Fall 2022 Data Science Intern Challenge

Ahmed Cheema

Question 1:

On Shopify, we have exactly 100 sneaker shops, and each of these shops sells only one model of shoe. We want to do some analysis of the average order value (AOV). When we look at orders data over a 30 day window, we naively calculate an AOV of $3145.13. Given that we know these shops are selling sneakers, a relatively affordable item, something seems wrong with our analysis.

a. Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

   The problem with the calculation of $3145.13 is that it simply calculates the mean of order_amount while ignoring total_items. In other words, a bulk purchase of 1000 items will be treated the same as a purchase of a single item.

b. What metric would you report for this dataset?

   I would report the sum of order_amount divided by the sum of total_items.

c. What is its value?

   The value is equal to $357.92.

Program (Python):

```
import pandas as pd

df = pd.read_csv('shopifySneaker.csv')

print("The average order value is equal to
${:.2f}.".format(df.order_amount.sum()/df.total_items.sum()))
```

Question 2:

a. **How many orders were shipped by Speedy Express in total?**

54 orders were shipped by Speedy Express in total.

```
SELECT ShipperName,COUNT(*) FROM (
SELECT Orders.OrderID, Shippers.ShipperName, Orders.ShipperID
FROM Orders
INNER JOIN Shippers ON Orders.ShipperID=Shippers.ShipperID
)
GROUP BY ShipperName
```

b. **What is the last name of the employee with the most orders?**

Peacock is the last name of the employee with the most orders with a total of 40 orders.

```
SELECT LastName,COUNT(*) FROM (
SELECT Orders.OrderID, Employees.LastName, Orders.EmployeeID
FROM Orders
INNER JOIN Employees ON Orders.EmployeeID=Employees.EmployeeID
)
GROUP BY LastName
ORDER BY COUNT(*) DESC
```

c. **What product was ordered the most by customers in Germany?**

Boston Crab Meat was ordered the most by customers in Germany with a total of 160 orders.

```
SELECT ProductName,SUM(Quantity)
FROM OrderDetails
   INNER JOIN Orders
   ON OrderDetails.OrderID = Orders.OrderID
   INNER JOIN Customers
   ON Orders.CustomerID = Customers.CustomerID
   INNER JOIN Products
   ON OrderDetails.ProductID = Products.ProductID
   WHERE Country == "Germany"
GROUP BY ProductName
ORDER BY SUM(Quantity) DESC
```