Q1.

library(dplyr)

repo <- read.csv("2019 Winter Data Science Intern Challenge Data Set - Sheet1.csv")

glimpse(repo) #see our data in more details

sum(is.na(repo)) #Check for missing values

summary(repo$order\_amount) #Get the 5 number summary for our data

mean(repo$order\_amount) #mean for data

#Mean = 3145.128 = our AOV avg order value. OUr AOV is based on mean so extreme values can affect our precision for AOV

aov <- repo %>% group\_by(shop\_id) %>% summarize(shop\_avg = sum(order\_amount/sum(total\_items))) #create a new data containing shops and order amount

aov %>% arrange(shop\_avg) #ascending order for avergae AOV of shop

#no extreme values

aov %>% arrange(desc(shop\_avg)) #descending order for avergae AOV of shop

#extreme value for shop 78. Since sneakers are affordable the value is very high which is $25725

#Need to examine the data for that shop only

filter(repo, shop\_id == 78) #we can notice that order\_amount for all orders of shop 78 the $ value is too high for a sneaker price. The price might contain decimal place that might have forgotten, so for eg for order amount $25725 the actual order would be worth $257.25

aov <- aov %>% mutate(shop\_avg = case\_when(shop\_avg == 25725 ~ 257.25,

TRUE ~ shop\_avg)) #need to convert avg of shop

glimpse(aov)

Q2.a.

Select Shippers.ShipperName, Count(Orders.OrderID) AS NumberOfOrders FROM Orders LEFT JOIN Shippers ON Orders.ShipperID = Shippers.ShipperID GROUP BY ShipperName;

Background pattern

Description automatically generated

Q2.B.

Select Employees.LastName, Count(Orders.OrderID) As TotalOrders From Orders Left Join Employees ON Orders.EmployeeID = Employees.EmployeeID Group by LastName Order by TotalOrders Desc;

Background pattern

Description automatically generated

Q2.c.

Select Products.ProductName, Customers.Country,

Count(Orders.OrderID) As TotalOrders From Customers

Left Join Orders On Customers.CustomerID = Orders.CustomerID

Left Join OrderDetails On Orders.OrderID = OrderDetails.OrderID

Left Join Products On OrderDetails.ProductID = Products.ProductID

Where Country = 'Germany' Group by ProductName Order by TotalOrders Desc;

Table

Description automatically generated