Question 1

a) Looking at the histogram of the data, there are outliers that are orders of magnitude greater than the order amounts of the vast majority of orders.
b) The simplest metric for data with a few outliers that substantially change the mean would be the median.
c) The median order amount is $284

```
import pandas as pd

# load data into pandas dataframe
csv_dir = "C:/Users/matth/Downloads/2019 Winter Data Science Intern
     Challenge Data Set - Sheet1.csv"
data_set = pd.read_csv(csv_dir)

# check average to verify reported results
print(data_set["order_amount"].mean())

# create histogram to get feel for data
histogram = data_set.hist(column="order_amount")

# get median as there are a some outliers that are orders of magnitude
# greater than the vast majority of orders
print(data_set["order_amount"].median())

# median is $284 which is more reasonable
# query the data to pull outliers that are greater than 10x the median
print(data_set.query("order_amount > 2840"))
```

Question 2

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

   54 orders

   SELECT COUNT(ShipperName) FROM Orders
   INNER JOIN Shippers ON Orders.ShipperID=Shippers.ShipperID
   WHERE ShipperName="Speedy Express"

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

   Peacock

   SELECT LastName, MAX(employeeCount)
   FROM (SELECT LastName, COUNT(Orders.EmployeeID) AS employeeCount FROM Orders
   JOIN Employees ON Orders.EmployeeID=Employees.EmployeeID
   GROUP BY Orders.EmployeeID)

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

Boston Crab Meat

```
SELECT ProductName, MAX(germanyProductQuant) FROM
(SELECT Products.ProductID, Products.ProductName, SUM(Quantity) AS
germanyProductQuant FROM (((OrderDetails
INNER JOIN Orders ON OrderDetails.OrderID=Orders.OrderID)
INNER JOIN Customers ON Orders.CustomerID = Customers.CustomerID)
INNER JOIN Products ON OrderDetails.ProductID=Products.ProductID)
WHERE Country="Germany"
GROUP BY Products.ProductID)
```