

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.

Ans: Incorrect AOV calculation is most likely due to use of count() function in calculation of total numbers of items. count() function will only count the number of rows, we need to use sum() function to get the total number of items.

- b. What metric would you report for this dataset?

Ans: To get the accurate AOV we need to get total number of orders and total number of items and then get AOV by dividing them.

```
total_order_amount = df['order_amount'].sum()
total_items = df['total_items'].sum()
aov = total_order_amount / total_items
```

```
print("Average Order Value: $%.2f" % (aov))
```

- c. What is its value?

\$357.92

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

Query -

```
SELECT count(*) AS Orders
FROM Orders
JOIN Shippers
ON Orders.ShipperID == Shippers.ShipperID
WHERE ShipperName == "Speedy Express"
```

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

Answer - Peacock

Query -

```
SELECT LastName FROM Orders
LEFT JOIN Employees
ON Orders.EmployeeID == Employees.EmployeeID
GROUP BY Orders.EmployeeID
ORDER BY Count(*) DESC
LIMIT 1
```

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

Answer -

Product Name : Camembert Pierrot

Product ID : 60

Query -

```
SELECT Products.ProductID, Products.ProductName, (COUNT(*) *
OrderDetails.Quantity) AS TotalOrders
FROM Orders, OrderDetails
JOIN Customers
ON Orders.CustomerID == Customers.CustomerID
JOIN Products
ON OrderDetails.ProductID == Products.ProductID
WHERE Country == "Germany"
GROUP BY Products.ProductID
ORDER BY TotalOrders DESC
LIMIT 1
```

