



Housing Price Regressor

Seattle, WA



The Business Objective



- The aims of this project are two-fold
 - To produce as accurate a home pricing model as is possible for the City of Seattle
 - To leverage model interpretability to provide actionable insights on what contributes to higher property values

Target Segment



- The Target Audience are small value-add investors and developers In the City of Seattle.
- The tool aims to provide accurate assessments of price as well as leverageable information on what features contribute most to higher-priced housing





The Data

- The data was initially collected from the Redfin website through its direct download option on November 31, 2023.
- Includes:
 - All Sales in City of Seattle Last 12 Months
 - Basic Vital Information on Each Property
 - Beds
 - Baths
 - Square Footage
 - Lot Size
 - Location
 - Submarket
 - Price
 - Property Type
 - Date of Sale
 - Etc.

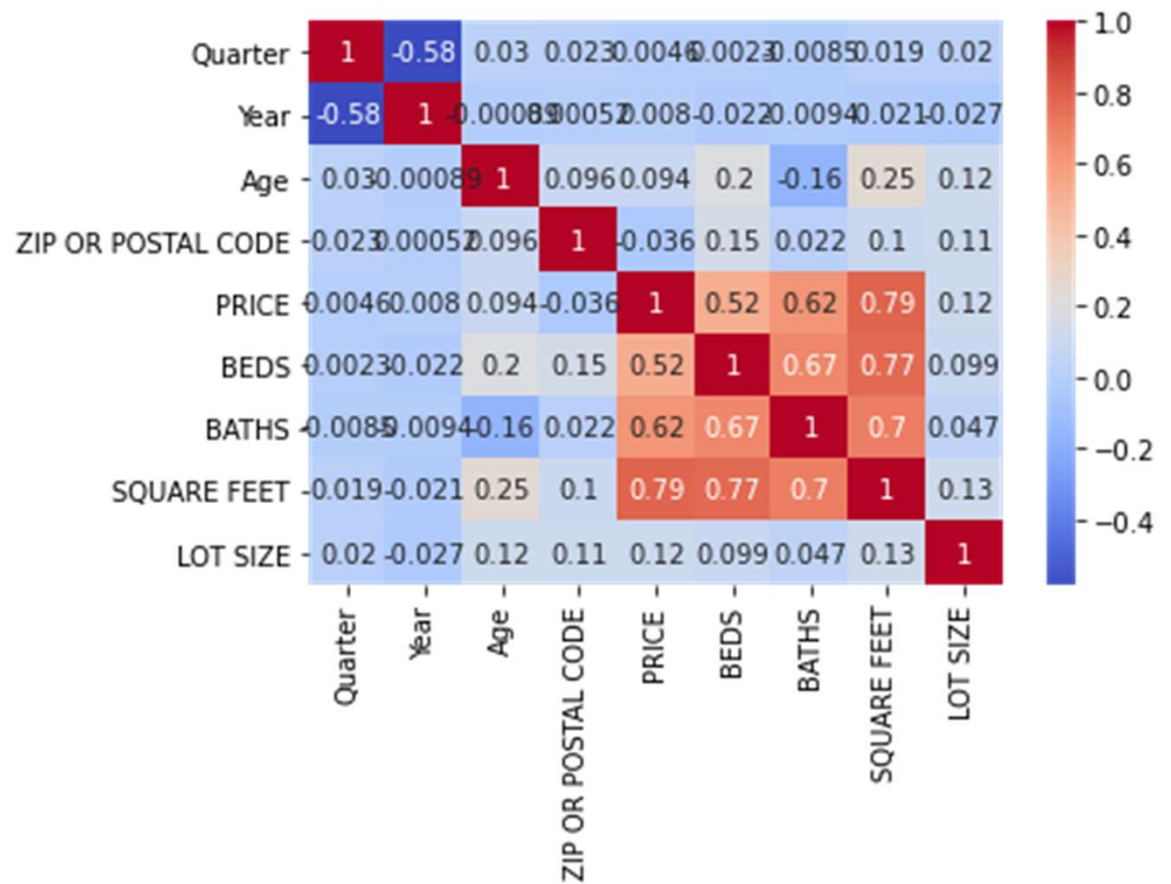


The Data



- Custom Web Scraper
 - BeautifulSoup and Selenium were Utilized
 - Provides Detailed Realtor/Listing Descriptions of every property
 - Provides unique previously untapped features through which to train later regression models

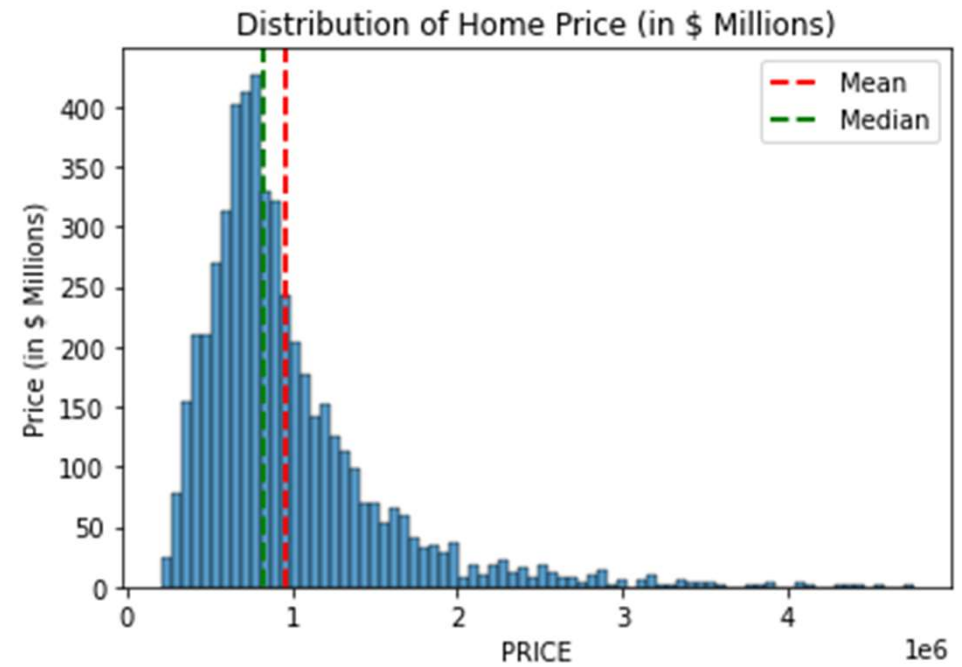
Correlation Heatmap



Price

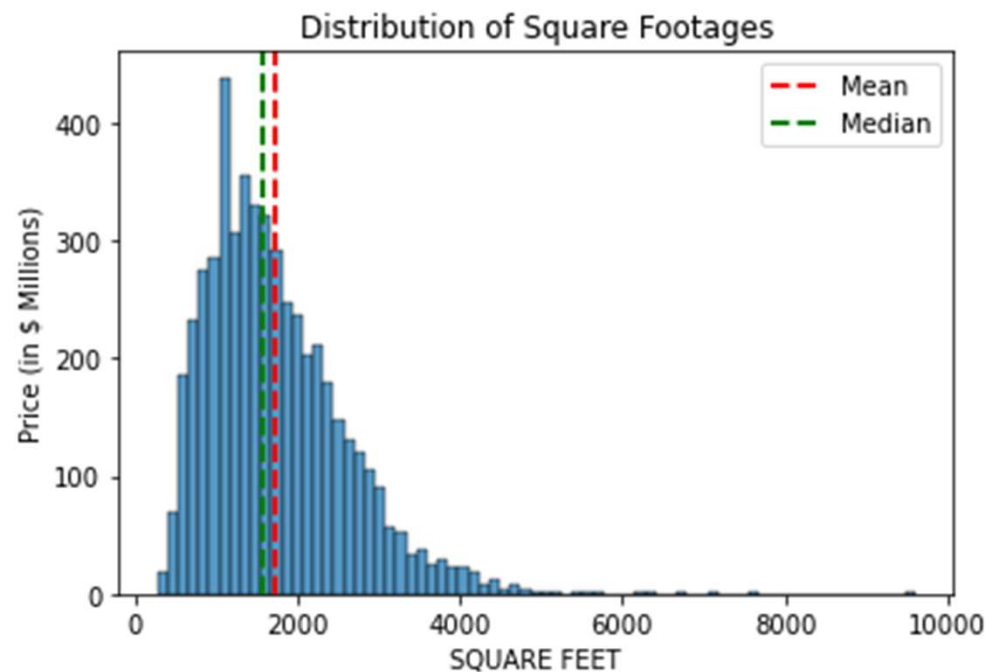
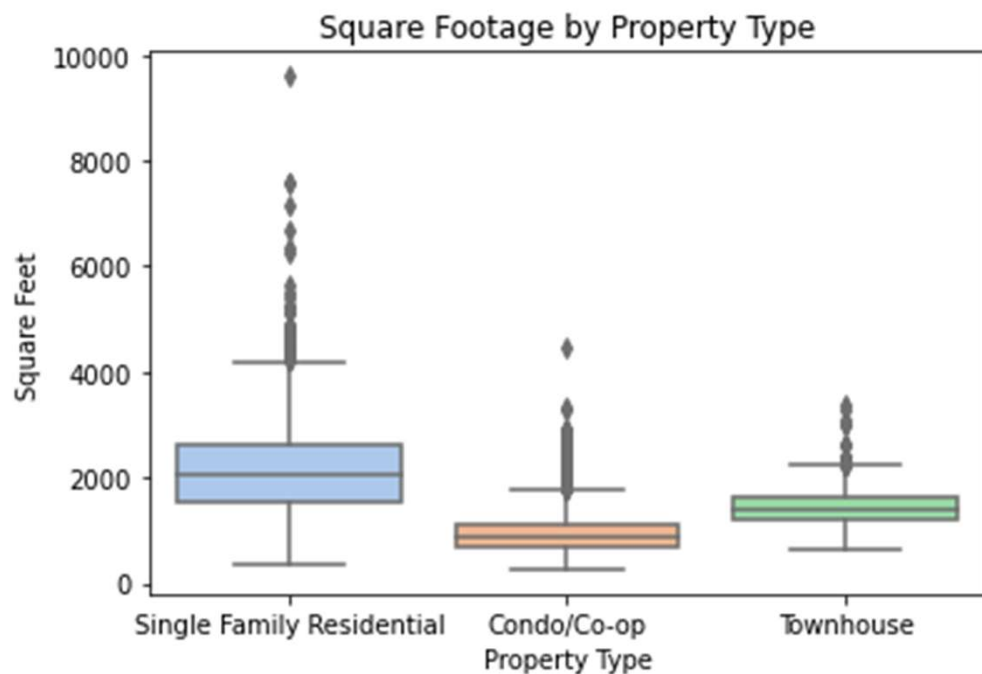
Median: \$820K

Mean: \$958K



Square Feet

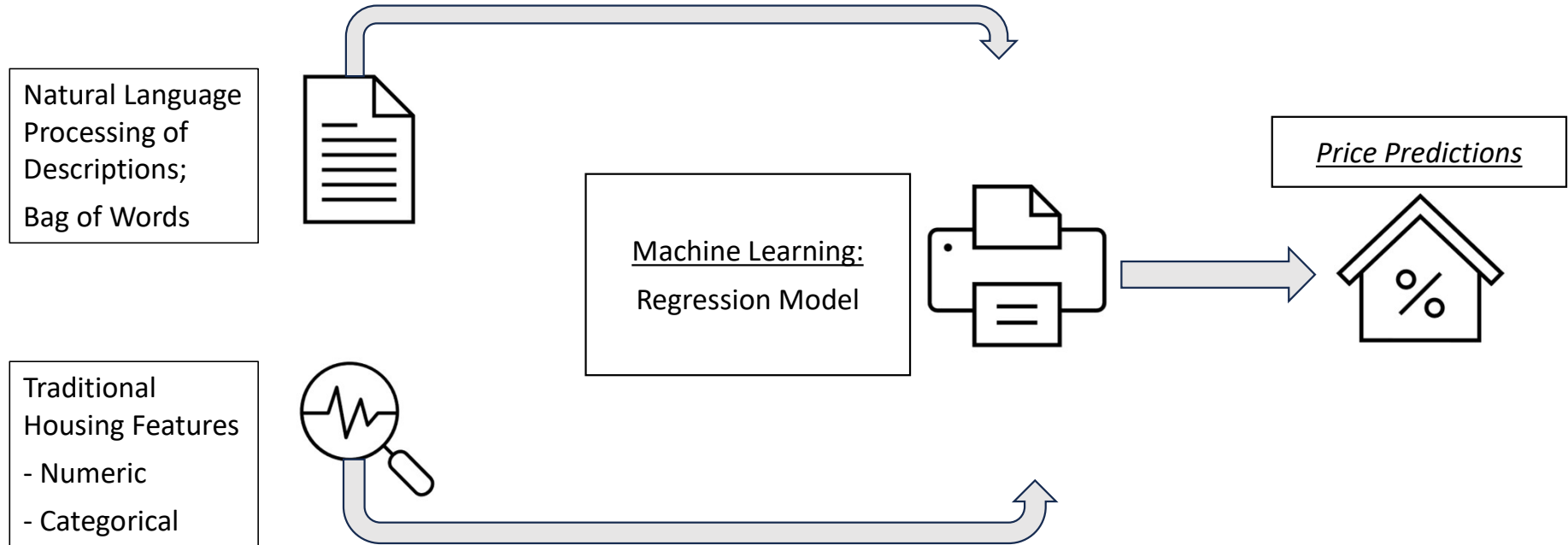
Median: 1,568 SF
Mean: 1,729 SF



Redfin Home Description Word Cloud



The Model Approach



The Metrics

Mean Absolute Error Percentage

Median Absolute Error Percentage

Regression Housing Models

- 1: Decision Tree Classifier (Untuned)
- 2: Random Forest Regressor (Tuned)
- 3: Support Vector Regressor (Tuned)
- 4: XGBoost Regressor (Tuned)
 - Mean Absolute Percentage Error: 11.64%
 - Median Absolute Percentage Error: 7.20%

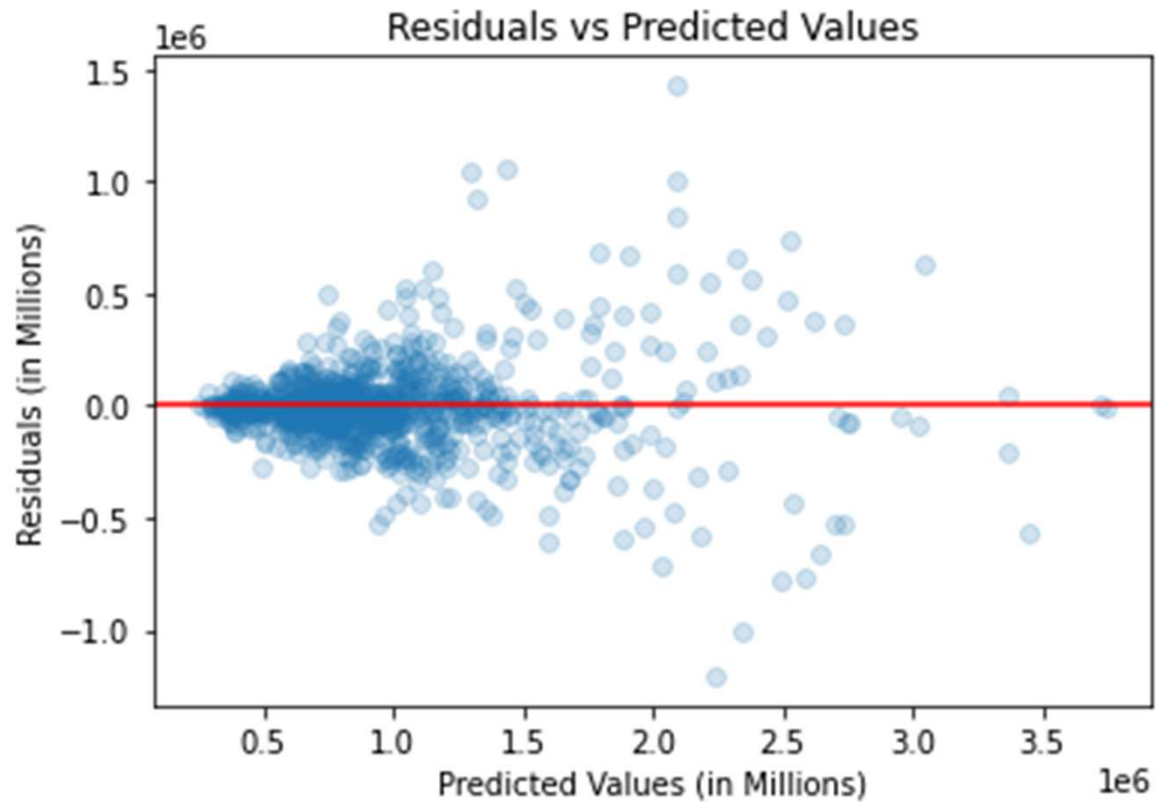


Best Model XGBoost: Hyperparameters

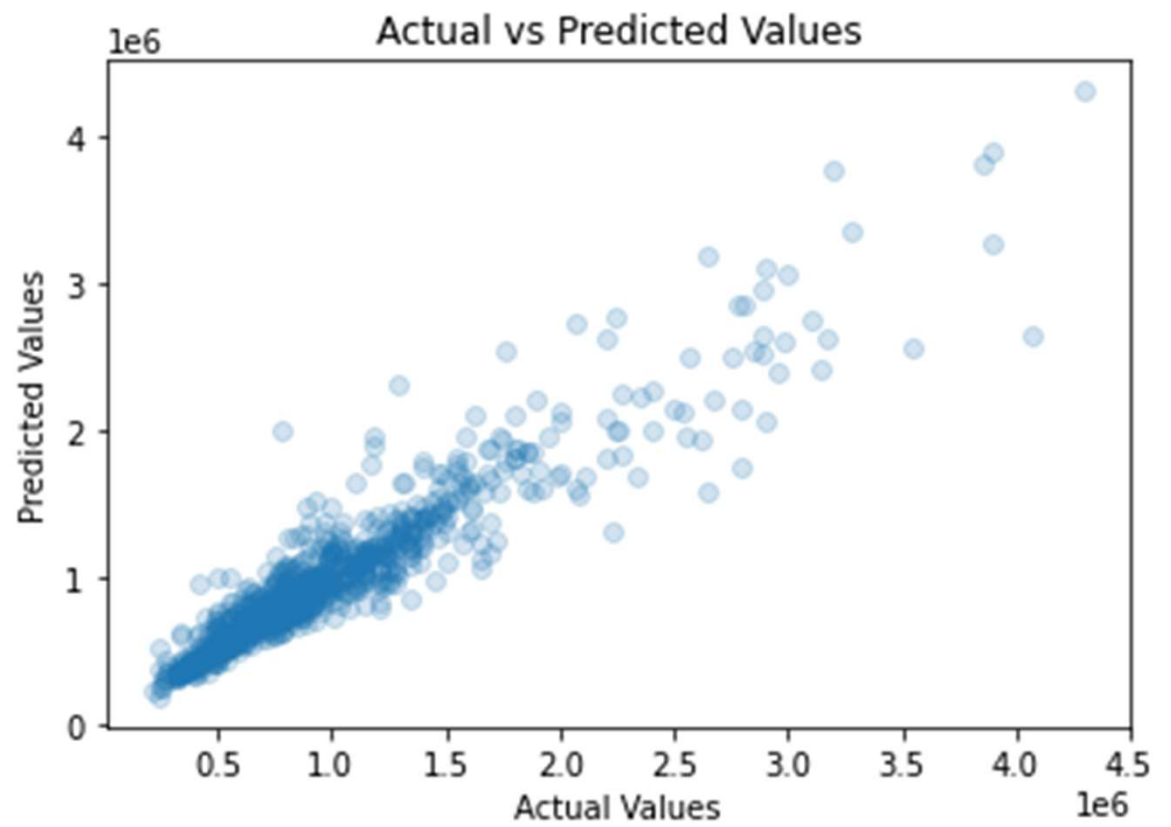
- 80% of Samples of Features Per Tree
- Max Depth of 10
- Minimum Child Weight of 4
- 500 Total Estimators
- 60% of Samples Per Tree.



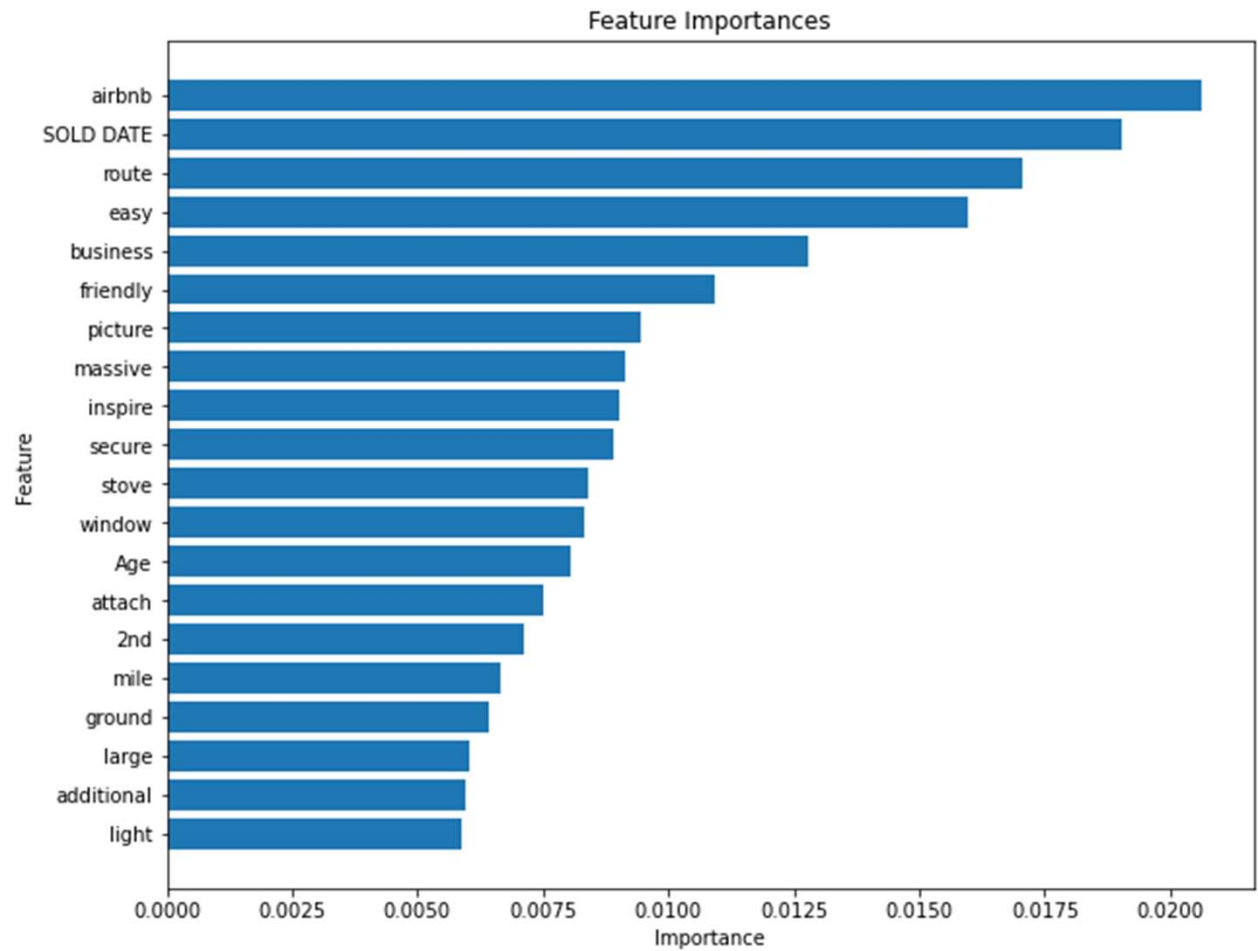
Best Model:
XGBoost
Regressor
Residuals




Best Model:
XGBoost
Regressor
Actual vs.
Predicted



Feature Importance's (XGBoost)





Model Performance Compared to Industry Benchmark

XGBoostRegressor:
Median Absolute
Percentage Error:
7.02%

Zestimate: Median
Absolute
Percentage Error:
2.11%

Conclusion:

- To Be Updated after Bigrams update