Raj Bajaj Shopify 2022 Data Science Internship Challenge

Q1.)

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

Ans. There are certain problems in our dataset that are skewing the observation. For example, shop 42 has 17 transactions of 2000 units each for $704,000 which heavily skews the data. There are several ways to evaluate the data:

- Use z-score to find the anomalies in data.
    - We will implement a z-score to see how each transaction differs from the norm. Then, we can filter out anomalies and see what went wrong with our analysis.
    - Looking at the fact that we need to have anomalies in our data, then we can use a different idea or a different method to find the Average Order Value (AOV). We can see that the third quartile was near the value.
- Removing anomalies after looking and analyzing data
    - We would need to analyze our data to find the values that skew our results. One way of doing so is using a box and whisker plot.
    - Look at each necessary data column, which is order_amount, total_items here.

b.) What metric would you report for this dataset?

Ans. As if we would just look at the plane data and see the mean or Average Order Value (AOV), then we would have our mean AOV of $3145.12 which is far from what it should be. There is a tendency to mean that any abnormal value can skew the result. Then for this problem, the third quartile result or 75% return is better to judge.

Another method to report this data is looking at the abnormalities with help of a z-score. Z-score is used here because it is the most common way to find anomalies. The score will help us with many irregular things.

Finally, if we are given permission to remove large anomalies from the data, then we can find the mean and give the result.

c.) What is its value?

Ans. The value depends on the way we report the data. If we want to report the data as it is formed, then we can report the 75% value which is the $390 Average Order Value.

We remove anomalies is permissible, then the Average Order Value can come out to be $300.16.

Q2.)

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

SQL –

```
SELECT Count(OrderID) as "Count of Speedy Express"
FROM Orders orders
JOIN Shippers shippers
ON orders.ShipperID = shippers.ShipperID
WHERE ShipperName = 'Speedy Express'
```

Ans. In total **54** orders were shipped by Speedy Express.

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

SQL –

```
SELECT LastName as "Employee with most order", Count(*) as "Number of orders"
FROM Orders orders
JOIN Employees employees
ON orders.EmployeeID = employees.EmployeeID
GROUP BY orders.EmployeeID
ORDER BY Count(*) DESC
LIMIT 1
```

Ans. Peacock was the name of the employee with most orders.

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

SQL –

```
SELECT products.ProductName as "Name of Product", SUM(order_detail.Quantity) as
"Number of Products"
FROM OrderDetails order_detail
LEFT JOIN Orders orders
ON orders.OrderID = order_detail.OrderID
LEFT JOIN Customers customers
ON orders.CustomerID = customers.CustomerID
LEFT JOIN Products products
ON products.ProductID = order_detail.ProductID
WHERE customers.Country = "Germany"
GROUP BY products.ProductID
ORDER BY SUM(order_detail.Quantity) DESC
LIMIT 1
```

Ans. Boston Crab Meat is the product ordered by most customers in Germany.