Logan Samples

4/28/2025

Short Analysis of Electronic Retailer Product Data

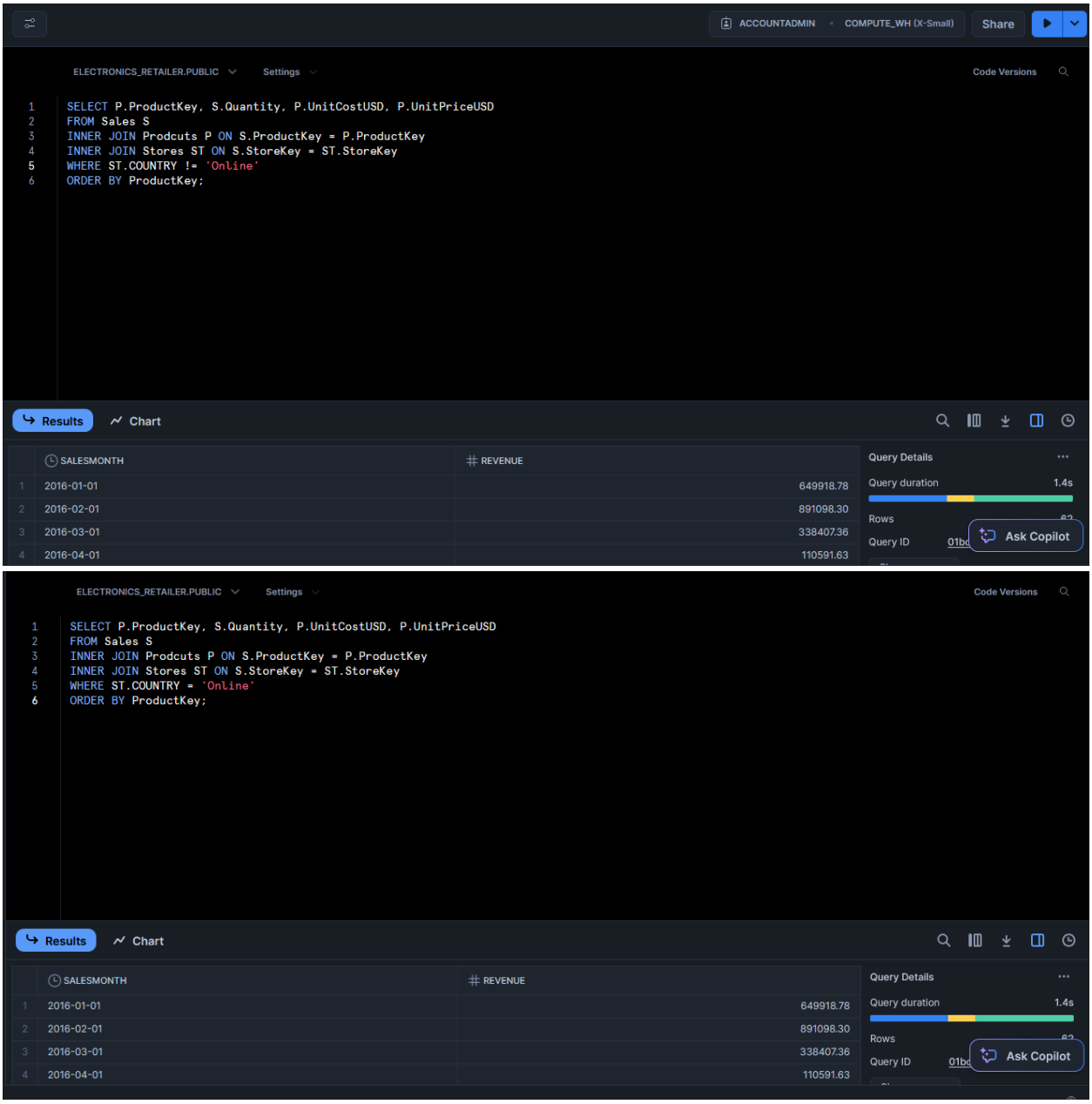
In my analysis, I was tasked with finding the top items an electronic retailer sold along with the countries that ordered the most units using data from MavenAnalytics. This is a large dataset containing five tables (Customers, Exchange Rates, Products, Sales, and Stores). This data was then imported into snowflake to conduct SQL queries.

I was able to find that computers were by far the best-selling product at 44151 units (22.3% of total sales) with TV and video products selling the least at only 11236 units (5.6% of total sales).

While examining which countries bought the most product. The United States is the biggest contributor by ordering 106,407 units. France is the lowest contributor only buying 5,385 units. From this it can be established that the United States is the strongest customer base followed by the the United Kingdom.

Through this analysis it can be concluded that computers are by far the best-selling product are computer products while the United States was the largest consumer. While computers outsold the other products by a large margin, it is important to note that computers may have lower margins and higher volume compared to the other products. Historically cheaper shipping cost within the United States may have provided more opportunities within that market.

Reference Pictures

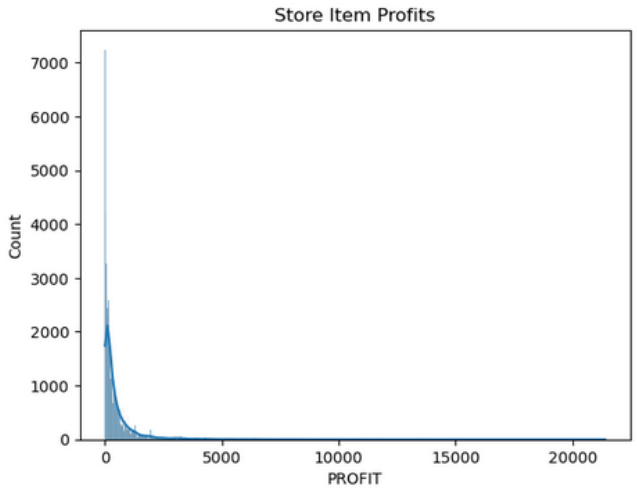


```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import mannwhitneyu
from scipy.stats import levene


Store_date_df = pd.read_csv('StoreSales.csv')
Online_data_df = pd.read_csv('OnlineStore.csv')
```

```
Store_data_df['PROFIT'] = (Store_data_df['UNITPRICEUSD'] - Store_data_df['UNITCOSTUSD']) * Store_data_df['QUANTITY']
```

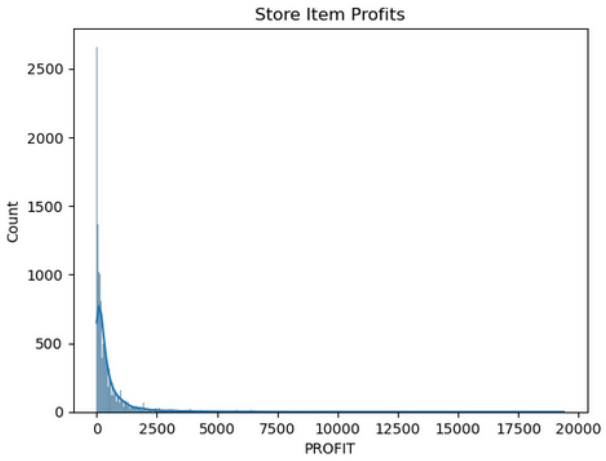
```
sns.histplot(Store_data_df['PROFIT'], kde=True)
plt.title("Store Item Profits")
plt.show()
```




```
Online_data_df['PROFIT'] = (Online_data_df['UNITPRICEUSD'] - Online_data_df['UNITCOSTUSD']) * Online_data_df['QUANTITY']
```

[102]:
```
Online_data_df['PROFIT'] = (Online_data_df['UNITPRICEUSD'] - Online_data_df['UNITCOSTUSD']) * Online_data_df['QUANTITY']
```

[103]:
```
sns.histplot(Online_data_df['PROFIT'], kde=True)
plt.title("Store Item Profits")
plt.show()
```



[107]:
```
stat, p = mannwhitneyu(Store_data_df['PROFIT'], Online_data_df['PROFIT'], alternative='two-sided')
print(f"U-statistic={stat:.3f}, p-value={p:.3f}")
```

```
U-statistic=331004953.000, p-value=0.044
```

[111]:
```
print("Store Median Profit:", Store_data_df['PROFIT'].median())
print("Online Median Profit:", Online_data_df['PROFIT'].median())
```

```
Store Median Profit: 195.52
Online Median Profit: 192.84
```

[123]:
```
def cliffs_delta(x, y):
    nx = len(x)
    ny = len(y)
    total = 0
    for xi in x:
        total += np.sum(xi > y) - np.sum(xi < y)
    return total / (nx * ny)

delta = cliffs_delta(Store_data_df['PROFIT'], Online_data_df['PROFIT'])

print(f"Cliff's Delta: {delta:.3f}")
```

```
Cliff's Delta: 0.011
```