Fall 2021 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.

Doing a quick plot of the order amount column, we can see that doing average on the order_amount column is a problem because there are some really high outliers, especially a bunch at $700,000.

b. What metric would you report for this dataset?

A better evaluation could be average item value. Or break down the order value by total items.

c. What is its value?

The average item value without any filtering is: $ 387.74
After removing outliers, the average item value is: $ 152.48

Please see the notebook for code, visualization and process.

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?

After examining the data, we will need the Orders and Shippers table to answer this question.

```
SELECT COUNT(*)
FROM Orders O
INNER JOIN Shippers S ON (O.ShipperID = S.ShipperID)
WHERE S.ShipperName = 'Speedy Express'
```

54 orders were shipped by Speedy Express.

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

```
SELECT LastName
FROM Orders O
INNER JOIN Employees E ON (O.EmployeeID = E.EmployeeID)
GROUP BY O.EmployeeID
ORDER BY COUNT(*) DESC
LIMIT 1
```

The employee with the last name Peacock had the most orders.

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

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

Boston Crab Meat was ordered the most by quantities at 160 pieces.