Summer 2022 Data Science Intern Challenge Solution

Solution by: Akash Dave

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.

- **The problem with calculating simple average or mean is that it is susceptible to outliers which can deviate the result and as the given dataset is right-skewed having a lot of outliers I would not suggest using mean to evaluate the data.**

b. What metric would you report for this dataset?

- **As there are many outlier in the dataset, I would suggest to use median instead of mean to know the average order value**

c. What is its value?

- **Before removing outliers or any cleaning of the data, the average order value (using median) is $284.0**

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?

- **54**
- **Query:-**

```
- SELECT count(OrderID)
- FROM Orders
- INNER JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID
- WHERE ShipperName is 'Speedy Express' ;
```

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

- **Peacock**
- **Query:-**

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

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

- **Steeleye Stout**
- **Query :-**

```
- Select od.ProductID, ProductName, Quantity
- From OrderDetails as od
- INNER JOIN Products as p on od.ProductID=p.ProductID
- WHERE  OrderID IN (
-           SELECT OrderID FROM Orders as o
-           INNER JOIN (
-           SELECT CustomerID FROM Customers
-         WHERE Country is 'Germany'
-      ) as g
-      ON o.CustomerID=g.CustomerID
- )
- ORDER BY Quantity DESC LIMIT 1
- ;
-
```