Fall 2022 Data Science Intern Challenge

Please complete the following questions, and provide your thought process/work. You can attach your work in a text file, link, etc. on the application page. Please ensure answers are easily visible for reviewers!

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. The average is being skewed by very large orders for one of your shops. They are selling 2000 units at a time where all other shops are selling 8 or less unit per order. I sorted the order value column in descending order to see the 18 large order for shop 42.
b. What metric would you report for this dataset? Because there are 18 orders that are large in value it is skewing the mean of the order value. I instead looked at the median value of the order value since 50% of the orders would be less, and 50% would be higher.
c. What is its value? The median value for the order value is a more reasonable $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.

a. How many orders were shipped by Speedy Express in total? Answer is 54

SQL Statement:

```
SELECT Count(OrderID) FROM [Orders] WHERE ShipperID = 1
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 1

Count(OrderID)
54

b. What is the last name of the employee with the most orders? Employee's last name with the most orders is Peacock.

SQL Statement:

```
SELECT Orders.EmployeeID, Employees.FirstName, Employees.LastName, count(Orders.OrderID)
FROM Orders JOIN Employees on Orders.EmployeeID = Employees.EmployeeID
GROUP BY Orders.EmployeeID
ORDER BY count(Orders.OrderID) DESC;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 9

EmployeeID	FirstName	LastName	count(Orders.OrderID)
4	Margaret	Peacock	40
3	Janet	Leverling	31
1	Nancy	Davolio	29
8	Laura	Callahan	27
2	Andrew	Fuller	20
6	Michael	Suyama	18
7	Robert	King	14
5	Steven	Buchanan	11
9	Anne	Dodsworth	6

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

The product ordered the most by customers in Germany was the Boston Crab meat.

SQL Statement:

```
SELECT OrderDetails.ProductID, Products.ProductName, sum(OrderDetails.Quantity), Customers.Country FROM [OrderDetails] JOIN Products ON OrderDetails.ProductID = Products.ProductID JOIN
Orders ON OrderDetails.OrderID = Orders.OrderID JOIN Customers ON Orders.CustomerID = Customers.CustomerID
WHERE Customers.Country = 'Germany'
Group By OrderDetails.ProductID
ORDER BY sum(OrderDetails.Quantity) DESC;
```

Edit the SQL Statement, and click "Run SQL" to see the result.

Run SQL »

Result:

Number of Records: 45

ProductID	ProductName	sum(OrderDetails.Quantity)	Country
40	Boston Crab Meat	160	Germany
31	Gorgonzola Telino	125	Germany
23	Tunnbröd	105	Germany
35	Steeleye Stout	100	Germany
19	Teatime Chocolate Biscuits	95	Germany
72	Mozzarella di Giovanni	86	Germany