

Summer 2022 Data Science Intern Challenge

Brandon Law



Question 1

Part A

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

Question 1

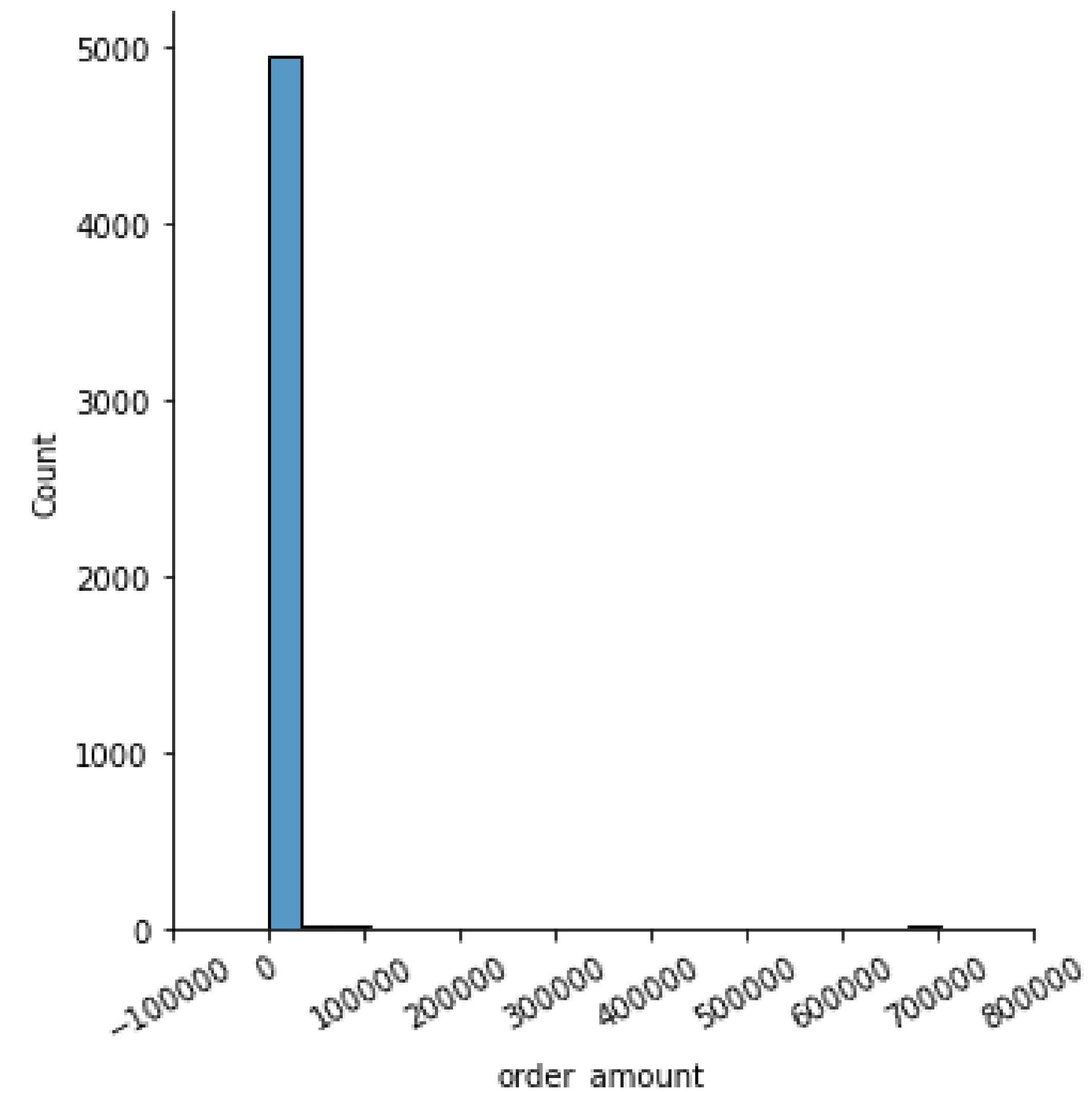
Part A

Part B/C

Considering that we are looking into AOV, I plotted the `order_amount` column into a normal distribution histogram and discovered **the data is right skewed**.

The outliers in **skewed data pulls the mean in the direction of the skew**.

This explains why the calculated AOV was \$3,145.13 for 100 shoe shops that sell just one shoe model.



Part B/C

**What metric would you report for
this dataset? What is its value?**

In this case, we would find the **mode** is a more impactful measure. Its value is not skewed by outliers and still provides a baseline average.

For the 100 shops, we can see the **most common dollar amount per order is \$153.**

The mode helps store owners optimize efforts that increase the number of customers, new and returning, to spend more than \$153 each visit.

Order Amount Measures:

Mean: \$3,145.128

Median: \$284

Mode: \$153

Question 2

Part A

**How many orders were shipped by
Speedy Express in total?**

Result: 54

Question 2

Part A

Part B

Part C

Started by finding the **ShipperID**
associated with **Speedy Express**
from the **Shippers** page.

```
SELECT COUNT(*)  
FROM Orders  
WHERE ShipperID == (  
    SELECT ShipperID  
    FROM Shippers  
    WHERE ShipperName == 'Speedy Express'  
)
```

Question 2

Part A

Part B

Part C

Started by finding the **ShipperID** associated with **Speedy Express** from the **Shippers** page.

Using the **ShipperID**, I filtered the list of orders that used **Speedy Express** as a shipping service.
Returned the count of that list.

```
SELECT COUNT(*)  
FROM Orders  
WHERE ShipperID == (  
    SELECT ShipperID  
    FROM Shippers  
    WHERE ShipperName == 'Speedy Express'  
)
```

Question 2

Part A

Part B

Part C

Started by finding the ShipperID associated with Speedy Express from the Shippers page.

Using the ShipperID, I filtered the list of orders that used Speedy Express as a shipping service.
Returned the count of that list.

```
SELECT COUNT(*)  
FROM Orders  
WHERE ShipperID == (  
    SELECT ShipperID  
    FROM Shippers  
    WHERE ShipperName == 'Speedy Express'  
)
```

Result: 54

Part B

**What is the last name of the employee
with the most orders?**

Result: Peacock

Question 2

Part A

Part B

Part C

I summed all of the orders
associated with each EmployeeID,
sorted in descending order and
returned the ID at the top of the list.

```
SELECT LastName  
FROM Employees  
WHERE EmployeeID == (  
    SELECT EmployeeID  
    FROM Orders  
    GROUP BY EmployeeID  
    ORDER BY COUNT(*) DESC  
    LIMIT 1  
)
```

Question 2

Part A

Part B

Part C

I summed all of the orders
associated with each EmployeeID,
sorted in descending order and
returned the ID at the top of the list.

I pulled the LastName that
corresponds to the EmployeeID
from the Employees page.

```
SELECT LastName  
FROM Employees  
WHERE EmployeeID == (  
    SELECT EmployeeID  
    FROM Orders  
    GROUP BY EmployeeID  
    ORDER BY COUNT(*) DESC  
    LIMIT 1  
)
```

Question 2

Part A

Part B

Part C

I summed all of the orders
associated with each EmployeeID,
sorted in descending order and
returned the ID at the top of the list.

I pulled the LastName that
corresponds to the EmployeeID
from the Employees page.

```
SELECT LastName
FROM Employees
WHERE EmployeeID == (
    SELECT EmployeeID
    FROM Orders
    GROUP BY EmployeeID
    ORDER BY COUNT(*) DESC
    LIMIT 1
)
```

Result: Peacock

Part C

What product was ordered the most by customers in Germany?

Result: Boston Crab Meat

Question 2

Part A Part B Part C

Started by filtering the list of **CustomerID**'s
that used an address in Germany.

```
SELECT ProductName
FROM Products
WHERE ProductID == (
    SELECT ProductID
    FROM OrderDetails
    WHERE OrderID in (
        SELECT OrderID
        FROM Orders
        WHERE CustomerID in (
            SELECT CustomerID
            → FROM Customers
            WHERE Country == 'Germany'
        )
    )
    GROUP BY ProductID
    ORDER BY SUM(Quantity) DESC
    LIMIT 1
)
```

Question 2

Part A Part B Part C

Started by filtering the list of CustomerID's that used an address in Germany.

I filtered all OrderID's that included CustomerID's from the previous list.

```
SELECT ProductName  
FROM Products  
WHERE ProductID == (  
    SELECT ProductID  
    FROM OrderDetails  
    WHERE OrderID in (  
        SELECT OrderID  
        FROM Orders  
        WHERE CustomerID in (  
            SELECT CustomerID  
            FROM Customers  
            WHERE Country == 'Germany'  
        )  
    )  
    GROUP BY ProductID  
    ORDER BY SUM(Quantity) DESC  
    LIMIT 1  
)
```

Question 2

Part A Part B Part C

Started by filtering the list of CustomerID's that used an address in Germany.

I filtered all OrderID's that included CustomerID's from the previous list.

From OrderDetails, I filtered the orders that match the OrderID from the previous list, grouped the orders by ProductID and summed the quantities. Sorting the list in descending order and returned the first ID.

```
SELECT ProductName  
FROM Products  
WHERE ProductID == (  
    SELECT ProductID  
    FROM OrderDetails  
    WHERE OrderID in (  
        SELECT OrderID  
        FROM Orders  
        WHERE CustomerID in (  
            SELECT CustomerID  
            FROM Customers  
            WHERE Country == 'Germany'  
        )  
    )  
    GROUP BY ProductID  
    ORDER BY SUM(Quantity) DESC  
    LIMIT 1  
)
```

Question 2

Part A Part B Part C

Started by filtering the list of CustomerID's that used an address in Germany.

I filtered all OrderID's that included CustomerID's from the previous list.

From OrderDetails, I filtered the orders that match the OrderID from the previous list, grouped the orders by ProductID and summed the quantities. Sorting the list in descending order and returned the first ID.

From the Products page, I selected the ProductName that matches the ProductID.

```
SELECT ProductName
FROM Products
WHERE ProductID == (
    SELECT ProductID
    FROM OrderDetails
    WHERE OrderID in (
        SELECT OrderID
        FROM Orders
        WHERE CustomerID in (
            SELECT CustomerID
            FROM Customers
            WHERE Country == 'Germany'
        )
    )
    GROUP BY ProductID
    ORDER BY SUM(Quantity) DESC
    LIMIT 1
)
```

Question 2

Part A Part B Part C

Started by filtering the list of CustomerID's that used an address in Germany.

I filtered all OrderID's that included CustomerID's from the previous list.

From OrderDetails, I filtered the orders that match the OrderID from the previous list, grouped the orders by ProductID and summed the quantities. Sorting the list in descending order and returned the first ID.

From the Products page, I selected the ProductName that matches the ProductID.

Result: Boston Crab Meat

```
SELECT ProductName
FROM Products
WHERE ProductID == (
    SELECT ProductID
    FROM OrderDetails
    WHERE OrderID in (
        SELECT OrderID
        FROM Orders
        WHERE CustomerID in (
            SELECT CustomerID
            FROM Customers
            WHERE Country == 'Germany'
        )
    )
    GROUP BY ProductID
    ORDER BY SUM(Quantity) DESC
    LIMIT 1
)
```