# Estimating Housing Values With Crime Incident Reports

CMPE 255-01 Team8
Professor: David C. Anastasiu
Pavana Srinivasadeshika Achar, Aartee Kasliwal, Chia-Chin Wu

Project Github https://github.com/Aartee/housing-price-prediction

# Agenda

- Introduction and Motivation
- Datasets
- Data cleaning, pre-processing and feature selection
- Regression models used
- Evaluation
- Conclusion
- Demo

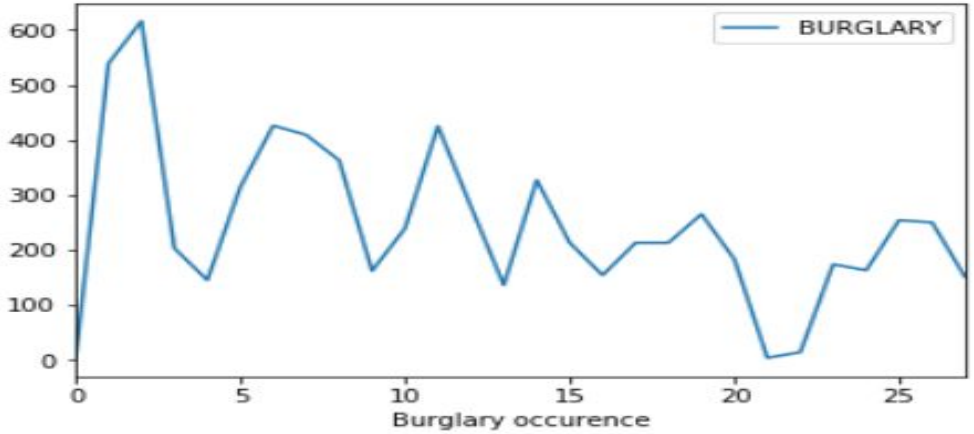
# Introduction

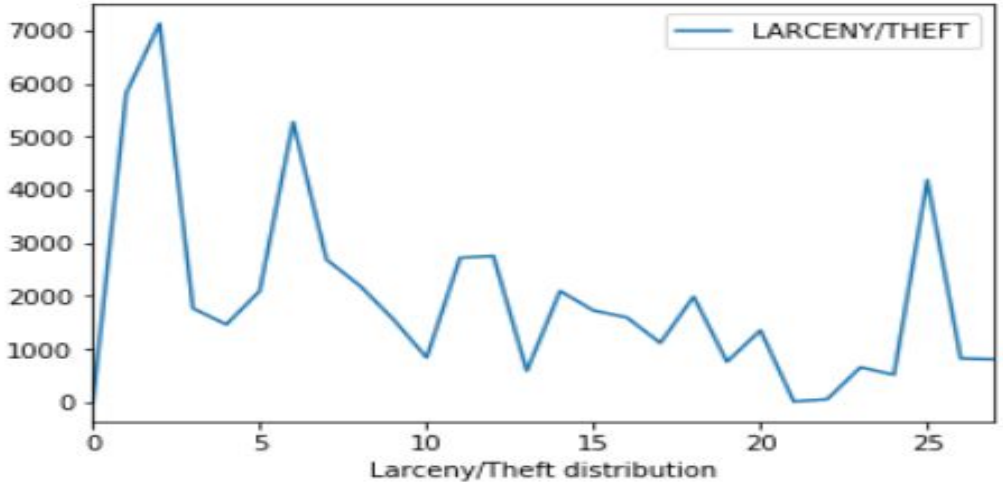
This project intends to find correlations of the criminal incidents with the house values and predict the price for a particular house based on the crime zone or zip code.

We have used San Francisco police department incidents dataset for criminal reports and Zillow's house value index dataset for San Francisco house values

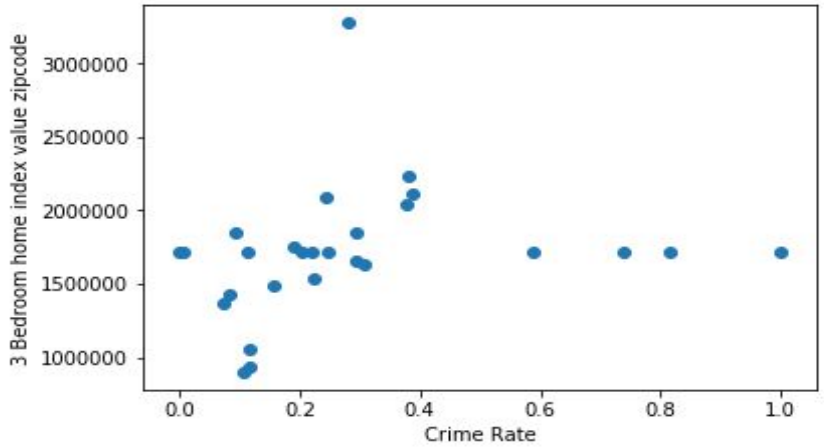
In this prototype, we plan to use the Zillow's dataset for training the module and test the model with listings from other agencies like Redfin and compare the prices.

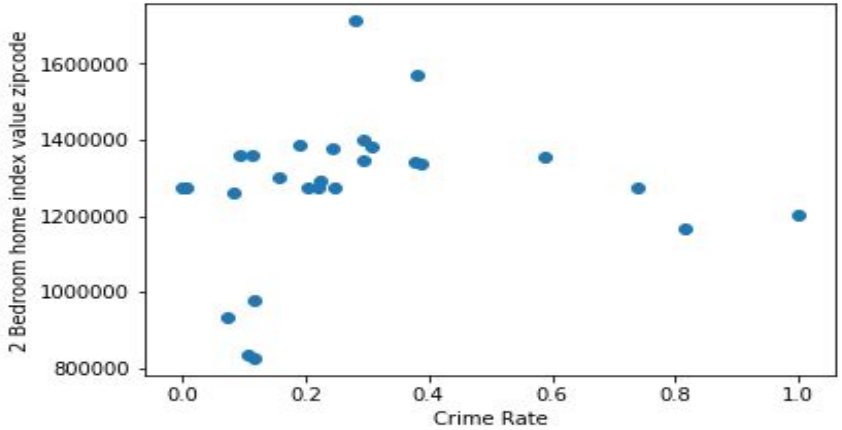
# Datasets


BURGLARY
600
500
400
300
200
100
0
0
5
10
15
20
25
Burglary occurence

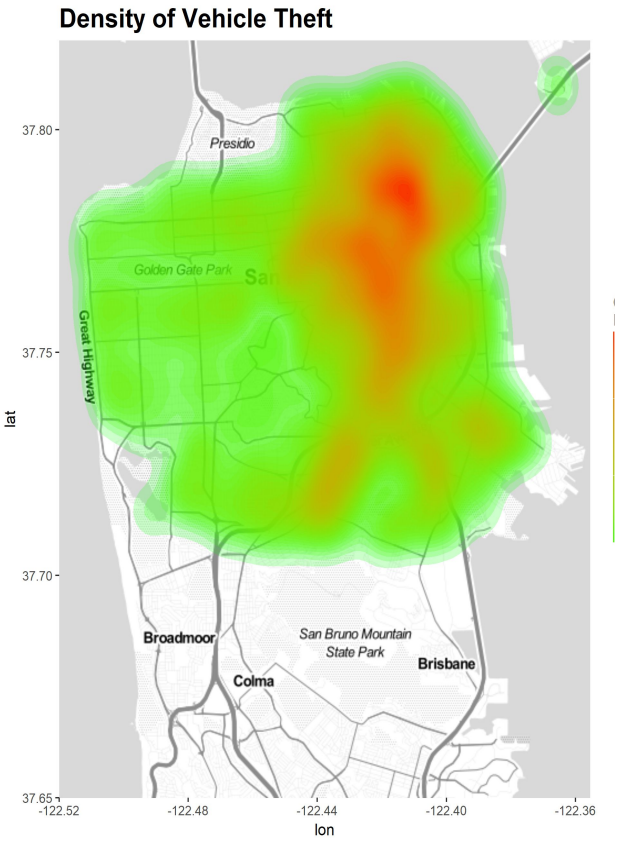

LARCENY/THEFT
7000
6000
5000
4000
3000
2000
1000
0
0
5
10
15
20
25
Larceny/Theft distribution

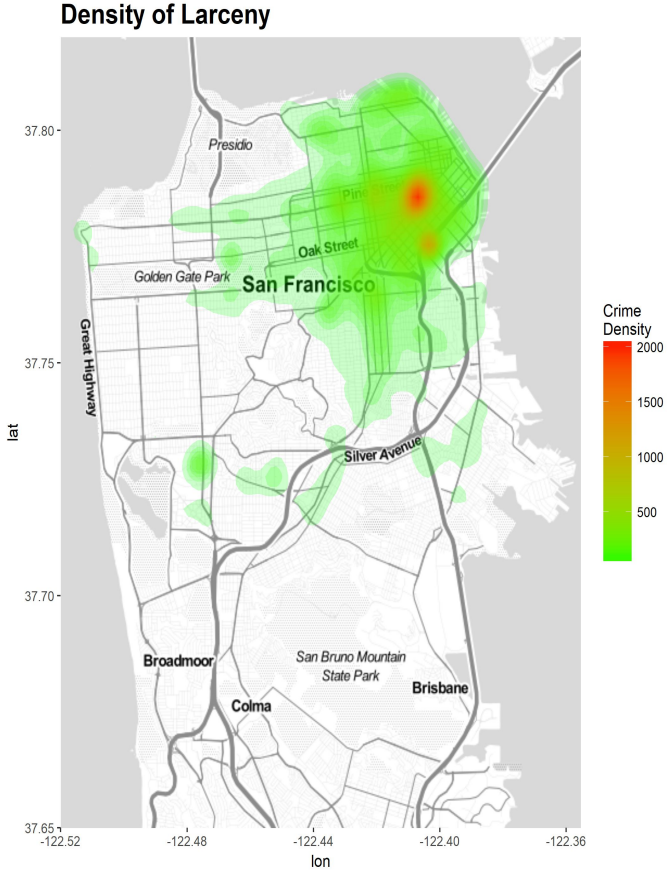
# Datasets


3 Bedroom home index value zipcode
3000000
2500000
2000000
1500000
1000000
0.0
0.2
0.4
0.6
0.8
1.0
Crime Rate

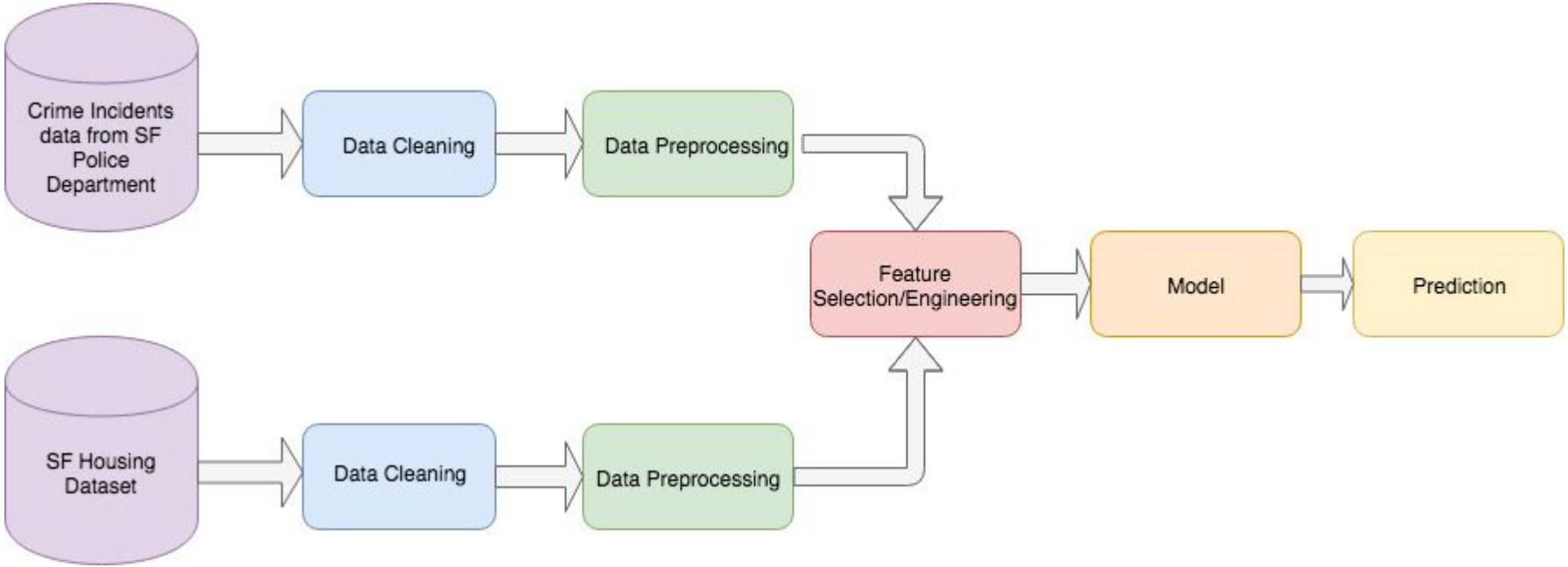

2 Bedroom home index value zipcode
1600000
1400000
1200000
1000000
800000
0.0
0.2
0.4
0.6
0.8
1.0
Crime Rate

# Visualization



# Model Flow Diagram



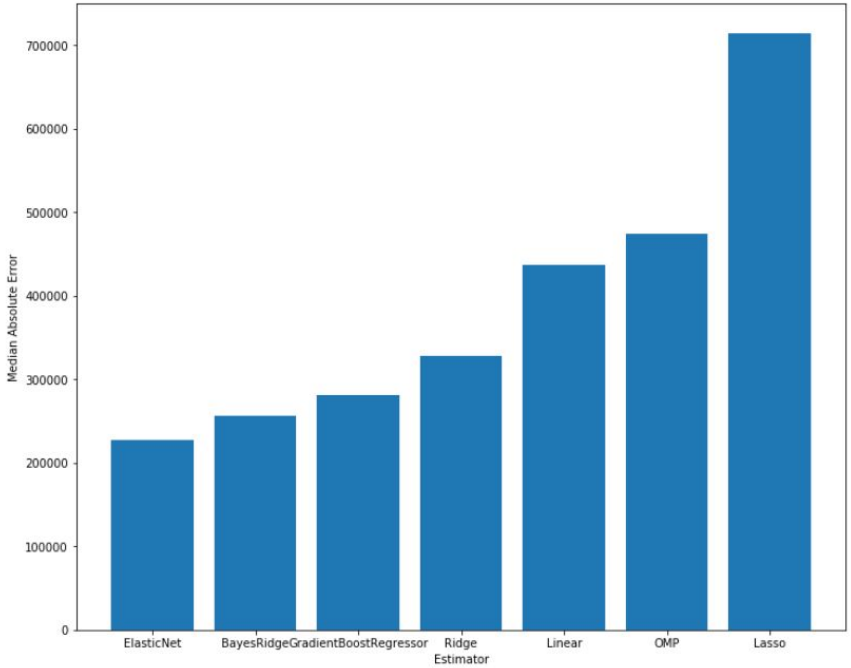
# Cleaning, Pre-Processing and Feature Selection

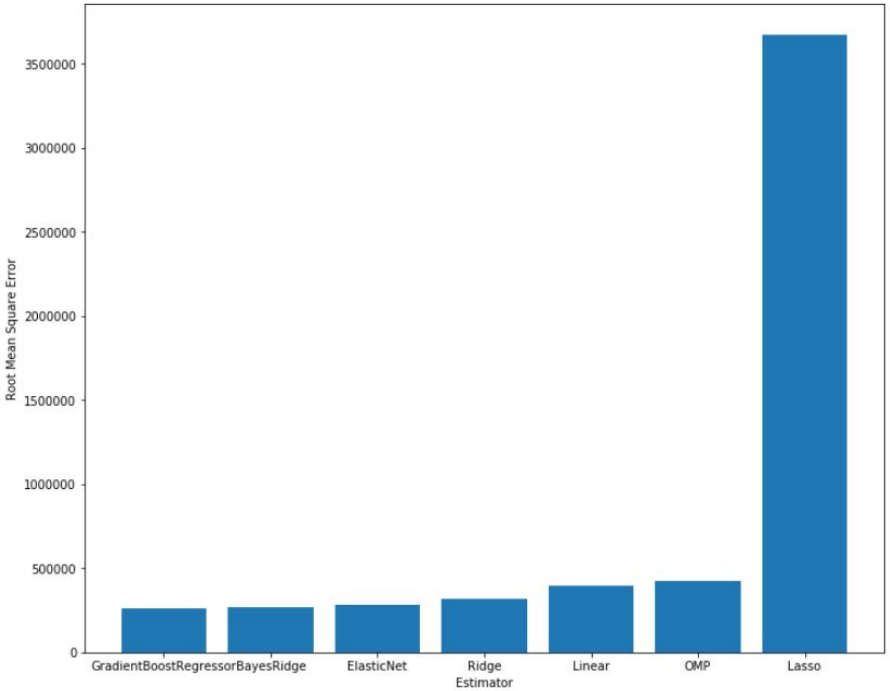
- The feature values for crime categories, latitude and longitude are extracted.
- We used uszipcode module's ZipcodeSearchEngine function to extract the zip code value for each of these location features.
- These rows are grouped by zip codes.
- The categories are aggregated for each group.
- This data :
    - is transformed to have the count features of crime categories associated with zipcode on the rows.
    - is further cleansed for removing NaN values.
    - The dataset is normalized  using min-max normalization technique so that each value is relative to its own column. This way a huge value in a particular feature will not be influencing the overall analysis.

# Regression Models Used

- Seven regression algorithms were implemented and evaluated to predict the housing values.
- The Gradient Boost Regressor, Linear Regressor and ELastic Net resulted with least mean absolute error value.
- other regression models used are:
    - Lasso
    - Linear Regressor
    - Ridgde Estimator
    - OMP Regressor
- The Gradient Boost Regressor was chosen to build the model.

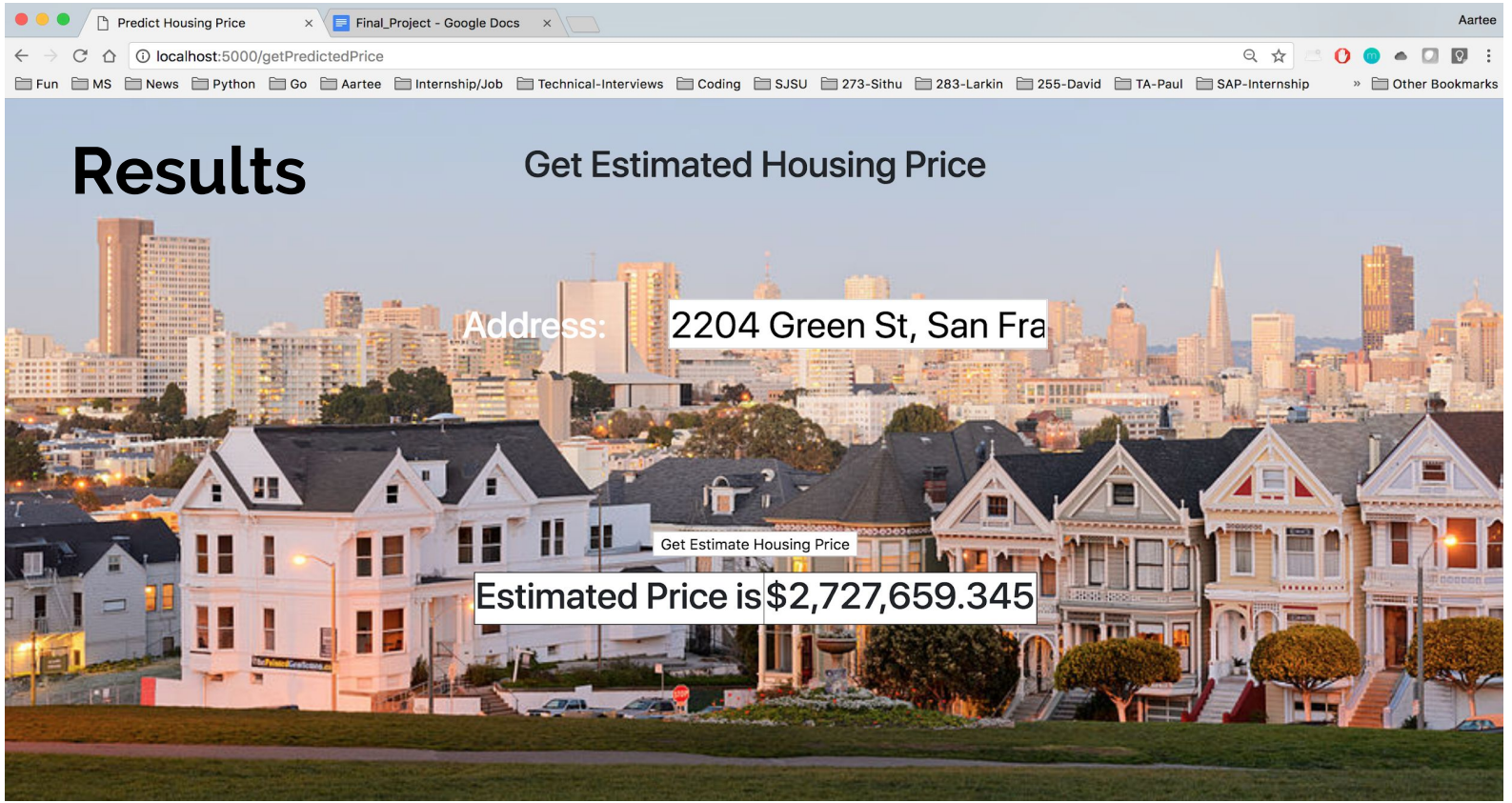
# Evaluation


Median Absolute Error
0
100000
200000
300000
400000
500000
600000
700000
ElasticNet
BayesRidge
GradientBoostRegressor
Ridge
Linear
OMP
Lasso
Estimator

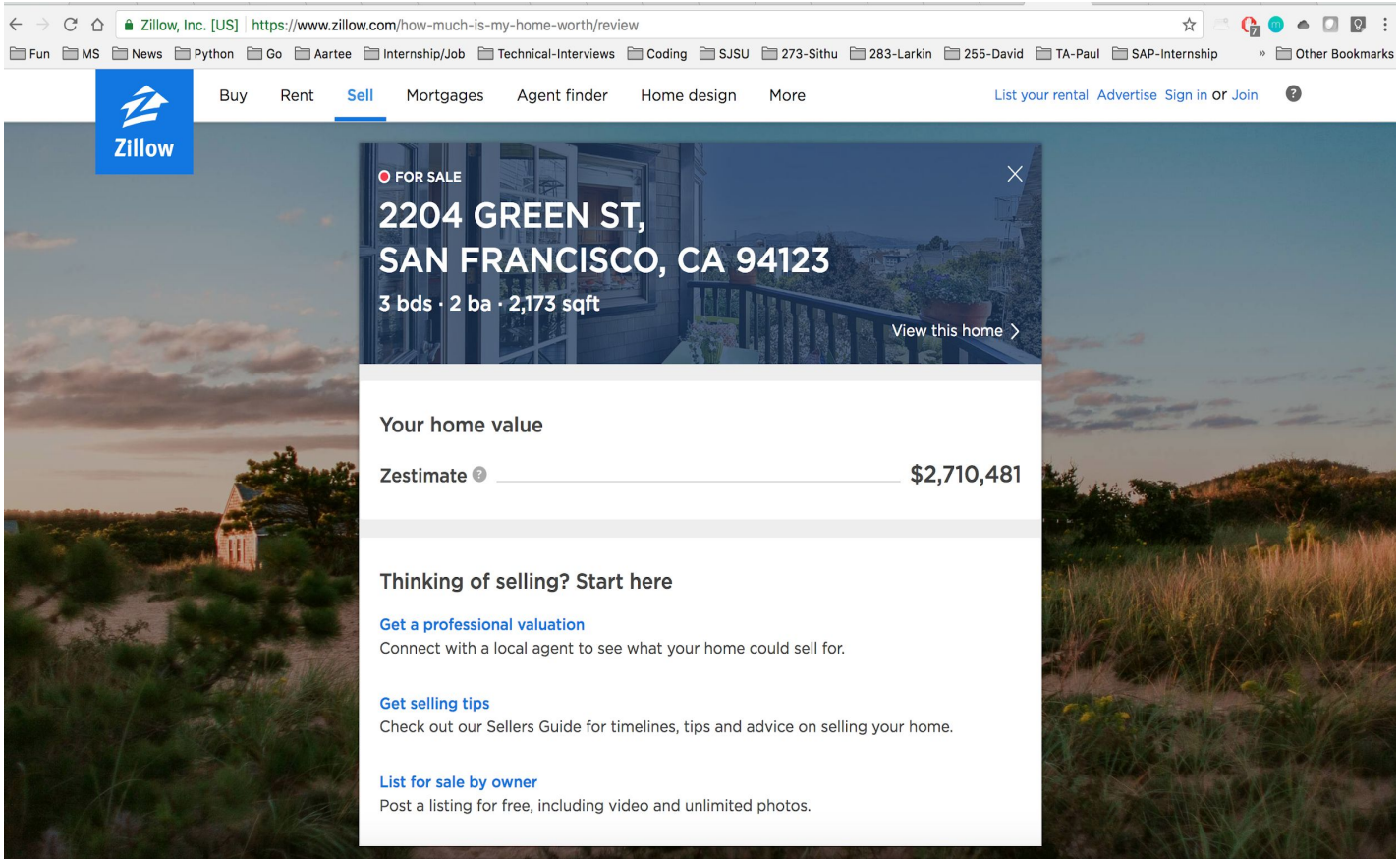

Root Mean Square Error
0
500000
1000000
1500000
2000000
2500000
3000000
3500000
GradientBoostRegressor
BayesRidge
ElasticNet
Ridge
Linear
OMP
Lasso
Estimator

# Evaluation

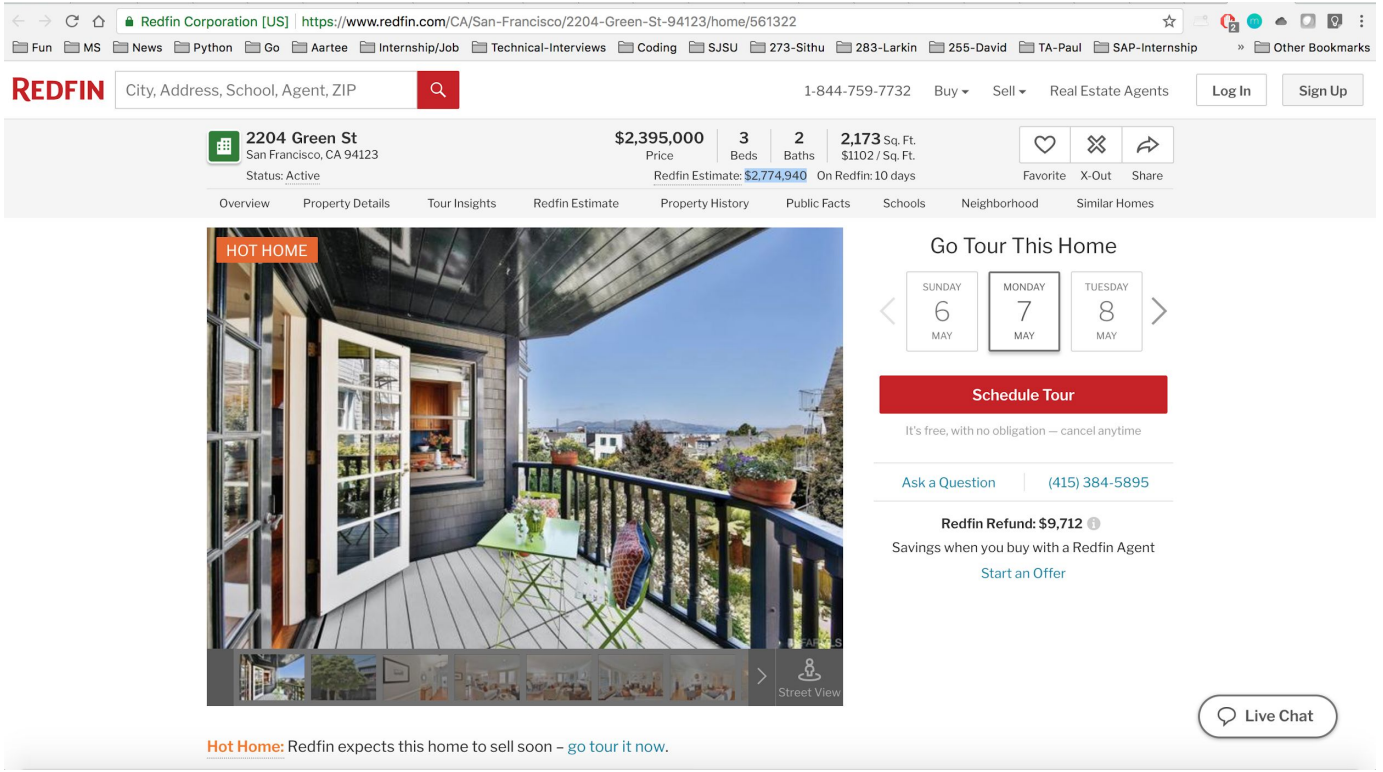
| Regression Model | RMSE | MAE |
|---|---|---|
| ElasticNet | $278,851.437 | $227,396.589 |
| BayesRidge | $267,999.561 | $255,789.832 |
| Gradient Boosted Regressor | $330,144.639 | $369,359.641 |
| Ridge Estimator | $318,868.649 | $327,853.379 |
| Linear | $398,856.522 | $437,108.744 |
| OMP | $422,020.877 | $473,896.455 |
| Lasso | $3,671,197.058 | $713,956.100 |


Predict Housing Price
Final_Project - Google Docs
Aartee
localhost:5000/getPredictedPrice
Fun
MS
News
Python
Go
Aartee
Internship/Job
Technical-Interviews
Coding
SJSU
273-Sithu
283-Larkin
255-David
TA-Paul
SAP-Internship
Other Bookmarks
Results
Get Estimated Housing Price
Address:
2204 Green St, San Fra
Get Estimate Housing Price
Estimated Price is $2,727,659.345

# Results


Zillow, Inc. [US] https://www.zillow.com/how-much-is-my-home-worth/review
Fun MS News Python Go Aartee Internship/Job Technical-Interviews Coding SJSU 273-Sithu 283-Larkin 255-David TA-Paul SAP-Internship Other Bookmarks
Zillow
Buy Rent Sell Mortgages Agent finder Home design More
List your rental Advertise Sign in or Join
FOR SALE
2204 GREEN ST,
SAN FRANCISCO, CA 94123
3 bds · 2 ba · 2,173 sqft
View this home
Your home value
Zestimate $2,710,481
Thinking of selling? Start here
Get a professional valuation
Connect with a local agent to see what your home could sell for.
Get selling tips
Check out our Sellers Guide for timelines, tips and advice on selling your home.
List for sale by owner
Post a listing for free, including video and unlimited photos.

# Results


Redfin Corporation [US] https://www.redfin.com/CA/San-Francisco/2204-Green-St-94123/home/561322
Fun MS News Python Go Aartee Internship/Job Technical-Interviews Coding SJSU 273-Sithu 283-Larkin 255-David TA-Paul SAP-Internship Other Bookmarks
REDFIN
City, Address, School, Agent, ZIP
1-844-759-7732
Buy
Sell
Real Estate Agents
Log In
Sign Up
2204 Green St
San Francisco, CA 94123
Status: Active
$2,395,000
Price
3
Beds
2
Baths
2,173 Sq. Ft.
$1102 / Sq. Ft.
Redfin Estimate: $2,774,940
On Redfin: 10 days
Favorite
X-Out
Share
Overview
Property Details
Tour Insights
Redfin Estimate
Property History
Public Facts
Schools
Neighborhood
Similar Homes
HOT HOME
Street View
Go Tour This Home
SUNDAY 6 MAY
MONDAY 7 MAY
TUESDAY 8 MAY
Schedule Tour
It's free, with no obligation — cancel anytime
Ask a Question
(415) 384-5895
Redfin Refund: $9,712
Savings when you buy with a Redfin Agent
Start an Offer
Live Chat
Hot Home: Redfin expects this home to sell soon – go tour it now.

# Conclusion

This project is a predictive model to solve a regression problem of predicting house sales value based on the crime rates of the neighborhood.

We used San Francisco crime incidents dataset for the criminal records dataset and the Zillow house value dataset for housing data set.

We performed data cleaning, data preprocessing tasks before applying the model. We trained and evaluated a handful of regression models to choose the best applicable algorithm.

We used a basic front end web page to accept the address of the house and predicted the rate for the house. The results were compared with other real estate agency websites like Redfin and Zillow

# References


[1] "Linear Regression Example — scikit-learn 0.19.1 documentation." http://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html. Accessed 6 May. 2018.
[2] "1.11. Ensemble methods — scikit-learn 0.19.1 documentation." http://scikit-learn.org/stable/modules/ensemble.html. Accessed 6 May. 2018.
[3] "Plot Ridge coefficients as a function of the regularization — scikit-learn ...." http://scikit-learn.org/stable/auto_examples/linear_model/plot_ridge_path.html. Accessed 6 May. 2018.
[4] "Orthogonal Matching Pursuit — scikit-learn 0.19.1 documentation." http://scikit-learn.org/stable/auto_examples/linear_model/plot_omp.html. Accessed 6 May. 2018.
[5] "Bayesian Ridge Regression — scikit-learn 0.19.1 documentation." http://scikit-learn.org/stable/auto_examples/linear_model/plot_bayesian_ridge.html. Accessed 6 May. 2018.
[6] "Lasso and Elastic Net — scikit-learn 0.19.1 documentation." http://scikit-learn.org/stable/auto_examples/linear_model/plot_lasso_coordinate_descent_path.html. Accessed 6 May. 2018.