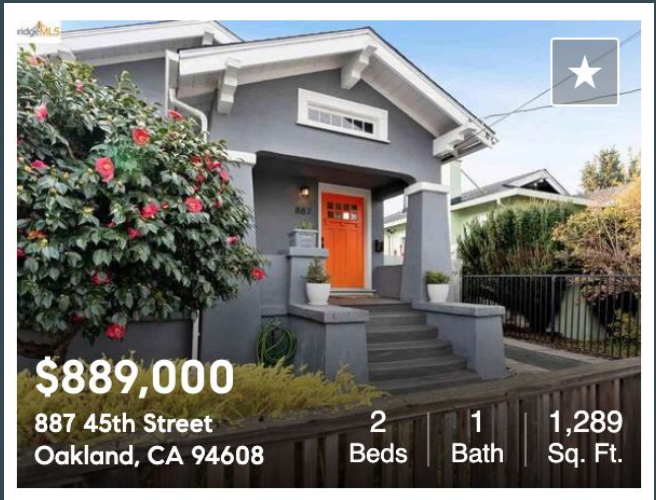
predicting Bay Area rent prices

Emma-Claire McCarthy

The **rent prediction model** will be utilized to make rental income predictions on for-sale properties.

Target User: rental property investors



url → **Rent Prediction Model** → predicted gross rental income

Data Sources

COMPASS

Walk Score®

Tools

Beautiful Soup

Scikit-learn

Pandas

Numpy

Seaborn

Features

bedrooms

bathrooms

square footage

year built

has address 2

walk score

city

parking

laundry

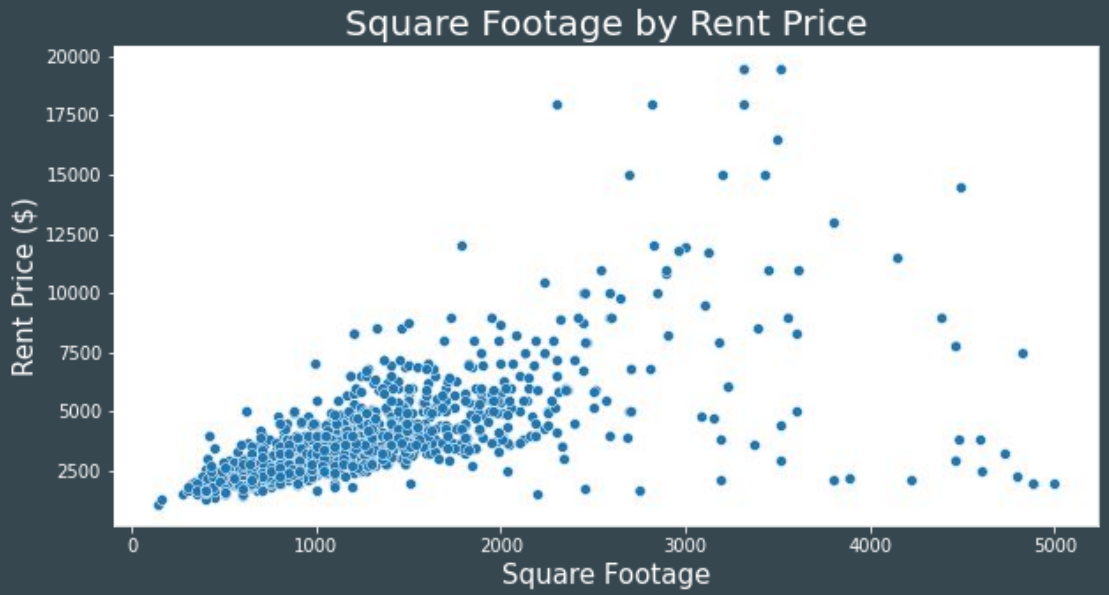
unit type

Model v1 Design and Results

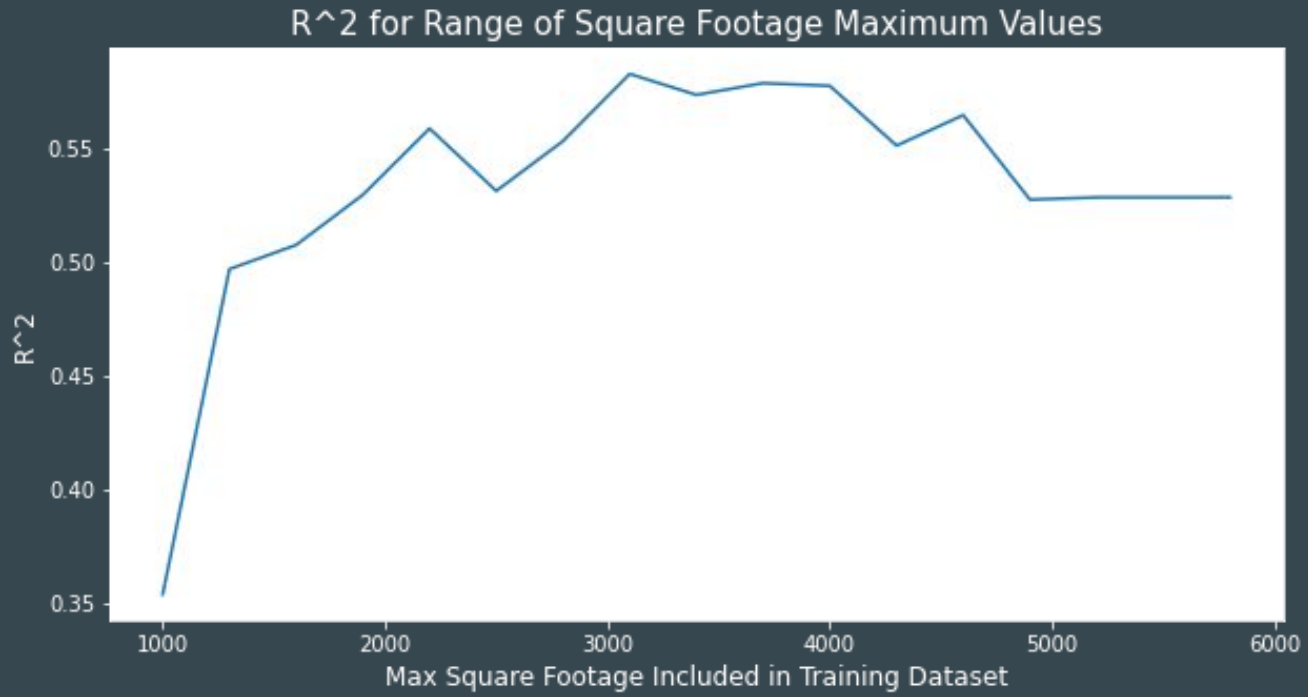
- 10 features → **OneHotEncoder()** → 72 features
- **Feature Engineering:** interaction terms (N=7), polynomial terms (N=2)
- **Feature Consolidation:** “condos” = “townhouses”, etc.
- **No Ridge/LASSO**

5-Fold Cross Validation Mean R^2: 0.567 ± 0.063

Square footage data above 2500 ft^2 is sparse and deviates from the apparent linear relationship with price.



Setting a constraint on model inputs: square footage must be below 3100 ft²



Model v2 design with maximum square footage constraint increased the R^2 by ~0.1

- Exclude all data points where square footage > 3100
- 10 features → **OneHotEncoder()** → 72 features
- **Feature Engineering:** interaction terms (N=7), polynomial terms (N=2)
- **Feature Consolidation:** “condos” = “townhouses”, etc.

5-Fold Cross Validation Mean R^2: 0.666 ± 0.037

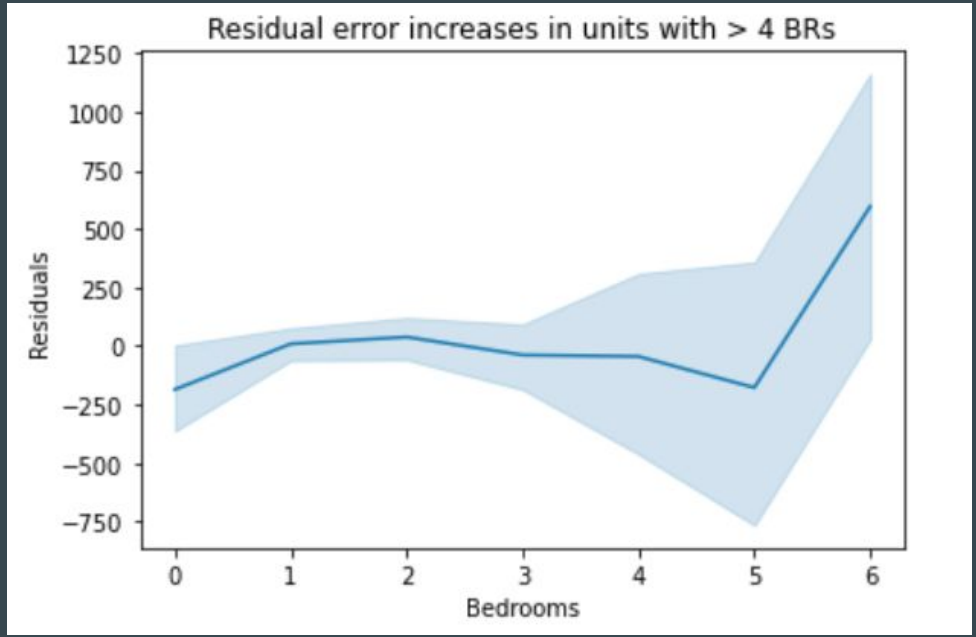
Applying RidgeCV and Evaluating Against Test Set

- Exclude all data points where square footage > 3100
- Model v2 preprocessing and feature engineering
- RidgeCV (alpha_ = 0.028)

Evaluation Against Test Set
R^2 = 0.697
RMSE = $967

Future Considerations

- improve predictions on large rental units
- expand geographical scope to include all bay area counties

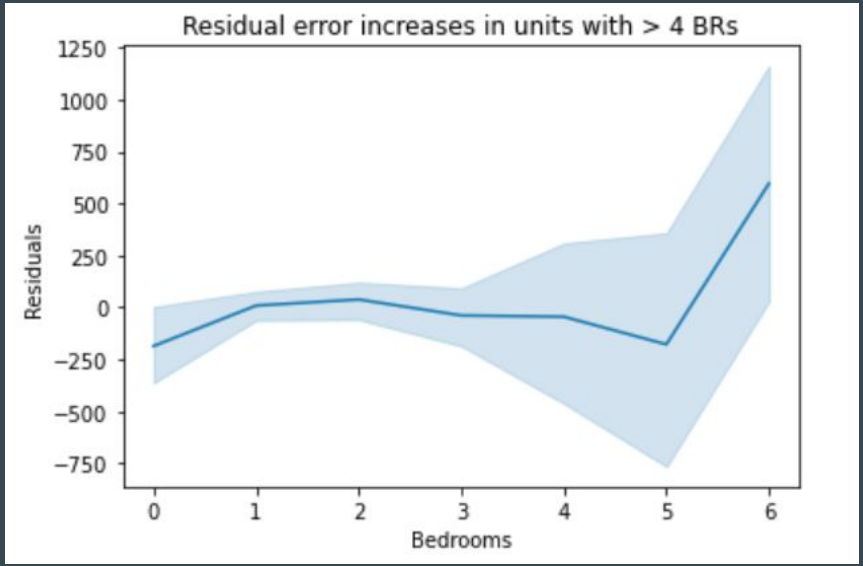


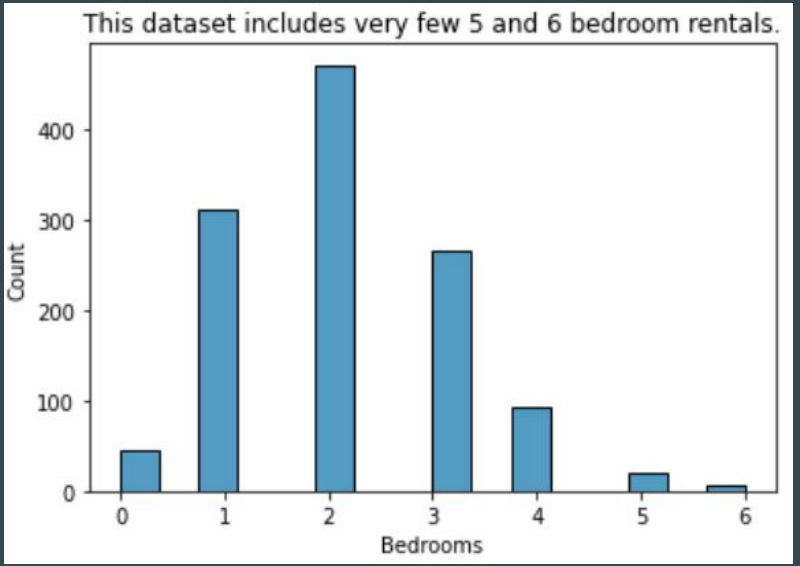
Thank You!

...

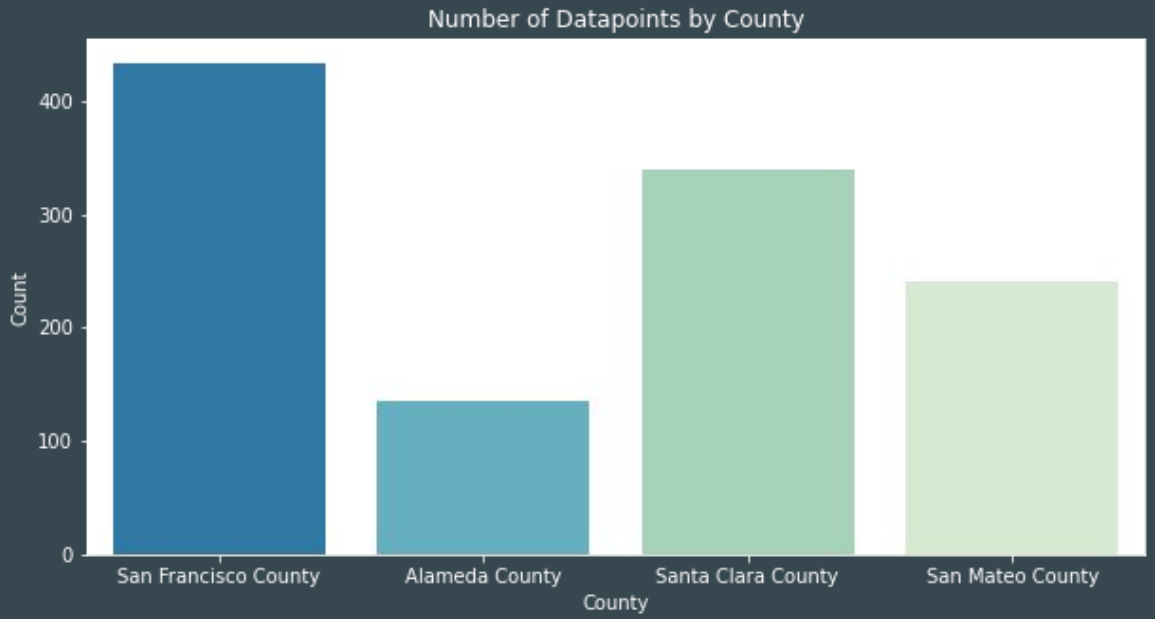
Appendix

High Error on Large Rental Units

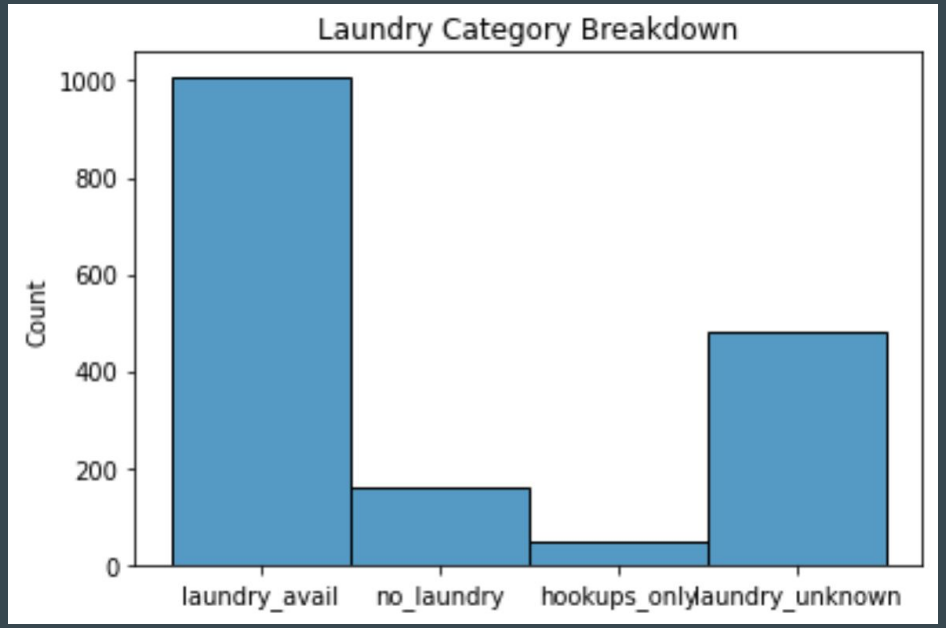




The scraped dataset includes data from four bay area counties.



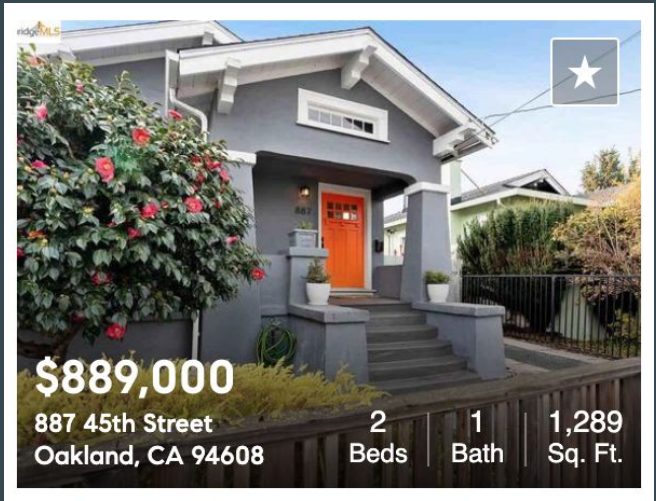
Laundry type was difficult to parse due to the variety of text inputs in the laundry field.



odlaundry:hookups
rslaundry:dryer,w
telaundry:dryer,c
itlaundryandanatt
telaundry:communi
s.laundrylocatedo
utlaundry:cabinet
umlaundry:nonepoo
itlaundry-1design
nelaundry:otherpo
nelaundry:yeslaun
telaundryroomwith
tslaundry:nonedis
sslaundryroom.gle
edlaundryapplianc
etlaundry:dryer,w

The **rent prediction model** will be utilized to make rental income predictions on for-sale properties.

Target User: rental property investors



url → **Rent Prediction Model** → predicted net rental income

predicted net rental income = (predicted) rent - (scraped) HOA - (user-input) insurance - (scraped) taxes - (user-input) upkeep