PRODUCT DEMAND PREDICTION WITH MACHINE LEARNING

PHASE 3 : DATA DEVELOPMENT AND PREPROCESSING

DATA DEVELOPMENT :

Data development for product demand prediction using machine learning entails a systematic process of collecting, cleaning, and preparing historical sales and relevant data. This information forms the foundation for creating and training predictive models. Feature engineering plays a crucial role, enabling the extraction of valuable insights from the data. Exploratory Data Analysis (EDA) provides essential context and helps guide feature selection. Subsequently, the dataset is divided into training and testing sets to facilitate model evaluation. The choice of machine learning algorithm depends on the nature of the data and the prediction task. Once the model is trained, it undergoes rigorous evaluation to ensure its accuracy and effectiveness in forecasting product demand. Deployment in a production environment, monitoring, and ongoing maintenance are key steps to ensure the model's real-world applicability. Additionally, considering external factors, ethical considerations, and regulations is essential. Demand prediction models should be seen as dynamic tools that evolve over time, leveraging both historical and real-time data to optimize forecasting accuracy and meet the changing demands of the market.

DATA PREPROCESSING :

Data preprocessing is a critical phase in developing machine learning models for product demand prediction. This step involves transforming and cleaning the raw data to make it suitable for analysis and modeling. To begin, the collected data, which typically includes historical sales records and relevant features, must be cleaned by handling missing values, outliers, and inconsistencies. Categorical variables may need encoding, and date and time information should be properly formatted. Feature scaling and normalization may be necessary to ensure that data is on a consistent scale, particularly if you're using algorithms sensitive to the magnitude of features. Moreover, data splitting is essential to partition the dataset into training and testing sets. This separation allows you to evaluate the model's performance on unseen data and ensures it can generalize effectively.

Proper data preprocessing not only enhances the model's predictive accuracy but also simplifies the model-building process. It sets the stage for model training and evaluation, ultimately facilitating the development of robust machine learning models for product demand prediction.

DATA DEVELOPMENT WITH PYTHON :

1. **ID :**

Building an effective product demand prediction model with machine learning involves several coding steps. You can use Python, along with popular libraries like Pandas, NumPy, Scikit-Learn, and potentially specialized libraries for time series forecasting, such as Prophet or ARIMA.
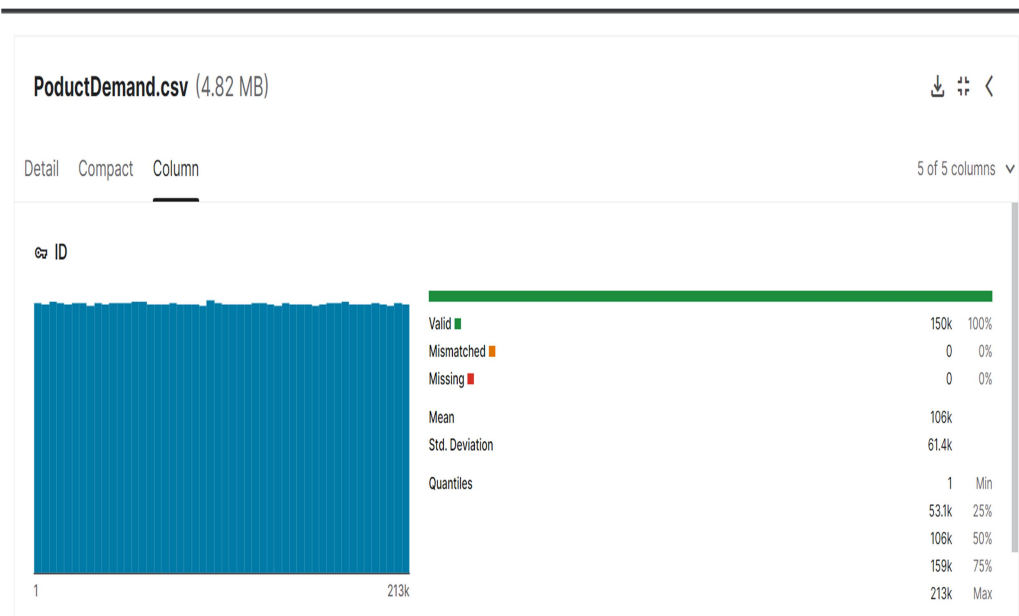

2. **STORE ID :**

Sales Data Table: Create a table to store historical sales data. This table can include columns like store_id, product_id, date, and quantity_sold. The store_id will represent the unique identifier for each store.
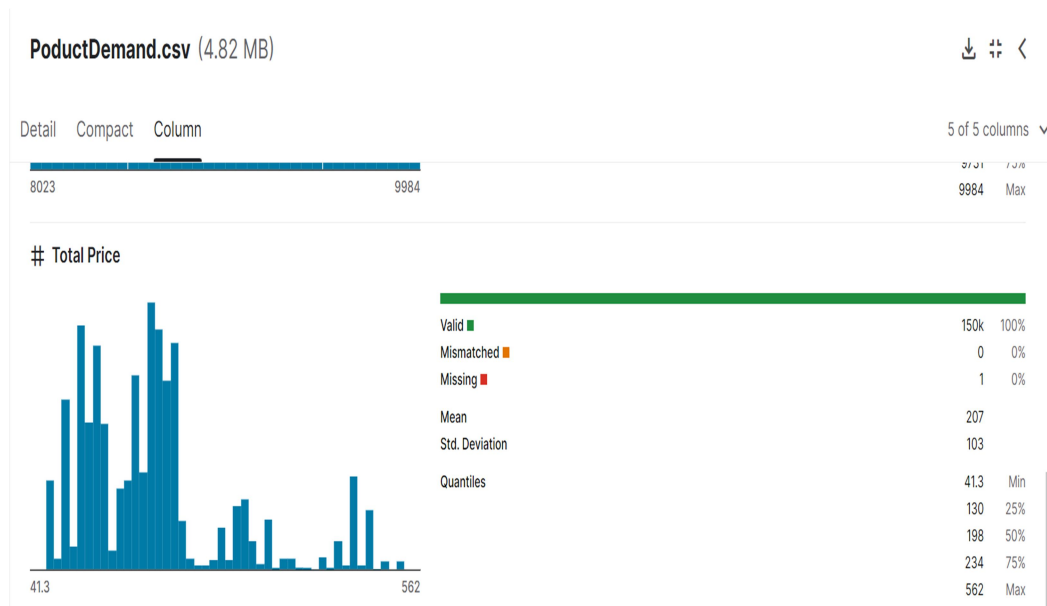
Product Information Table: Store detailed product information in a separate table. Include attributes like product_id, product_name, category, price, and any other relevant product features.

Store Information Table: If necessary, create a store information table with details about each store, such as location, size, or any unique characteristics that might impact demand.
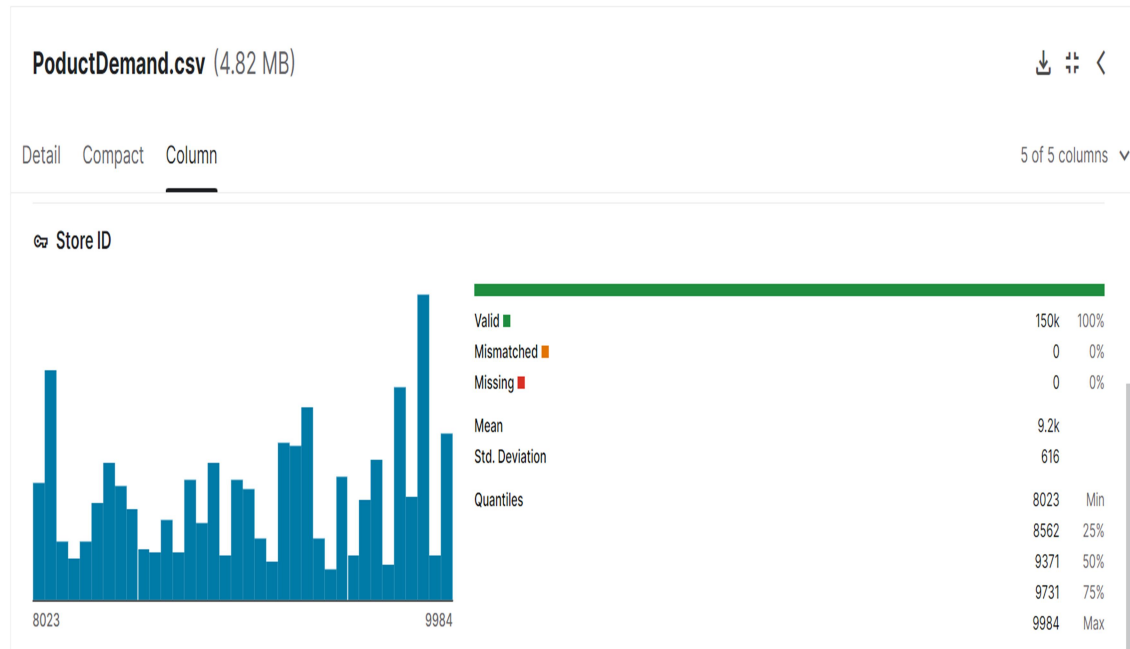
3. **TOTAL PRICE :**

        In the context of product demand prediction with machine learning, the "total price" often refers to the predicted revenue or sales value, considering both the predicted demand and the product's price.In summary, the total price for product demand prediction involves calculating the expected revenue by multiplying the predicted demand with the predicted product price. Accurate predictions can help businesses optimize pricing strategies and inventory management.

8023   9984

9984   Max

\# Total Price

Valid ■	150k	100%
Mismatched ■	0	0%
Missing ■	1	0%
Mean	207	
Std. Deviation	103	
Quantiles	41.3	Min
	130	25%
	198	50%
	234	75%
	562	Max

41.3   562

4. BASE PRICE :

In product demand prediction with machine learning, the "base price" is a fundamental factor that significantly influences a product's sales dynamics. The process begins with data collection, where historical sales data and relevant features, including the base price, are gathered. This data undergoes preprocessing, encompassing tasks such as handling missing values, outliers, and ensuring data consistency. Feature engineering is pivotal, and the base price is a critical feature. As the dataset is divided into training and testing sets, machine learning models like regression or time series forecasting are employed, considering the base price as a key input to predict demand accurately.

PoductDemand.csv (4.82 MB)

Detail Compact **Column**

5 of 5 columns ∨

 # Base Price

Valid ■	150k	100%
Mismatched ■	0	0%
Missing ■	0	0%
Mean	219	
Std. Deviation	111	
Quantiles	61.3	Min
	133	25%
	206	50%
	234	75%
	562	Max

61.3   562
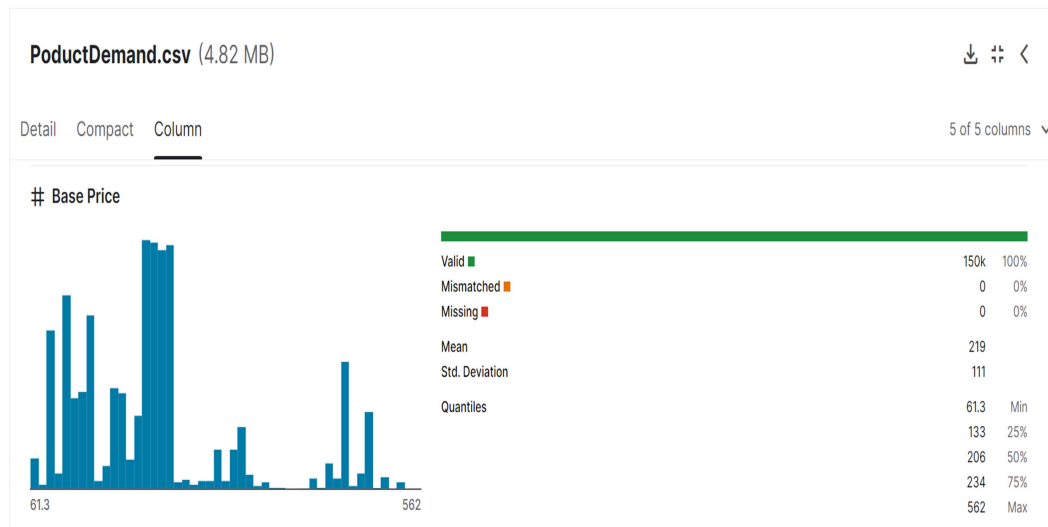
5. UNITS SOLD :

In the realm of product demand prediction using machine learning, forecasting the number of "units sold" is a pivotal goal for businesses seeking to optimize inventory, pricing, and supply chain strategies. The process begins with data preparation, involving the collection of historical sales data, as well as additional factors such as product attributes, pricing, and marketing efforts. The data is then subjected to preprocessing, addressing issues like missing values and outliers, and transformed into a suitable format for analysis.

Feature engineering plays a crucial role in creating predictive features that aid in understanding and forecasting demand. For "units sold" predictions, temporal features, seasonality patterns, and historical sales data are commonly incorporated. Machine learning models, including time series models, regression algorithms, or even advanced deep learning models, are employed to make predictions based on these features.

Training the model on historical data allows it to learn patterns and relationships between the predictors and the number of units sold. Model evaluation is conducted using metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to assess prediction accuracy.

Visualizations, such as time series plots and forecasts, help in interpreting the model's performance. They offer insights into sales trends and fluctuations, which can be essential for inventory management and demand forecasting.

Ultimately, the machine learning model, once deployed, provides a tool for real-time or future demand prediction, enabling businesses to make informed decisions about production, pricing, and inventory control. Regular monitoring of the model's performance and continuous refinement based on new data and changing market condition necessary to ensure that "units sold" predictions remain accurate and valuable for business decision-making.

PoductDemand.csv (4.82 MB)

Detail   Compact   **Column**   1 of 5 columns ⌄

\# **Units Sold**

Valid ■	150k	100%
Mismatched ■	0	0%
Missing ■	0	0%
Mean	51.7	
Std. Deviation	60.2	
Quantiles	1	Min
	20	25%
	35	50%
	62	75%
	2876	Max

1   2876