**STT 811**

**Homework 2**

**Due Friday, January 27, at 11:59:59 pm**

1. The purpose of this problem is to create a modeling dataset for customer churn in the next year, based on the customer history dataset. In ICA 3, you performed the row augmentation to get missing rows added to the data. To get to the modeling dataset:
   a. Create the target field based on the strategy given in the slides.
   b. Create the following additional input variables that would be used to make predictions:
      i. Customer age
      ii. Number of years as a customer
      iii. Using the complaints dataset, create a count of complaints made each year and join it to the dataset (be careful with what year you join!)

2. The purpose of this problem is to create a modeling dataset for predicting which orders will order which products, based on the order history dataset. In ICA 3, you performed the row augmentation. Now
   a. Create the target field based on the strategy given in the slides.
   b. Create the following additional input variables that would be used to make predictions. All of these x variables should be from the standpoint of the prior month.
      i. Number of months since the last order
      ii. Number of months out of the previous 12 months that an order was placed
      iii. Average quantity ordered per order

3. With the OJ dataset, calculate some additional X's and join them back to the original dataset; specifically:
   a. the average price for each brand by store
   b. The previous week's fraction of customers who bought Minute Maid, by store (note: you will lose the first week when you use this as an X)