**Winter 2021 Data Science Intern Challenge**

**1.a.** Mean or Average is a good parameter for analysis only when the dataset is symmetrical like a normal distribution. However, for this dataset, mean ($3145.13) is much greater than the median ($284) thus, the distribution is right-skewed. The distributions which are highly skewed have mean further away from the center.

Therefore, for skewed distributions, using mean or average order value as in this case would lead to erroneous conclusions due to the outliers which have a significant impact on the mean.

**1.b** When a distribution is skewed, median is a better measure of central tendency than mean. Outliers and skewed data have a much smaller effect on median.

Thus, A better way of doing this analysis would be calculating the median instead since unlike mean, median doesn’t depend on all the values and would still work well with a few outliers.

**1.c.** Median value of this dataset is $284

**2.a.** Total orders shipped by Speedy Express = **54**

Select Count(o.OrderID) as OrdersShipped

from Orders as o

inner join

(Select \* from Shippers

where upper(trim(ShipperName)) = 'SPEEDY EXPRESS') as s

on o.ShipperID = s.ShipperID

A screenshot of a cell phone

Description automatically generated

**2.b.** Last name of the Employee with the most orders is **Peacock**.

SELECT LastName from

(SELECT LastName, Max(totcount) as maxcount

from

(SELECT o.EmployeeID, Count(o.OrderId) as totcount, LastName, FirstName

FROM Employees as e

inner join Orders as o

on e.EmployeeID = o.EmployeeID

group by o.EmployeeID

) as tab

) as tab2

A screenshot of a cell phone

Description automatically generated

**2.c.** The product ordered the most by customers in Germany is **Boston Crab Meat**.

Select ProductName from

(Select ProductName, Max(totcount) as Most

from

(Select ProductName, Sum(Quantity) as totcount

from (Select \* from Customers where upper(trim(Country)) = 'GERMANY') as c

inner join orders as o

on c.CustomerID = o.CustomerID

inner join OrderDetails as od

on o.OrderID = od.OrderID

inner join Products as p

on od.ProductID = p.ProductID

) as tab

group by ProductName

) tab2

A screenshot of a cell phone

Description automatically generated