

```
In [45]: from IPython.display import Image, display
display(Image(filename='Project.1.jpg'))
```



```
In [1]: pip install pandas
```

Requirement already satisfied: pandas in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (2.3.3)
Requirement already satisfied: numpy>=1.26.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas) (2.3.4)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Note: you may need to restart the kernel to use updated packages.

```
In [2]: pip install numpy
```

Requirement already satisfied: numpy in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (2.3.4)
Note: you may need to restart the kernel to use updated packages.

```
In [4]: pip install matplotlib
```

Requirement already satisfied: matplotlib in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (3.10.7)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cyclor>=0.10 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (1.4.9)
Requirement already satisfied: numpy>=1.23 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (2.3.4)
Requirement already satisfied: packaging>=20.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (12.0.0)
Requirement already satisfied: pyparsing>=3 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (3.2.5)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.17.0)
Note: you may need to restart the kernel to use updated packages.

In [5]: `pip install seaborn`

Collecting seaborn

Using cached seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from seaborn) (2.3.4)
Requirement already satisfied: pandas>=1.2 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from seaborn) (2.3.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from seaborn) (3.10.7)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.3)
Requirement already satisfied: cycler>=0.10 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.9)
Requirement already satisfied: packaging>=20.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (25.0)
Requirement already satisfied: pillow>=8 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (12.0.0)
Requirement already satisfied: pyparsing>=3 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.5)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from pandas>=1.2->seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.17.0)
Using cached seaborn-0.13.2-py3-none-any.whl (294 kB)
Installing collected packages: seaborn
Successfully installed seaborn-0.13.2
Note: you may need to restart the kernel to use updated packages.

```
In [10]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

```
In [ ]: Customer review Project
```

```
In [17]: df = pd.read_csv("Customer_Review.csv")
df.head()
```

```
Out[17]:
```

	Serial Number	Age	Gender	Review	Education	Purchased
0	1	56	Male	Poor	PG	Yes
1	2	46	Female	Good	School	Yes
2	3	32	Female	Good	PG	No
3	4	25	Female	Average	PG	Yes
4	5	38	Male	Poor	UG	No

```
In [ ]: Getting features type
```

```
In [18]: df.dtypes
```

```
Out[18]: Serial Number    int64
Age                int64
Gender            object
Review            object
Education          object
Purchased          object
dtype: object
```

```
In [ ]: Print columns
```

```
In [19]: df.columns
```

```
Out[19]: Index(['Serial Number', 'Age', 'Gender', 'Review', 'Education', 'Purchased'], dtype='object')
```

```
In [ ]: count of missing values
```

```
In [20]: df.isnull().sum()
```

```
Out[20]: Serial Number    0
Age                0
Gender            0
Review            0
Education          0
Purchased          0
dtype: int64
```

```
In [ ]: Deleting columns that have 30% missing values
```

```
In [25]: threshold = 30 # percentage
```

```
# Calculate percentage of missing values per column
missing_percentage = df.isnull().sum() / len(df) * 100

# Drop columns above threshold
columns_to_drop = missing_percentage[missing_percentage > threshold].index
df = df.drop(columns=columns_to_drop)
print(f"✅ Dropped columns with >{threshold}% missing: {list(columns_to_drop)}")
```

```

# Fill numeric columns with median
num_cols = df.select_dtypes(include=np.number).columns
for col in num_cols:
    df[col] = df[col].fillna(df[col].median())

# Fill categorical columns with mode
cat_cols = df.select_dtypes(include='object').columns
for col in cat_cols:
    df[col] = df[col].fillna(df[col].mode()[0])

print("\nMissing values after handling:\n", df.isnull().sum())

```

✅ Dropped columns with >30% missing: []

Missing values after handling:

```

Serial Number    0
Age              0
Gender           0
Review           0
Education        0
Purchased        0
dtype: int64

```

In []: Handle Outliers

```

In [32]: def remove_outliers(df, cols):
    Parameters:
        df (DataFrame): The pandas DataFrame
        cols (list): List of numeric columns to check for outliers

    Returns:
        df_cleaned: DataFrame with outliers removed
        df_cleaned = df.copy()

    for col in cols:
        Q1 = df_cleaned[col].quantile(0.25)
        Q3 = df_cleaned[col].quantile(0.75)
        IQR = Q3 - Q1
        lower_bound = Q1 - 1.5 * IQR
        upper_bound = Q3 + 1.5 * IQR

        # Keep only rows within the bounds
        df_cleaned = df_cleaned[(df_cleaned[col] >= lower_bound) & (df_cleaned[col]

    return df_cleaned

numeric_columns = df.select_dtypes(include=np.number).columns
df = remove_outliers(df, numeric_columns)

print("✅ Shape after outlier removal:", df.shape)

```

✅ Shape after outlier removal: (100, 6)

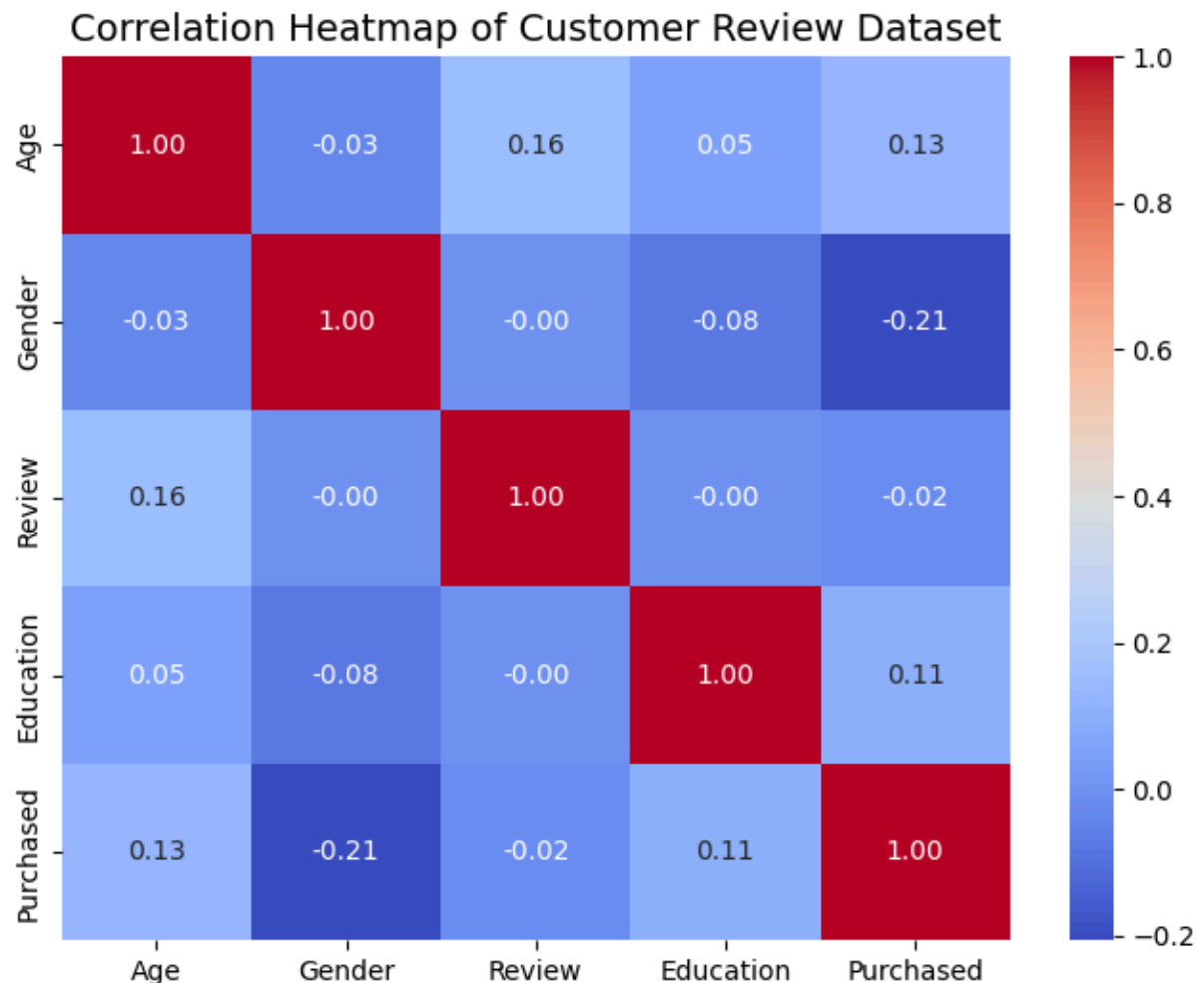
In []: Data Visualization
Correlation heatmap

```
In [38]: pip install scikit-learn
```

Requirement already satisfied: scikit-learn in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (1.7.2)
Requirement already satisfied: numpy>=1.22.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from scikit-learn) (2.3.4)
Requirement already satisfied: scipy>=1.8.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from scikit-learn) (1.16.3)
Requirement already satisfied: joblib>=1.2.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from scikit-learn) (1.5.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in c:\users\waleed.abdelgawad\appdata\local\anaconda3\envs\code1\lib\site-packages (from scikit-learn) (3.6.0)
Note: you may need to restart the kernel to use updated packages.

```
In [39]: from sklearn.preprocessing import LabelEncoder
```

```
In [41]: plt.figure(figsize=(8,6))  
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt=".2f")  
plt.title("Correlation Heatmap of Customer Review Dataset", fontsize=14)  
plt.show()
```



```
In [ ]: BAR Chart
```

```
In [42]: plt.figure(figsize=(6,4))  
data['Gender'].value_counts().plot(kind='bar', color='skyblue', edgecolor='black')
```

```
plt.title("Number of Customers by Gender")  
plt.xlabel("Gender")  
plt.ylabel("Count")  
plt.show()
```

