

Project Proposal- Linear Regression/Web Scraping

Question/need:

- I plan to build a model that aims to predict house prices in Staten Island and Brooklyn based on metrics such as number of bedrooms/bathrooms, square footage, lot size, heating/cooling type, parking etc.

Data Description:

- Data will be scraped from Zillow to include roughly 1500 rows of houses (before cleaning). An individual sample/unit of analysis in this project is therefore a house listing. Each house listing will be evaluated based on the aforementioned features.

Tools:

- The tools I plan on using are BeautifulSoup, Sklearn, StatsModels, Pandas, Matplotlib, Seaborn, Numpy

MVP Goal:

- An MVP for this project would be a jupyter notebook showcasing a working linear regression model that uses select highly correlated features to predict the value of a home