

This is the directory of my second capstone report. The project focuses on building machine learning models to forecast New York City property valuations.

The key files are:

- Capstone Project 2 Proposal
- Milestone Report
- Milestone Report 2
- Capstone Report Final

These lay out my goals, process, and accomplishments.

If you're interested in digging into my code and seeing how I did it, check out the Jupyter notebooks.

Data cleaning takes the raw property valuation data and cleans it.

Data visualization, Graphing, and Mapping contain the code used to visualize the data. Mapping works specifically with the Python module Folium.

Machine Learning is where I train my random forest models on the data.

randfor_functions contains a number of simple python programs for working with Scikitlearn random forests. I use these when working in the Machine Learning file.

The raw data and a data dictionary can be downloaded from NYC OpenData. I've linked to that in my milestone reports, but I can't upload a 6GB file to GitHub. Sorry!