Predicting Airbnb Prices

Natural Chan

Design

- Purpose:

  Holiday Weekend in San Francisco

- Goal:

  - Predict how much a listing should go for

 - Predict how much an individual should expect to spend on their stay

Data

- Scrape a list of zip codes within 20 miles
- Use these zip codes to create a list of queries
- Search each query on Airbnb and gather as much information as we can on each listing

- Over 20000 listings
- 13 features

Data Preprocessing and Analysis

- Removed duplicate listings
- Removed erroneous observations
- Encoded categorical features
- Removed features an individual would not have

Distribution of Prices
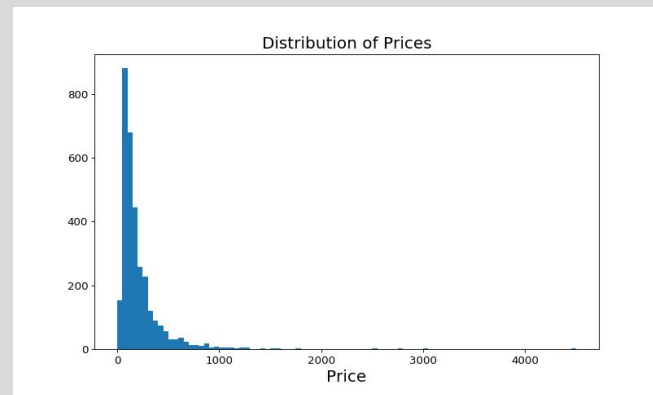
	title	listing_type	city	zip_code	numerical_descriptors	amenities	dates	rating	num_reviews	discounted	original_value	price_per_night	cleaning_fee	service_fee
7194	Queen Anne Hotel, Victorian King	room in boutique hotel	Western Addition/NOPA	94133	2 guests - 1 bedroom - 1 bed - 1 private bath	Wifi	Dec 3 – 5	4.45	22	False	NaN	179.0	0.0	0.0
919	Dorm Bed In Friendly Hostel Community #13	hostel beds	Financial District	94104	3 guests - 1 bedroom - 8 beds - 1 shared bath	Wifi - Kitchen - Washer	Dec 24 – 26	4.70	102	False	NaN	53.0	0.0	0.0
12400	Sunny enclosed backyard in quiet St w easy par...	entire rental unit	Paradise Park	94649	6 guests - 3 bedrooms - 3 beds - 2 baths	Wifi - Kitchen - Free parking - Washer	Dec 24 – 26	4.50	4	False	NaN	300.0	120.0	102.0
19524	Private room &bath /shared driveway	private room	Daly City	94044	1 guest - 1 bedroom - 1 bed - 1 private bath	Wifi - Free parking	Dec 17 – 19	NaN	0	True	74.0	38.0	40.0	0.0
19734	Affordable 1Bed/1Bath w/ Separated Living Room	entire rental unit	South San Francisco	94066	2 guests - 1 bedroom - 1 bed - 1 bath	Wifi - Washer	Dec 31 – Jan 2	4.26	39	False	NaN	69.0	60.0	28.0
2498	Private Guest Suite with Private Bathroom and ...	private room	Portola	94110	3 guests - 1 private bedroom - 3 beds - 1 priv...	Wifi - Self check-in	Dec 17 – 19	4.85	172	False	NaN	129.0	59.0	45.0
9890	Charming 2 Bedroom Apartment	entire rental unit	Dogtown	94609	4 guests - 2 bedrooms - 2 beds - 1 bath	Wifi - Kitchen - Washer	Dec 3 – 5	4.78	73	False	NaN	144.0	120.0	58.0
11640	Cottage in Temescal	entire rental unit	Shafter	94619	2 guests - 1 bedroom - 1 bed - 1 bath	Wifi - Kitchen	Dec 24 – 26	4.84	55	False	NaN	130.0	100.0	51.0
1873	Awesome location in Union Square at THE DONATELLO	hotel room	Downtown	94108	4 guests - Studio - 2 beds - 1 private bath	Wifi - Self check-in	Dec 24 – 26	5.00	7	False	NaN	374.0	0.0	106.0
4675	Quiet Getaway. Centrally located in Inner Rich...	private room	Richmond District	94121	2 guests - 1 bedroom - 1 bed - 1 bath	Wifi - Kitchen	Dec 24 – 26	NaN	0	False	NaN	62.0	75.0	28.0

Base Model

OLS with all 209 features

- **Adj. R2**: 0.569
- **Test R2**: -1.97e19

Actual vs Predicted Prices

Residual Plot

QQ-plot
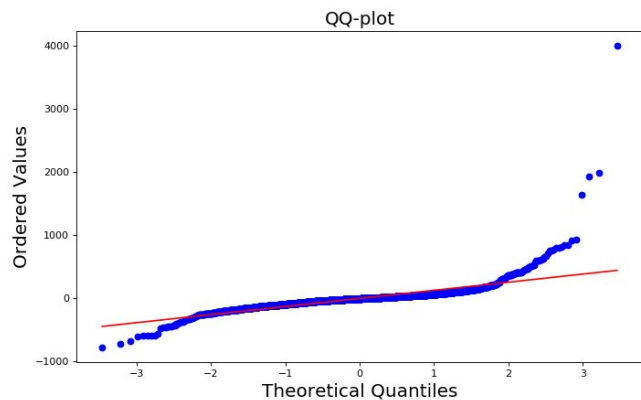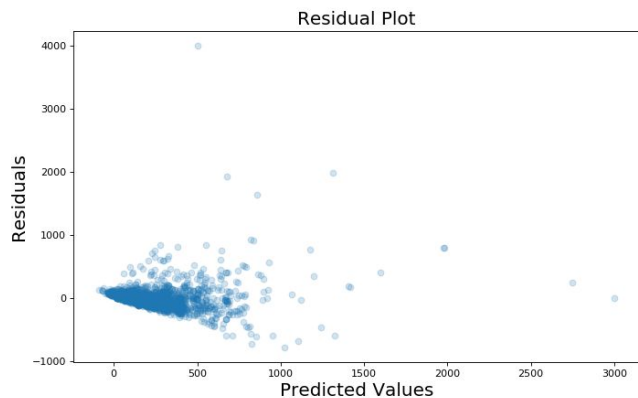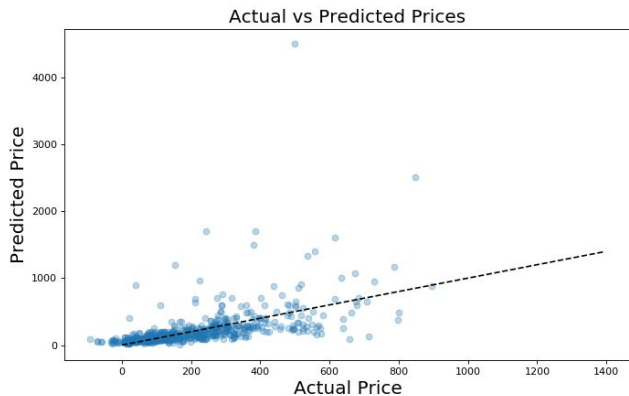
Feature Selection

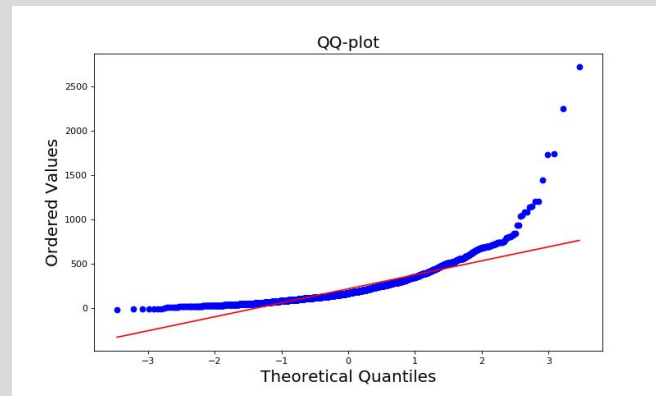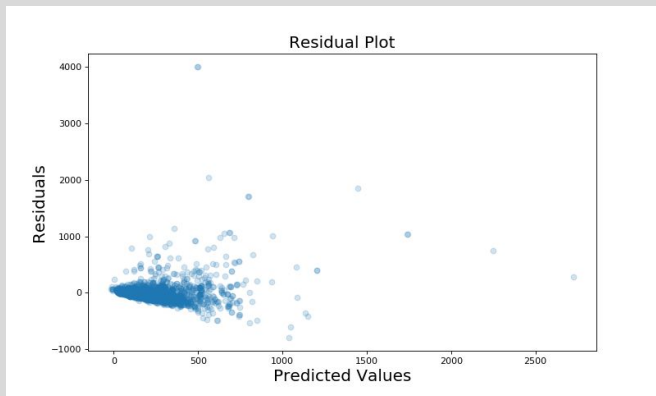OLS after removing any features with low significance or causing multicollinearity

- Down to 47 features
- **Adj. R2**: 0.446
- **Test R2**: 0.374
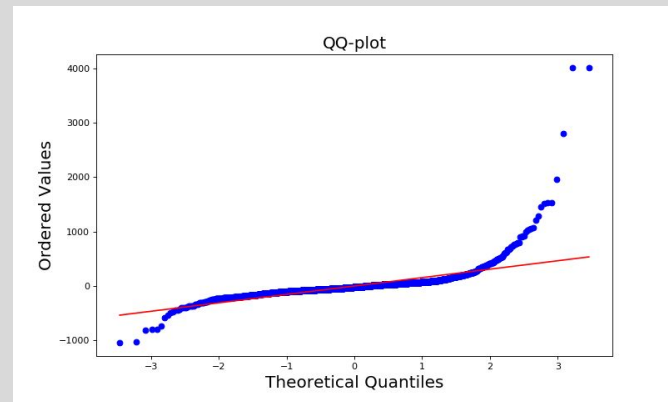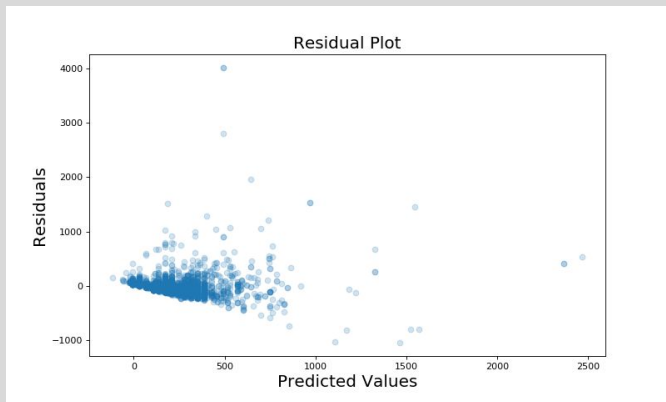
Regularization

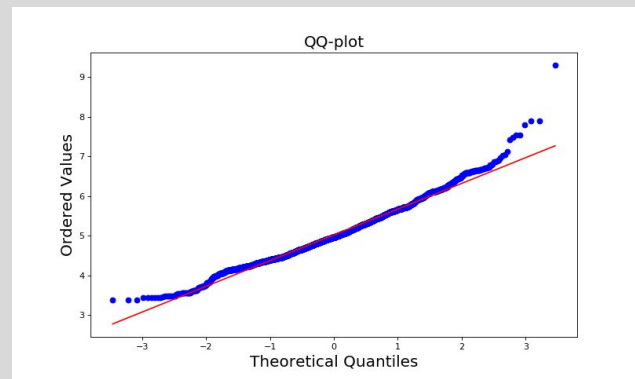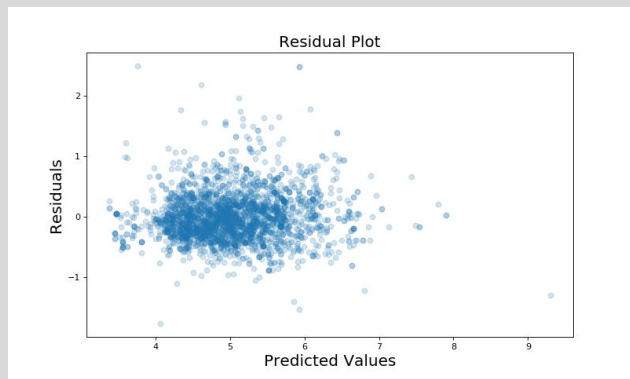Elastic Net with remaining 47 features

- **Adj. R2**: 0.487
- **Test R2**: 0.492
- **alpha**: 0.406
- **L1 Ratio**: 0.1

Transformation

Elastic Net with a log transformation on the target variable

- **Adj. R2**: 0.734
- **Test R2**: 0.716
- **alpha**: 0.038
- **L1 Ratio**: 0.1

Results

Variables with Increase in Price

- # of Guests: 17.17%
- # of Bedrooms: 12.16%
- # of Bathrooms: 11.34%
- Fisherman's Wharf: 8.16%
- Entire Residential Home: 7.63%
- Western Addition/NOPA: 7.43%
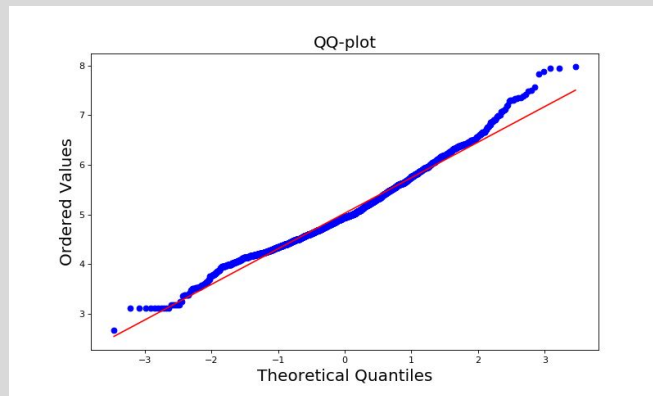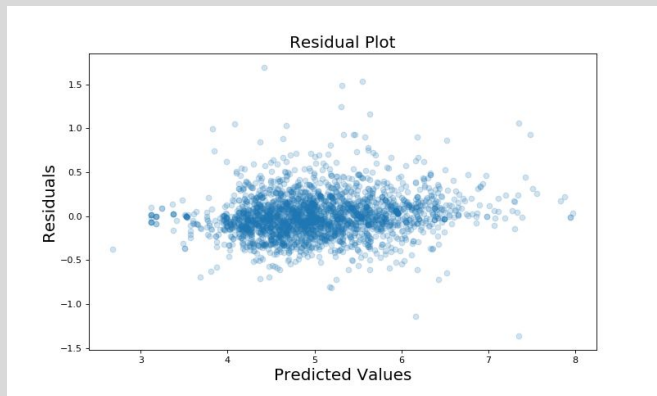- Marina: 7.00%
- Chinatown: 6.77%

Variables with Decrease in Price

- Private Room: -14.93%
- Shared Room: -13.43%
- Hostel Beds: -8.00%
- Shared Baths: -5.76%
- Oakland: -5.18%
- Daly City: -4.04%
- Elmhurst: -3.71%
- South San Francisco: -3.33%

XGBoost

XGBoost to see if we can further improve on our results

- **Adj. R2**: 0.940
- **Test R2**: 0.780

Actual vs Predicted Prices

Residual Plot

QQ-plot

Future Work

- Use listing location to create new features such as tourist attractions within the vicinity
- Observe more dates to see effect on price
- Gather data on hotels and create a separate model for comparison
- Hypertune our parameters for XGB by using cross validation