Name: Meet Gandhi

Part A

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.

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

Ans:

The naive calculation considered in the above question, considers the AOV as the sum of all order_amount over the total order's ie: 5000 which is (15725640/5000) giving us the value of $3145.13.

The problem with this approach is it considers the total order_amount for each orders with the number of transections happening irrespective to the total_items bought in a transection. By not considering the total_items in the formula the AOV is largely affected, for example when a regular customer comes to the shop they might buy one or two sneakers (at max, 8 here) similar to majority of the transections in the database. But when there is a wholesale purchase the total_items for one transection may go to 2000 sneakers where the order_amount increases exponentially distrupting the AOV as examplified in the order_id 16 and 16 other wholesale purchase. Also there are some outliers where a single order with a single item costs 25725$ which also plays an factor for distrupting AOV.

A better way:

1. We can use median instead of mean for average order value

2. Remove bulk orders and outliers and compute AOV with mean

2. What metric would you report for this dataset?

Ans:

Using mean for calculating AOV is not able to perform well as mean is not robust to outliers but using median performs better in such cases thus calculating the Median order value instead of average order value.

3. What is its value?

- Calculating the median after removing the outliers : $ 302.58
- Calculating the mead without removing the outliers: $ 284

Part B

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

Query: SELECT count() AS NumberOfOrders FROM [Orders] where ShipperID = (select ShipperID from Shippers where ShipperName like 'Speedy Express')

Ans: 54

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

Query: SELECT LastName FROM [Employees] where EmployeeID = (SELECT EmployeeID FROM orders GROUP by EmployeeID ORDER BY count(EmployeeID) desc limit 1)

Ans: Peacock

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

Query: SELECT ProductName FROM [Products] where ProductID = (SELECT ProductID FROM [OrderDetails] where OrderID in (SELECT OrderID FROM [Orders] where CustomerID in (SELECT CustomerID FROM [Customers] where country like "Germany")) group by ProductID order by count(*) desc limit 1)

Ans: Gorgonzola Telino