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.

Solution Approach 1: After doing some exploratory data analysis, we have all the products with average price of 152.47 while we have an outlier product sold at Shop_id 78 with cost of 25725 and while all product lie between 90-325 this product seem to be not part of sneaker so after removing this anomaly we got AOV of $2717.36 and if we remove bulk sales too then we have AOV of $302.58 which is quite realistic

Solution Approach 2: Another approach I understand that we can divide our products in to 3 broad categories as: common products (value 90-352), High Value products (value 25725) and bulk sale (product sold in large quantity) which make AOV for each category more realistic as: Common products now have AOV of $302.58 while bulk and high value products have AOV of $704000 & $49213.04 respectively.

b. What metric would you report for this dataset?

Solution Approach: I will report customer lifetime value, shop wise average sales and weekly sales for 3 categories described above.

c. What is its value?

Solution Approach:
Customer lifetime value will help us segment different groups of customers like frequent buyers, bulk buyers and high end product buyers.
Shop wise avg sales will help us understand the demand of certain type of sneakers as each shop sale only 1 model and its weekly trend this will help us model and forecast for future sales and keep our inventory accordingly
As described on answer a., we can also model and forecast demand for different categories like bulk buyers, High Value products and common products.

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 orders were shipped by Speedy Express

Query:

```
SELECT a.ShipperID, b.ShipperName, count(a.orderID) as order_count FROM Orders as a
left join Shippers as b on a.ShipperID = b.ShipperID
Where ShipperName = 'Speedy Express'
group by a.ShipperID
```

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

Peacock

Query:

```
SELECT a.EmployeeID, b.LastName, count(a.orderID) as order_count FROM Orders as a
left join Employees as b on a.EmployeeID = b.EmployeeID
group by a.EmployeeID
order by order_count desc
limit 1
```

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

Boston Crab Meat with 4 orders 160 in quantity, followed by Gorgonzola Telino with 5 orders and 125 in quantity.

Query:

```
WITH customer_details (orderID, customerID, customername, country)
AS
(select a.orderID, a.customerID, b.customername, b.country
  FROM Orders as a
  left join customers as b on a.customerID = b.customerID ),

  product_details (OrderDetailID, orderID, ProdictID, Quantity, ProductName)
AS
(select a.OrderDetailID, a.orderID, a.ProductID, a.Quantity, b.ProductName
  FROM OrderDetails as a
  left join products as b on a.ProductID = b.ProductID )

Select a.ProductName, count(a.orderDetailID) as order_count, sum(a.quantity) as quantity_count
from product_details as a
left join customer_details as b on a.orderID = b.orderID
where b.country = 'Germany'
group by a.ProductName
order by quantity_count desc
```