

Winter 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.
 - Just calculating average order value for all the stores includes outliers which might show higher/lower average for all the stores.
 - Looking at AOV also didn't take into account the number of shoes sold in each order, increasing the AOV linearly.
 - Another way we can evaluate this data is by finding the median order value which is the majority of AOV or average value per item.
- b. What metric would you report for this dataset?
 - Average order value per item.
- c. What is its value?

\$387.74. The analysis is done in the Notebook in this GitHub repo.

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?

```
SELECT COUNT(DISTINCT OrderID) FROM Orders
WHERE ShipperID = 1;
```

Total orders that were shipped by Speedy Express is 54.

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

```
SELECT Employees.LastName, COUNT(Orders.OrderID) FROM Employees
INNER JOIN Orders ON Employees.EmployeeID = Orders.EmployeeID
GROUP BY Employees.LastName
ORDER BY COUNT(Orders.OrderID) DESC
LIMIT 1;
```

Peacock - 40

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

```
SELECT OrderDetails.ProductID, Products.ProductName, COUNT(*)
FROM Orders
LEFT JOIN Customers ON Orders.CustomerID=Customers.CustomerID
LEFT JOIN OrderDetails ON OrderDetails.OrderID=Orders.OrderID
LEFT JOIN Products on Products.ProductID=OrderDetails.ProductID
WHERE Country = 'Germany'
GROUP BY OrderDetails.ProductID
ORDER BY 3 Desc
LIMIT 1;
```

The product that was ordered the most by customers in Germany is Gorgonzola Telino.