Capstone Project 2 Proposal

NYC Property Valuation and Tax Data

Problem Statement: Property valuation and assessments determine property tax, exceptions and abatements. They are related to, but not exact arbiters of, final property sale prices.

Who Might Care: A method to better predict property valuations and the effect of factors such as location, easement, and building class will benefit anyone involved in the real estate and housing markets. Potential buyers might also be interested in being able to predict the taxes they will have to pay and how the valuation will change over time.

Dataset: The City of New York Department of Finance, through NYC OpenData, provided a dataset containing property valuations dating from 2010 through 2019. This dataset tracks over 100 different qualities, from locational data like borough code and block range to property details like year constructed and lot frontage and financial details such as the current tax classification, tentative and final assessments of land and total value. The dataset tracks 5.7 million valuations.

Problem: How can we predict property valuation prices ahead of time? What factors most influence the valuation price? Tracking property valuations provides a more robust dataset than tracking market price sales because the data are updated annually and assessed by certified appraisers. We expect them to be more consistent in their valuations than interested buyers. The annual updates mean that properties that are infrequently sold will still be represented in the dataset.

Outcomes: This project aims to deliver a model that predicts property valuations based on associated property data. This model should be transparent about which factors influence property valuation in which manner and direction.

Known issues: Property valuation is not an exact match for final property price. Although this number is still directly relevant for assessing taxes and related fees/exceptions, other factors

influence the final sale price of a property. Additionally, with the current pandemic, all bets are off: It's very difficult to predict what the market will be doing over the coming months. It should be emphasized that this model is predicting valuations, not market prices. The dataset is also much larger and more dimensional than the last capstone's dataset: Much more cleaning is required before serious work can begin.