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!

Rishabh Kansal, rk793@cornell.edu

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.
> b. What metric would you report for this dataset?
> c. What is its value?

The answers to these are on Github, in the R-markdown file.

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?**

They shipped 54 orders.

```
SELECT COUNT(*) FROM [Orders]
WHERE ShipperID = 1
```

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

LastName = Peacock has the most orders.

```
SELECT E.LastName as last_name, COUNT(O.OrderID) as num_orders
FROM Employees E join Orders O ON E.EmployeeID = O.EmployeeID
GROUP BY O.EmployeeID
ORDER BY num_orders DESC;
```

Number of Records: 9

last_name	num_orders
Peacock	40
Leverling	31
Davolio	29
Callahan	27
Fuller	20
Suyama	18
King	14
Buchanan	11
Dodsworth	6

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

Product number 31 was ordered the most (by number of orders) by customers of Germany.

```
CREATE VIEW CUSTOMERS_GERMANY AS
SELECT CustomerID FROM [Customers]
WHERE Country = "Germany";

CREATE VIEW ORDERS_GERMANY AS
SELECT * FROM [Orders]
WHERE CustomerID in CUSTOMERS_GERMANY

SELECT OD.ProductID, COUNT(O.OrderID) as num_orders
FROM OrderDetails OD JOIN ORDERS_GERMANY O on OD.OrderID = O.OrderID
GROUP BY OD.ProductID
ORDER BY num_orders DESC;
```

Result:

Number of Records: 45

ProductID	num_orders
31	5
76	4
40	4
72	3
36	3

Doing the same analysis, by gross amount of a particular product being ordered. We can easily tweak the third block of the above code and get:

```
SELECT ProductID, SUM(Quantity) as amount
FROM OrderDetails OD JOIN ORDERS_GERMANY O on OD.OrderID = O.OrderID
GROUP BY OD.ProductID
ORDER BY amount DESC
```

We can see that going by quantity, german customers have ordered the most units of porductID 40.

Result:

Number of Records: 45

ProductID	amount
40	160
31	125
23	105
35	100