Fall 2022 Data Science Intern Challenge

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.

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

- Order amount for shop_id 78 is very high
- 17 orders and 2000 items are being purchased for shop_id 42. This could be a mistake, this made AOV value to be very high, SO need to fixed

Question 1b: What metric would you report for this dataset?

- I tried mean and median to see which will be more accurate value.
- median would be best option and not end of having skewed data

Question 1c: What is its value?

- Tried both mean and median but, MEDIAN would be best option giving 284.0 for all value of the products.

SQL part:

Question 2: For this question you'll need to use SQL. <u>Follow this link</u> 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.

Question 2a. How many orders were shipped by Speedy Express in total?

```
SELECT S.ShipperName, COUNT(O.OrderID) AS Number_of_orders
FROM Orders AS O
LEFT JOIN Shippers AS S
ON O.ShipperID = S.ShipperID
GROUP BY ShipperName
HAVING ShipperName = 'Speedy Express';
```

Output: Speedy Express = 54

```
SELECT S.ShipperName, COUNT(O.OrderID) AS Number_of_orders
FROM Orders AS O
LEFT JOIN Shippers AS S
ON  O.ShipperID = S.ShipperID
GROUP BY ShipperName
HAVING ShipperName = 'Speedy Express';
```

Output:

Result:

Number of Records: 1

ShipperName	Number_of_orders
Speedy Express	54

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

```
SELECT E.LastName, Count(*) AS number_of_orders
FROM Orders AS O
JOIN Employees AS E
ON E.EmployeeID = O.EmployeeID
GROUP BY E.EmployeeID
ORDER BY number_of_orders DESC
LIMIT 1;
```

Output: Peacock = 40

```
SELECT E.LastName, Count(*) AS number_of_orders
FROM Orders AS O JOIN Employees AS E
ON E.EmployeeID = O.EmployeeID
GROUP BY E.EmployeeID
ORDER BY number_of_orders DESC
LIMIT 1;
```

Output:

Result:

Number of Records: 1

LastName	number_of_orders
Peacock	40

Question 2c: What product was ordered the most by customers in Germany?

SELECT P.ProductName, SUM(OD.Quantity) AS total FROM OrderDetails OD
JOIN Products AS P ON P.ProductID=OD.ProductID
JOIN Orders AS O
ON O.OrderID=OD.OrderID
JOIN (SELECT CustomerID FROM Customers WHERE Country="Germany") AS C
ON C.CustomerID=O.CustomerID
GROUP BY OD.ProductID
ORDER BY total DESC LIMIT 1;

Output: Boston Crab Meat = 160

```
SELECT P.ProductName, SUM(OD.Quantity) AS total FROM OrderDetails OD
JOIN Products AS P ON P.ProductID=OD.ProductID
JOIN Orders AS O ON O.OrderID=OD.OrderID
JOIN (SELECT CustomerID FROM Customers WHERE Country="Germany") AS C ON C.CustomerID=O.CustomerID
GROUP BY OD.ProductID
ORDER BY total DESC LIMIT 1;
```

Output:

Result:

Number of Records: 1

ProductName	total
Boston Crab Meat	160