EDS DATA ANALYSIS USING NUMPY AND PANDAS

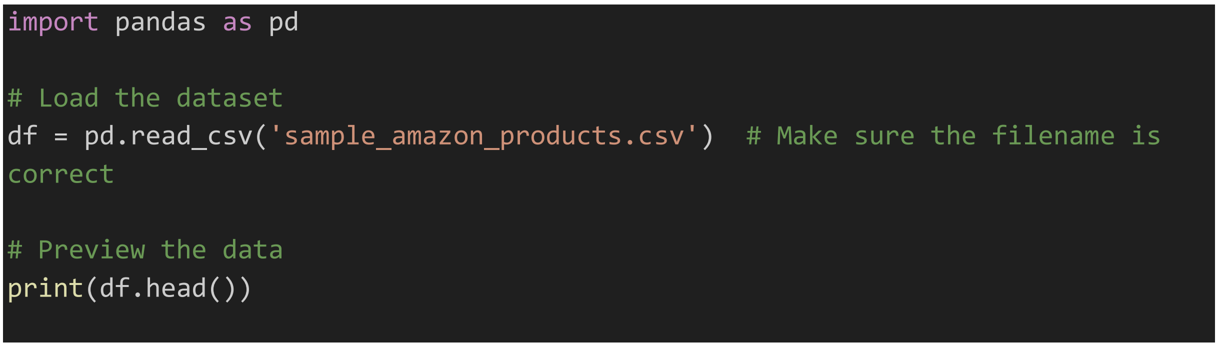
Name: Roshan shinde

PRN: 202401120022

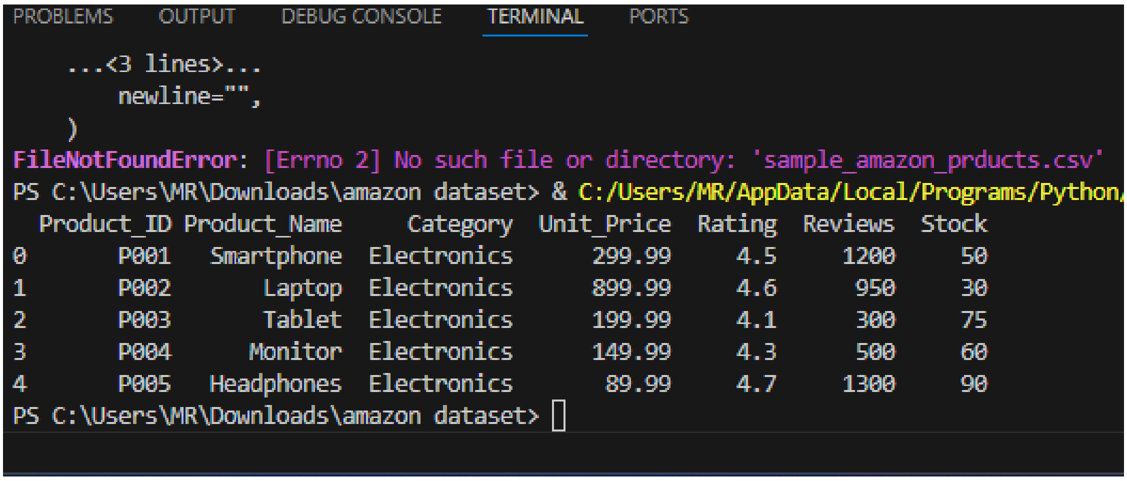
Roll No:CS8-18

Batch:CS8

1) Upload dataset

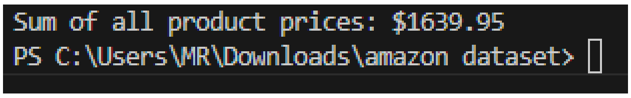


```
import pandas as pd

# Load the dataset
df = pd.read_csv('sample_amazon_products.csv')  # Make sure the filename is
correct

# Preview the data
print(df.head())
```



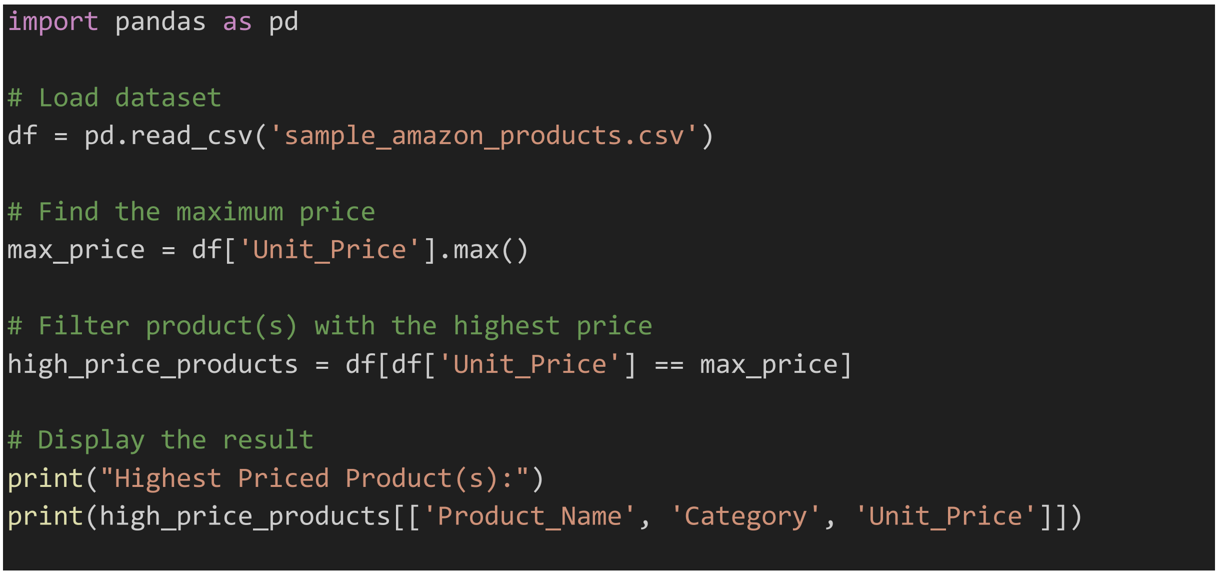
```
    ...<3 lines>...
        newline="",
    )
FileNotFoundError: [Errno 2] No such file or directory: 'sample_amazon_prducts.csv'
PS C:\Users\MR\Downloads\amazon dataset> & C:/Users/MR/AppData/Local/Programs/Python/
  Product_ID Product_Name     Category  Unit_Price  Rating  Reviews  Stock
0       P001   Smartphone  Electronics      299.99     4.5     1200     50
1       P002       Laptop  Electronics      899.99     4.6      950     30
2       P003       Tablet  Electronics      199.99     4.1      300     75
3       P004      Monitor  Electronics      149.99     4.3      500     60
4       P005   Headphones  Electronics       89.99     4.7     1300     90
PS C:\Users\MR\Downloads\amazon dataset> 
```

2) Calculate the sum of prices of all poducts

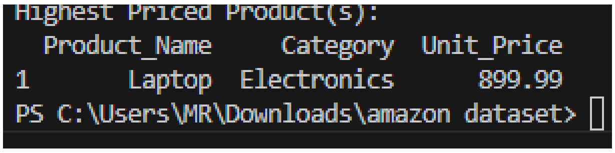


```
Sum of all product prices: $1639.95
PS C:\Users\MR\Downloads\amazon dataset> 
```

3) Maimum price of product:

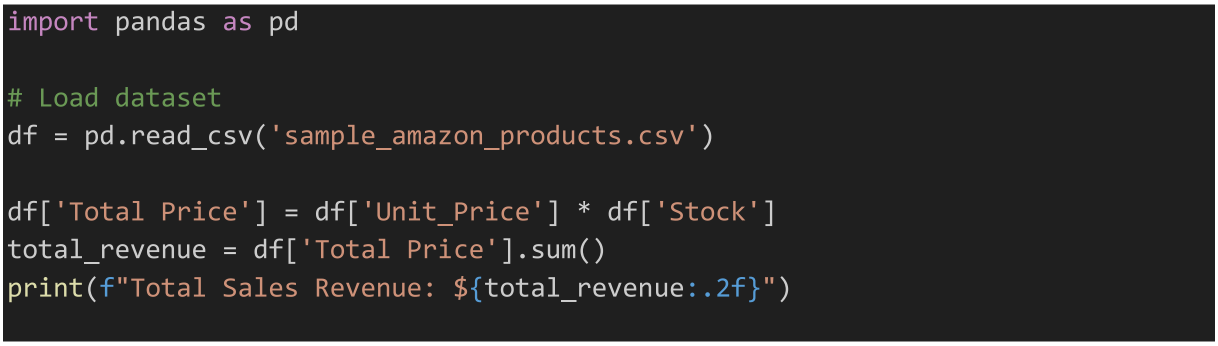


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

# Find the maximum price
max_price = df['Unit_Price'].max()

# Filter product(s) with the highest price
high_price_products = df[df['Unit_Price'] == max_price]

# Display the result
print("Highest Priced Product(s):")
print(high_price_products[['Product_Name', 'Category', 'Unit_Price']])
```

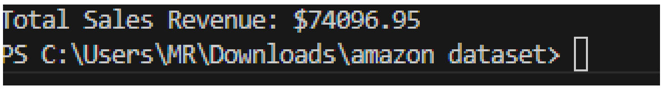


```
Highest Priced Product(s):
  Product_Name     Category  Unit_Price
1       Laptop  Electronics      899.99
PS C:\Users\MR\Downloads\amazon dataset> 
```

4) What is the total sales revenue?

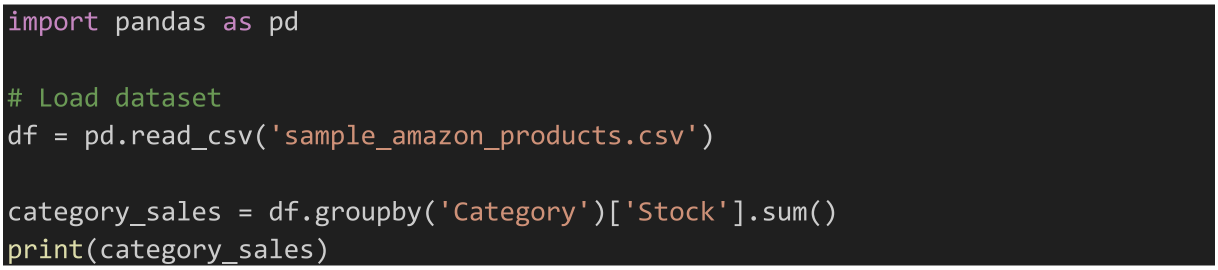


```
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

df['Total Price'] = df['Unit_Price'] * df['Stock']
total_revenue = df['Total Price'].sum()
print(f"Total Sales Revenue: ${total_revenue:.2f}")
```

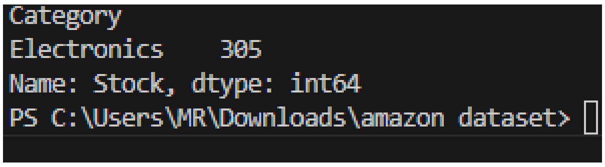


```
Total Sales Revenue: $74096.95
PS C:\Users\MR\Downloads\amazon dataset> 
```

4) What are the total units sold per category?

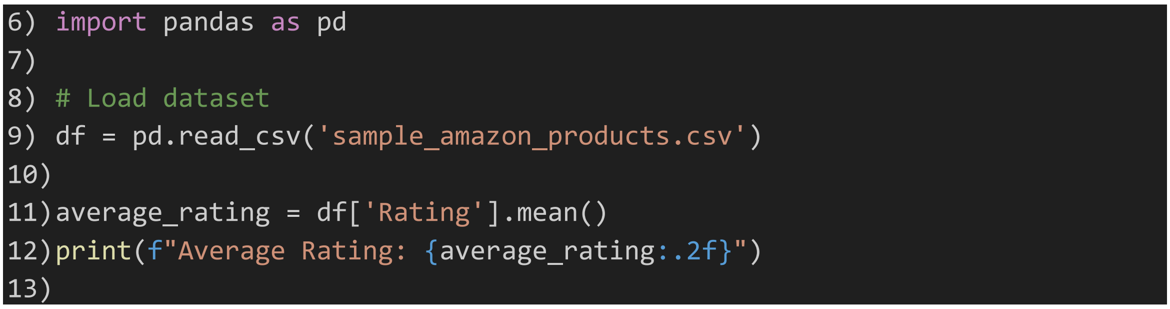


```
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

category_sales = df.groupby('Category')['Stock'].sum()
print(category_sales)
```

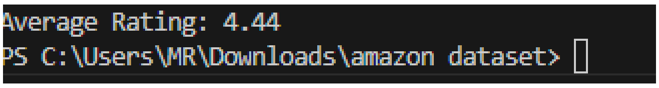


```
Category
Electronics    305
Name: Stock, dtype: int64
PS C:\Users\MR\Downloads\amazon dataset> 
```

5) What is the average product rating?

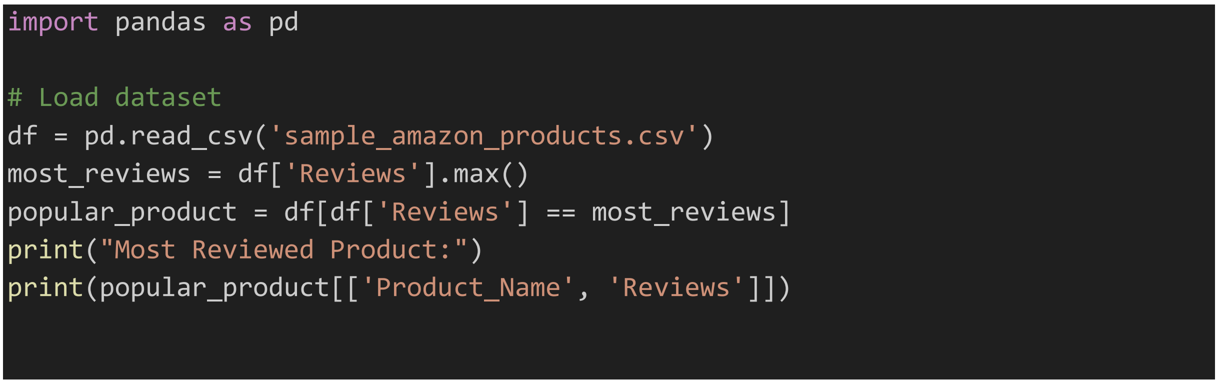


```
6) import pandas as pd
7)
8) # Load dataset
9) df = pd.read_csv('sample_amazon_products.csv')
10)
11)average_rating = df['Rating'].mean()
12)print(f"Average Rating: {average_rating:.2f}")
13)
```

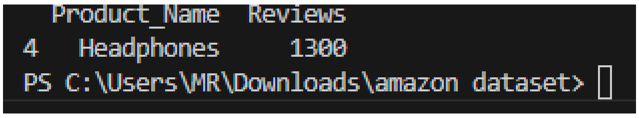


```
Average Rating: 4.44
PS C:\Users\MR\Downloads\amazon dataset>
```

6) Which product has the most reviews?

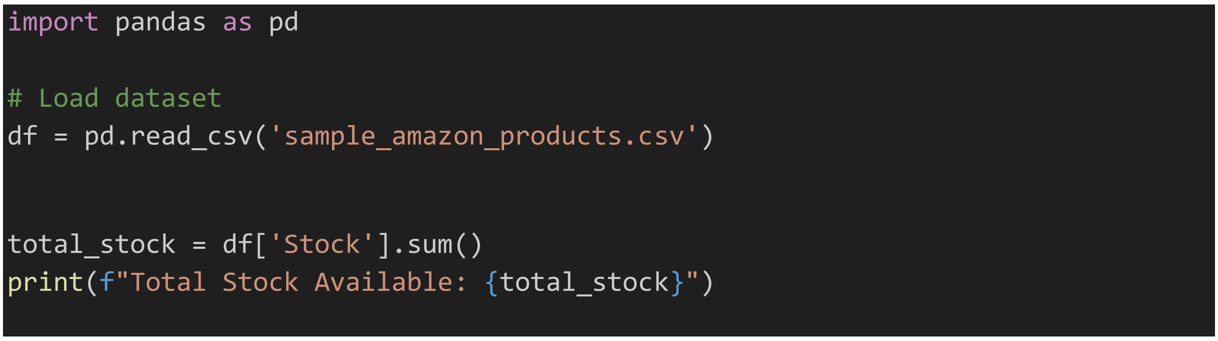


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')
most_reviews = df['Reviews'].max()
popular_product = df[df['Reviews'] == most_reviews]
print("Most Reviewed Product:")
print(popular_product[['Product_Name', 'Reviews']])
```

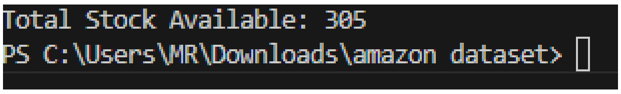


```
  Product_Name  Reviews
4   Headphones     1300
PS C:\Users\MR\Downloads\amazon dataset>
```

7) What is the total stock available across all products?

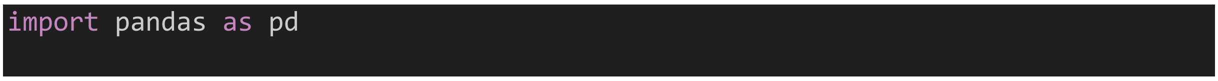


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


total_stock = df['Stock'].sum()
print(f"Total Stock Available: {total_stock}")
```

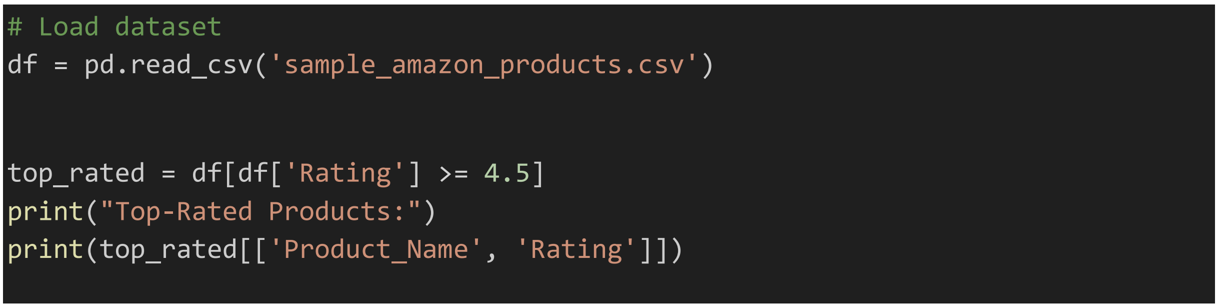


```
Total Stock Available: 305
PS C:\Users\MR\Downloads\amazon dataset>
```

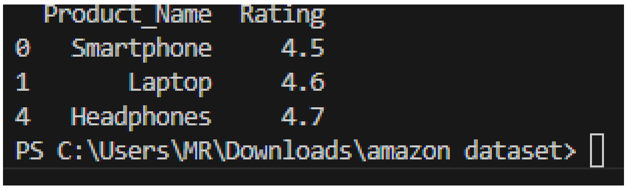
8) List all products with a rating of 4.5 or higher.



```python
import pandas as pd
```

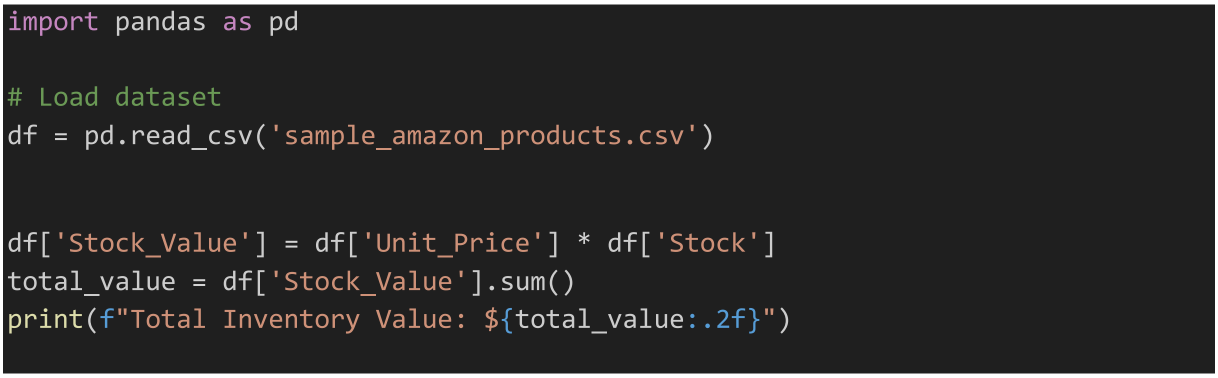


```python
# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


top_rated = df[df['Rating'] >= 4.5]
print("Top-Rated Products:")
print(top_rated[['Product_Name', 'Rating']])
```

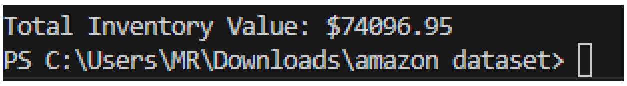


```
  Product_Name  Rating
0   Smartphone     4.5
1       Laptop     4.6
4   Headphones     4.7
PS C:\Users\MR\Downloads\amazon dataset> 
```

9) What is the total value of all products in stock (Stock × Unit_Price)?

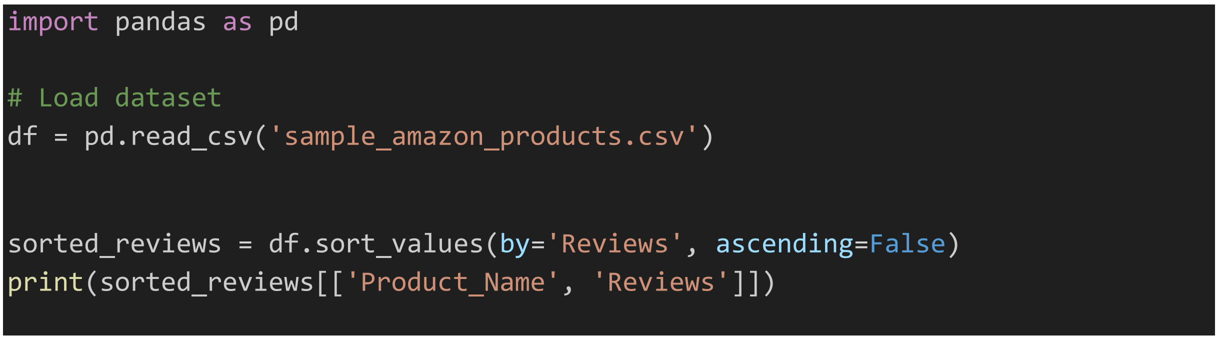


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


df['Stock_Value'] = df['Unit_Price'] * df['Stock']
total_value = df['Stock_Value'].sum()
print(f"Total Inventory Value: ${total_value:.2f}")
```

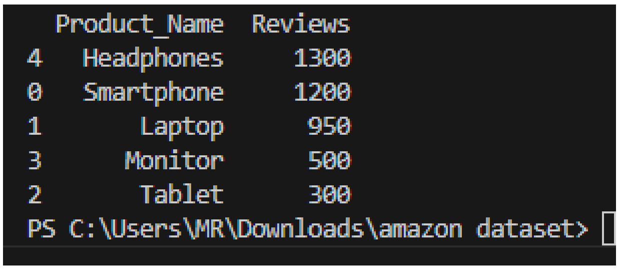


```
Total Inventory Value: $74096.95
PS C:\Users\MR\Downloads\amazon dataset> 
```

10) Sort products by number of reviews (descending).

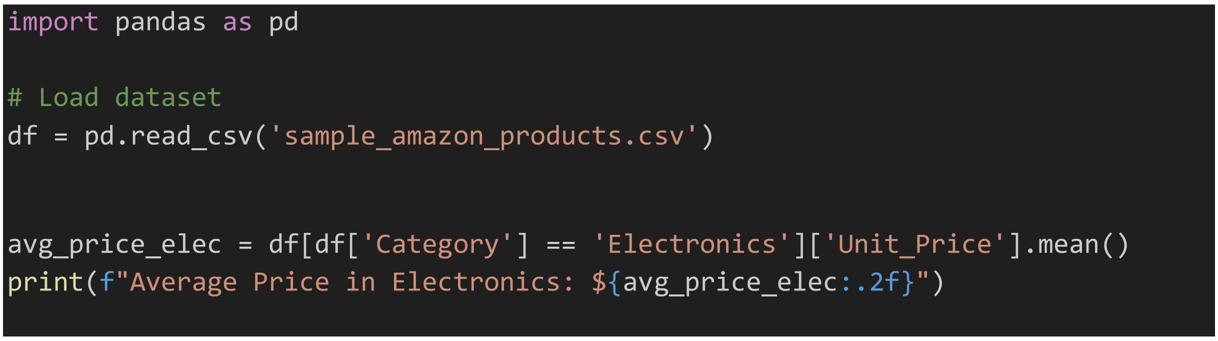


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


sorted_reviews = df.sort_values(by='Reviews', ascending=False)
print(sorted_reviews[['Product_Name', 'Reviews']])
```

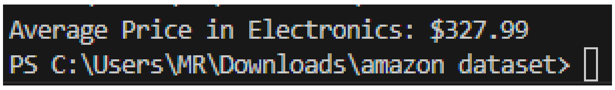


```
  Product_Name  Reviews
4   Headphones     1300
0   Smartphone     1200
1       Laptop      950
3      Monitor      500
2       Tablet      300
PS C:\Users\MR\Downloads\amazon dataset> 
```

11) What is the average price of products in the 'Electronics' category?

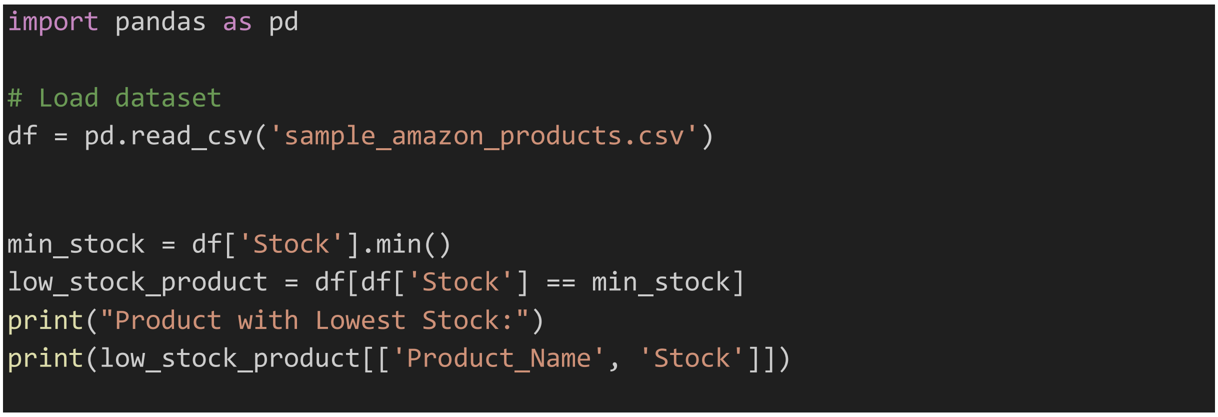


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


avg_price_elec = df[df['Category'] == 'Electronics']['Unit_Price'].mean()
print(f"Average Price in Electronics: ${avg_price_elec:.2f}")
```

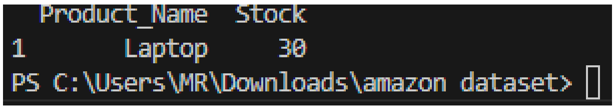


```
Average Price in Electronics: $327.99
PS C:\Users\MR\Downloads\amazon dataset> 
```

12) Which product has the lowest stock available?

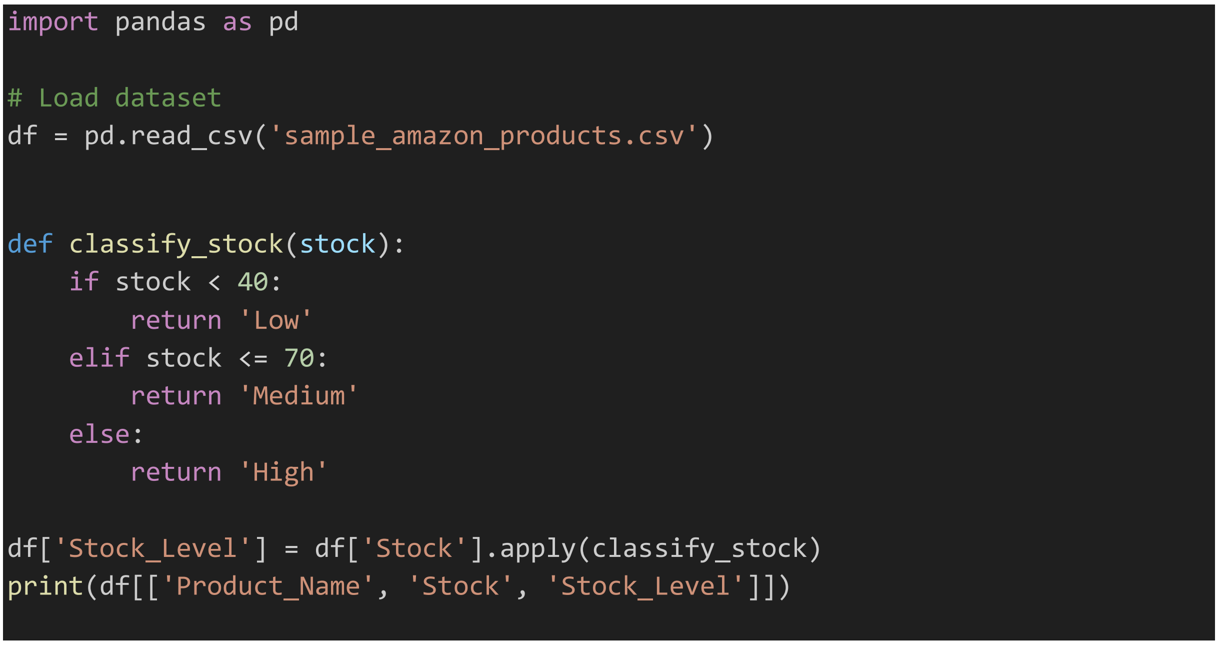


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


min_stock = df['Stock'].min()
low_stock_product = df[df['Stock'] == min_stock]
print("Product with Lowest Stock:")
print(low_stock_product[['Product_Name', 'Stock']])
```

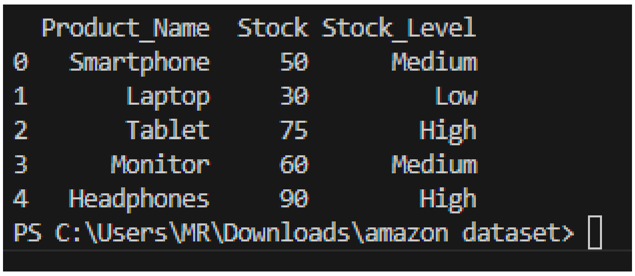


```
  Product_Name  Stock
1       Laptop     30
PS C:\Users\MR\Downloads\amazon dataset> 
```

13) Add a column to classify stock levels (Low, Medium, High).

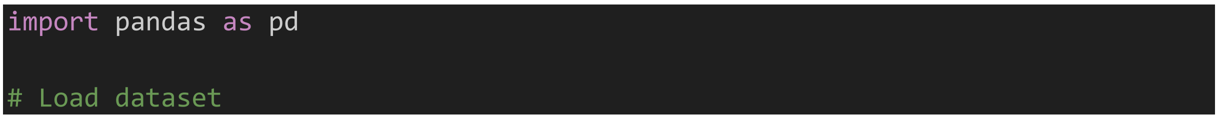


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')


def classify_stock(stock):
    if stock < 40:
        return 'Low'
    elif stock <= 70:
        return 'Medium'
    else:
        return 'High'

df['Stock_Level'] = df['Stock'].apply(classify_stock)
print(df[['Product_Name', 'Stock', 'Stock_Level']])
```

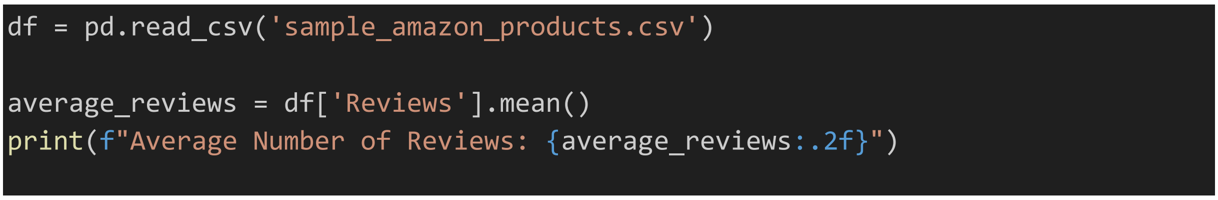


```
  Product_Name  Stock Stock_Level
0   Smartphone     50      Medium
1       Laptop     30         Low
2       Tablet     75        High
3      Monitor     60      Medium
4   Headphones     90        High
PS C:\Users\MR\Downloads\amazon dataset> 
```

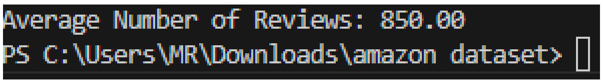
14) What is the average number of reviews per product?



```python
import pandas as pd

# Load dataset
```

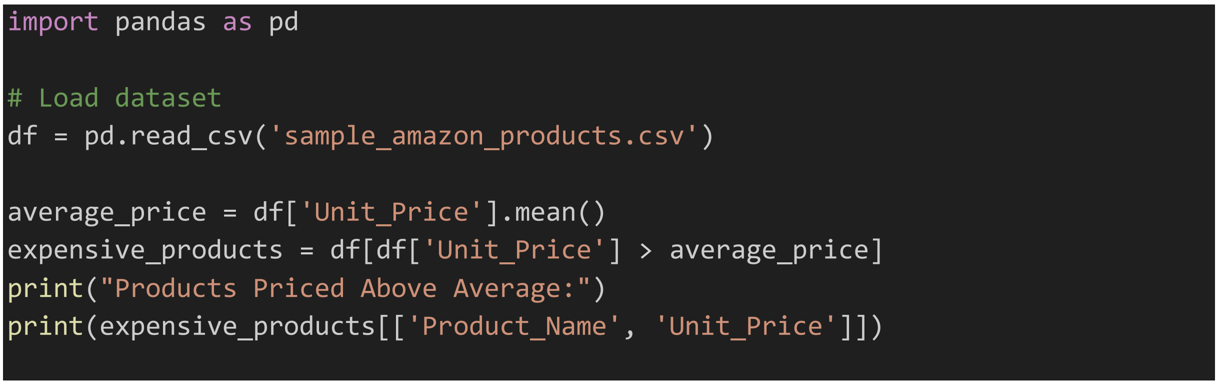


```python
df = pd.read_csv('sample_amazon_products.csv')

average_reviews = df['Reviews'].mean()
print(f"Average Number of Reviews: {average_reviews:.2f}")
```

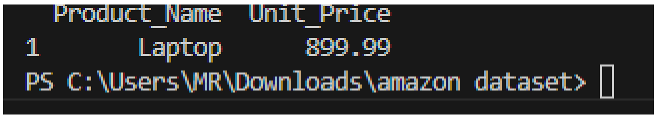


```
Average Number of Reviews: 850.00
PS C:\Users\MR\Downloads\amazon dataset> 
```

15) Identify products priced above the average unit price.

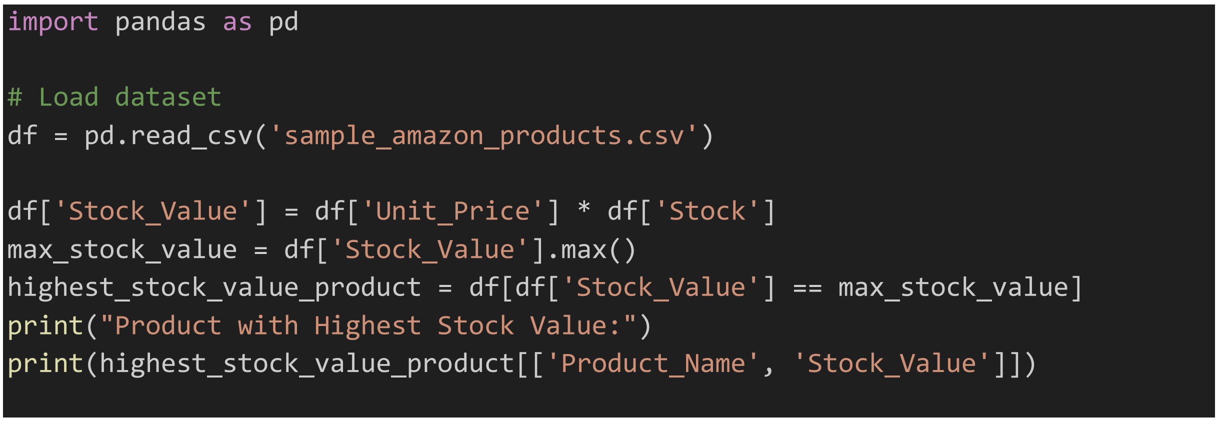


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

average_price = df['Unit_Price'].mean()
expensive_products = df[df['Unit_Price'] > average_price]
print("Products Priced Above Average:")
print(expensive_products[['Product_Name', 'Unit_Price']])
```

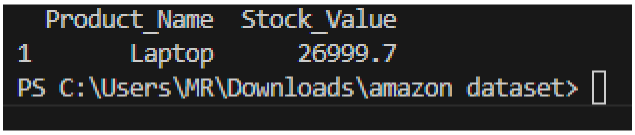


```
  Product_Name  Unit_Price
1       Laptop      899.99
PS C:\Users\MR\Downloads\amazon dataset> 
```

16) Which product has the highest stock value (Stock × Unit Price)?

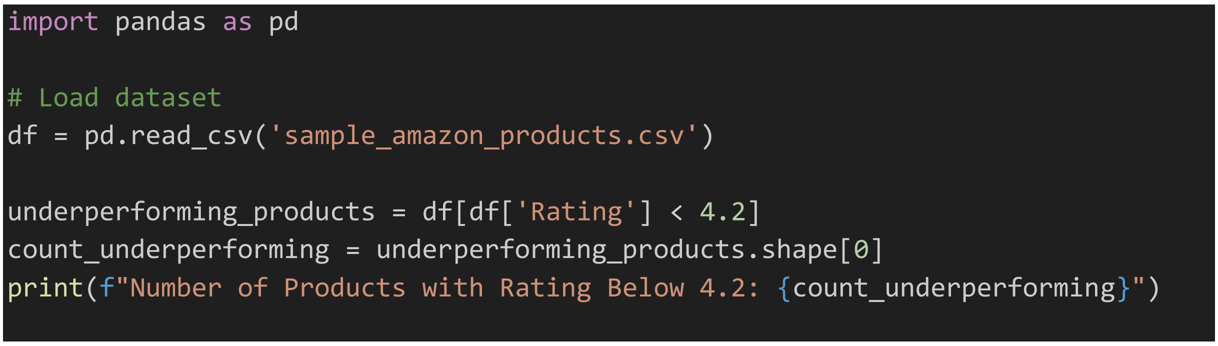


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

df['Stock_Value'] = df['Unit_Price'] * df['Stock']
max_stock_value = df['Stock_Value'].max()
highest_stock_value_product = df[df['Stock_Value'] == max_stock_value]
print("Product with Highest Stock Value:")
print(highest_stock_value_product[['Product_Name', 'Stock_Value']])
```

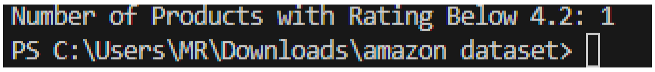


```
  Product_Name  Stock_Value
1       Laptop      26999.7
PS C:\Users\MR\Downloads\amazon dataset> 
```

17) Count how many products have a rating below 4.2

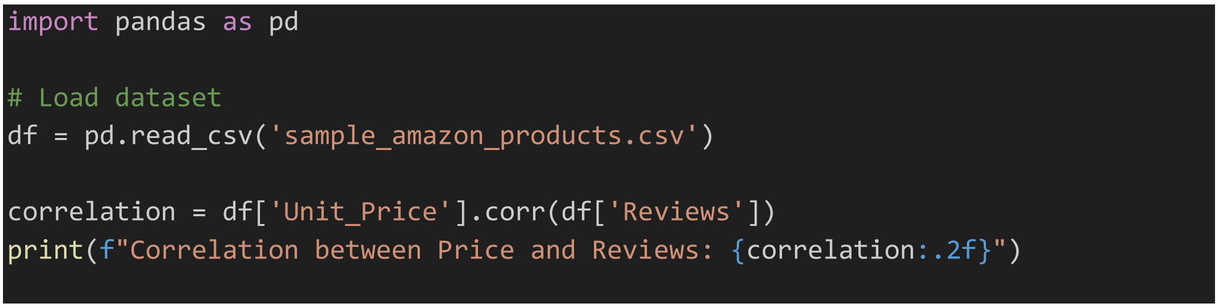


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

underperforming_products = df[df['Rating'] < 4.2]
count_underperforming = underperforming_products.shape[0]
print(f"Number of Products with Rating Below 4.2: {count_underperforming}")
```

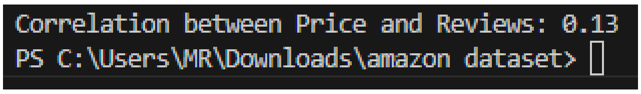


```
Number of Products with Rating Below 4.2: 1
PS C:\Users\MR\Downloads\amazon dataset> 
```

18) Calculate the correlation between price and number of reviews.

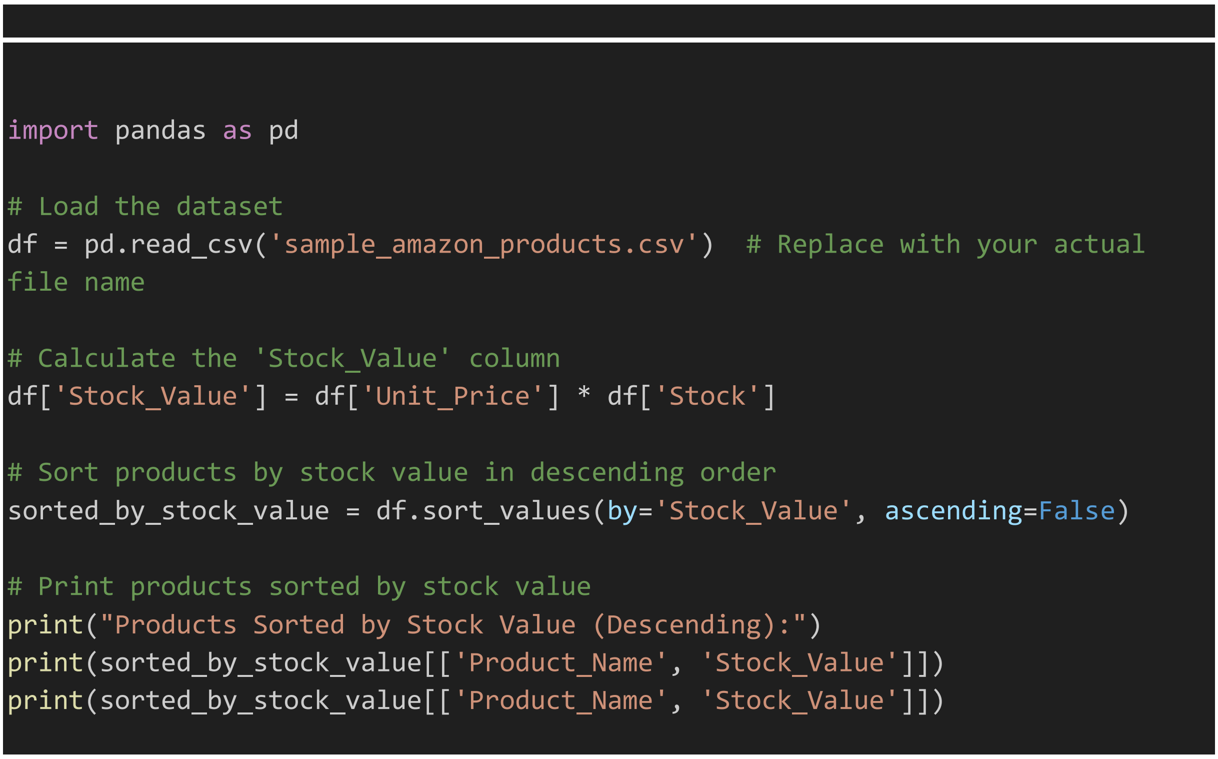


```python
import pandas as pd

# Load dataset
df = pd.read_csv('sample_amazon_products.csv')

correlation = df['Unit_Price'].corr(df['Reviews'])
print(f"Correlation between Price and Reviews: {correlation:.2f}")
```

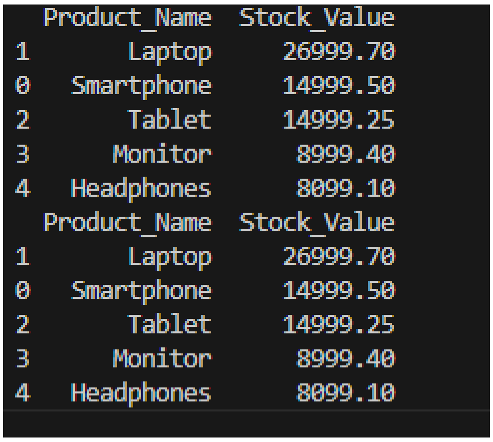


```
Correlation between Price and Reviews: 0.13
PS C:\Users\MR\Downloads\amazon dataset> 
```

19) Show products sorted by stock value (descending).



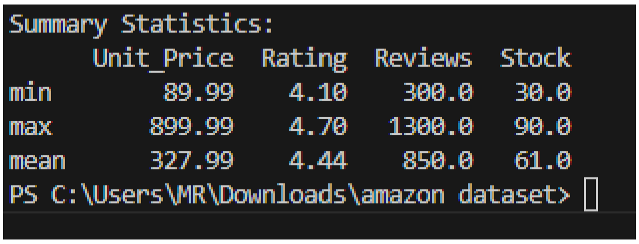
```python
import pandas as pd

# Load the dataset
df = pd.read_csv('sample_amazon_products.csv')  # Replace with your actual file name

# Calculate the 'Stock_Value' column
df['Stock_Value'] = df['Unit_Price'] * df['Stock']

# Sort products by stock value in descending order
sorted_by_stock_value = df.sort_values(by='Stock_Value', ascending=False)

# Print products sorted by stock value
print("Products Sorted by Stock Value (Descending):")
print(sorted_by_stock_value[['Product_Name', 'Stock_Value']])
print(sorted_by_stock_value[['Product_Name', 'Stock_Value']])
```



```
  Product_Name  Stock_Value
1       Laptop     26999.70
0   Smartphone     14999.50
2       Tablet     14999.25
3      Monitor      8999.40
4   Headphones      8099.10
  Product_Name  Stock_Value
1       Laptop     26999.70
0   Smartphone     14999.50
2       Tablet     14999.25
3      Monitor      8999.40
4   Headphones      8099.10
```

20) Create a summary DataFrame showing min, max, and mean for `Unit_Price`, `Rating`, `Reviews`, and `Stock`.

```
import pandas as pd

# Load the dataset
df = pd.read_csv('sample_amazon_products.csv')  # Replace with your actual
file name

summary = df[['Unit_Price', 'Rating', 'Reviews', 'Stock']].agg(['min', 'max',
'mean'])
print("Summary Statistics:")
print(summary)
```



```
Summary Statistics:
      Unit_Price  Rating  Reviews  Stock
min        89.99    4.10    300.0   30.0
max       899.99    4.70   1300.0   90.0
mean      327.99    4.44    850.0   61.0
PS C:\Users\MR\Downloads\amazon dataset>
```