

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.

Link of Jupyter notebook (Question 1):

https://github.com/robmanch/Shopify/blob/master/Question_1_Code.ipynb

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

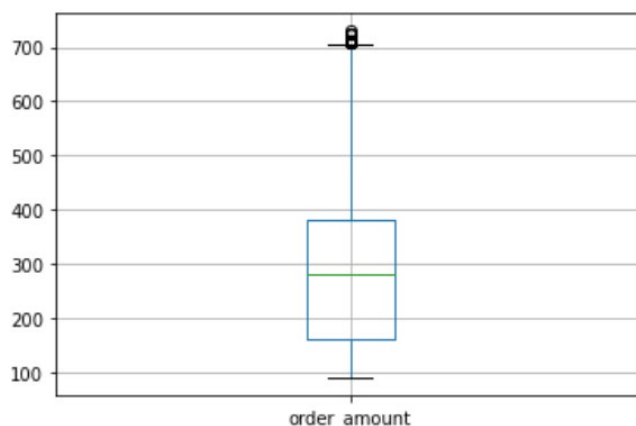
Answer:



It can be seen from the boxplot that this data contains **outliers**. Hence, the mean should not be used to calculate the AOV, as it is sensitive to the outliers. To evaluate better this data, outliers should be removed.

b. What metric would you report for this dataset?

Answer:



The above boxplot shows the distribution of “order_amount” after removing the outliers. But this data is right-skewed, so the median of this data (with no outliers) should be used to evaluate the AOV.

c. What is its value?

Answer: 280.0

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?

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

Answer: **54**

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

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

Answer: **Peacock**

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

```
SELECT ProductName
FROM Products
INNER JOIN OrderDetails ON Products.ProductID = OrderDetails.ProductID
INNER JOIN Orders ON Orders.OrderID = OrderDetails.OrderID
INNER JOIN Customers ON Orders.CustomerID = Customers.CustomerID
WHERE Country = 'Germany'
GROUP BY ProductName
ORDER BY COUNT(Orders.OrderID) DESC
LIMIT 1;
```

Answer: **Gorgonzola Telino**