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.
b. What metric would you report for this dataset?
c. What is its value?
Answers please check Github

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?
ANSWER: 54
SELECT COUNT(*)
FROM ORDERS O
JOIN SHIPPERS S
ON O.SHIPPERID = S.SHIPPERID
WHERE S.SHIPPERNAME =='Speedy Express'
b. What is the last name of the employee with the most orders?
ANSWER: Buchanan
SELECT LASTNAME
FROM EMPLOYEES E
JOIN (SELECT EMPLOYEEID, COUNT(*)
FROM ORDERS

```
ORDER BY COUNT(*) DESC
LIMIT 1
) A
ON E.EMPLOYEEID = A.EMPLOYEEID
```

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

```
SELECT PRODUCTNAME
FROM PRODUCTS P
JOIN
(SELECT PRODUCTID, COUNT(*)
FROM ORDERS O
JOIN CUSTOMERS C
ON O.CUSTOMERID = C.CUSTOMERID
JOIN ORDERDETAILS OD
ON O.ORDERID = OD.ORDERID
WHERE C.COUNTRY == 'Germany'
GROUP BY PRODUCTID
ORDER BY COUNT(*) DESC
LIMIT 1) A
ON P.PRODUCTID = A.PRODUCTID
```