# Stat 542 Project II Jinran Yang

## 1. Computer System

MacBook Pro (Retina, 13-inch, Mid 2014)

Processor:2.6 GHz Intel Core i5

Memory: 8 GB 1600 MHz DDR3

Software: macOS 10.14 (18A391)

## 2. Walmart Stores Forecasting

This project is about predicting the future weekly sales for each department in each of 45 Walmart stores located in different regions based on the historical data. And we are asked to provide three models prediction models.

The structure of the code is provided by one of the teaching assistant Josh, it helps me understanding the objective, the desired results, and the flow of code of this project.

Inspired by the Professor Liang, I choose ***snaive, tslm***, and ***stlf*** functions from the ***forecast*** package to build models.

### 2.1 Pre-processing

I replace all of the missing values of weekly sales with zeros.

### 2.2 Model

(1) Based on the suggestion of Professor Liang, I used seasonal naive, which sets each forecast to be equal to the last observed value from the same season of the year. Based on the code provided by Josh, I filter the time segment and then I spilt the data by department and store. Each segment contains the corresponding ***Weekly_Sales***.I created a time-series object for each set of observed sales values for a store department. And I set ***frequency*** equal to 52 (365/7) for the time series to indicate that the observations were made weekly. Then I run ***snaive*** function to forecast per store.

(2) Based on the suggestion of Professor Liang, I use ***tslm*** function which is used to fit linear models to time series including trend and seasonality components. The

preprocessing procedure is the same as the first model. Then I run ***tslm*** function with the formular ***tslm ~ trend + season*** where ***trend*** represents the time trend in the data and ***season*** represents week.

(3) My third model used the ***tslm*** for fold 1-6 and ***stlf*** functions for folds 7-10 from the forecast package. ***stlf*** function is a method applying a non-seasonal forecasting method to the seasonally adjusted data and re-seasonalizing using the last year of the seasonal component. Since ***stlf*** requires 2 years of data in the time series. So when the time series contained less than 2 years observations, I used ***tslm*** methods in fold 1-6 which is the same as model 2.

### 2.3 Result

| Fold | snaive | tslm | tslm_stlf |
|:---:|:---:|:---:|:---:|
| 1 | 2262.422 | 2042.401 | 2042.401 |
| 2 | 1787.081 | 1440.083 | 1440.083 |
| 3 | 1779.052 | 1434.716 | 1434.716 |
| 4 | 1716.117 | 1596.988 | 1596.988 |
| 5 | 2400.395 | 2327.638 | 2327.638 |
| 6 | 1696.900 | 1674.185 | 1674.185 |
| 7 | 2086.967 | 1718.577 | 1622.237 |
| 8 | 1750.283 | 1420.817 | 1348.015 |
| 9 | 1719.887 | 1430.801 | 1279.667 |
| 10 | 1680.956 | 1447.034 | 1245.370 |
| Overall Average: | 1888.006 | 1653.324 | 1601.130 |

```
> cat("Running time :",(proc.time() - start_time)[3],"\n")
Running time : 1361.923
```