

Shopify Data Analyst Intern Challenge Response

Ewan Simms

1.

a)

The mean is skewed by outliers of the bulk orders from user 607 at store 42 and the vastly more expensive sneakers sold at store 78. Furthermore, the distribution is right skewed, which will skew the mean higher, as demonstrated by the graph below (which even excludes the aforementioned outliers). Thus, the average is not an accurate measure of central tendency in this scenario.

Distribution of Order Amounts



b)

The **median** is a better measure of central tendency in this situation. It is less susceptible than the mean to the influence of outliers of this right skewed distribution.

c)

The median is **\$284**.

2.

a)

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

54 orders where shipped by Speedy Express

b)

```
SELECT LastName FROM Employees WHERE EmployeeID IN  
(SELECT EmployeeID FROM [Orders]  
GROUP BY EmployeeID ORDER BY COUNT(*) DESC LIMIT 1);
```

The employee with the last name **Peacock** has the most orders.

c)

```
SELECT ProductName FROM Products WHERE ProductID IN  
(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);
```

Boston Crab Meat is the most popular product for German customers.