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.

In your calculation you included the outliers in calculating the average, that's why the average value was high. However, if you remove the outliers, i.e, shop-id : 42, 17 orders for 2000 quantities/order, $11968000 as given in the below screenshot, the average order quantity & value would be 2 and ~$754, respectively.

There are multiple ways to do this calculation:

1) With average, figure out stddev to understand the data spread
2) Categorize the orders in bulk or regular based on the standard deviation
3) Apply pivot to understand the data & calculating the results

total_items	(All)		
Order Category	**Order Count**	**Average of total ordered items**	**Average of order_amount**
Bulk Order	17	2000.0	704000
Regular Order	4983	2.0	754
order >=2000 units is a bulk order			

b. What metric would you report for this dataset?

As given above.

c. What is its value?

Avg order quantity = 2
Average amount = $754

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?

SQL Statement:

```
SELECT count(OrderID) as OrderCount from Orders o inner join Shippers s on o.ShipperID = s.ShipperID where
o.ShipperID=1
```

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

Run SQL »

Result:

Number of Records: 1

OrderCount
54

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

SQL Statement:

```
SELECT [LastName] from (select count(OrderID) as [countMax], e.LastName FROM [Orders] o inner join Employees e on
o.EmployeeID = e.EmployeeID group by o.EmployeeID
order by count(OrderID) desc ) as ordercount limit 1
```

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

Run SQL »

Result:

Number of Records: 1

LastName
Peacock

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

SQL Statement:

```
SELECT p.ProductName from [Orders] o inner join OrderDetails od on o.OrderID=od.OrderID
inner join Products p on od.ProductID = p.ProductID where CustomerID in (select CustomerID FROM [Customers] where
Country='Germany') group by p.ProductID order by od.Quantity desc limit 1
```

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

Run SQL »

Result:

Number of Records: 1

ProductName
Steeleye Stout