Fall 2021 Data Science Intern Challenge

(ANSWERS)

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?

Answer 1:

a. The average order value is high because of unusual high volume transactions that occurred in the 30 day window being considered. To get a realistic AOV value, an option will be to remove outliers (Q3 + 1.5 interquartile range = $730.5) and re-calculating the AOV. Alternatively, it might be best to report the median value for this analysis.
b. We should report the median value or the adjusted mean value.
c. Median Order value = $284 or Adjusted AOV = $280

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?
b. What is the last name of the employee with the most orders?
c. What product was ordered the most by customers in Germany?

Answers:

a. Orders shipped by Speedy Express

```
SELECT a.ShipperName,b.NumOfOrders FROM Shippers a INNER
JOIN (SELECT ShipperID, Count(OrderDate) AS NumOfOrders FROM Orders
GROUP BY ShipperID) b
ON a.ShipperID = b.ShipperID
WHERE ShipperName = 'Speedy Express';
```

b. The last name of the employee with most orders?

```
SELECT b.LastName FROM
(SELECT EmployeeID, COUNT(OrderID) AS NumOfOrders FROM ORDERS
GROUP BY EmployeeID
ORDER BY NumOfOrders DESC) a INNER JOIN (SELECT EmployeeID, LastName FROM
Employees) b
ON a.EmployeeID = b.EmployeeID
LIMIT 1;
```

c. The product most ordered by customers in Germany

```
SELECT y.ProductName FROM
(SELECT a.CustomerID,b.OrderID FROM Customers a
INNER JOIN Orders b on a.CustomerID = b.CustomerID
WHERE a.Country = 'Germany') x
INNER JOIN
(SELECT a.OrderID, b.ProductName FROM OrderDetails a
INNER JOIN Products b
ON a.ProductID = b.ProductID) y
ON x.OrderID = y.OrderID
GROUP BY y.ProductName
ORDER by Count(y.ProductName) DESC
LIMIt 1;
```