**STT 811**

**In-Class Assignment 3**

1. Use the Customer History data and follow the steps to generate all necessary rows for the modeling dataset. You do not have to generate the target (yet).

years <- data.frame(year = seq(2005,2022))

customer\_history2 <- sqldf("SELECT \*

FROM customer\_history

INNER JOIN years

WHERE year <= Last\_Year

AND year >= First\_Year")

1. With the order\_history dataset,
   1. Use the pivot\_wider function to create separate columns by year.

order\_history2 <- sqldf("SELECT CustomerID, Product\_ID, Year, Month, sum(Quantity) as Quantity

FROM order\_history

GROUP BY customerID")

order\_wide <- pivot\_wider(order\_history2, names\_from = Year, values\_from = Quantity)

* 1. Use the the pivot\_longer function to transform the data generated in (a) to the similar form. Compare the number of rows to the original dataset.

order\_history2 <- sqldf("SELECT CustomerID, Product\_ID, Year, Month, sum(Quantity) as Quantity

FROM order\_history

GROUP BY customerID")

order\_wide <- pivot\_wider(order\_history2, names\_from = Year, values\_from = Quantity)

order\_long <- pivot\_longer(order\_wide, cols = c('CustomerID','Product\_ID','Month'), names\_to = 'Year', values\_to = 'Quantity')

order\_long