

San Francisco Apartment Pricing

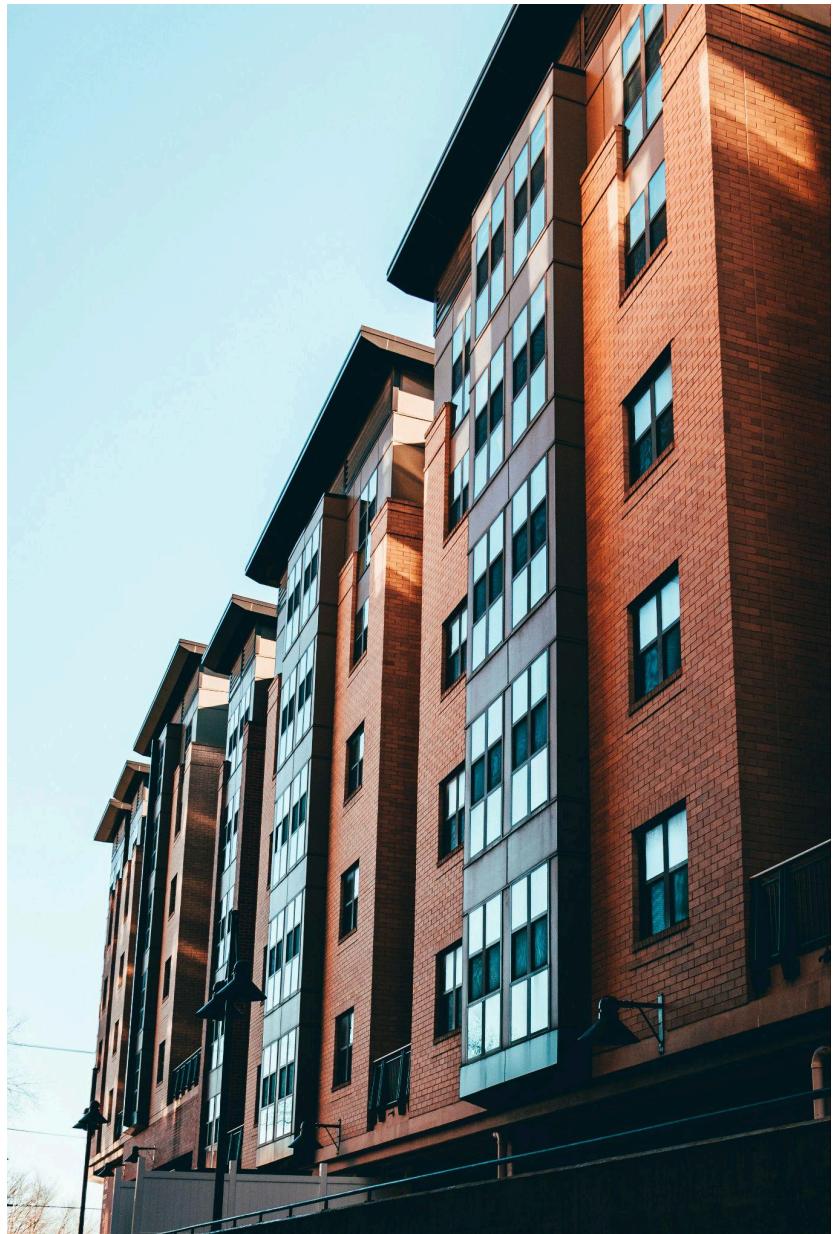
Final Report

Problem Statement

Bay Area Property Management has purchased a large apartment complex in the San Francisco metro area and is considering how to price its units. It wants to use a data science methodology and has asked me to model similar apartments in the area to find how much monthly rent it should charge its tenants for its different units.

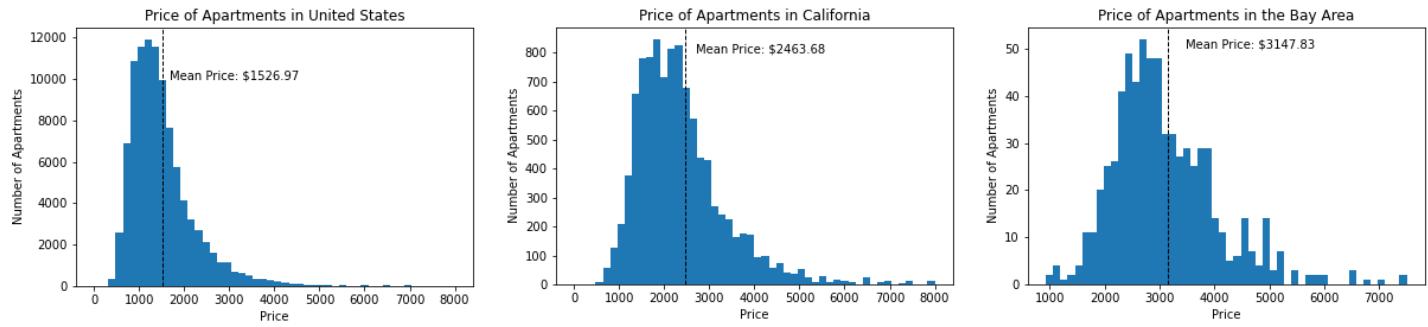
Data Wrangling

The dataset I found to answer this question was a csv of apartment classifieds provided by University of California Irvine. The original dataset contained over 100,000 instances across the United States from December 2018 to December 2019. It included features such as the title of the listing, the number of bedrooms, bathrooms, square footage, city, and state it is located in.

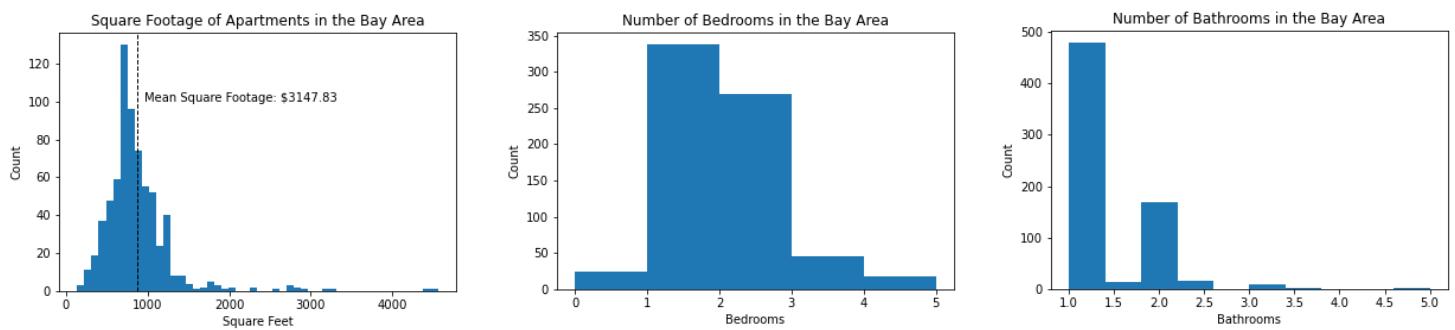


Exploratory Data Analysis

The first thing I wanted to see in this dataset was the difference between prices in the US, California, and the Bay Area. California and the Bay Area in particular is known to have one of the highest costs of living and rents so this will give some interesting context. After slicing the dataset to listings in California and then cities considered in the San Francisco Bay Area we can confirm this assumption. From here on I chose to only look at the Bay Area data as it makes sense to compare to the local market.



The next features I felt would be valuable to explore are the size of apartments in square feet, as well as the quantity of bedrooms and bathrooms. I believe these will be the most important indicators when it comes to modeling. It shows that the average unit has a square footage is around 3,000 ft² and has 1 or 2 bedrooms or bathrooms.



Preprocessing

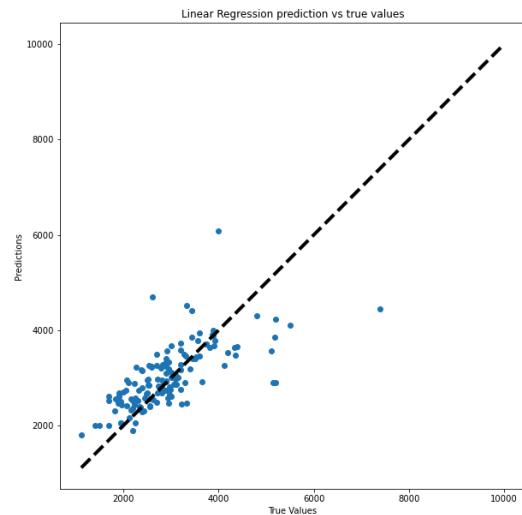
During the feature selection process I removed a couple of features I believed wouldn't be useful. Two of these were the title and body of the listing. These were bodies of text data that couldn't be encoded. I initially included latitude and longitude features in my models, but found it led to overfitting results that could not be generalized.

The training and testing set includes 5 features; The number of bedrooms in the unit, the number of bathrooms, whether dogs and/or cats where allowed, the square footage, and the city it is in. I then encoded the city name and pets features, and used a standard scaler to scale the square footage.

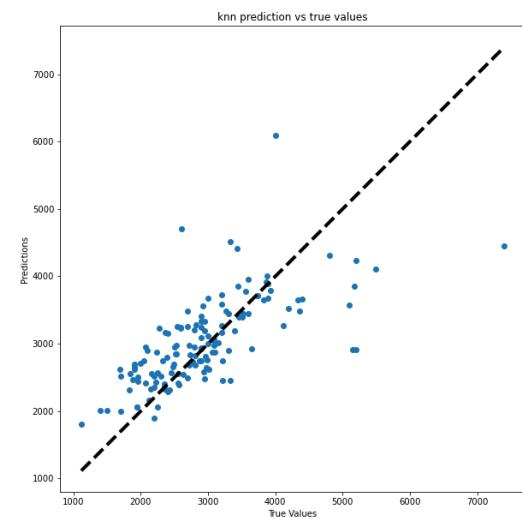
I then created a training and testing split with a test size of 20%.

Modeling

Before evaluating the models I chose to use the R^2 scoring metric due to my models being regression models. The first model I chose to train on this dataset was a linear regression model. This is a relatively simple model and will give a good baseline to judge my other models against. This model had an R^2 score of 0.273 which isn't great, but is better than nothing.

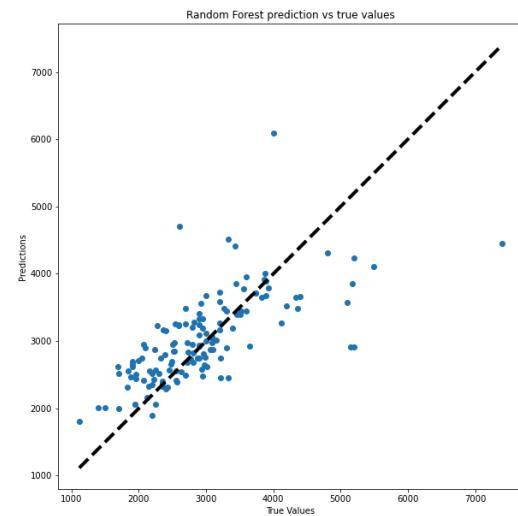


The next model I ran was a K-Nearest Neighbors model. I used a grid search for hyperparameter tuning of neighbors between 1 and 25 and the weighting to be 'uniform' and 'distance'. This model had an R^2 score of 0.507 and the best parameters were 3 neighbors and 'distance' weighting.



Nolan Long

I then created a Random Forest model. The Hyperparameters for the grid search included the number of estimators of [100, 200, 300, 1000] and the maximum number of features of each tree being ['auto', 'square root', 'log2']. This created an R^2 score of 0.554. The best hyperparameters were 1000 estimators and square root as the max features. Due to the interpretability of this type of model I was able to find the importances of each feature. The most important was square footage with a score of 0.43 followed by city name of 0.24, bathrooms 0.20, bedrooms .10, and pets with .03. This was my best performing model and the interpretability was valuable context to give the apartment complex I was advising which is why I chose to select this model for my predictions.



Final Predictions

Using the model I was able to give pricing predictions for its following types of units below. Note that the complex allows cats and dogs and is located in the city of San Francisco

Unit Name	Bedrooms	Bathrooms	Square Feet	Price
Studio	0	1	500	\$2,380
1br	1	1.5	750	\$3,418
2br	2	2	1,250	\$4,667