

Zillow's House Market Analysis

Group Lakers Team

Raaj Kiran Reddy Anumula

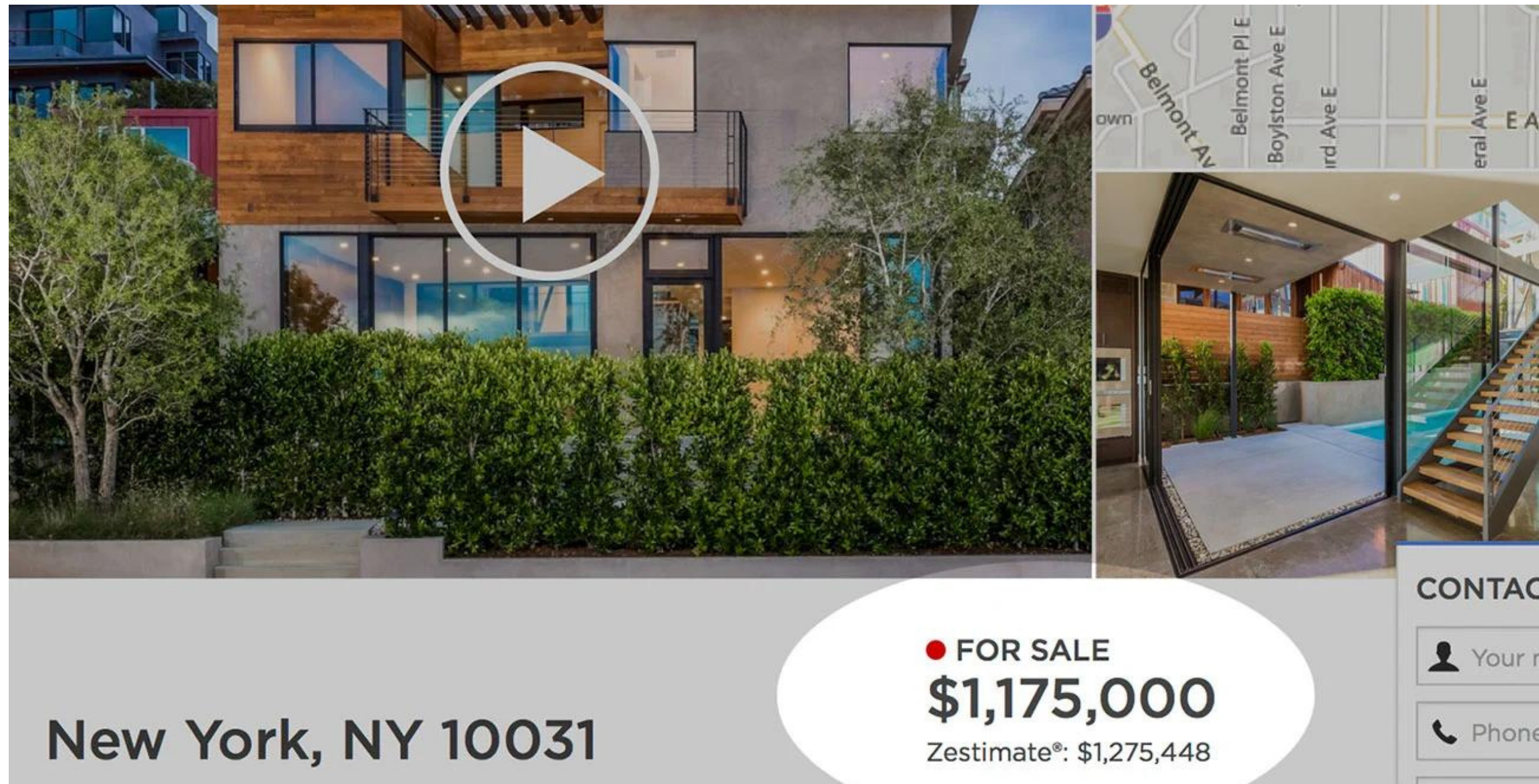
Lokesh Eravelli

Prasad Kalingi

Quan Gu

Understanding The Business Problem

$$\text{logerror} = \log(\text{Zestimate}) - \log(\text{SalePrice})$$



New York, NY 10031

● FOR SALE
\$1,175,000
Zestimate®: \$1,275,448

CONTACT

Your name

Phone

Data Collection

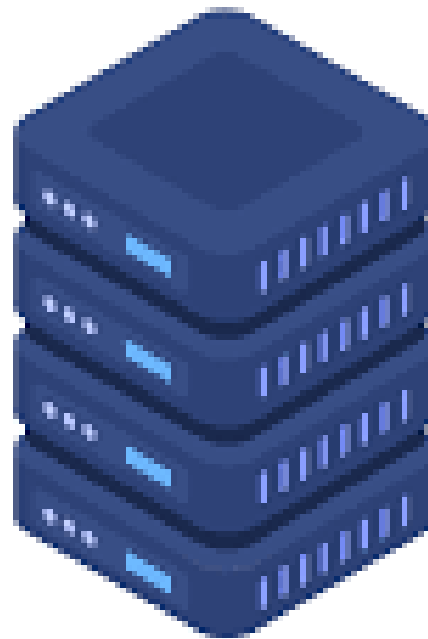


Web Scrapping

Zillow sales url



kaggle



competatiaonal data set

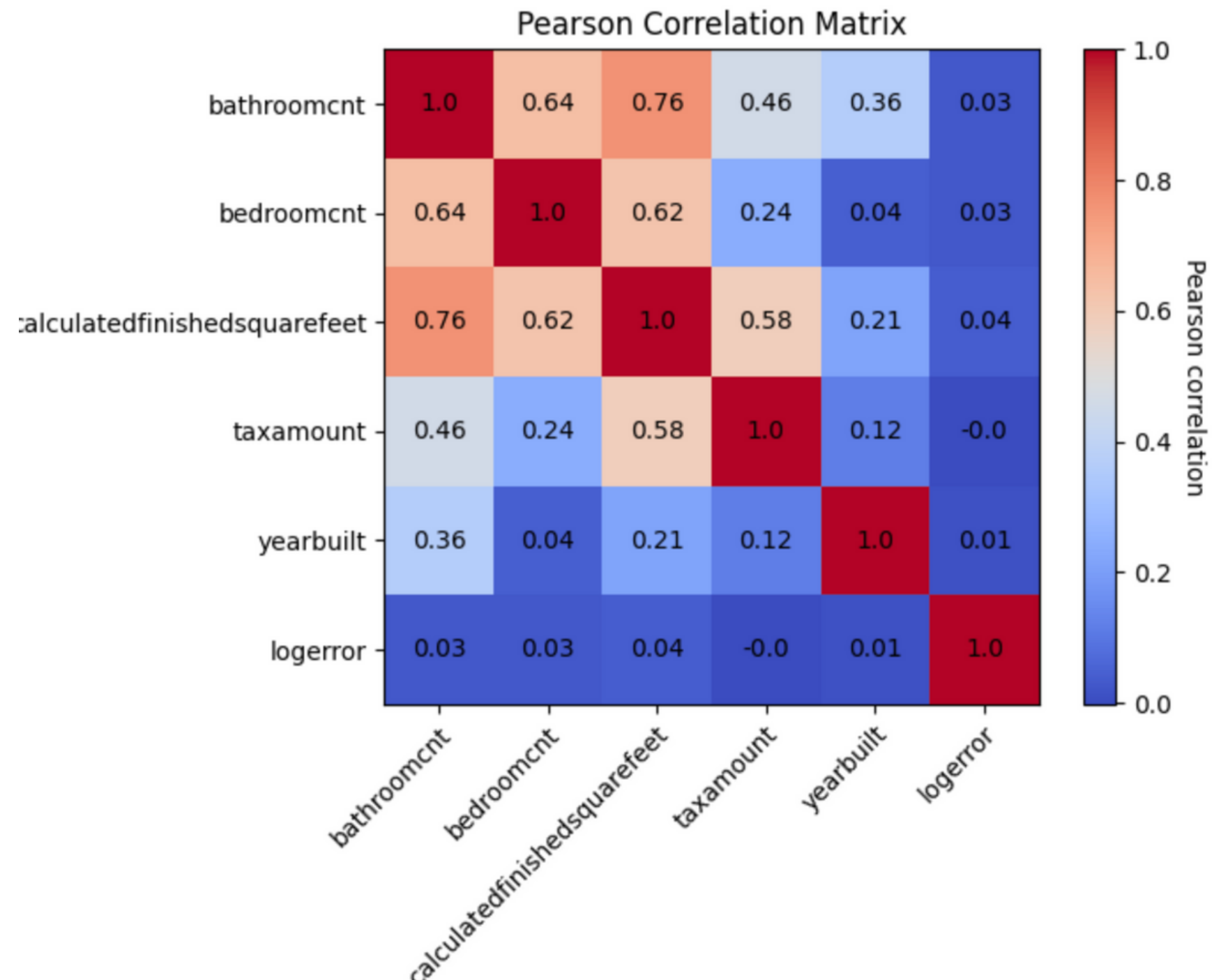


Data Preprocessing

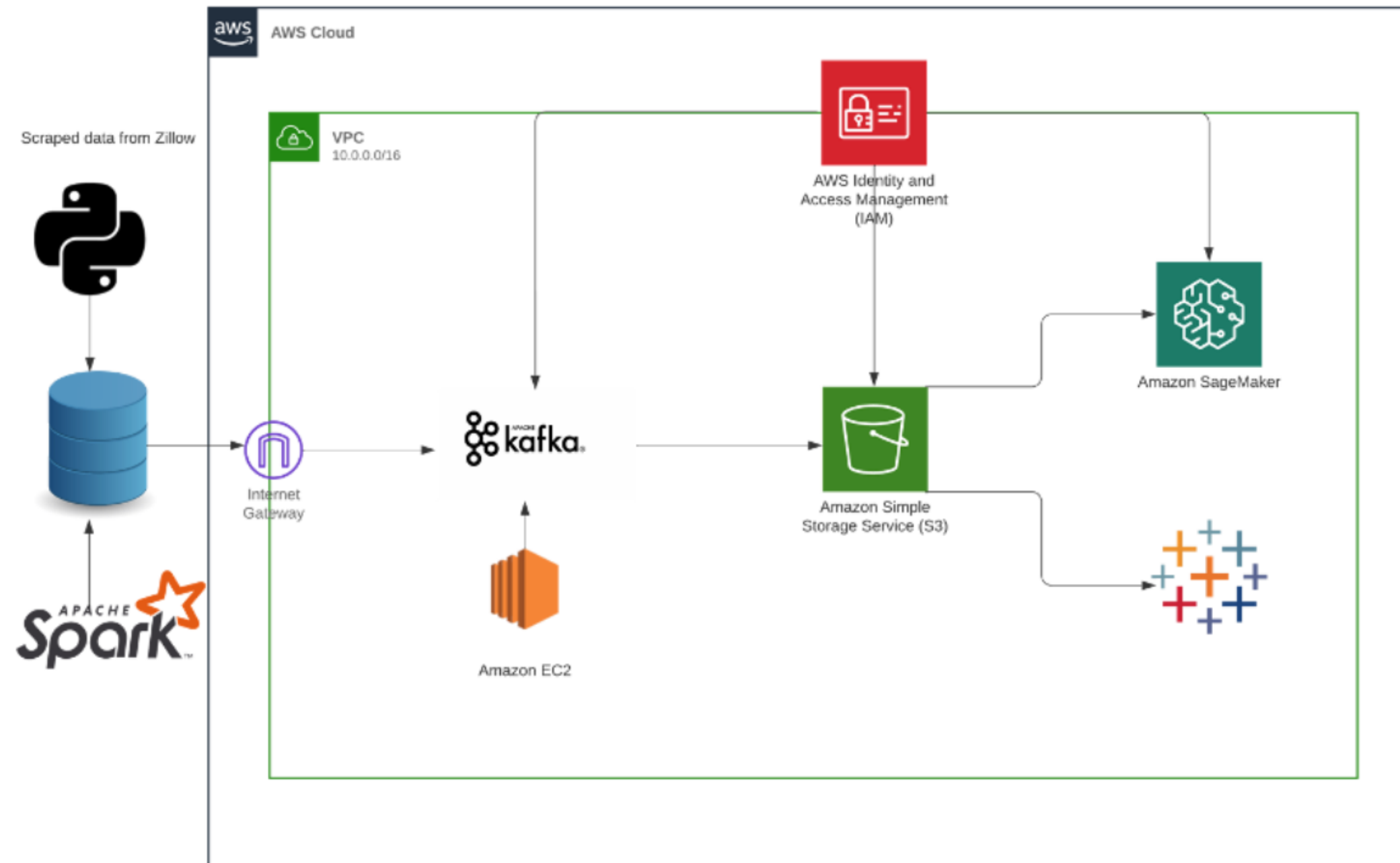


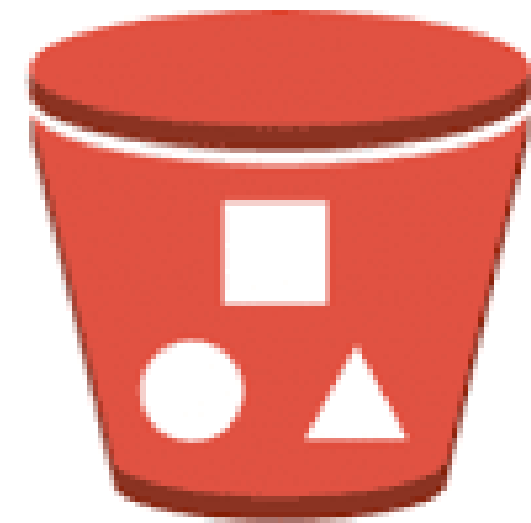
Handling Null values
calculating and dropping them(%>90)
Imputing them accordingly

Exploratory Data Analysis



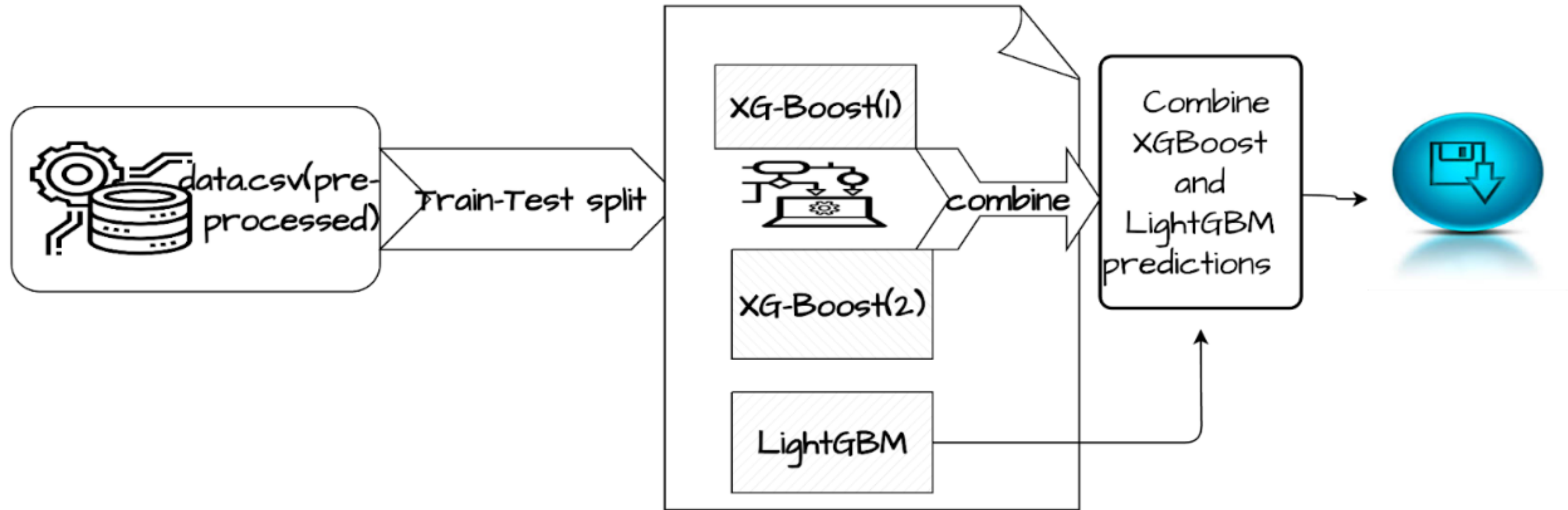
System architecture design





S3 Bucket

ML-Model architecture



LightGBM and XGBoost are two well-liked machine learning techniques for regression issues that are utilized as models in our code. A gradient boosting system called LightGBM makes use of tree-based learning techniques. Although XGBoost is a gradient boosting framework as well, it performs better due to a more advanced approach for managing missing values and outliers.

Logerror Predictions

Predicting with OLS and combining with XGB/LGB/baseline predicitions: ...

predict... 0

predict... 1

predict... 2

predict... 3

predict... 4

predict... 5

Combined XGB/LGB/baseline/OLS predictions:

| | ParcelId | 201610 | 201611 | 201612 | 201710 | 201711 | 201712 |
|---|----------|---------|---------|---------|---------|---------|---------|
| 0 | 10754147 | -0.0184 | -0.0184 | -0.0184 | -0.0184 | -0.0184 | -0.0184 |
| 1 | 10759547 | -0.0078 | -0.0078 | -0.0078 | -0.0078 | -0.0078 | -0.0078 |
| 2 | 10843547 | 0.0558 | 0.0558 | 0.0558 | 0.0558 | 0.0558 | 0.0558 |
| 3 | 10859147 | 0.0461 | 0.0461 | 0.0461 | 0.0461 | 0.0461 | 0.0461 |
| 4 | 10879947 | 0.0097 | 0.0097 | 0.0097 | 0.0097 | 0.0097 | 0.0097 |

Evaluation Results

Mean Absolute Error: 0.01

Mean Squared Error: 0.00

Root Mean Squared Error: 0.02

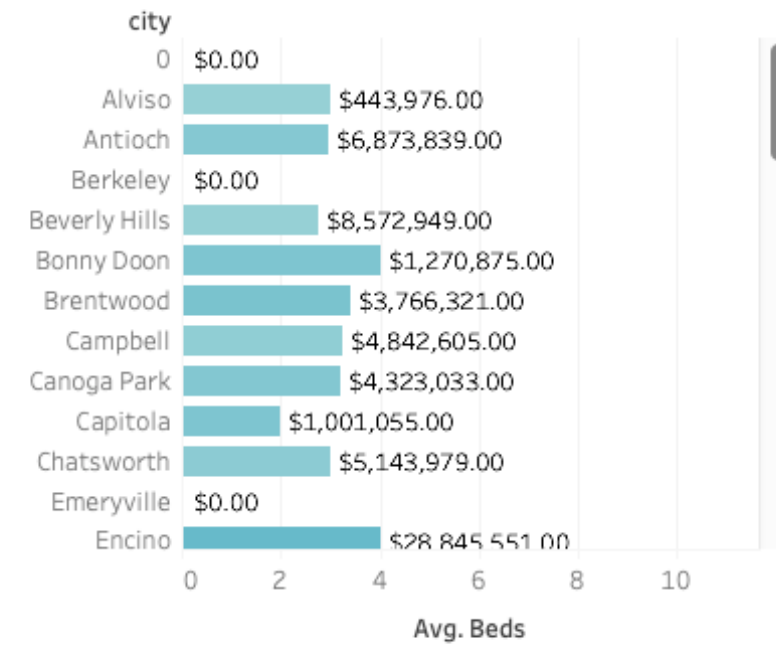
R-squared: 0.00

Explained Variance Score: 0.00

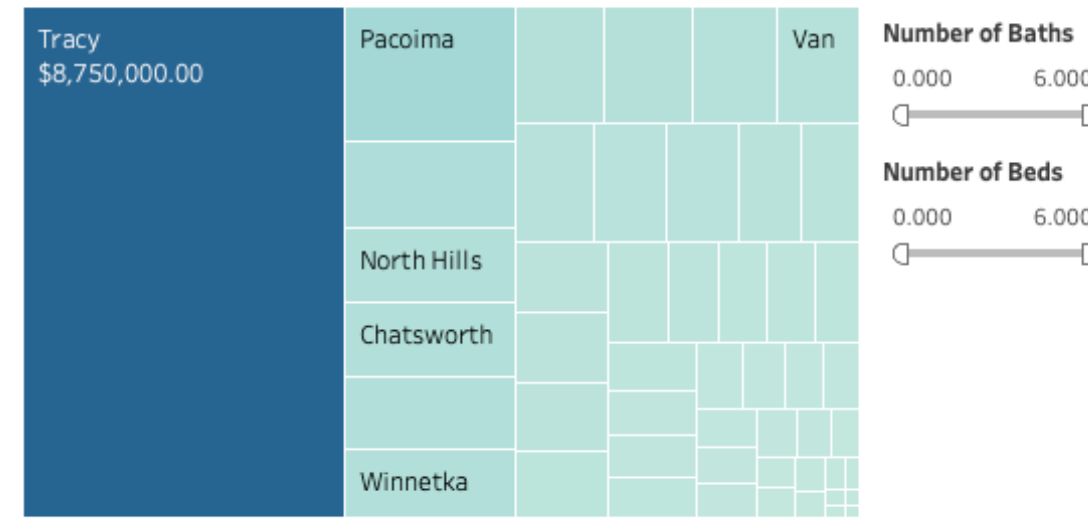
Median Absolute Error: 0.01

Visualizations

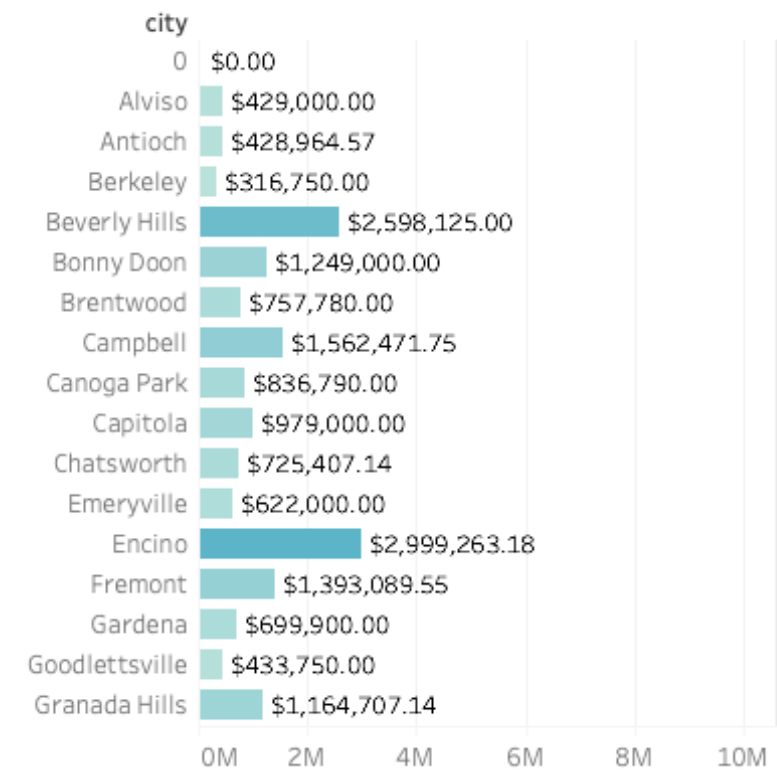
zestimate based on beds and baths



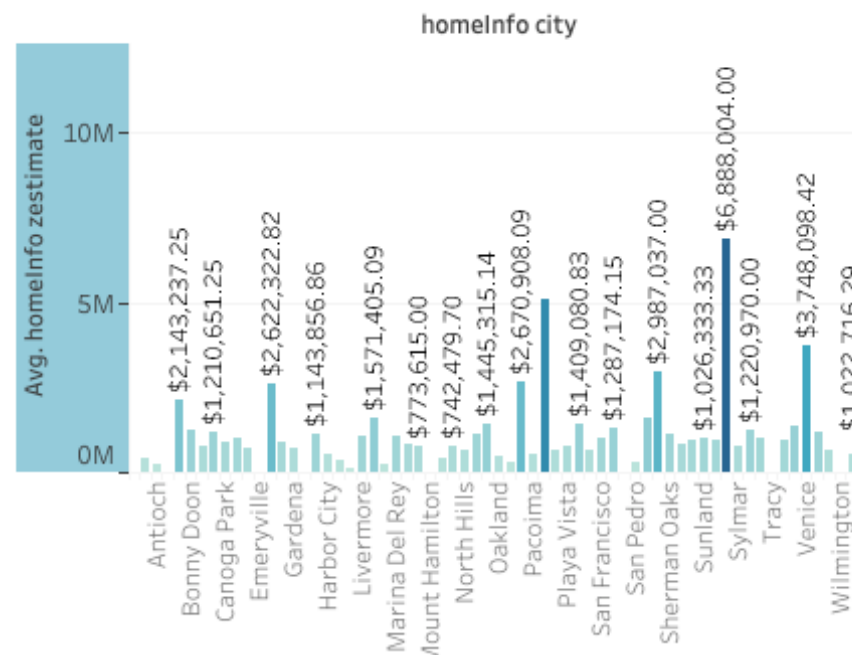
best deal (log error)



actual(sale) price



estimated zillow price



Visualizations



Future work and Scope

- our solution will continue to improve its capabilities to meet the diverse demands of real estate agents, buyers, and sellers.
- Incorporating a greater variety of data sources such as neighborhood features, economic data, and local infrastructure projects can increase the model's accuracy in forecasting property appraisals and providing deeper insights into market patterns.
- Improving the web platform's user interface with interactive elements, visualizations, and tailored suggestions will make it more accessible and meet customer needs.
- AI-driven solutions such as chatbots, natural language processing techniques, and predictive analytics can automate many real estate processes, anticipate market trends, and provide tailored recommendations to increase efficiency and profitability in the sector.

