In [4]:

```
# ============================
#  2: Exploratory Data Analysis (EDA)
# ============================

# 1. Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os

plt.style.use('seaborn-v0_8')
pd.set_option('display.max_columns', None)

# 2. Load Dataset
csv_files = [f for f in os.listdir() if f.endswith('.csv')]
if not csv_files:
    raise FileNotFoundError("No CSV file found. Please add your dataset.")
else:
    print("Dataset Loaded:", csv_files[0])
    df = pd.read_csv(csv_files[0])

# 3. Basic Info
print("\nShape:", df.shape)
print("\nColumns:", df.columns.tolist())

print("\nInfo:")
df.info()

print("\nDescribe:")
print(df.describe())

# 4. Missing & Duplicate Check
print("\nMissing Values:\n", df.isnull().sum())
print("\nDuplicate Rows:", df.duplicated().sum())

# 5. Univariate Analysis
num_cols = df.select_dtypes(include=np.number).columns
cat_cols = df.select_dtypes(exclude=np.number).columns

# 6. Correlation
if len(num_cols) > 1:
    print("\nCorrelation Matrix:")
    print(df[num_cols].corr())

# 7. Outlier Detection
Q1 = df[num_cols].quantile(0.25)
Q3 = df[num_cols].quantile(0.75)
IQR = Q3 - Q1
outliers = ((df[num_cols] < (Q1 - 1.5 * IQR)) | (df[num_cols] > (Q3 + 1.5 * IQR))).sum()
print("\nOutlier Count per Column:")
print(outliers)


print("\nEDA Completed Successfully")
print("- Checked structure, missing data, duplicates, distributions, and correlations")
print("- Dataset ready for preprocessing or visualization")
```

```
Dataset Loaded: flipkart_laptop_data.csv

Shape: (120, 3)

Columns: ['Product Name', 'Price', 'Rating']

Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 3 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Product Name  120 non-null    object 
 1   Price         120 non-null    object 
 2   Rating        120 non-null    float64
dtypes: float64(1), object(2)
memory usage: 2.9+ KB

Describe:
           Rating
count  120.000000
mean     4.269167
std      0.221453
min      3.300000
25%      4.100000
50%      4.300000
75%      4.400000
max      4.900000

Missing Values:
 Product Name    0
Price            0
Rating           0
dtype: int64

Duplicate Rows: 36

Outlier Count per Column:
Rating    2
dtype: int64

EDA Completed Successfully
- Checked structure, missing data, duplicates, distributions, and correlations
- Dataset ready for preprocessing or visualization
```

In [ ]: