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. **It comes out that there are several rows with anomalous values. In particular the shop with the id:42, inserted many equal values with order_amount of 704000, total_items 2000, at 4 am. This leads me to think that there is a problem in the computer system of this shop**
b. What metric would you report for this dataset? **In these cases it is preferable to use as metric the median instead of the mean, that is useful in case of dataset with evident outilers**
c. What is its value? **The value of the median of this dataset for the column "order_amount" 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.

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

```
SELECT count(*)
FROM Orders inner join Shippers on Shippers.ShipperID = Orders.ShipperID
WHERE ShipperName like "Speedy Express"
```

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

```
SELECT  count(OrderID), LastName
FROM Orders inner join Employees on Employees.EmployeeID =
Orders.EmployeeID
GROUP BY  LastName
ORDER BY count(OrderID) desc
LIMIT 1
```

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

```
SELECT
  Products.ProductName,
  Products.ProductID,
  Count(*)
FROM
  (
    (
      Products
      inner join OrderDetails on Products.ProductID = OrderDetails.ProductID
    ) as OrdersDetails_with_Products
    inner join Orders on Orders.OrderID = OrdersDetails_with_Products.OrderID
  ) as OrdersDetails_with_Products_and_Orders
  inner join Customers on Customers.CustomerID =
OrdersDetails_with_Products_and_Orders.CustomerID
where
  Country like "Germany"
group by
  Products.ProductID
Order by
  Count(*) desc
limit
  1
```