Regression Analysis: BucketKart Sales Prediction Case Study

Dataset: BucketKart.csv

Context

A sales forecast is a prediction of future sales revenue based on historical data, industry trends, and the status of the current sales pipeline. Businesses use the sales forecast to estimate weekly, monthly, quarterly, and annual sales totals. It is extremely important for a company to make an accurate sales forecast as it adds value across an organization and helps the different verticals to chalk out their future course of action.

Forecasting helps an organization plan its sales operations by region and provides valuable insights to the supply chain team regarding the procurement of goods and materials. An accurate sales forecast process has many benefits which include improved decision-making about the future and reduction of sales pipeline and forecast risks. Moreover, it helps to reduce the time spent in planning territory coverage and establish benchmarks that can be used to assess trends in the future.

Objectives

BucketKart is an organization that owns a chain of supermarkets and food marts providing a wide range of products. They want to predict the future sales revenue of its different outlets so that they can strategize their sales operation across different tier cities and plan their inventory accordingly.

To achieve this purpose, BucketKart has hired a data science firm, shared the sales records of its various outlets for the previous quarter, and asked the firm to come up with a suitable model to predict the total sales of the stores for the upcoming quarter.

Data Description

The data contains the different attributes of the various products and stores. The detailed data dictionary is given below.

- Product_Id - unique identifier of each product, each identifier having two letters at the beginning followed by a number.
- Product_Weight - the weight of each product
- Product_Sugar_Content - sugar content of each product like low sugar, regular, and no sugar
- Product_Allocated_Area - the ratio of the allocated display area of each product to the total display area of all the products in a store
- Product_Type - broad category for each product like meat, snack foods, hard drinks, dairy, canned, soft drinks, health and hygiene, baking goods, bread, breakfast, frozen foods, fruits and vegetables, household, seafood, starchy foods, others
- Product_MRP - maximum retail price of each product
- Store_Id - unique identifier of each store
- Store_Establishment_Year - the year in which the store was established

- Store_Size - the size of the store depending on sq. feet like high, medium, and low
- Store_Location_City_Type - the type of city in which the store is located like Tier 1, Tier 2, and Tier 3. Tier 1 consists of cities where the standard of living is comparatively higher than its Tier 2 and Tier 3 counterparts.
- Store_Type - the type of store depending on the products that are being sold there like Departmental Store, Supermarket Type 1, Supermarket Type 2, and Food Mart
- Product_Store_Sales_Total - total revenue generated by the sale of that particular product in that particular store