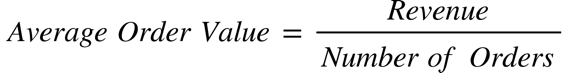
Question 1.

1. First, to understand the question we must look at the formula to compute the average order value.



Revenue is calculated based on profits – cost but in this dataset, we are only given profits within the total\_amount column. Using the formula above, we can see that the value $3145.13 was found based on just computing mean of the total\_amount column in the dataset.

Using EDA, we can view what is creating such an anomaly in the average order value. Based on the dataframe, we can see that there are large datapoints that is skewing the average higher.

1. The mean as a measurement of central tendency is one of three ways that gives an insight towards our users. However, in this dataset using the mean is not a viable method as there are multiple large outliers within the dataset. When this occurs, we should look towards using the median since it isn’t easily affected by large outliers.
2. The median order value of this single shoe across 100 stores during the time frame of March is $284

Question 2.

1. First, I looked at the Shippers table which provides the names of each shipper and their associated ID. After, I looked at the OrderDetails table but noticed that this did not provide any relevant information to answer the question. I then looked at the Orders table which has the OrderID and the ShipperID. Once finding this useful information, I selected ShipperName and counted the ShipperID to find how much product each company has shipped out. I joined the two tables based on the common ShipperID and filtered out the results based on Speed Express to find that they have shipped out 54 orders.

SELECT Shippers.ShipperName, COUNT(Orders.ShipperID) as TotalShipped

FROM (Orders

INNER JOIN Shippers ON Shippers.ShipperID = Orders.ShipperID)

GROUP BY Shippers.ShipperName

HAVING ShipperName = 'Speedy Express';

1. To solve this question, I first looked at the Employees Table to find the Employee ID. I noticed that this table is important as it linked the EmployeeID to the person’s Last name. Secondly, I looked at the Orders table to see that it shares an EmployeeID column with the Employees table. These two pieces are important to answer the question so therefore I join the two tables together. I used count function on the EmployeeID from the Orders table to see how many sales each employees had. Lastly, I order the table to descending order to see that Peacock has the most sales out of all employees with a total of 40.

SELECT Orders.EmployeeID, Employees.LastName, COUNT(Orders.EmployeeID) as TotalSales

FROM (Orders

INNER JOIN Employees ON Employees.EmployeeID = Orders.EmployeeID)

GROUP BY Orders.EmployeeID, Employees.LastName

ORDER BY COUNT(Orders.EmployeeID) DESC;

1. Using the same methods above, I first check the tables for relevant information to answer each question. What I found was that the customers table had the CustomersID and Country column which is useful to determine germany’s sales. The CustomerID column is also found in the Orders table which links each customer order with their OrderID. Next I found that the OrderDetails table shares the OrderID column with the Orders table and a productID columns which is useful for identifying the products sold. Lastly, I looked at the Products table to find that it shares a common (productID) column with OrderDetails and links these productIDs to the products name. I selected the necessary columns between all three tables and join these tables together to filtered for products sold in Germany. The most sold product in Germany is the Gorgonzola Telino.

SELECT Products.ProductName, Customers.Country, Count(Products.ProductID) as TotalSales

FROM (((Products

INNER JOIN OrderDetails ON Products.ProductID = OrderDetails.ProductID)

INNER JOIN Orders On OrderDetails.OrderID = Orders.OrderID)

INNER JOIN Customers On Orders.CustomerID = Customers.CustomerID)

GROUP BY Products.ProductName, Customers.Country

HAVING Country = 'Germany'

ORDER BY 3 DESC;