Winter 2021 Data Science Intern Challenge

ANDI OSIKA 602-214-8360 andioiksa@gmail.com

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.
    i. Outliers could be corrupting the data.
    ii. A quick inspection finds 17 orders for $704K and 2000 items each.

    iii. In addition, there are 17 orders for $25,725 each - for one item and
        1. 14 orders for double for two items definitely skewing the metric in addition to two more orders that contain multiples of $25,725 -
        2. 1 for $154,350 (6 items @ $25,725) and
        3. 1 for $102900 (4 items @ $25,725)

    Clearly there's a pattern around the $25,725 value that needs to be investigated.

b. What metric would you report for this dataset?
    If you adjusted your AOV by removing the outliers above you lose an acceptable 1.2% of the data. You'd use an Adjusted Average Order Value
c. What is its value?
    The AAOV is $302.58

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: 196

Shipper ID 1 = Speedy Express : Query Below:

```
SELECT COUNT(ShipperID = 1) FROM [Orders];
```

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

Answer: Peacock - 40 orders

Query Below:

```
SELECT COUNT(O.OrderID) as OrderCount, E.LastName FROM [Orders] AS O
JOIN [Employees] AS E WHERE O.EmployeeID == E.EmployeeID
GROUP BY(LastName)
ORDER BY OrderCount
```

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

Answer: Boston Crab Meat

Query Below:

```
SELECT SUM(od.Quantity) as TotalPurchased, p.ProductName FROM
[OrderDetails] od JOIN  [Orders] o
ON od.OrderID = o.OrderID
JOIN [Products] p
on od.ProductID = p.ProductID
JOIN [Customers] c
ON o.CustomerID = c.CustomerID
WHERE [Country] = 'Germany'
GROUP BY p.ProductID
ORDER BY (TotalPurchased)
```