Summer 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.

By creating a new column at the end of the initial data set, we can divide the total order value by the number of items bought. We can then take an average of that new column. Instead of calculating the AOV, we can simply calculate the average item value.

b. What metric would you report for this dataset?

The metric would simply be AIV, short for average item value.

c. What is its value?

The AIV for the required data set is about $387.74

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 orders**

```
SELECT ShipperID, COUNT(*) FROM Orders GROUP BY ShipperID;
```

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

```
SELECT *, COUNT(*) AS Num_orders FROM Orders JOIN Employees ON Orders.EmployeeID = Employees.EmployeeID
GROUP BY 3 ORDER BY COUNT(*) desc LIMIT 1;
```

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

```
CREATE TABLE CustomerPerOrder AS SELECT * FROM Customers JOIN Orders ON Customers.CustomerID = Orders.CustomerID;

CREATE TABLE BriefOrders AS SELECT CustomerID, Country, OrderID FROM CustomerPerOrder;

CREATE TABLE OrdersCountry AS SELECT * FROM BriefOrders JOIN OrderDetails ON BriefOrders.OrderID = OrderDetails.OrderID WHERE Country = 'Germany';

SELECT OrderID, ProductID, Quantity FROM [OrdersCountry] GROUP BY ProductID ORDER BY Quantity desc;

SELECT * FROM Products WHERE ProductID = 35;
```