Shopify Data Science Intern Challenge

Shamini Wadhwani

R Markdown

#1a

```
library(readxl)
Data <- read_excel("2019 Winter Data Science Intern Challenge Data Set.xlsx")
sum(Data$order_amount)/nrow(Data)
```

```
## [1] 3145.128
```

The calculation is currently not taking into account the total number of items purchased with each order. Currently, the calculation is only taking into account the number of orders.

#b In order to find the Average Order Value, the calculation should take into account the revenue/the number of items sold. Hence, the metric total_items should be used to calculate the Average Order Value with the following formula:

```
sum(Data$order_amount)/sum(Data$total_items)
```

```
## [1] 357.9215
```

#c We can determine from this that the average order value is $357.9125

#2a Answer: 54 Query:

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

#2b Answer: Peacock Query:

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

#2c Answer: Gorgonzola Telino Query:

```
SELECT ProductName FROM Products
WHERE ProductID = (SELECT TOP 1 ProductID FROM (SELECT COUNT(*), ProductID FROM
(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 COUNT(*) DESC));
```