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.**

Based on the analysis conducted, what could be wrong with your calculation is that it does not consider the different types of customers. Specifically, most customers place orders with less than ten items, but just a few orders recorded with 2,000 items will increase the average order value dramatically.

A better way to evaluate this data, would be to, first, investigate the total item distribution. A visual scatter plot and a numerical distribution will clearly show that there are a few major outlying orders with 2,000 items in each order. Due to these outliers, I would choose to analyze the orders with eight items or less separate from those orders with 2,000 items in the order. For scaling purposes, I will consider orders with less than 100 items a small order and orders with 100 items or more a large order. Two separate average order value metrics should be calculated because these two groups of customers have differing buying patterns.

b. **What metric would you report for this dataset?**

I would report the average order value for all orders with less than 100 items, as well as an average order value for all orders with 100 items or more.

c. **What is its value?**

SMALL ORDERS (<100 ITEMS) - AVG ORDER VALUE: $754.09
LARGE ORDERS (≥100 ITEMS) - AVG ORDER VALUE: $704000.00

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?**

```
SELECT COUNT(OrderID) orders_shipped
FROM Orders
        JOIN Shippers using (ShipperID)
WHERE ShipperName = 'Speedy Express';
```

Answer: 54

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

```
SELECT LastName
FROM Employees
        JOIN Orders USING (EmployeeID)
GROUP BY LastName
ORDER BY COUNT(OrderID) DESC
LIMIT 1;
```

Answer: Peacock

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

```
SELECT ProductName
FROM Products JOIN OrderDetails using (ProductID)
        JOIN Orders using (OrderID)
   JOIN Customers using (CustomerID)
WHERE Country = 'Germany'
GROUP BY ProductName
ORDER BY COUNT(ProductID) DESC
LIMIT 1;
```

Answer: Gorgonzola Telino