Alok Kommajesula  
Metro Ridership

**Background**

The goal of this project is to look at the ridership of the DC metrorail and what indicators are significant in determining if an individual will use the service.

Indicators observed include:

* Day of the week
* Average weekly price of a gallon of gasoline
* Amount of federal benefits a rider is eligible for through the Smartrip program
* Amount of precipitation for a day
* Mean temperature for a day

**Hypothesis**

I hypothesis that the change in federal benefits to the Smartrip program as well as day of the week will have the highest impact on ridership.

**Data**

The Metro ridership data was made available through the Washington Metropolitan Area Transit Authority (WMATA) as a CSV file for download.

The Gas Prices data was acquired through through the U.S Energy Information Agency as a CSV file for download.

The Weather data was found through weather underground and was also available for download as CSV files.

Pre-processing steps:

All the data was in 3 separate data frames so the prep work involved merging them together based on the date. Once they were combined, it allowed for creating of other columns including a dummy variable for a weekday or weekend.

Exploring the Data

After cleaning the data to only span from January 1st 2004 to December 31st 2014, there were some interesting characteristics about the data. The correlation matrix showed the following:

          gasprice      Temp  Smartrip  Ridership   Weekday

gasprice   1.000000  0.311724  0.190010   0.061584  0.001228

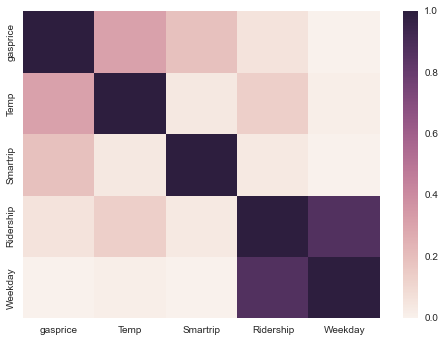
Temp       0.311724  1.000000  0.038693   0.141178  0.012887

Smartrip   0.190010  0.038693  1.000000   0.033364  0.000806

Ridership  0.061584  0.141178  0.033364   1.000000  0.871093

Weekday    0.001228  0.012887  0.000806   0.871093  1.000000

The heat map showed the following:



**Features:**

I choose these features to go with because I believed that weather and cost would be the biggest considerations an individual would take into consideration when deciding to use the metro. I hope to add labor indicators including unemployment and median income to account for market fluctuations which could impact the use of public transportation.

Details of your modeling process, including how you selected your models and validated them

**Challenges and Successes**

I have been able to successfully start the regression process but have not yet been able to produce any meaningful results. In the coming weeks, I will flesh this out and go more in depth with these indicators.

**Business applications**

As of now the test data is only looking at data from 2004 to 2013 and may produce a good model to predict daily ridership. This will allow the WMATA to forecast daily metro usage and make the necessary decisions to accommodate the riders.