

Forecasting Metro Ridership

Lena Nguyen | March 21, 2015

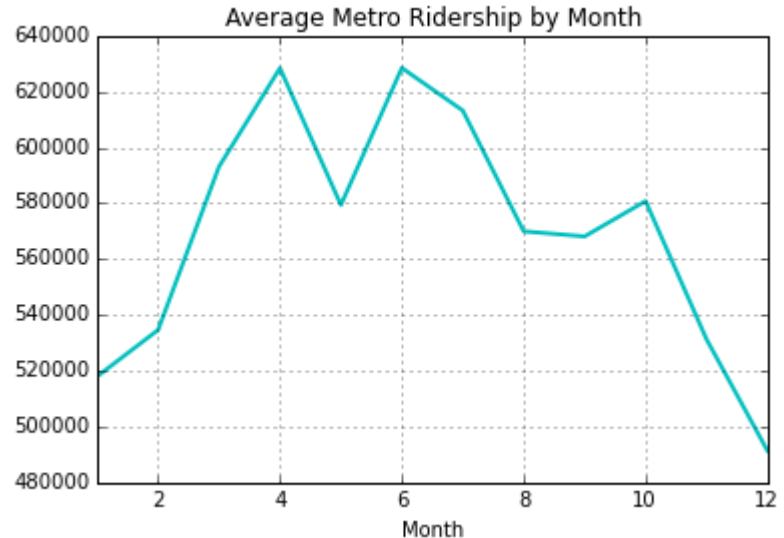
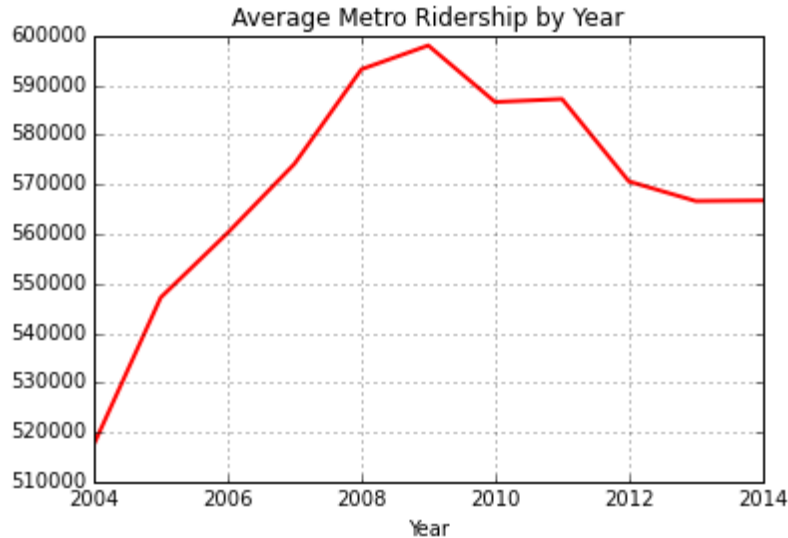
The Goal

To forecast Metro ridership based on different input variables (Regression)

Historic Metro Ridership Data

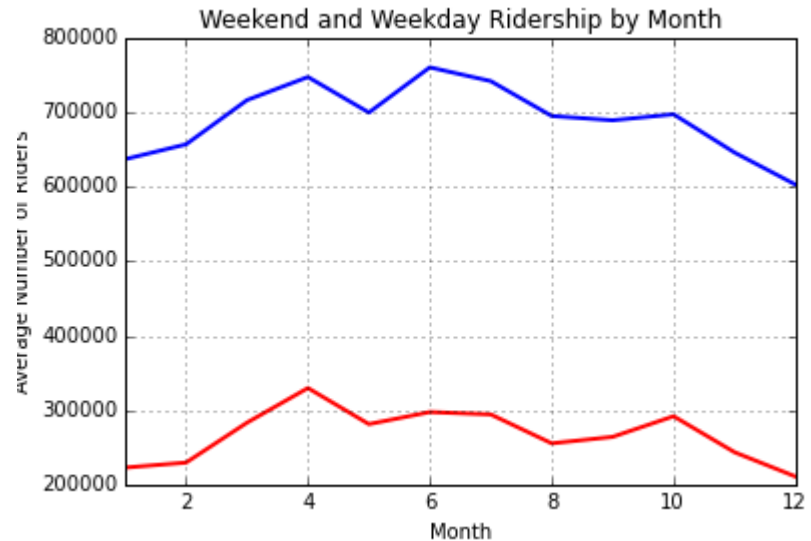
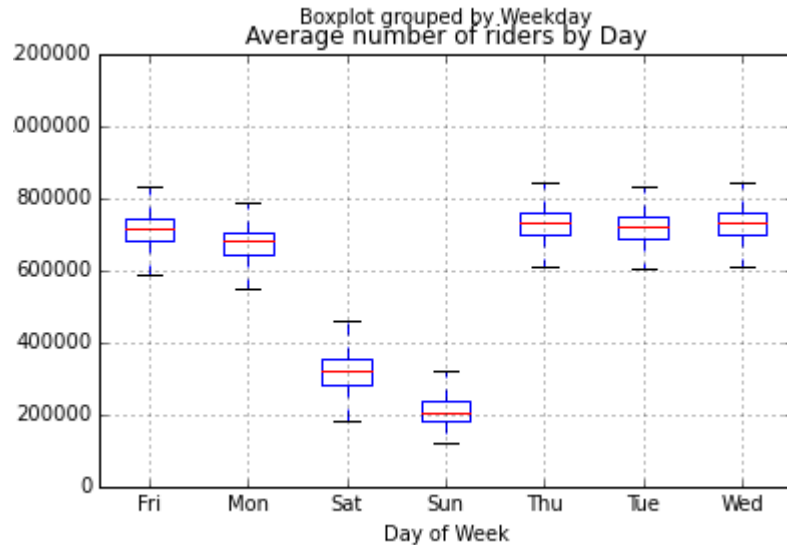
- Open Data DC: daily ridership data from 2004-2014, but not disaggregated by station
- Plan It Metro website: Daily ridership data with Entrance/Exit Station data, but only for certain months (ie May 2012, May 2013)
- WMATA website: Annual ridership data by station, but all in PDF (what a pain!)

Data Visualization



Using the metro ridership data from Open Data DC

Data Visualization



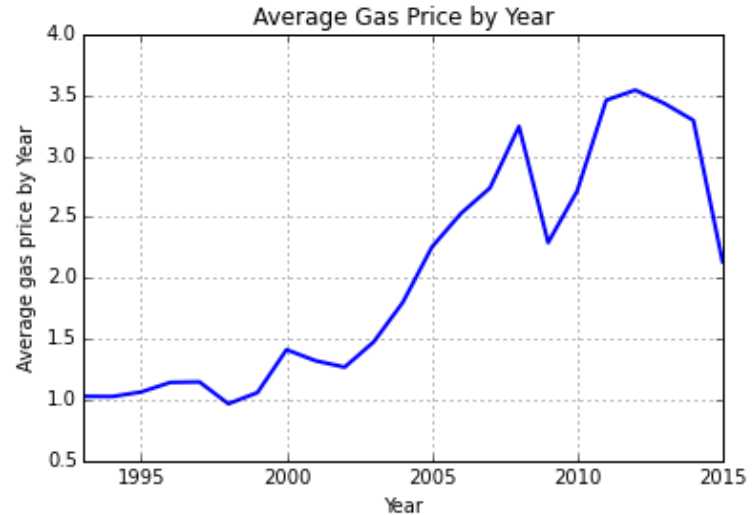
In graph above, blue line is weekday and red line is weekend

Input variables

- Gas prices
- Climate/Weather
- Employment
- Tourist Numbers
- Reliability of Metro

Gas Prices

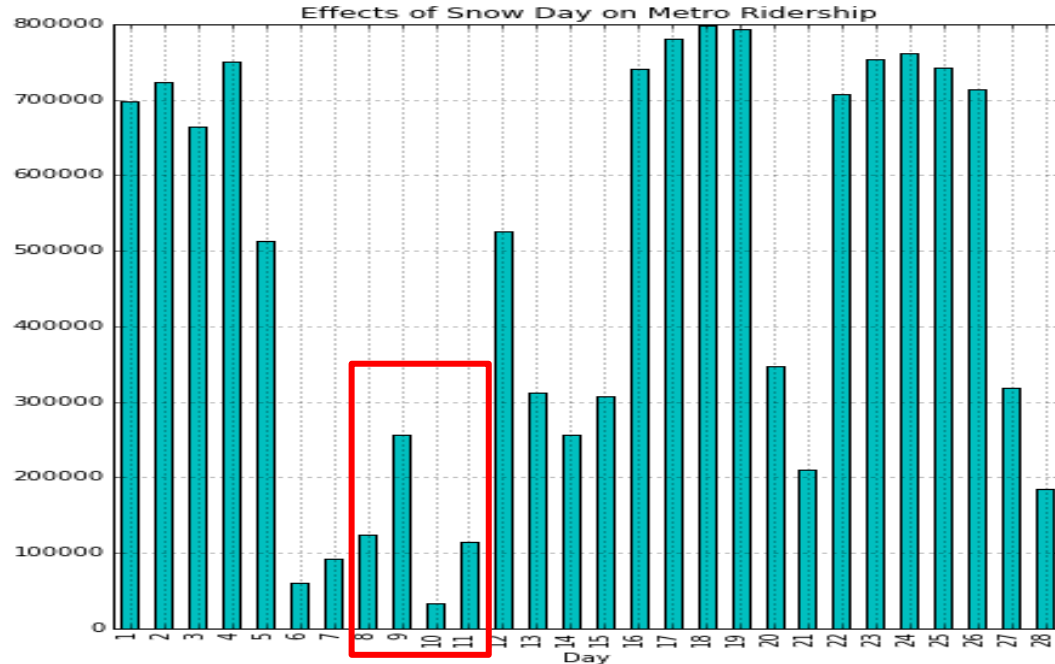
- Historic gas prices available for whole country and by US region on US Energy Information Administration website
- People would rather drive than take Metro if it's cheaper



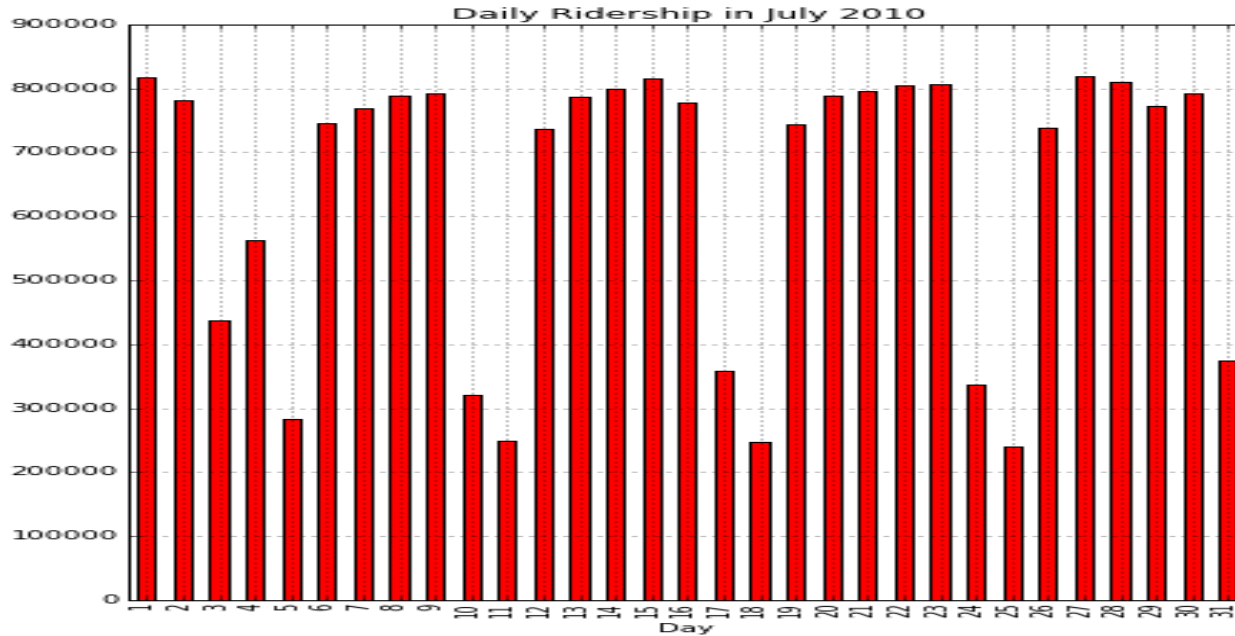
Climate/Weather

- Data available on the NOAA website
- Tourists less likely to visit during cold weather
- People are less likely to go places during bad weather
- Snow day means fewer people are going to work so much lower ridership

Winter Ridership (February 2010)



Summer Ridership (July 2010)



Still to be done...

- Tourist numbers
 - White House Visitor numbers?
 - Smithsonian attendance only annually
- Employment
 - US Bureau of Labor Statistics has labour force data for DC metro area
- Reliability of Metro
 - Twitter @Metrorailinfo posts outages/delays
- More input variables?
- Visualize the Origin/Destination/Time Period data from Plan It Metro