

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?

Solution 1 :

Collab Notebook Link : -

<https://colab.research.google.com/drive/1A8FZ9-H3XJ0DFvuGyRuX3-e3UFLfxs2I?usp=sharing>

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.

Solution 2 :

- a. How many orders were shipped by Speedy Express in total?

Query :

```
SELECT s.ShipperName,count(*)
FROM Orders o
JOIN Shippers s
      ON s.ShipperID = o.ShipperID
GROUP BY s.ShipperName
      HAVING s.ShipperName = 'Speedy Express';
```

[illegible]

```
SELECT count(*)
FROM Orders o
WHERE o.ShipperID =
      ( SELECT s.ShipperID
        FROM Shippers s
        WHERE s.ShipperName = 'Speedy Express');
```

Result : 54

Number of Records: 1

count(*)	ShipperName
54	Speedy Express

Note : Since we need to find the value of total orders by Speedy Express, option 2 is a better query as it does not involve JOIN, thus less computation.

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

Query:

```
SELECT MAX(OrderCount) as Max, lastname  
FROM  
    (SELECT e.lastname, count(*) as OrderCount  
    FROM Orders o  
    JOIN Employees e  
        ON e.EmployeeID = o.EmployeeID  
    GROUP BY e.EmployeeID);
```

Result: Peacock

Number of Records: 1

lastname	Max
Peacock	40

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

Query:

```
SELECT ProductName,max(countProd)
FROM
    (SELECT p.ProductName, count(*) as countProd
    FROM Orders o
    JOIN Customers c
        ON o.CustomerId = c.CustomerId
    JOIN OrderDetails d
        ON o.orderID = d.orderID
    JOIN Products p
        ON p.productID = d.productID
    WHERE c.country = "Germany"
    GROUP BY d.productID)
```

[illegible]

```
SELECT * FROM
(SELECT p.ProductName, count(*) as countProd
FROM Orders o
JOIN Customers c
      ON o.CustomerId = c.CustomerId
JOIN OrderDetails d
      ON o.orderID = d.orderID
JOIN Products p
      ON p.productID = d.productID
WHERE c.country = "Germany"
GROUP BY d.productID
ORDER BY countProd DESC)
LIMIT 1
```

Result: Gorgonzola Telino

Number of Records: 1

ProductName	max(countProd)
Gorgonzola Telino	5