Question 1

Think about what could be going wrong with our calculation. Think about a better way to evaluate this data.

1) We are analyzing the affordability of the shoes with wrong dataset and metrics, and this caused our analysis is wrong. There are dataset and features that skews the observation and calculation. What we can do includes feature engineering (Jupyter section 1), removing outliers to re-evaluate average order value (Jupyter section 2) then use a new metrics (Jupyter section 3).
2) Feature engineering (Jupyter section 1):
    a) Our issue was to judge the affordability based on the total order amount which is incorrect. From our analysis, we can see that one consumer has more than 1 purchase per order. Therefore, we added a new feature order_per_item.
3) Remove anomalies to re-evaluate average order value again (Jupyter section 2).
    a) The average order will be skewed because shop 78 is selling $25,725 per order. We should take out shop 78.
    b) From our analysis, we can also see that shop 42 is selling 2000 items comparing with some stores are selling 1 or 2 items. This skews the observation.
    c) After removing the outlier shop 42 and 78, we get an average order value of $300. This amount makes more sense with the fact that some customers purchase more than 1 item per visit. The average number of items is about 2 items after we removed outliers.
4) Use new metrics (Jupyter section 3)
        i) We implemented a z-score to see how each transaction differs from the norm. The shops that lays away from the average are shop 42 and 78. This reflected what we have analyzed previously.

What metric would you report for this dataset?

5) It is often incorrect to judge and analyze the dataset based on the average. From the previous analysis, we found that large number can skew the average and make it not reasonable. In the last section of the observations, we can use percentiles with a 25% median and 75% to judge the entire dataset. In the describe() function, I have added a z-score of each order to determine if they are within the limits of a typical order. The z-score measures the number of standard deviations away from the average. This number can capture irregular values for instance, a value that is far away from average.

What is its value?

    a) We have already found the median and percentiles could help us better judge the observations.
    b) A z-score indicates how far a data point is away from its average. Therefore, we found shop 48 and 78 caused issued with a larger z-scores comparing with other stores. To conclude, using percentiles and median is helpful and

it provides the same conclude as we observed previously by finding stores that have large orders manually. After removing the outlier shop 42 and 78, we get an average order value of $300. This amount makes more sense with the fact that some customers purchase more than 1 item per visit. The average number of items is about 2 items after we removed outliers.

Question 2

a. How many orders were shipped by Speedy Express in total?

```
SELECT COUNT(Orders.ShipperID), Shippers.ShipperName
FROM Orders
INNER JOIN Shippers ON Orders.ShipperID=Shippers.ShipperID where Shippers.ShipperID=1;
```

From the result of the query, there are 54 orders shipped by Speedy Express.

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

```
SELECT Employees.LastName, count(Orders.OrderID), Employees.EmployeeID FROM Employees
Inner Join Orders on Orders.EmployeeID = Employees.EmployeeID
group by Employees.EmployeeID
order by count(Orders.OrderID) DESC limit 1;
```

Peacock is the last name of the employee with the most orders with 40 orders.

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

```
SELECT p.ProductName, SUM(Quantity) AS Total_Quantity
FROM Orders AS o, OrderDetails AS od, Customers AS c, Products AS p
WHERE c.Country = "Germany" AND od.OrderID = o.OrderID AND od.ProductID = p.ProductID AND
c.CustomerID = o.CustomerID
GROUP BY p.ProductID
ORDER BY Total_Quantity DESC
LIMIT 1;
```

Boston Crab Meat is the product that was ordered most by customers in Germany with 160 orders.