**PRODUCT DEMAND PREDICTION WITH MACHINE LEARNING**

INTRODUCTION:

Product demand prediction with machine learning is the process of leveraging advanced algorithms and data analysis techniques to forecast the future demand for specific products. This field has gained significant prominence due to its ability to provide more precise and data-driven insights into consumer behavior and market trends.

MAIN OBJECTIVES:

The main objective of product demand prediction with machine learning is to accurately forecast the future demand for a particular product or set of products. This helps businesses and organizations in several ways.

PROCESS PHASES:

1. Data Collection

2. Data Preprocessing

3. Feature Engineering

4. Visualization

5. Interpretation of Results

PHASE 1: DATA COLLECTION

● The data collection process in product demand prediction with

machine learning is a crucial initial step to build accurate

predictive models.

● Collecting the demand of the products should be included here

● Huge volumes of data are needed for the Data analysis process.

● Data collection contains data like store ID, Base price , Total price ,

Unit sold.

PHASE 2 : DATA PREPROCESSING

● After collecting the data that should be well prepared and clean

for analysis.

● Handling the missing values, outliers, and data inconsistencies

are should be including here.

● It transform the data through scaling, encoding categorical

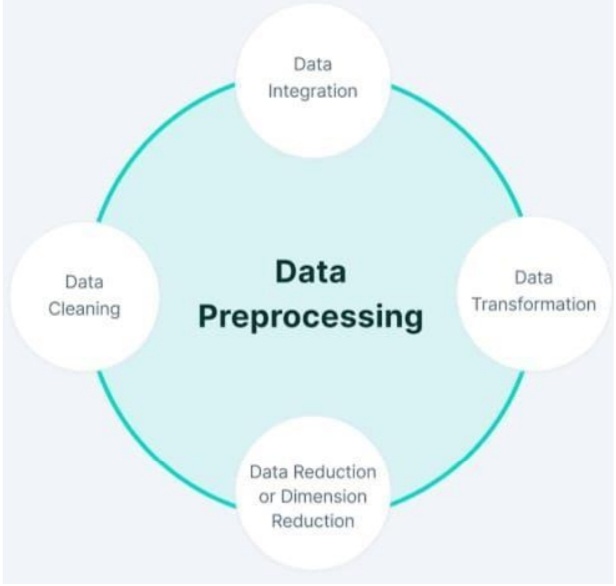
variables, and feature engineering.

● It integrate the data from different sources.

● Data cleaning, Data integration, Data transformation and

Dimension reduction are the important factors in Data

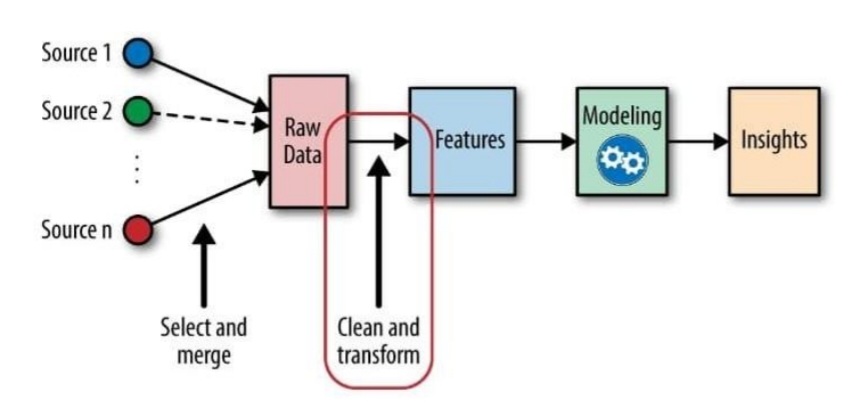
Preprocessing.



**PHASE 3 : FEATURE ENGINEERING**

● Create relevant features that will be used in the machine learning model. Features could include product attributes, time-related features , and variables related to external factors affecting demand.

● It reduce the dimensionality in case of the necessary situation.



**PHASE 4 : VISUALIZATION**

**●**  Visualization is one of the key concept in data science which can

be used for give the pictorial or virtual representation about the data.

● Several plots are used for visualize the customer segments.

● Plots example:

1. Bar Chart

2. Scatter Plot

3. Pie Plot

4. Line Plot

5. Histogram

● In python, Matplotlib library used for the visualization

● Using these charts we can easily visualize the product demand

**PHASE 5 : INTERPRETATION OF RESULT**

● The goal of this phase is to interpret the product demand and predict

according to the data provided.

● It identifies the distinguishing characteristic of each segment.

● Profile each Product based on their Demand such as store ID , base

price etc

**ACCORDING TO THE DATASET**

According to the dataset the product demand prediction contains

**● the product id;**

**● store id;**

**● total price at which product was sold;**

**● base price at which product was sold;**

**● Units sold (quantity demanded);**

The basic idea behind predict is to predict the demand between base price and total price along with unit sold

**DATASET**: https://www.kaggle.com/datasets/chakradharmattapalli/product-demand-prediction-with-machine-learning

**PROGRAM:**

Let us start by importing necessary libraries and the dataset we need for the task of product demand prediction

**import pandas as pd**

**import numpy as np**

**import plotly.express as px**

**import seaborn as sns**

**import matplotlib.pyplot as plt**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn.tree import DecisionTreeRegressor**

LOADING THE DATA:

**data = pd.read\_csv("c://documents/productdemand.csv")**

**data.head()**

Identify the shape of dataset

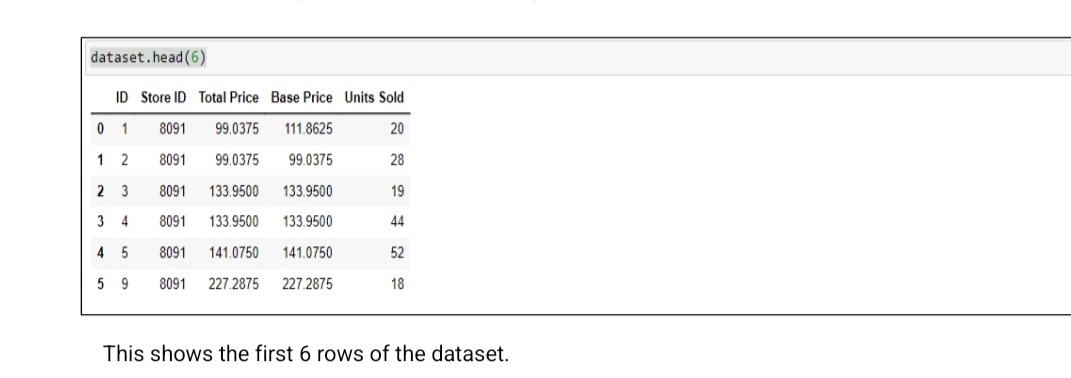


The dataset contains 150150 rows and 5 columns

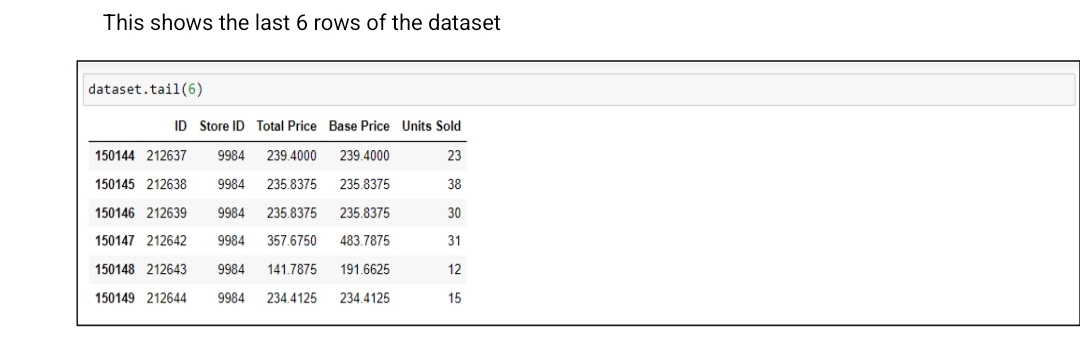
To check the head and tail of the dataset use the following

● dataset.head(contains the no. of rows)

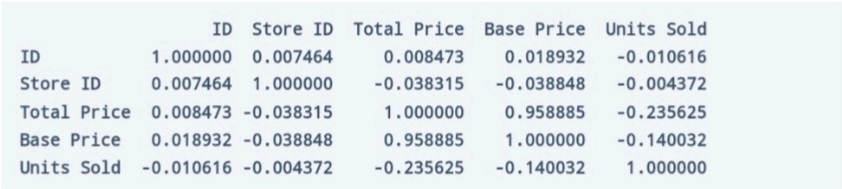
● dataset.tail(contains the no. of rows)



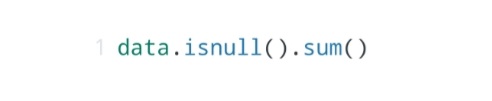
.

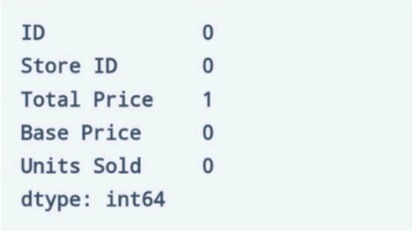


We have the data of Product ID , Store ID , Total Price , Base Price , Units Sold in our dataset.There are 5 columns and 150150 rows.



Now lets have at whether this dataset contains null value or not

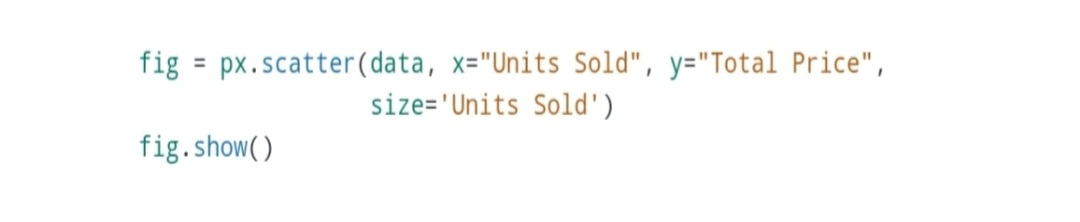


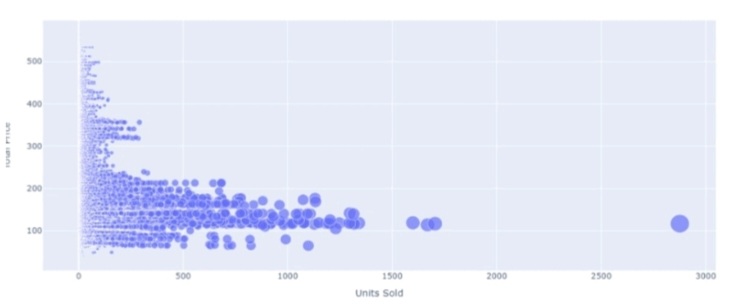


So the dataset has only one missing value in the Total Price column, We remove that entire row for now:



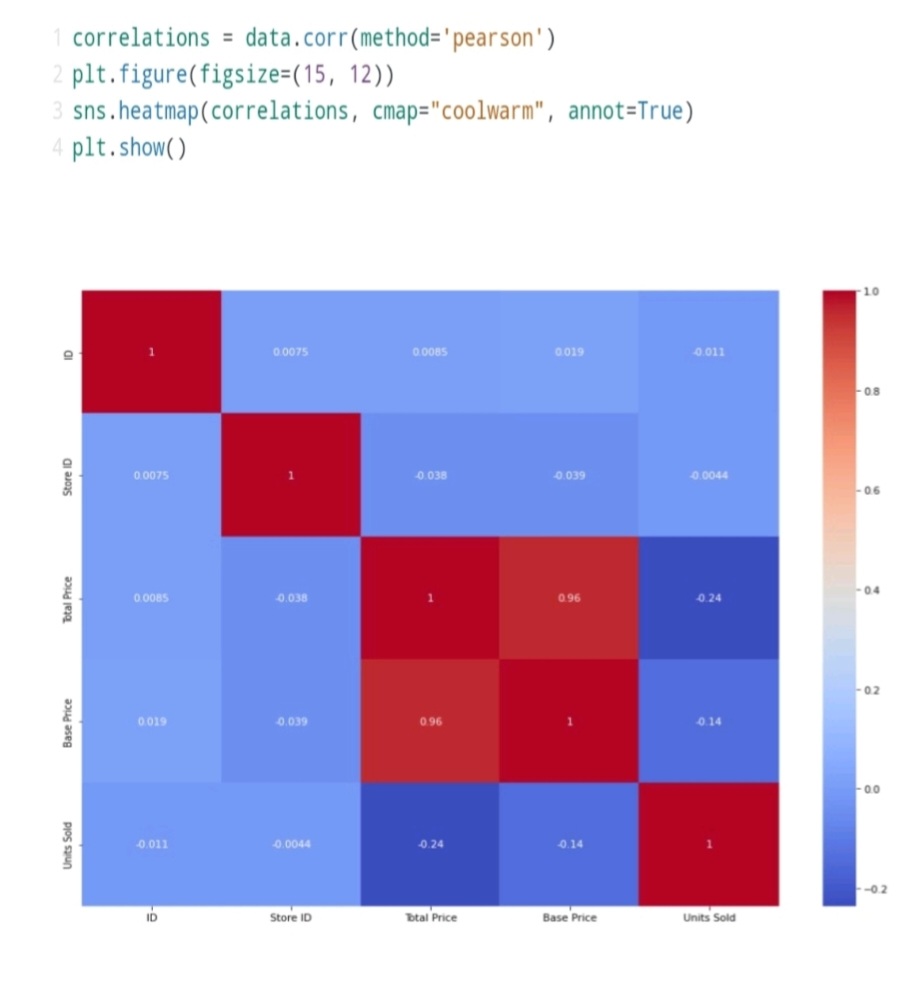
Let us now analyze the relationship between the price and the demand for the product. Here we use a scatter plot to see how the demand for the product varies with the price change



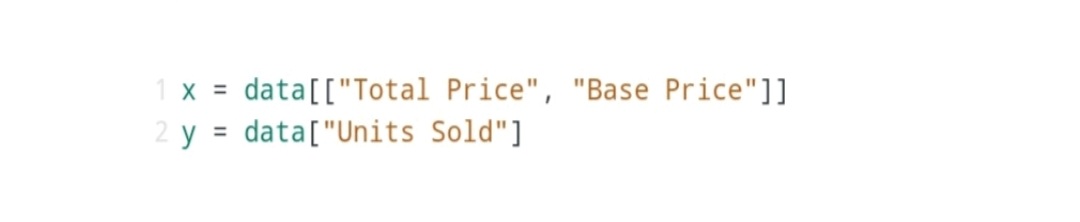


We can see that most of the data points show the sales of the product is increasing as the price is decreasing with some exceptions. Now let’s have a look at the correlation between the features of the dataset:

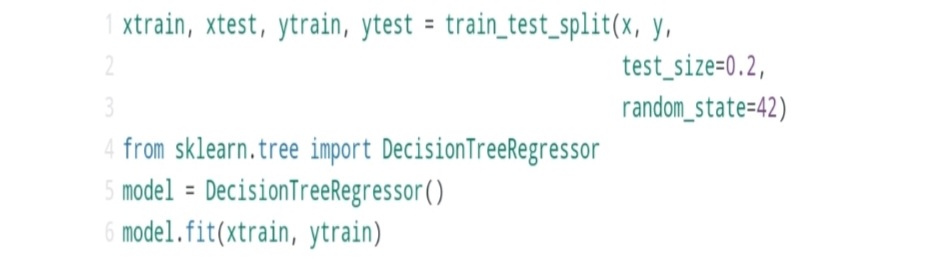




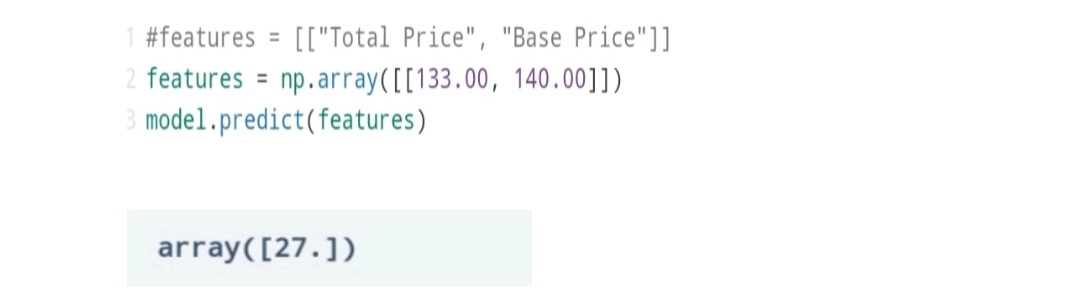
Now let’s move to the task of training a machine learning model to predict the demand for the product at different prices. We choose the Total Price and the Base Price column as the features to train the model, and the Units Sold column as labels for the model:



Now let’s split the data into training and test sets and use the decision tree regression algorithm to train our model



Now let’s split the data into training and test sets and use the decision tree regression algorithm to train our model



**DATA PREPROCESSING**

IMPORT LIBRARY: Import the required libraries such as PANDAS , NUMPY, MATPLOTLIB, and Sklearn

NumPy : It is used for mathematical and numerical functions

PANDAS : It is a analysis tool used for data manipulation

MATPLOTLIB : It is used for data visualization

SKLEARN :  To Implement Machine Learning model and Statistical Modeling

**Handle Missing Data:**

Missing Data can affect the performance of your machine learning models, and bias the conclusions derived from all the statistical analysis on the data.

So, it is necessary to handle the missing data before training our Machine Learning model.

Missing data can be handled in following ways:

○●Deleting rows

The row with the missing data can be removed if it have 70 – 75% of missing values

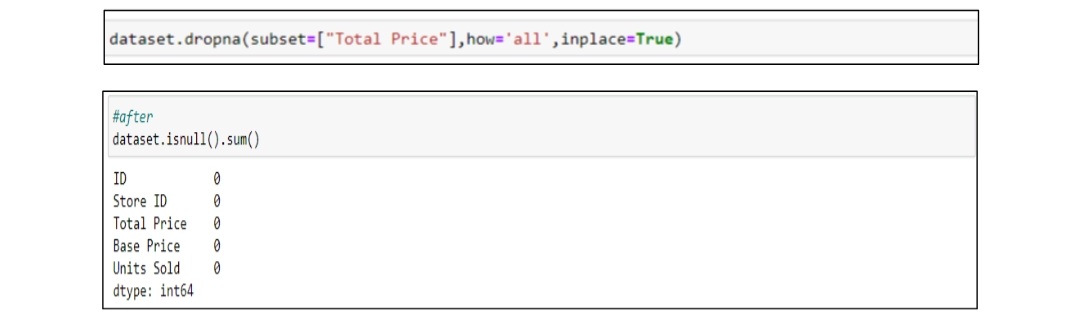
○● Replacing with Mean / Median / Mode

i) We can calculate the mean, median or mode of the feature and replace it with the missing values

ii) Replacing with the above three approximations are a statistical approach of handling the missing values

It seems to be 1 missing value in 'TOTAL PRICE' in our data set

We have to remove the value by a method called dropna() method



Now there are no missing value in the dataset.

**Splitting the Dataset**:

The dataset should be split into Train data and Test data for the following reasons

■ To estimate the performance of machine learning algorithms that are applicable for prediction-based Algorithms/Applications

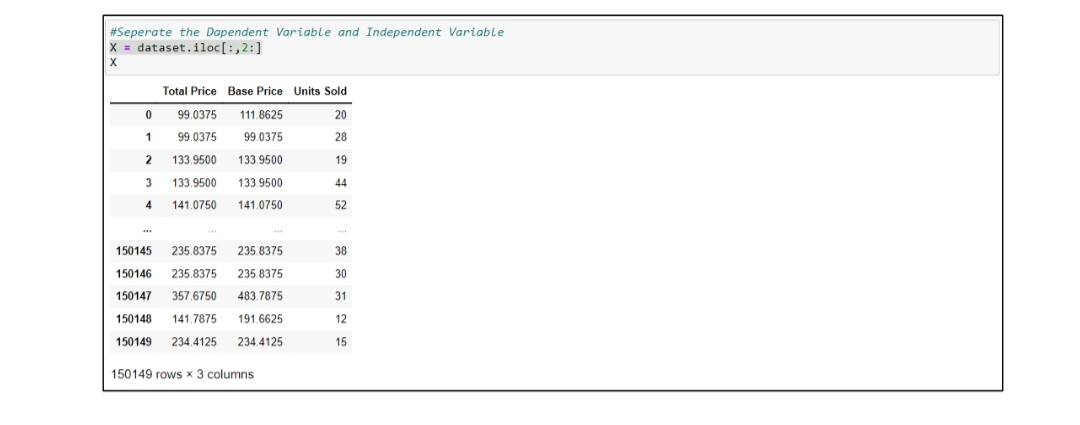
■ To measure the accuracy of our model

Before splitting our dataset we should separate the Dependent variable and Independent variable

In our dataset the Dependent variable are ‘Total Price’ , ‘Base Price’ , ‘Units Sold’ and the Independent variable is ‘Product ID’.

Dependent Variable - The dependent variable is  the variable about which predictions or explanations are being sought.

Independent Variable - Independent variables (also referred to as Features) are the input for the prediction in our model which is not dependent on other variable.

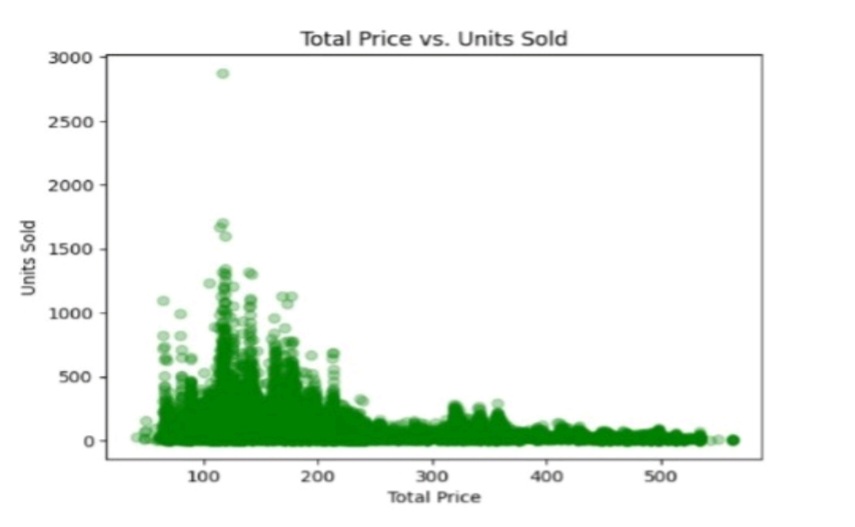


**VISUALIZATION**

■ Scatterplot Visualization

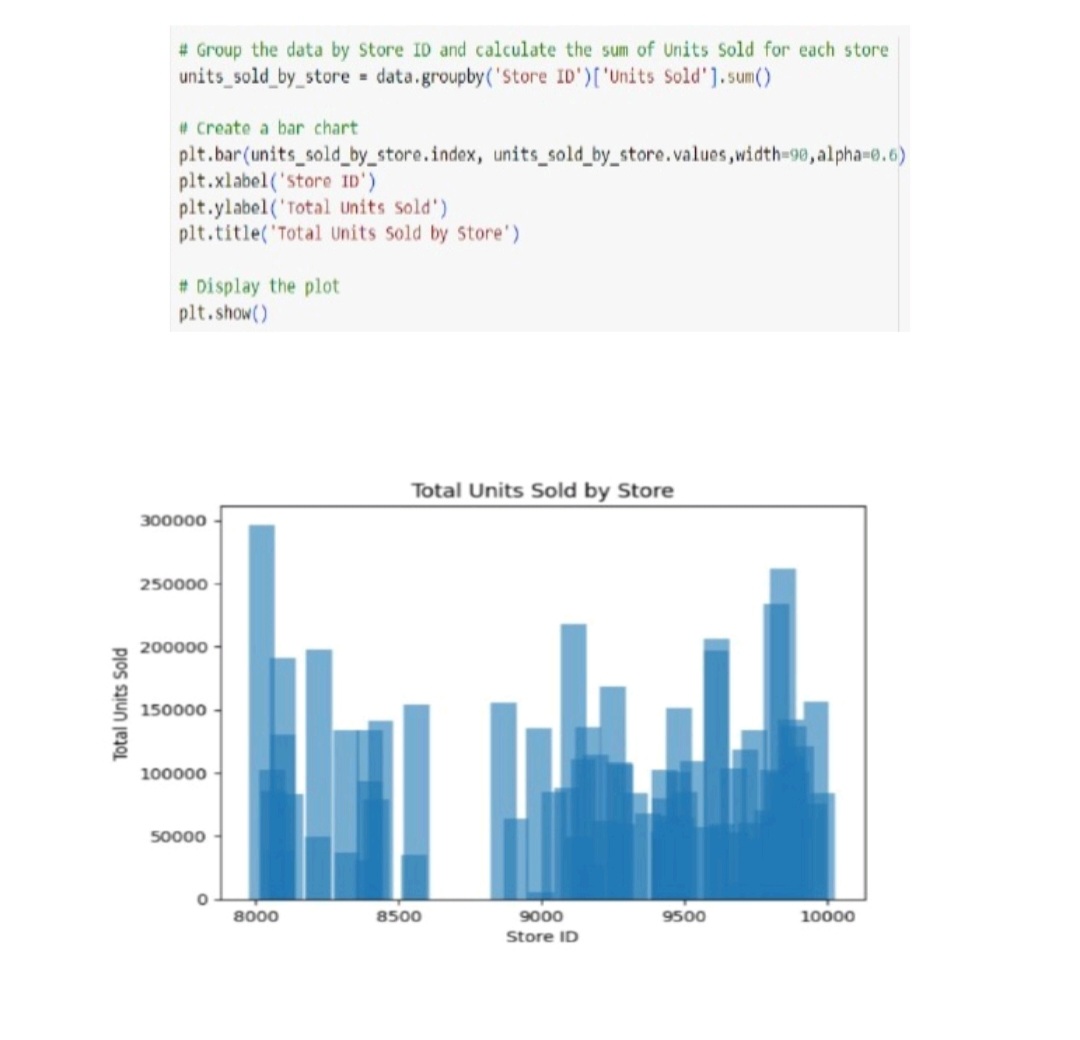
The follwing contains the visualization of dataset using scatterplot





■ Barchat visualization

The following contains the visualizatio of bar chart



**MODEL EVALUATION:**

For Model Evaluation we use multiple models to check which one performs well and best in data. Because the linear regression did not perform well enough

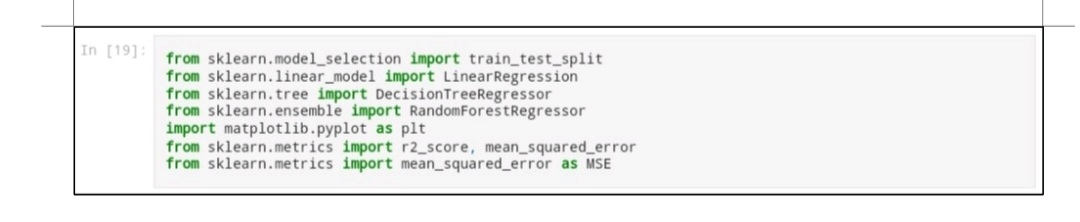
Here we use methods like DecisionTree , RandomForest , LinearRegression , XGboost.

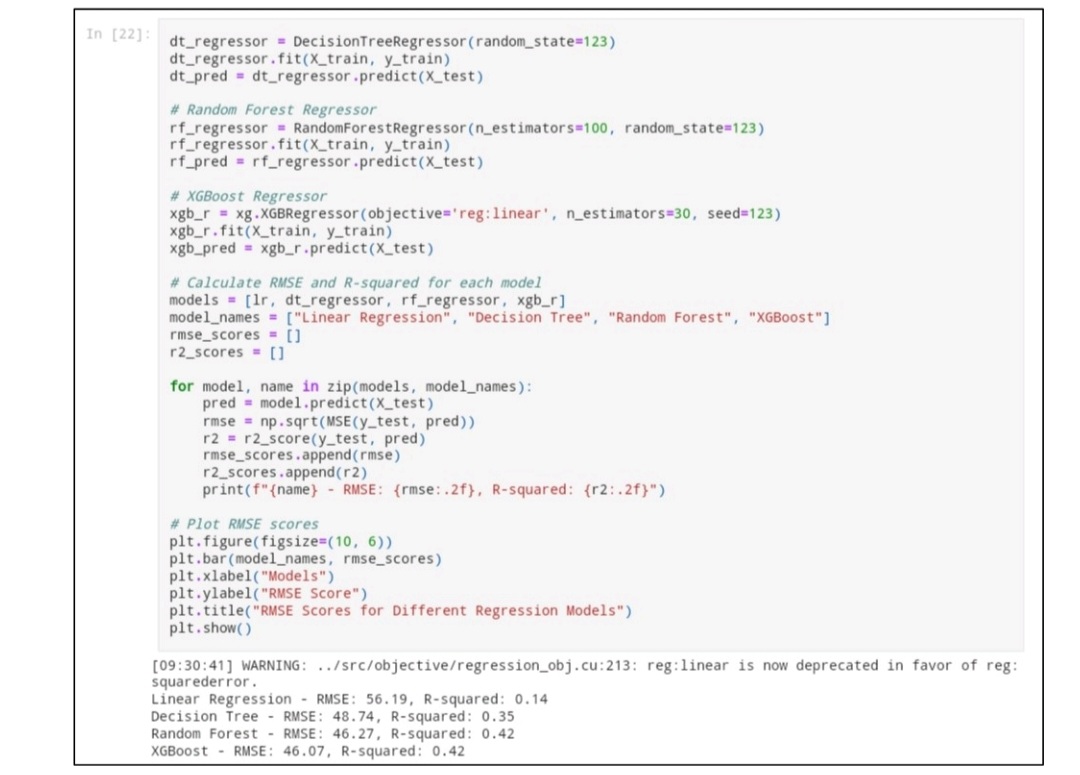
LinearRegression:It is used to predict the value of a variable of a variable based on the value of another variable.

RandomForest: It is a commonly used machine learning algorithm used for classification and regression.

DecisionTree: It is a tree-like decision support tool, displaying decisions and their outcomes.

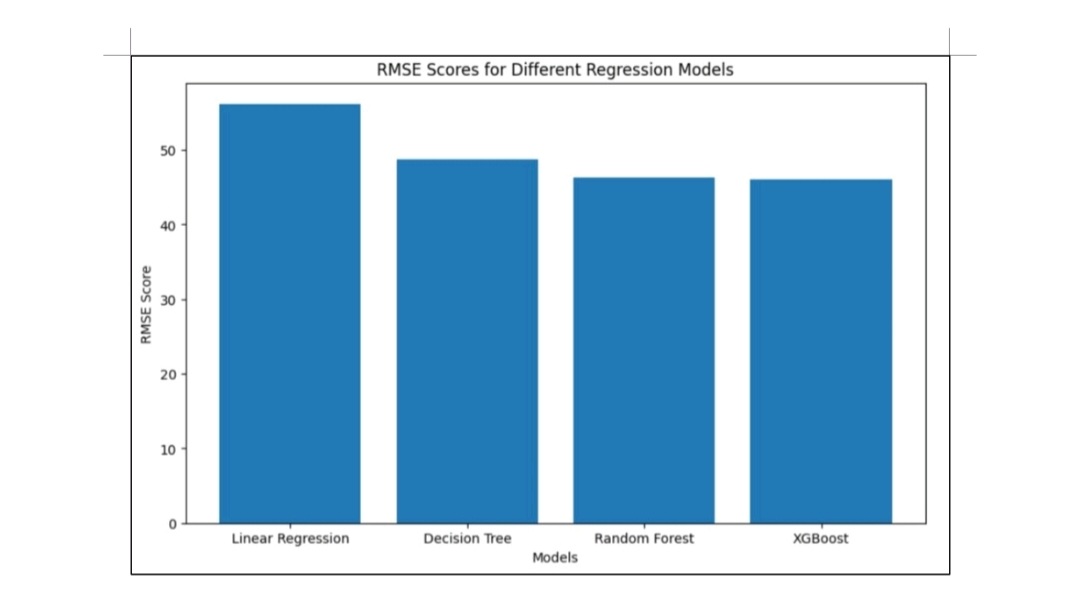
XGboost: It is a machine learning algorithm that uses an ensemble of decision trees and gradient boosting to make predictions.





BAR CHART:

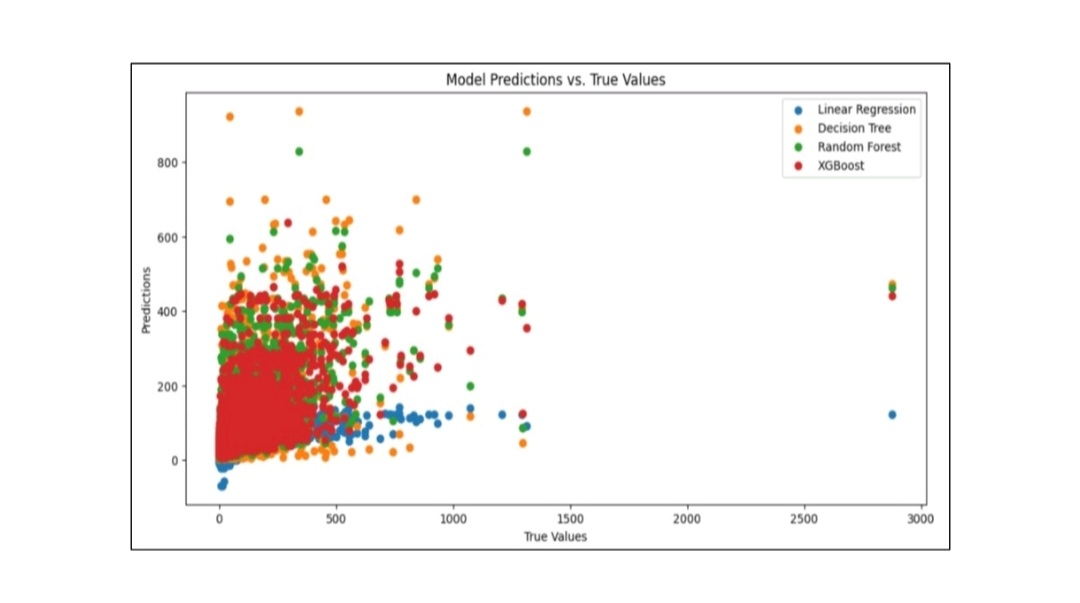
The bar chart represents the difference between regression models and their RMSE





SCATTERPLOT:

The Scatter Plot represents the Actual predicts vs the Model Predictions and how their values differ



**CONCLUSION:**

In conclusion, product demand prediction with machine learning is a powerful tool for businesses seeking to optimize their operations and meet customer needs more effectively.

By leveraging advanced machine learning techniques, historical data, and relevant features, organizations can develop accurate predictive models to anticipate future demand for their products.

This predictive capability has far-reaching implications, from improved inventory management and supply chain optimization to better production planning and targeted marketing strategies. It enables businesses to stay agile and responsive in a dynamic market environment, ultimately leading to enhanced customer satisfaction and a competitive edge.

As machine learning continues to advance, product demand prediction will become an increasingly vital component of business strategy, allowing companies to adapt to changing market conditions and make data-driven decisions that drive success.

TEAM LEADER : Santhoshkumar K

NM ID : au723721243044