



Predicting House Prices for homes in Brooklyn and Staten Island

Objective: Predict the price of a newly built home, given some features, using Zillow.com

+ Tools Used:

BeautifulSoup

matplotlib

The logo for seaborn, featuring a circular icon with a bar chart and a scatter plot.

seaborn

The logo for scikit-learn, featuring a blue circle and an orange circle with the word "scikit" above "learn".

scikit
learn

The logo for pandas, featuring a stylized grid icon.

pandas

Methodology:

Scraped Zillow for houses for sale in Staten Island and Brooklyn
Gathered other census features from census.gov in .csv format

Target

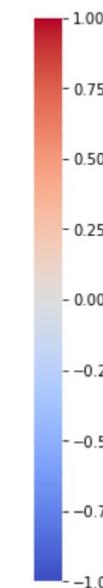
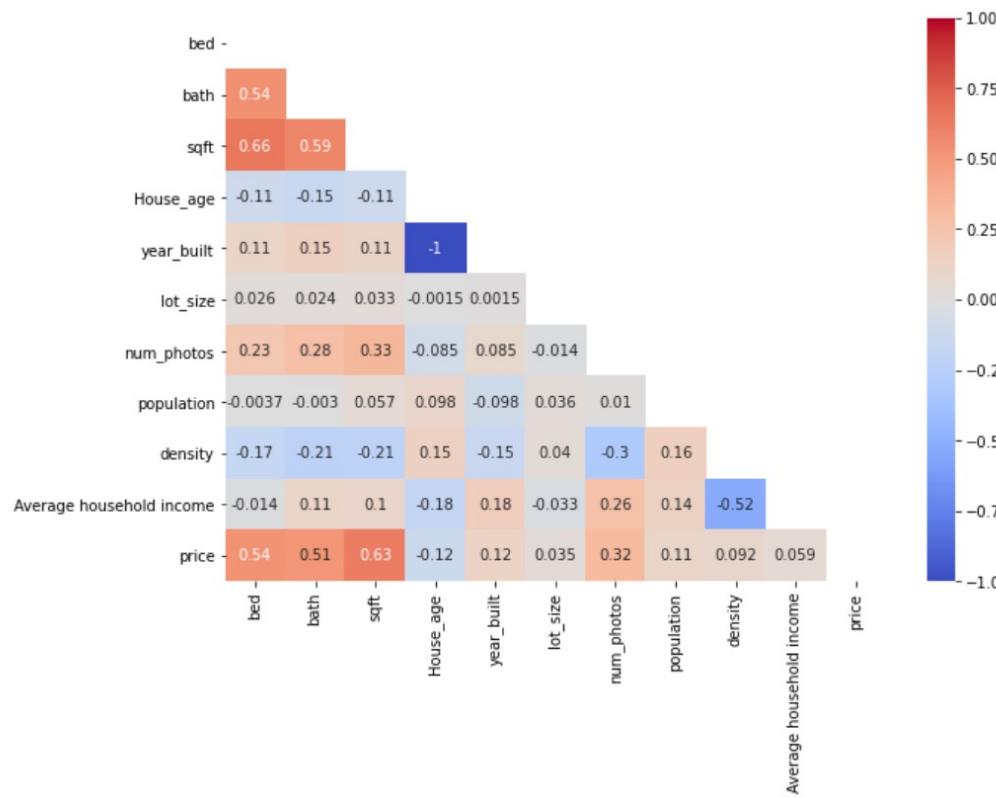
Price

Features

Bed
Bath
Sqft
House_type
Year_built
Heating
Cooling
parking
Lot_size
Num_photos
Address with zip
Average household income
zip

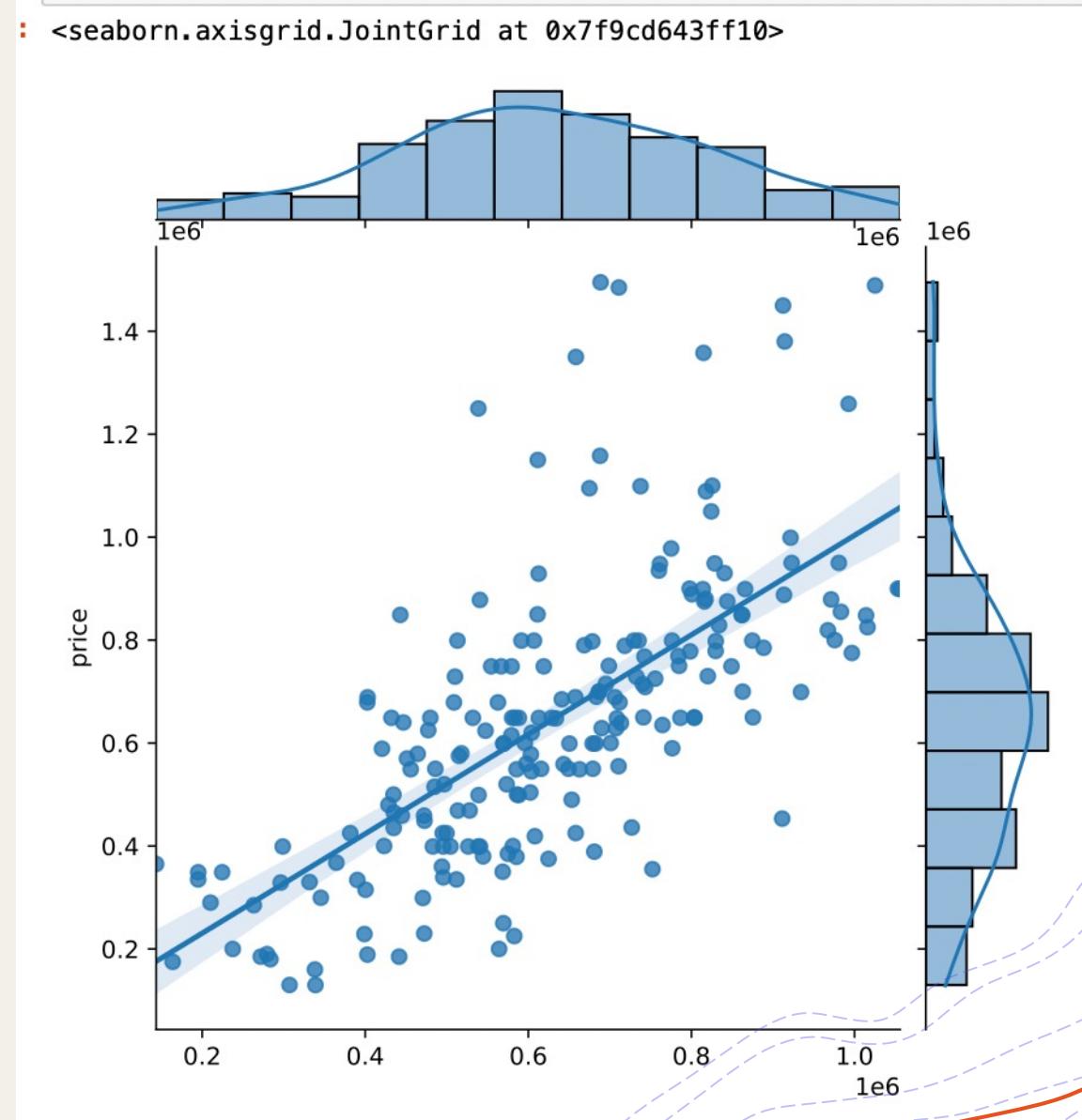
Total Data:

(1565, 15)



Baseline Linear Regression Model:

Metric	Value
R^2	0.726445
RMSE	167761.58
MAE	141648.12



Final Model

+ Polynomial Regression with Lasso Regularization:

Metric	Value
R^2	0.6067308
RMSE	182949.52
MAE	120517.40

Conclusion:

- + Model seems to predict better for lower priced homes
- + Unsurprisingly feature engineering and poly features with bed, bath and sqft had the greatest impact on improving my model
- + However, census data didn't do much for my model

Future Work

- + Find a better way to incorporate categorical values
- + Bring in more features like walk score, transit score, view
- + Time based model