Shopify Fall 2021 Data Science Intern Challenge

Peter L’Oiseau

4/28/2021

## Question 1: Shoe Sales

### A

##### The issue in the understanding of this metric is that AOV is the average price of all orders and not shoes which an order can comprise of many. There are bulk orders of shoes (2000 pairs) which greatly skews this metric. If one divides the sum of order\_amount by the sum of total\_items, then we have a measure of the average price of a shoe sold.

### B

##### I think a better way to understand the central tendacy of this data set is to take the median value for the average shoe price on an order. The bulk orders from shop\_id 42 skew the order\_amount column greatly but even when divided by the number of shoes sold, there is still a big outlier. shop\_id 78 sells a shoe which is priced in excess of $25,000 where the next highest selling shoe is $352. Using the median of the average price of a shoe sold on an order will give a better metric for understanding the most common sales. Note, we assume that each shoe on an order is the same price but this is granted because each order comes from one shop which only sells one model of shoe. This is different than the median price of a shoe however because the bulk sales are given as much weight as any other order. I think this is fair because it gives insight into the most common type of orders made but a median and mean of all shoes sold are also valuable metrics to understand along with the context of bulk and premier orders.

### C

##### The value of the median average shoe price on an order is $153 a much more afordable and common purchase type for consumers.

## Question 2: SQL

### A Query:

### SELECT COUNT(ord.OrderID) AS Total

### FROM Orders AS ord

### INNER JOIN Shippers AS shp ON ord.ShipperID=shp.ShipperID

### WHERE shp.ShipperName="Speedy Express";

### Output:

#### Total: 54

### B Query:

##### SELECT TOP 1 emp.LastName, COUNT(ord.OrderID) AS Total

##### FROM Employees AS emp

##### INNER JOIN Orders AS ord ON emp.EmployeeID=ord.EmployeeID

##### GROUP BY emp.EmployeeID, emp.LastName

##### ORDER BY COUNT(ord.OrderID) DESC;

### Output:

##### LastName: Peacock Total: 40

###### Note question requires last name but prompt requests numerical answer so both are included

### C Query:

##### SELECT TOP 1 prod.ProductName, prodsum.Total FROM

##### (SELECT ord.ProductID, SUM(ord.Quantity) AS Total FROM

##### (SELECT ord.OrderID, ord.CustomerID, odets.ProductID, odets.Quantity

##### FROM OrderDetails AS odets

##### INNER JOIN Orders AS ord ON odets.OrderID=ord.OrderID) AS ord

##### INNER JOIN Customers AS cust ON ord.CustomerID=cust.CustomerID

##### WHERE cust.Country="Germany"

##### GROUP BY ord.ProductID) AS prodsum

##### INNER JOIN Products AS prod ON prodsum.ProductID=prod.ProductID

##### ORDER BY prodsum.Total DESC;

### Output:

##### ProductName: Boston Crab Meat Total: 160

###### Note question requires product name but prompt requests numerical answer so both are included