PRODUCT DEMAND PREDICTION WITH MACHINE LEARNINGS

Phase-3

Development part-1

In this part you will begin building your project by loading and preprocessing the dataset.

Begin building the product demand prediction model by loading and preprocessing the dataset.

Collect and preprocess the historical sales data and external factors for analysis.

Certainly, in this section, we'll cover the initial steps of building a product demand prediction model by loading and preprocessing the dataset. This process includes collecting, cleaning, and preparing the historical sales data and external factors for analysis. Below are the steps you can follow:

1. Data Collection:

a. Historical Sales Data:

* Collect historical sales data for the product(s) you want to predict demand for. This data should include information such as date, product identifier, quantity sold, and any other relevant sales-related features. You might have this data in a CSV, Excel, or a database. For example, you can use Pandas in Python to read data from a CSV file:

import numpy as np

import pandas as pd

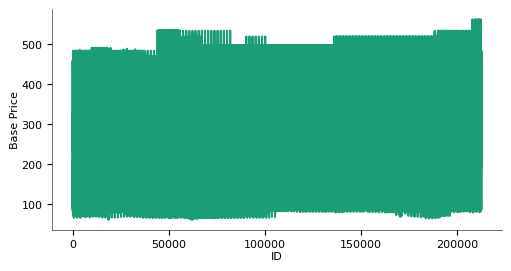
df=pd.read\_csv("/content/drive/MyDrive/PoductDemand.csv")

df

output

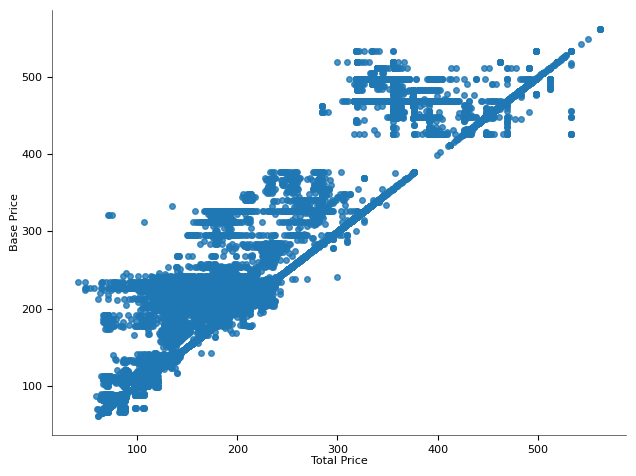
| **ID** | **Store ID** | **Total Price** | **Base Price** | **Units Sold** |
| --- | --- | --- | --- | --- |
| **0** | 1 | 8091 | 99.0375 | 111.8625 | 20 |
| **1** | 2 | 8091 | 99.0375 | 99.0375 | 28 |
| **2** | 3 | 8091 | 133.9500 | 133.9500 | 19 |
| **3** | 4 | 8091 | 133.9500 | 133.9500 | 44 |
| **4** | 5 | 8091 | 141.0750 | 141.0750 | 52 |
| **...** | ... | ... | ... | ... | ... |
| **150145** | 212638 | 9984 | 235.8375 | 235.8375 | 38 |
| **150146** | 212639 | 9984 | 235.8375 | 235.8375 | 30 |
| **150147** | 212642 | 9984 | 357.6750 | 483.7875 | 31 |
| **150148** | 212643 | 9984 | 141.7875 | 191.6625 | 12 |
| **150149** | 212644 | 9984 | 234.4125 | 234.4125 | 15 |

150150 rows × 5 columns



sales\_data['Date'] = pd.to\_datetime(sales\_data['Date'])

sales\_data.set\_index('Date', inplace=True)



b. External Factors:

* Identify and collect external factors that might influence product demand. These could include economic indicators, weather data, holiday calendars, marketing campaign schedules, and more. Ensure that you have data for the same time periods as your sales data.

Data Cleaning and Preprocessing:

a. Handling Missing Values:

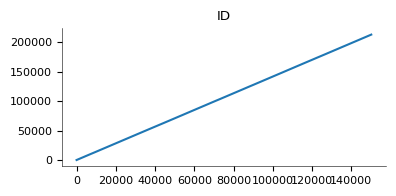
* Check for missing values in your datasets and decide how to handle them. You can choose to impute missing values or remove rows/columns with missing data.

b. Dealing with Duplicates:

* Check for and remove duplicate entries in your data to avoid data redundancy.

c. Date Formatting:

* Ensure that date columns are in the correct format. You may need to parse dates, set them as the index, and create additional date-related features (e.g., day of the week, month, year).



d. Feature Engineering:

* Create new features that may be relevant for demand prediction. For example, you can add lag features (e.g., previous month's sales), moving averages, or calculate the day of the week.

e. Data Integration:

* Merge the sales data with external factors using a common timestamp or date column. This integration will enable you to analyze how external factors affect demand.

f. Categorical Encoding:

* If your data contains categorical variables (e.g., product categories), you may need to encode them. You can use one-hot encoding or label encoding depending on the nature of the categorical data.

g. Outlier Handling:

* Identify and handle outliers in the data. You can use statistical methods or domain knowledge to decide whether to remove, transform, or impute outliers.

Data Exploration:

* Visualize your data to understand its distribution and relationships between variables. This step can help you identify patterns and insights that will inform your modeling approach.

Data Splitting:

* Split your preprocessed data into training, validation, and test sets. The training data will be used for model training, validation for hyperparameter tuning, and the test data for final model evaluation.

Conclusion

* With the historical sales data and external factors preprocessed and ready, you can proceed to select and develop the appropriate machine learning model for demand prediction, as explained in the previous response.