FUTURESKILLS AI BOOTCAMP ASSIGNMENT 2

Problem Statement:

Clean and preprocess a retail sales dataset by calculating revenue, detecting and removing outliers in price, and normalizing numerical columns.

This block will import essential libraries for data processing and handling.

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from google.colab import files
```

We will then proceed to load the file in a pandas dataframe.

```
df = pd.read_csv('/content/Sales.csv')  # Update the path if needed
df.head()  # Display the first few rows
```

	ProductID	ProductCategory	Price	QuantitySold	Revenue
0	101	Electronics	1500	5	6500
1	102	Electronics	2000	3	6000
2	103	Furniture	500	10	5000
3	104	Clothing	50	100	4000
4	105	Electronics	10000	1	10000

Next steps: Generate code with df View recommended plots New interactive sheet

The value of the Revenue Column should contain the product of the Price and QuantitySold columns

```
df['Revenue'] = df['Price'] * df['QuantitySold']
```

We will use the Interquartile Range (IQR) method to remove outliers in the Price column

```
Q1 = df['Price'].quantile(0.25)
Q3 = df['Price'].quantile(0.75)
IQR = Q3 - Q1
```

```
# Define lower and upper bounds
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Remove outliers
df = df[(df['Price'] >= lower_bound) & (df['Price'] <= upper_bound)]
```

Scales Price and Revenue between 0 and 1.

```
scaler = MinMaxScaler()
df[['Price', 'Revenue']] = scaler.fit_transform(df[['Price', 'Revenue']])
```

The dataset below is the resultant after all of these operations

```
df.to_csv('/content/Cleaned_Sales.csv', index=False)
print("Cleaned dataset saved successfully!")
```

```
Cleaned dataset saved successfully!
```

The dataset below is the resultant that we will end up with after all of these operations.

```
files.download('/content/Cleaned_Sales.csv')
```