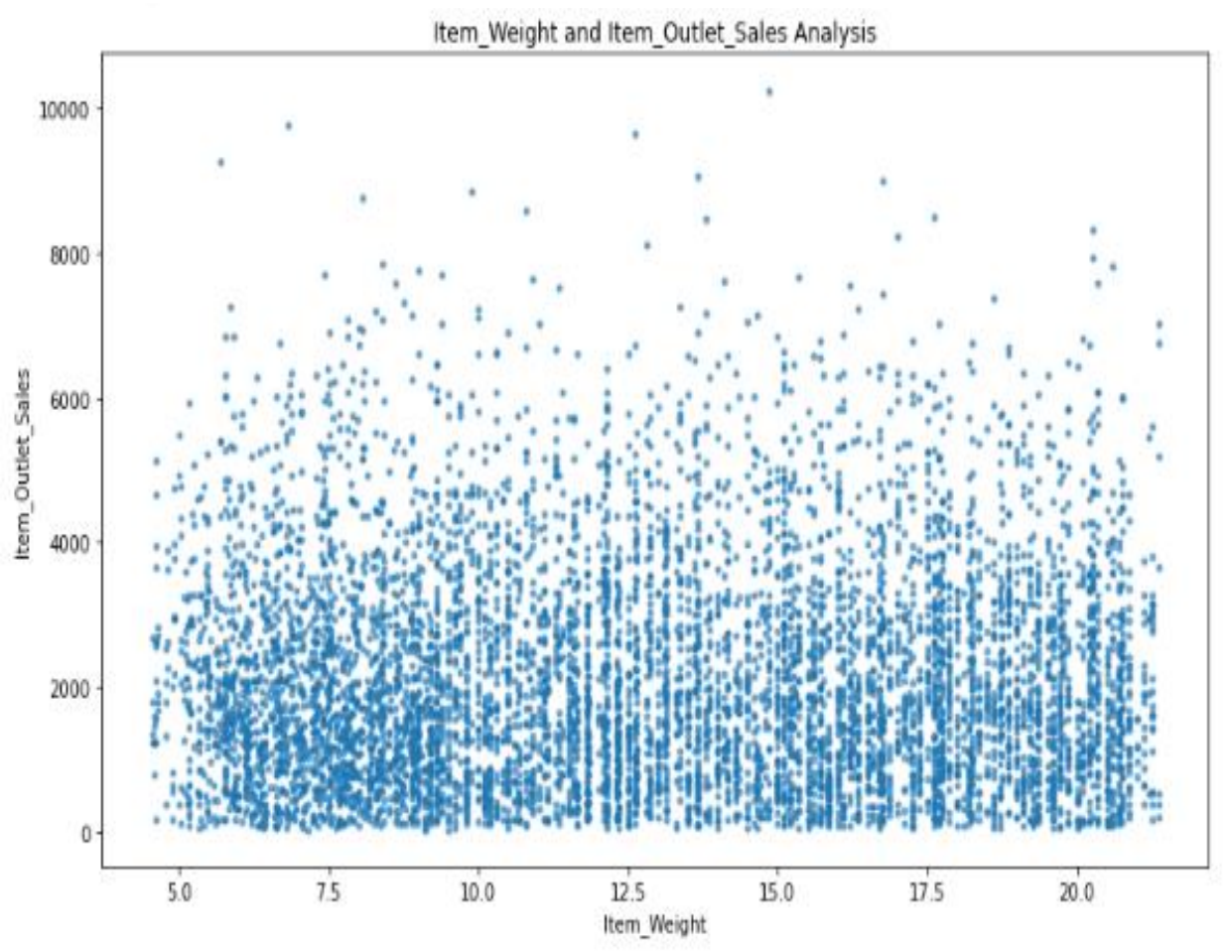

Outlet Types
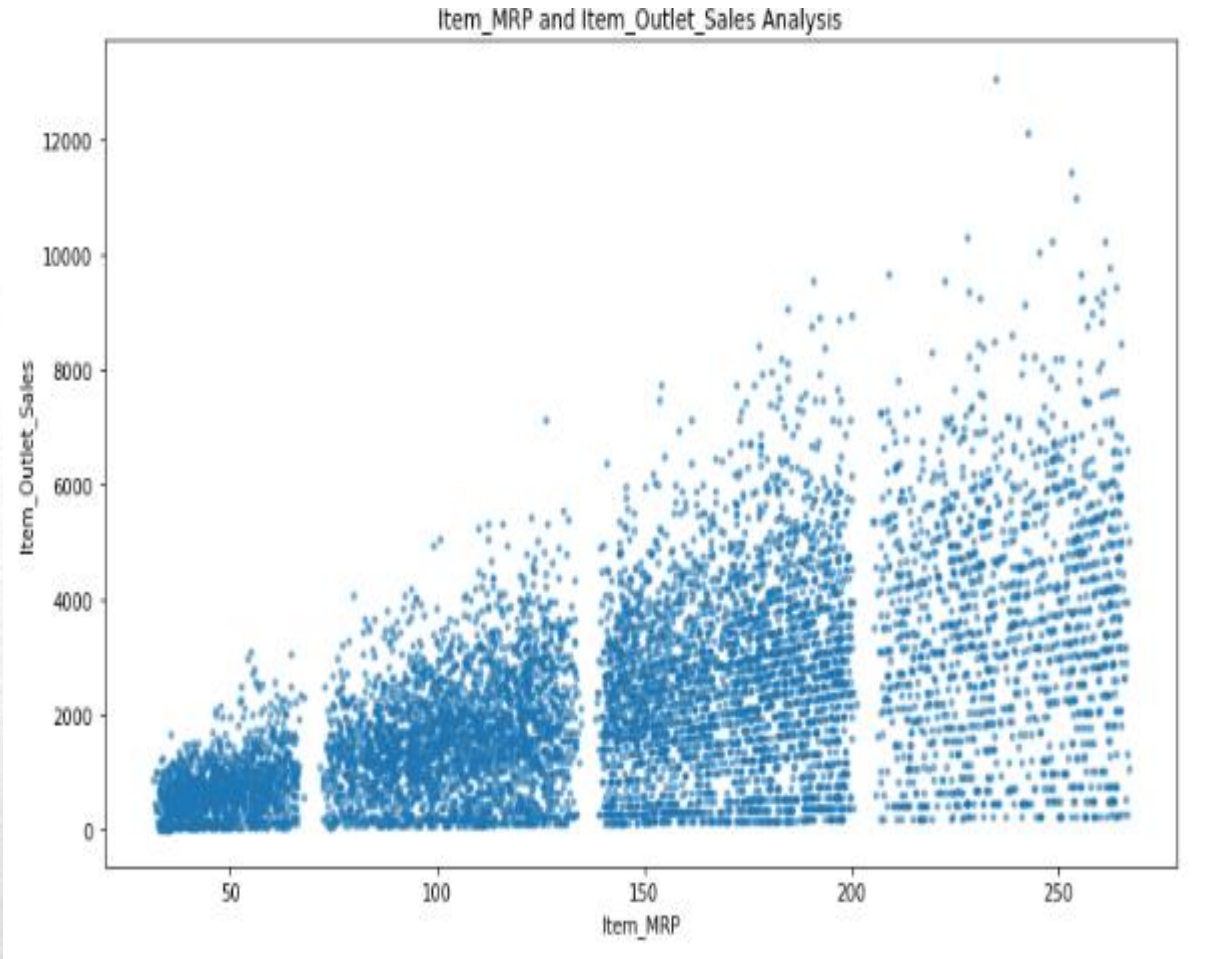Supermarket type 1 is with highest freq.

Product categories available in BigMart
Fruits & veg with highest and Seafood with lowest quantity

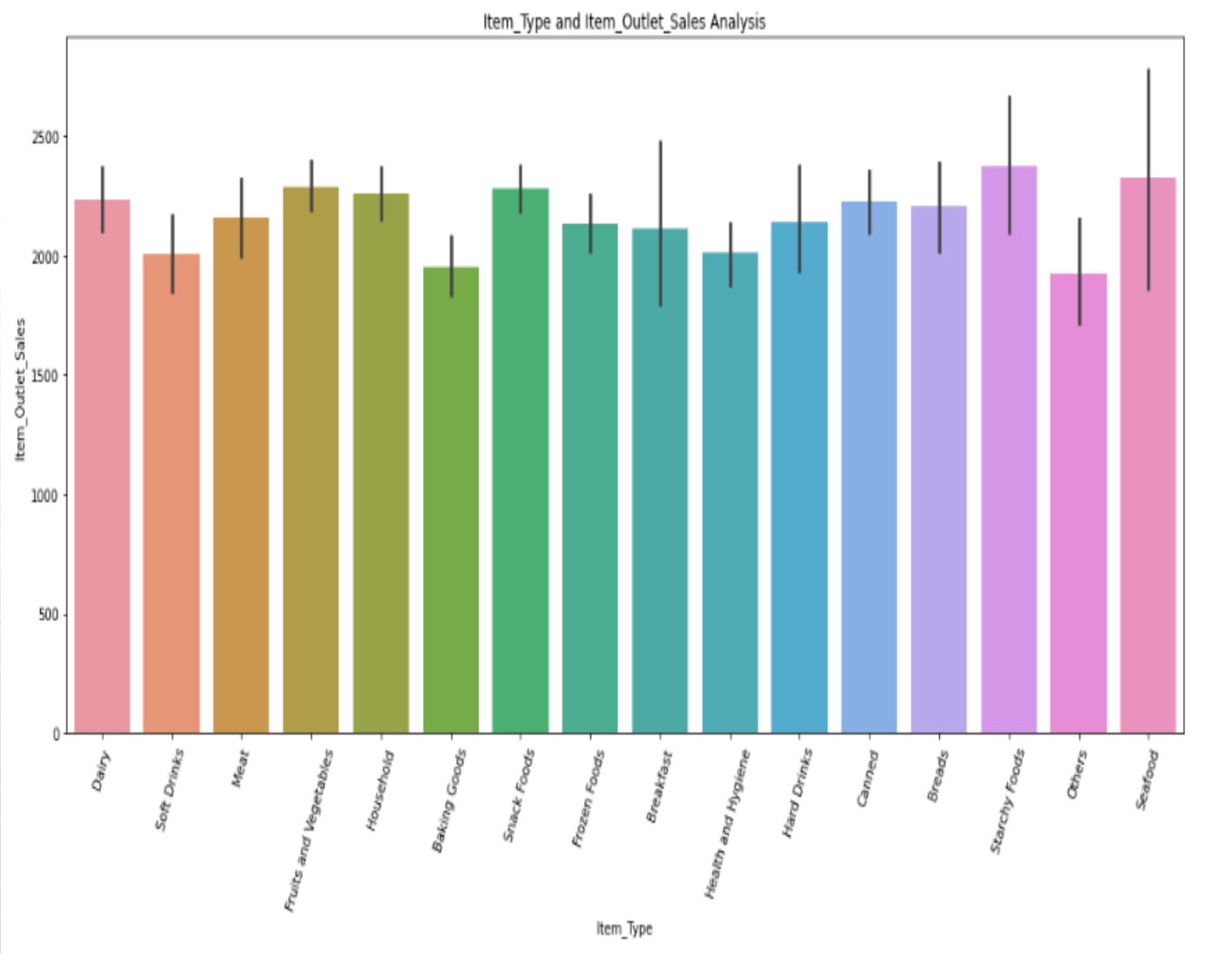# EXPLORATORY DATA ANALYSIS (Bivariate Analysis)
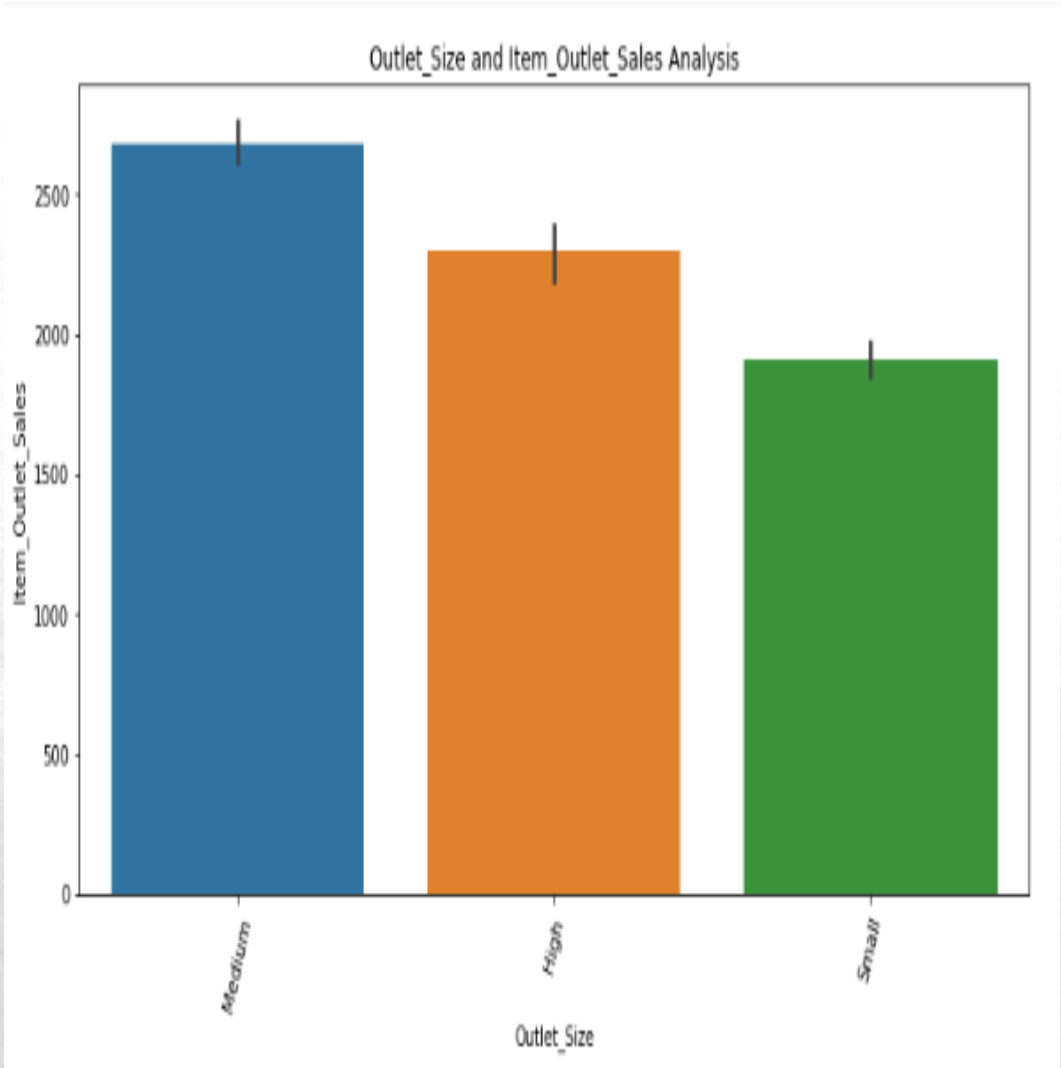

Item Weight Vs Outlet

Item MRP Vs Outlet Sales

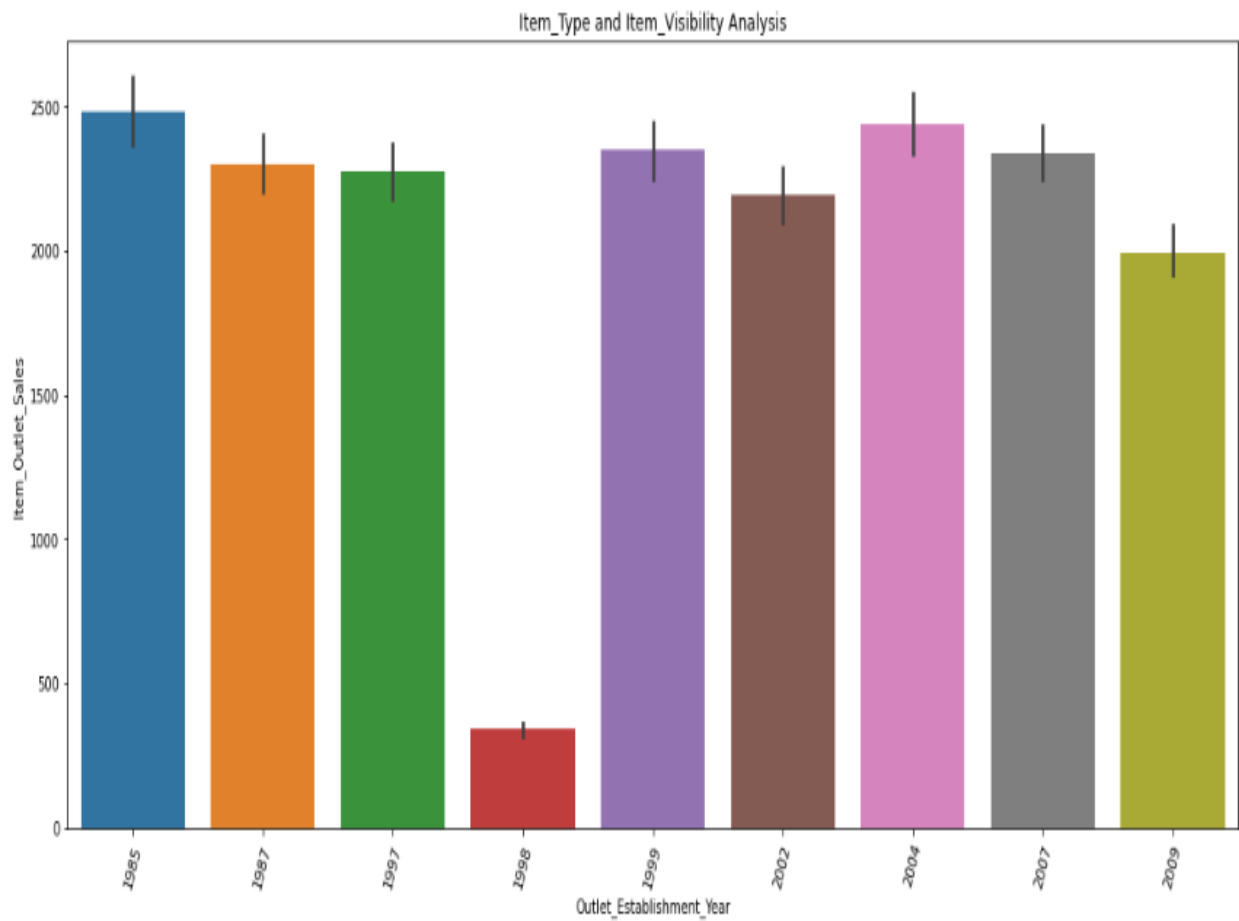# EXPLORATORY DATA ANALYSIS (Bivariate Analysis)

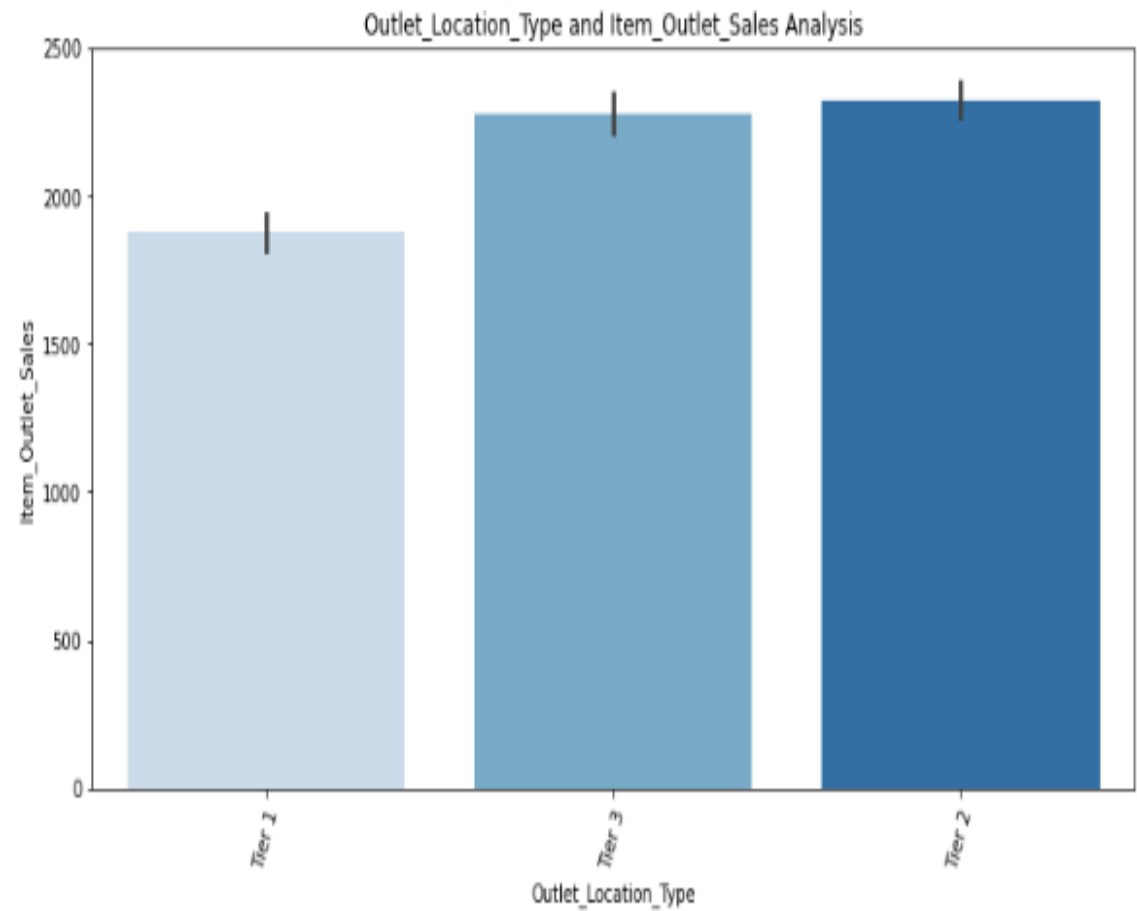
Item Type Vs Outlet Sale

Outlet Size Vs Outlet Sale

EXPLORATORY DATA ANALYSIS (Bivariate Analysis)

Outlet Establishment Year Vs Outlet Sales

Outlet Location Type Vs Outlet Sales

Let's take a look at Models and Evaluation Metrics part :

MODELS AND EVALUATION METRICS

Evaluation Metrics Used : RMSE (Root Mean Square Error)

MODEL NAME	RMSE SCORE OBTAINED
Linear Regression:	1099.5948
RandomForestRegressor:	1131.9948
XGBRegressor:	1095.9810

RECOMMENDATIONS

- From the analysis we can see that outlets with medium size have higher sales than outlet with large size. Bigmart should investigate on this and do the appropriate actions to increase the sale at outlets with larger size.

- We can see that the sale is almost same for all categories of products (i.e. – There is no dominance of any specific category on outlet sales). Hence should have more stock procured for items like 'Soft Drinks', 'Households', 'Hard Drinks' which are having longer usable period

- Programs should be carried out frequently like giving discounts, payback points, buys x get y free to attract the customers.

- Membership facility should be provided the regular customers which will have some extra benefits than other customers.

- Also, from the data analysis we noticed that outlet sales do not have significant impact of features like 'Establishment Year' and 'Item Visibility' which should not be the case. This should be investigated more as, its should be the case like, older the outlet, higher should be sale since a greater number of customers are known about the outlet and offers on the floor. Also, Items required on daily basis should have higher visibility. E.g.-   Milk, bread, biscuits should be placed near the entrance.

- Last but not the least, a proper customer service helpdesk to solve the issues faced by the customer and feedback system should be there in each outlet to get an idea on customer experience and suggestions to serve better and increase the sales by attracting more customers.

Thank You