Name - Jay Patel

Email - [jayp81200@gmail.com](mailto:jayp81200@gmail.com)

Summer 2022 Data Science Intern Challenge

Question 1: Given some sample data, write a program to answer the following: click here to access the required data set 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.

b. What metric would you report for this dataset?

c. What is its value?

Solution Problem 1 (Check Out Code [Here](https://github.com/jaypatel-31/Shopify-Summer-2022-Data-Science-Intern-Challenge/blob/main/Summer%202022%20Data%20Science%20Intern%20Challenge.ipynb))

a)Average order value (AOV) is the average amount of money each customer spends per transaction with your store. We can calculate average order value using this simple formula:

average order value = Total revenue / number of orders i.e

AOV = Sum of all order amounts / Total number of orders

In our case , it can be calculated as

Total\_amount\_of\_all\_orders = df.order\_amount.sum()

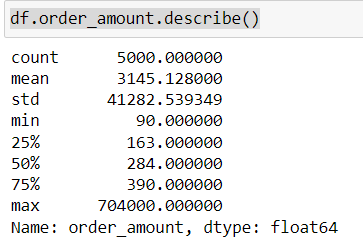
Total\_items\_ordered = df.total\_items.sum()

AOV = Total\_amount\_of\_all\_orders / Total\_items\_ordered

= 15725640 / 43936

= $357.92152221412965

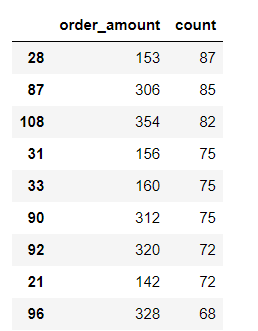
Lets understand what went wrong and how AOV of $3145.13 was wrongly calculated.



This means $3145.13 is actually the mean of all the order amounts.

A likely reason for the wrong calculation is that the author might have calculated the total\_items using count() function rather than using sum().Count() will only return the number of rows but sum() will return the total number of items.

b) I will suggest the median of the dataset as a metric.



This is the table in decreasing order wrt to count of different unique order amounts.Order of 153$ was ordered most(Mode).But counts are close to each other.There mode will not give any significance.Hence,if not AOV then median will be a important metric.

c) Its value is $284.

Question 2: For this question you’ll need to use SQL. Follow this link to access the data set required for the challenge. Please use queries to answer the following questions. Paste your queries along with your final numerical answers below.

1. How many orders were shipped by Speedy Express in total?
2. What is the last name of the employee with the most orders?
3. What product was ordered the most by customers in Germany?

Solution Problem 2

a)

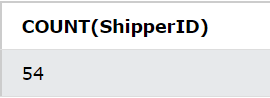
//Method 1 : Speedy Express has ID 1 which can help us getting number of orders without join

SELECT

COUNT(ShipperID)

FROM Orders

WHERE ShipperID == 1;



//Method 2 : Using Left Join

SELECT COUNT(OrderID)

FROM Orders

LEFT JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID

WHERE Shippers.ShipperName = 'Speedy Express'

There are 54 orders which were shipped by speedy express.



b)

SELECT LastName

FROM Employees

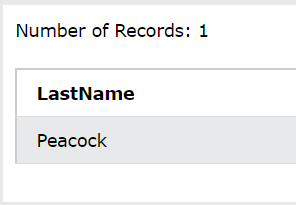
LEFT JOIN Orders

ON Employees.EmployeeID = Orders.EmployeeID

GROUP BY Orders.EmployeeID

ORDER BY COUNT(\*) DESC

LIMIT 1



Peacock is the last name of the employee with the most orders.

c)

SELECT [Products].ProductName, SUM(OrderDetails.Quantity) AS Total\_orders, Customers.Country

FROM [Products]

JOIN [OrderDetails] ON [OrderDetails].ProductID = [Products].ProductID

JOIN [Orders] ON [Orders].OrderID = [OrderDetails].OrderID

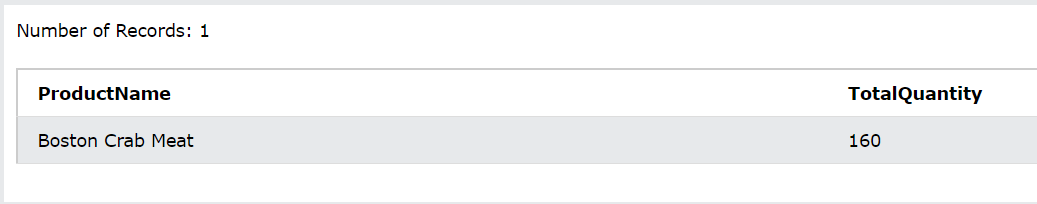
JOIN [Customers] ON [Customers].CustomerID = [Orders].CustomerID

WHERE [Customers].Country = "Germany"

GROUP BY [Products].ProductName

ORDER BY Total\_orders DESC

LIMIT 1;



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