Product Demand Prediction with Machine Learning

Data Preprocessing

Steps:

Data Import:

The code imports the Pandas library and reads a CSV dataset called "ProductDemand.csv" located on the user's desktop.

Data Exploration:

The initial dataset is displayed using data.head(), and the code checks for missing values in the dataset using data.isnull().sum().

Handling Missing Values:

Missing values in the "Total Price" column are filled with the mean value of that column to ensure data completeness.

Profit Percentage Calculation:

The code calculates the profit percentage for each product based on the "Base Price" and "Total Price" and adds this as a new column.

Demand Score Calculation:

A demand score is calculated by combining "Units Sold" and "Profit Percentage" with defined weights.

High Demand Classification:

Products are classified as "High Demand" or not based on whether their demand score exceeds a specified threshold (in this case, a threshold of 50).

Data Summary:

The code prints summary statistics of the dataset using data.describe().

Data Export:

Finally, the preprocessed dataset is saved to an Excel file called "ProductDemandupdate.xlsx" on the user's desktop, excluding the index column.

Program:

```python
import pandas as pd

#importing Dataset
data = pd.read_csv("C:/Users/91735/Desktop/archive/PoductDemand.csv")
print("ProductDemand Dataset\n",data.head())

# Check for missing values
print("Missing values in the Dataset\n",data.isnull().sum())

# Fill missing values in the 'Total Price' column with the mean value
mean_total_price = data['Total Price'].mean()
data['Total Price'].fillna(mean_total_price, inplace=True)

# Check for missing values
print("After filling missing values\n",data.isnull().sum())

#Calculate Profit Percentage
data['Profit Percentage'] = ((data['Total Price'] - data['Base Price']) / data['Base Price']) * 100
print("After Profit Percentage Manipulation\n",data.head())

# Define the weights for Units Sold and Profit Percentage
weight_units_sold = 0.6
weight_profit_percentage = 0.4

# Calculate the demand score
data['Demand Score'] = (weight_units_sold * data['Units Sold']) + (weight_profit_percentage * data['Profit Percentage'])
print("After Demand score Manipulation\n",data.head())

#Calculate the High Demand
threshold = 50  # Can adjust this threshold value as needed
data['High Demand'] = (data['Demand Score'] >= threshold).astype(int)
print("After High Demand Manipulation\n",data.head())
print(data.describe())

# Preprocessed dataset
data.to_excel("C:/Users/91735/Desktop/archive/PoductDemandupdate.xlsx", index=False)
print("Your preprocessed dataset is in Desktop/archive/PoductDemandpreprocessed.xlsx")
```

Output:

ProductDemand Dataset

	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

Missing values in the Dataset

ID   0
Store ID   0
Total Price   1
Base Price   0
Units Sold   0
dtype: int64

After filling missing values

ID   0
Store ID   0
Total Price   0
Base Price   0
Units Sold   0
dtype: int64

After Profit Percentage Manipulation

	ID	Store ID	Total Price	Base Price	Units Sold	Profit Percentage
0	1	8091	99.0375	111.8625	20	-11.464968
1	2	8091	99.0375	99.0375	28	0.000000
2	3	8091	133.9500	133.9500	19	0.000000
3	4	8091	133.9500	133.9500	44	0.000000
4	5	8091	141.0750	141.0750	52	0.000000

After Demand score Manipulation

	ID	Store ID	Total Price	...	Units Sold	Profit Percentage	Demand Score
0	1	8091	99.0375	...	20	-11.464968	7.414013
1	2	8091	99.0375	...	28	0.000000	16.800000
2	3	8091	133.9500	...	19	0.000000	11.400000
3	4	8091	133.9500	...	44	0.000000	26.400000
4	5	8091	141.0750	...	52	0.000000	31.200000

[5 rows x 7 columns]

After High Demand Manipulation

	ID	Store ID	Total Price	...	Profit Percentage	Demand Score	High Demand
0	1	8091	99.0375	...	-11.464968	7.414013	0
1	2	8091	99.0375	...	0.000000	16.800000	0
2	3	8091	133.9500	...	0.000000	11.400000	0
3	4	8091	133.9500	...	0.000000	26.400000	0
4	5	8091	141.0750	...	0.000000	31.200000	0

[5 rows x 8 columns]

	ID	Store ID	...	Demand Score	High Demand
count	150150.000000	150150.000000	...	150150.000000	150150.000000
mean	106271.555504	9199.422511	...	29.047939	0.142537
std	61386.037861	615.591445	...	34.832555	0.349602
min	1.000000	8023.000000	...	-27.532523	0.000000
25%	53111.250000	8562.000000	...	10.800000	0.000000
50%	106226.500000	9371.000000	...	19.800000	0.000000
75%	159452.750000	9731.000000	...	35.400000	0.000000
max	212644.000000	9984.000000	...	1706.164263	1.000000

[8 rows x 8 columns]

Your preprocessed dataset is in Desktop/archive/PoductDemandpreprocessed.xlsx