1. Project Title: Holiday Steals and Summer Blowouts: Does conventional wisdom about retail sales hold true?
2. Introduction
   1. As graduate students on tight budgets, we are constantly looking for the best deals on clothing. Clothing stores often have sales, but it’s hard to predict when those sales will take place.
   2. Conventional wisdom suggests that factors such as temperature, macroeconomic fluctuations, and holidays are reasonable predictors of retail sales.
   3. In this project, we gathered data on these types of factors and utilized basic modeling and cross validation techniques to see if they were able to predict future retail sales.
   4. With the data and methods we used, we were unable to find associations between typical “conventional wisdom” factors and retail sales at four stores: Express, LOFT, Gap, and J Crew.
3. Data
   1. Sales: To obtain a proxy for sales that are occurring at these stores, we used the Twitter API to obtain data from their accounts and used keyword analysis to identify days where they had sales. We attempted to select stores that had comparable styles of clothing and types of consumers. We did not distinguish between online and in-store sales.
   2. Weather: We used the WeatherUnderground API to obtain temperature and precipitation data from 4 major cities in different regions of the United States: LA, New York, Chicago, and Houston.
   3. Unemployment: We obtained monthly unemployment statistics from the US Bureau of Labor Statistics
   4. Stock prices: We obtained daily stock prices for the stores that are publically listed from Yahoo Finance.
   5. Holidays: We obtained US Holiday data from the R timeDate package.
4. Exploring the data
   1. Plot: Sales over time
      1. The plot for sales over time shows that Express and Loft have more sales per month compared to Gap and JCrew. In addition to having more sales, Express and Loft have greater variation in the number of sales per month.
   2. Plot: Sales by Temperature & Holiday Over Time
      1. As shown in the below plot, there is no clear relationship between sales, average temperature, and holidays over time.
   3. Plot: Sales by Stock Price Over Time
      1. As shown in the below plots, across stores, there is no clear relationship between stock price and sales over time.
   4. Plot: Histograms for Time From Last Sale
      1. The histograms below show the frequency of sales across stores. Among the four stores, Express has sales most frequently, while JCrew has sales least frequently.
5. Predictions
   1. We split the data from each store into Train and Test datasets.
   2. We fit both an OLS and a logistic regression model to the training sets of each of the stores to test model fit and specification details
   3. We ran each model on the test sets to check the accuracy of our predictions
6. Results
   1. Both our OLS and our logistic regression models almost always predict that there is less than a 50% chance of a sale. This means that we can accurately predict when there is no sale, but our models aren’t meaningful for predicting when sales will happen.
   2. As a result, accuracy metrics do not fully communicate the appropriateness of our models for predicting the outcome of interest. This is likely because our predictor variables jointly have very limited predictive power on our outcome of interest “sale”.
   3. Overall, these results likely point to the fact that factors outside of the predictors we have included in our model are those that are more strongly predictive of sales.
   4. It is also possible that the factors we included in our models impact sales with time lag, which is not captured in the simple linear models that we have used.
   5. Given the data and methods we used, we were unable to find significant associations between “conventional wisdom” factors and retail sales at four stores: Express, LOFT, Gap, and J Crew.
   6. Some next steps we could take would be to consider other factors as predictors and to work with time lag models for longitudinal data to capture possible delayed reactions in the data.