**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")
```

2. With the order_history dataset,
   a. 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)
```

   b. 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
```