California Rental Value Estimation

•••

Data Scientist: Ivy Lai

Overview

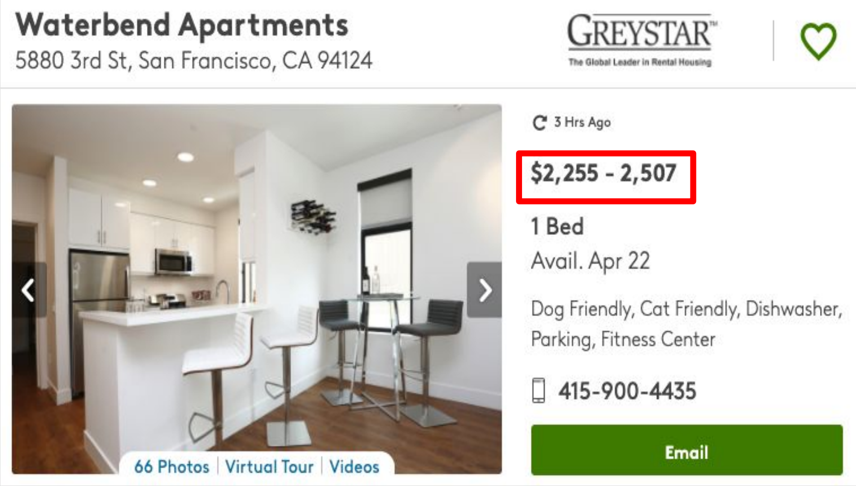
- **Goal:** Predict California rental value based on apartment features.
- **Data Collection:** Web-scraped over 150 cities, 4,000 properties and 15,000 apartment listings from apartments.com.

Features

- Rental Price *(Numeric)*
- Number of Beds *(Numeric)*
- Number of Baths *(Numeric)*
- SQ-FT *(Numeric)*
- City *(Categorical)*
- Walkscore *(Numeric)*
- Onsite Parking *(Binary)*
- WiFi *(Binary)*
- Washer/Dryer *(Binary)*
- Minimum Lease Length *(Numeric)*
- Number of Kitchen Features *(Numeric)*
- Air Conditioning *(Binary)*
- Elevator *(Binary)*
- Pool *(Binary)*
- Fitness Center *(Binary)*
- Allow Pets *(Binary)*
- Security System *(Binary)*
- Furnished *(Binary)*

Important Note

- Half of the rental prices were listed as a range instead of an exact number.
- Prices specified as a range will be converted to its average value.
    - E.g. $1,000 - $1,200 will become $1,100



Waterbend Apartments
5880 3rd St, San Francisco, CA 94124

$2,255 - 2,507

1 Bed
Avail. Apr 22

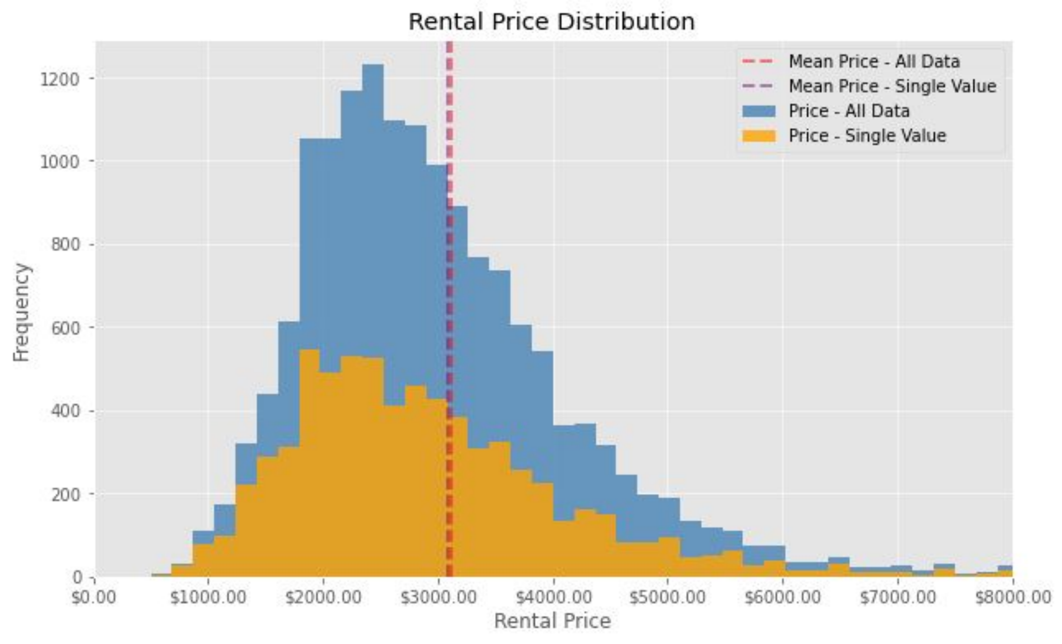
Dog Friendly, Cat Friendly, Dishwasher, Parking, Fitness Center

415-900-4435

GREYSTAR
The Global Leader in Rental Housing
3 Hrs Ago
66 Photos | Virtual Tour | Videos
Email

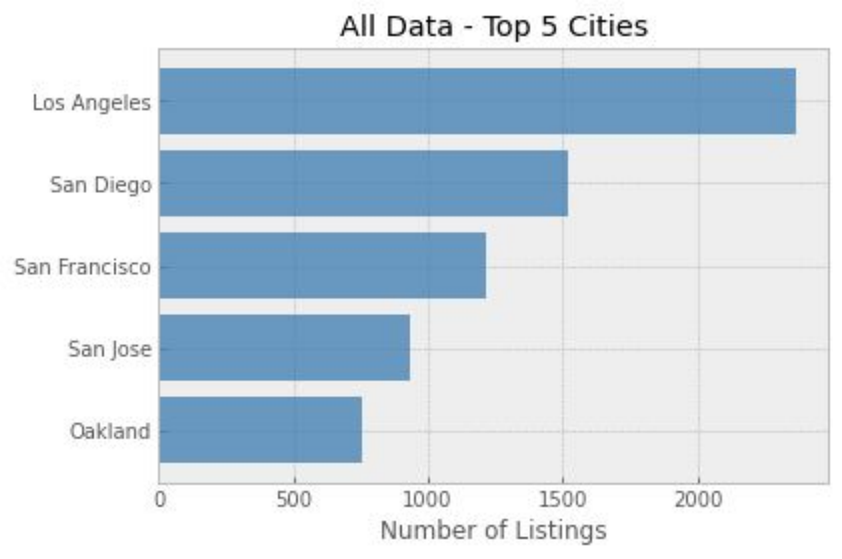
Rental Price Distribution

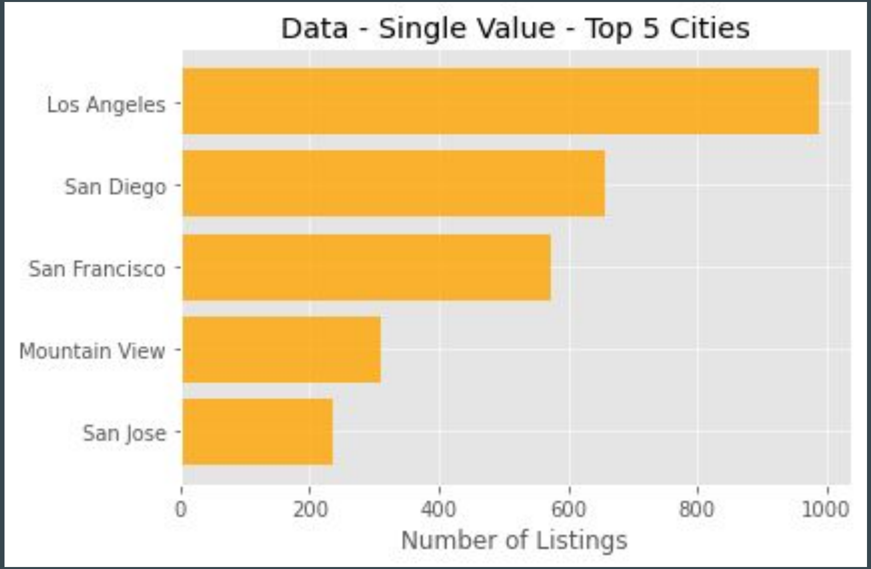
- The whole dataset (with the range converted) and the data with single values have similar distribution.
  - Right-skewed
  - Mean price = $3,100



Cities

- About 40% of the listings for both datasets came from 5 cities.
- Dummy variables created for the top three cities: LA, SD and SF.





Model Selection

- Evaluate different models using the Root Mean Square Error (RMSE).
- Baseline: predicting the mean price.

	Baseline	Linear Regression	Ridge	Decision Tree	Random Forest	Gradient Boosting
All Data (Train)	1872	1430	1433	1440	1096	1116
All Data (Test)	1786	1436	1436	1381	1012	1017
Data - Single Value (Test)	2104	1569	1570	1530	1215	1251

Conclusion & Next Steps

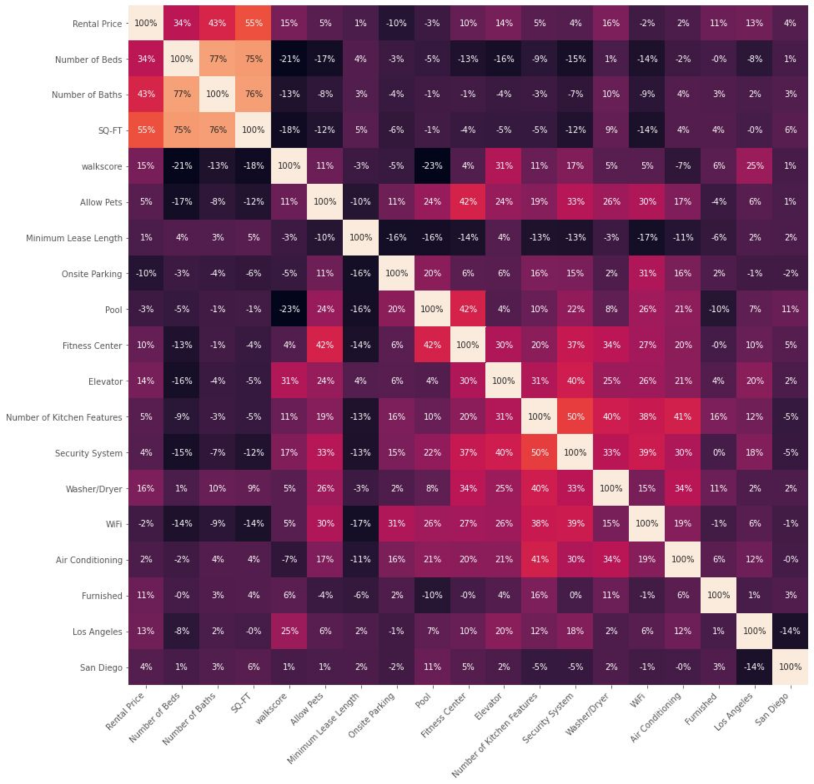
- Successfully created a prediction model that reduces RMSE by 42% compared to the baseline.
- Inconclusive evidence to support the approach of averaging the prices.
- To do - Collect more data for longer period of time.

Thank You!

Contact me:

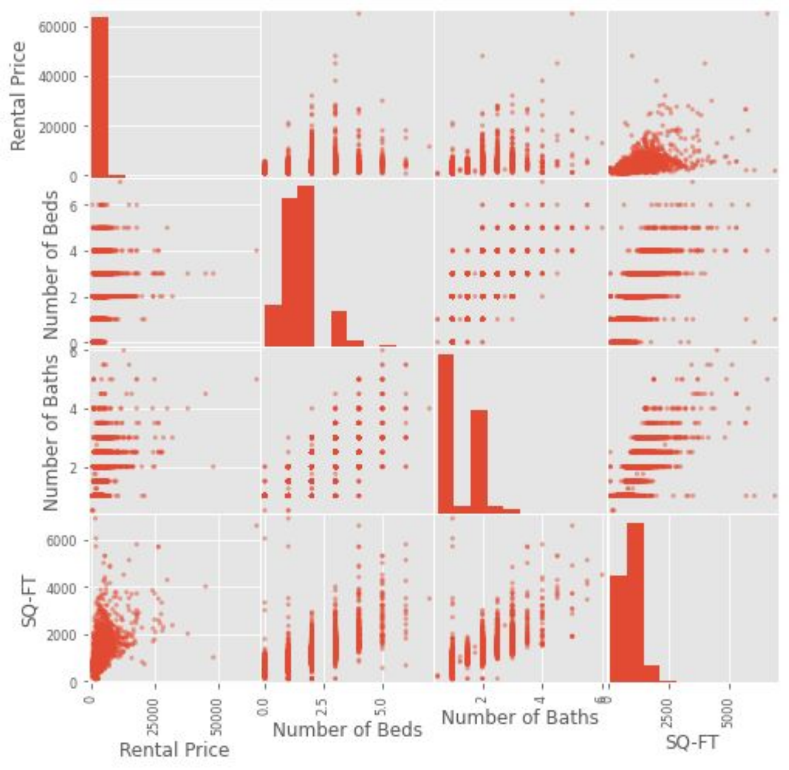
- Email
- Github

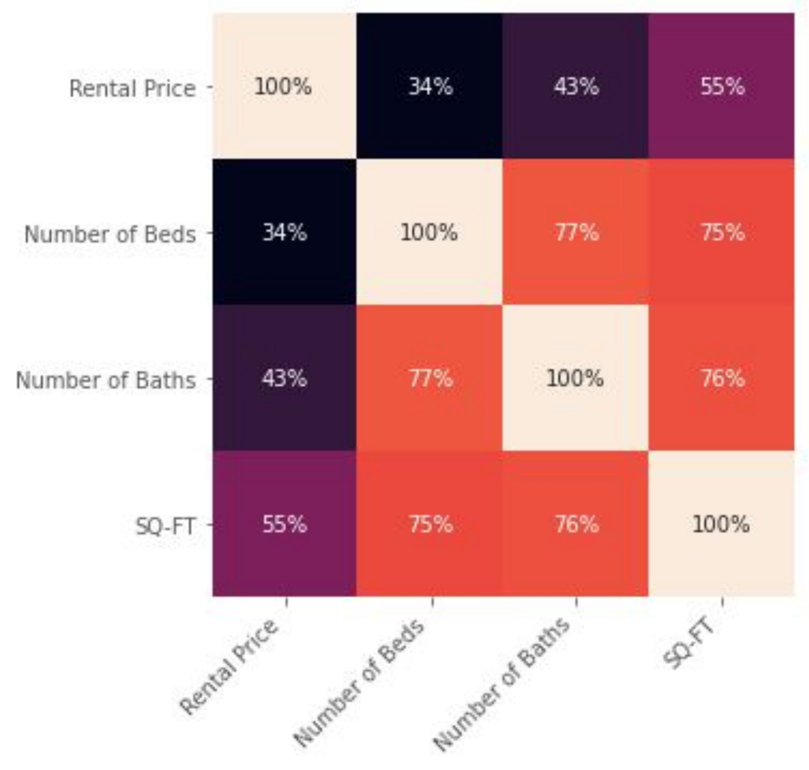
Appendix 1: Feature Correlation Matrix



	Rental Price	Number of Beds	Number of Baths	SQ-FT	walkscore	Allow Pets	Minimum Lease Length	Onsite Parking	Pool	Fitness Center	Elevator	Number of Kitchen Features	Security System	Washer/Dryer	WiFi	Air Conditioning	Furnished	Los Angeles	San Diego
Rental Price	100%	34%	43%	55%	15%	5%	1%	-10%	-3%	10%	14%	5%	4%	16%	-2%	2%	11%	13%	4%
Number of Beds	34%	100%	77%	75%	-21%	-17%	4%	-3%	-5%	-13%	-16%	-9%	-15%	1%	-14%	-2%	-0%	-8%	1%
Number of Baths	43%	77%	100%	76%	-13%	-8%	3%	-4%	-1%	-1%	-4%	-3%	-7%	10%	-9%	4%	3%	2%	3%
SQ-FT	55%	75%	76%	100%	-18%	-12%	5%	-6%	-1%	-4%	-5%	-5%	-12%	9%	-14%	4%	4%	-0%	6%
walkscore	15%	-21%	-13%	-18%	100%	11%	-3%	-5%	-23%	4%	31%	11%	17%	5%	5%	-7%	6%	25%	1%
Allow Pets	5%	-17%	-8%	-12%	11%	100%	-10%	11%	24%	42%	24%	19%	33%	26%	30%	17%	-4%	6%	1%
Minimum Lease Length	1%	4%	3%	5%	-3%	-10%	100%	-16%	-16%	-14%	4%	-13%	-13%	-3%	-17%	-11%	-6%	2%	2%
Onsite Parking	-10%	-3%	-4%	-6%	-5%	11%	-16%	100%	20%	6%	6%	16%	15%	2%	31%	16%	2%	-1%	-2%
Pool	-3%	-5%	-1%	-1%	-23%	24%	-16%	20%	100%	42%	4%	10%	22%	8%	26%	21%	-10%	7%	11%
Fitness Center	10%	-13%	-1%	-4%	4%	42%	-14%	6%	42%	100%	30%	20%	37%	34%	27%	20%	-0%	10%	5%
Elevator	14%	-16%	-4%	-5%	31%	24%	4%	6%	4%	30%	100%	31%	40%	25%	26%	21%	4%	20%	2%
Number of Kitchen Features	5%	-9%	-3%	-5%	11%	19%	-13%	16%	10%	20%	31%	100%	50%	40%	38%	41%	16%	12%	-5%
Security System	4%	-15%	-7%	-12%	17%	33%	-13%	15%	22%	37%	40%	50%	100%	33%	39%	30%	0%	18%	-5%
Washer/Dryer	16%	1%	10%	9%	5%	26%	-3%	2%	8%	34%	25%	40%	33%	100%	15%	34%	11%	2%	2%
WiFi	-2%	-14%	-9%	-14%	5%	30%	-17%	31%	26%	27%	26%	38%	39%	15%	100%	19%	-1%	6%	-1%
Air Conditioning	2%	-2%	4%	4%	-7%	17%	-11%	16%	21%	20%	21%	41%	30%	34%	19%	100%	6%	12%	-0%
Furnished	11%	-0%	3%	4%	6%	-4%	-6%	2%	-10%	-0%	4%	16%	0%	11%	-1%	6%	100%	1%	3%
Los Angeles	13%	-8%	2%	-0%	25%	6%	2%	-1%	7%	10%	20%	12%	18%	2%	6%	12%	1%	100%	-14%
San Diego	4%	1%	3%	6%	1%	1%	2%	-2%	11%	5%	2%	-5%	-5%	2%	-1%	-0%	3%	-14%	100%

Appendix 2: Price, Bed, Bath, SQ-FT

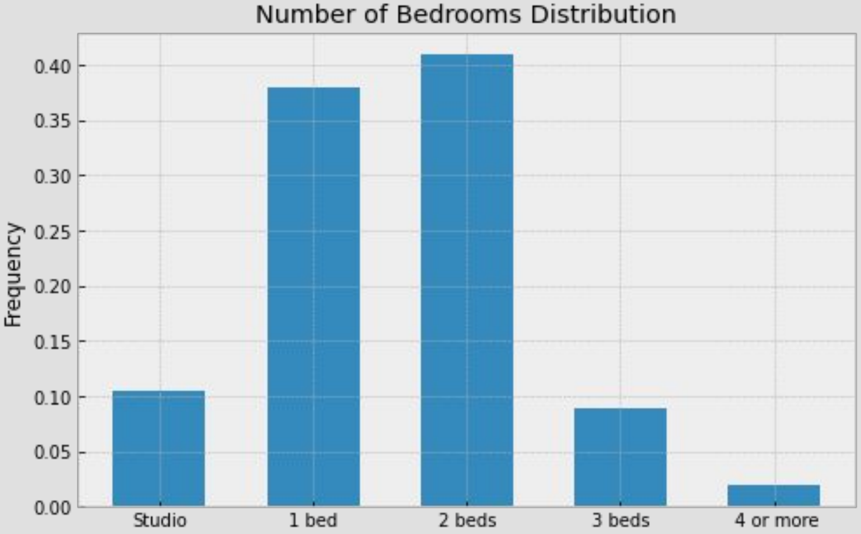


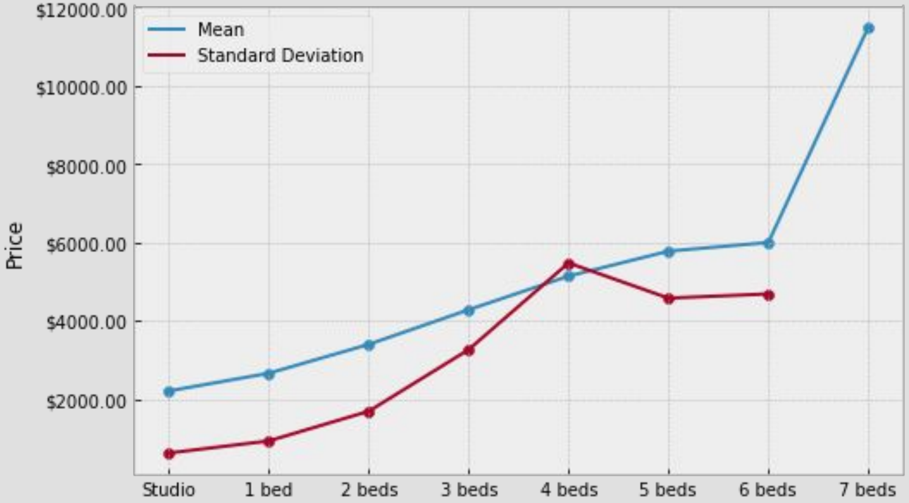


	Rental Price	Number of Beds	Number of Baths	SQ-FT
Rental Price	100%	34%	43%	55%
Number of Beds	34%	100%	77%	75%
Number of Baths	43%	77%	100%	76%
SQ-FT	55%	75%	76%	100%

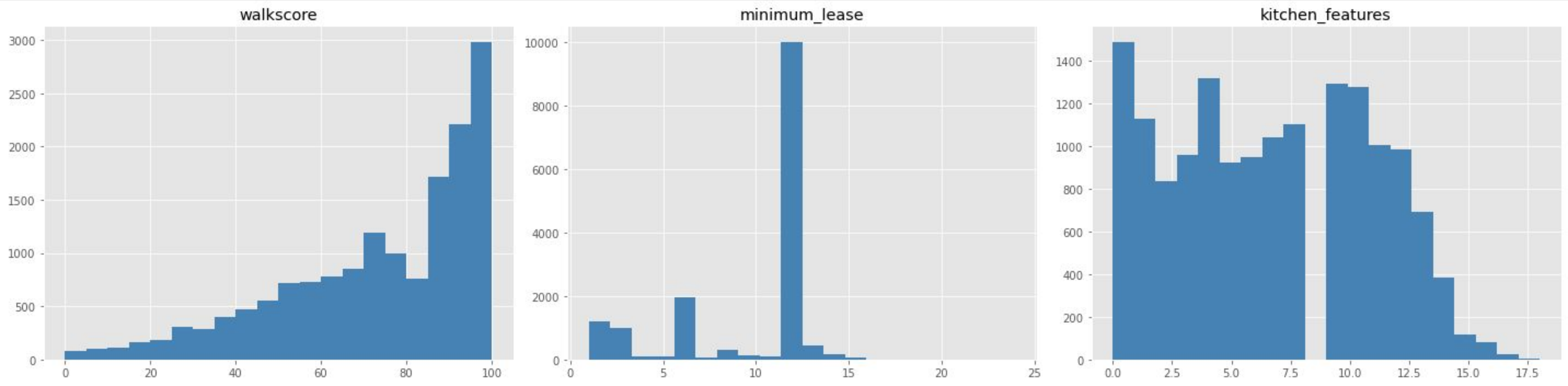
Appendix 3: Number of Bedrooms

- 98% of the listings have 3 or less bedrooms.
- Standard deviation of rental price is much higher with 4 or more bedrooms

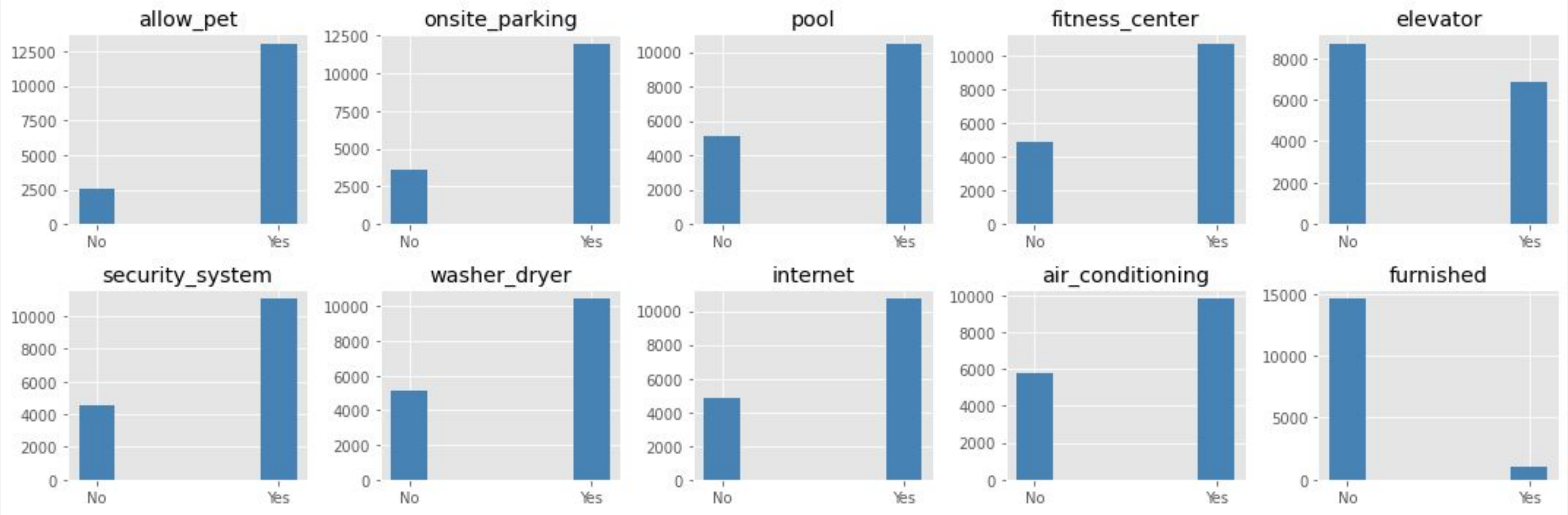




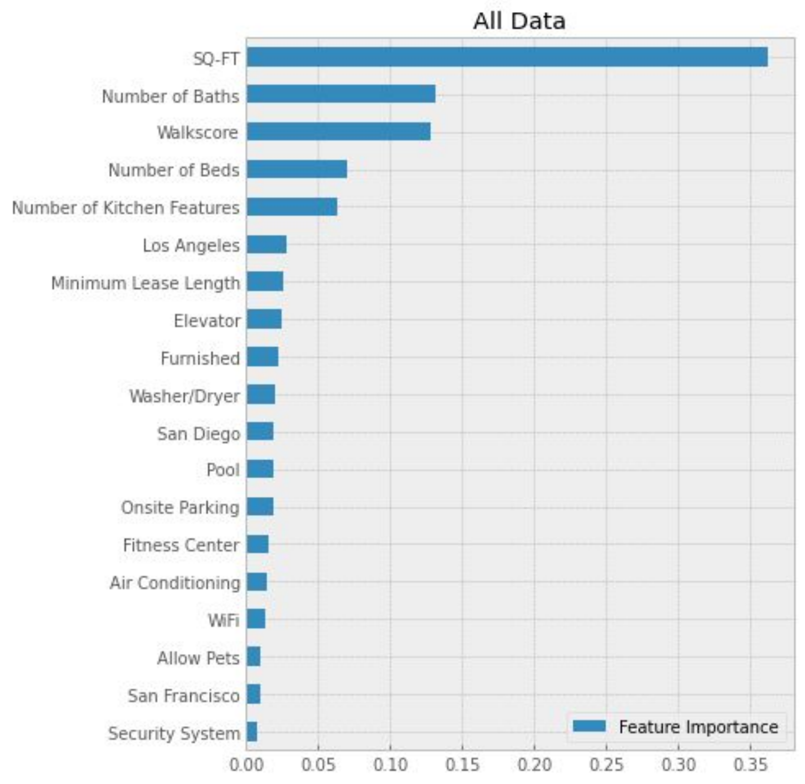
Appendix 4: Numeric Features Distributions

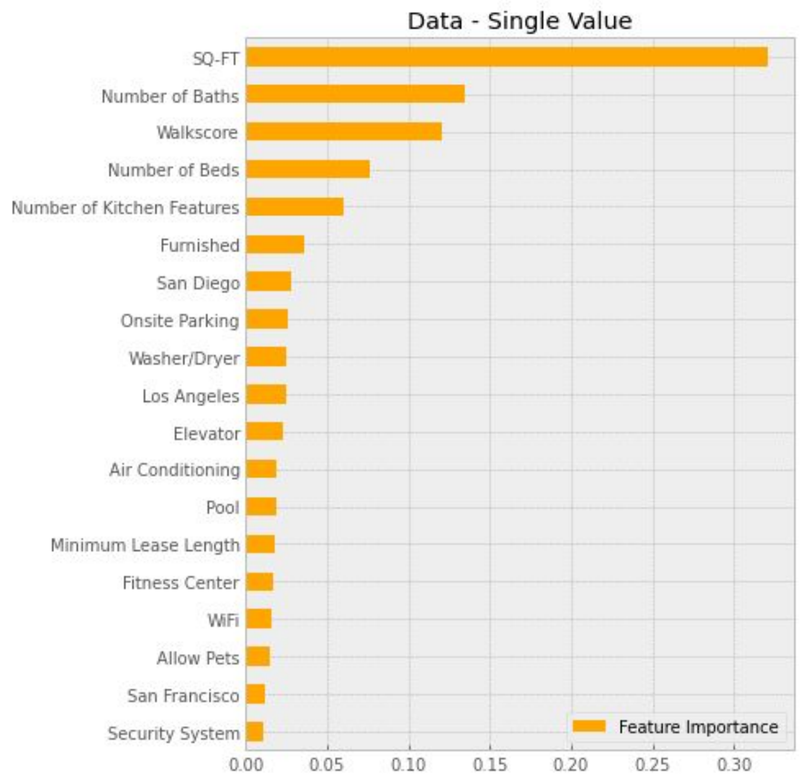


Appendix 5: Binary Features Distributions

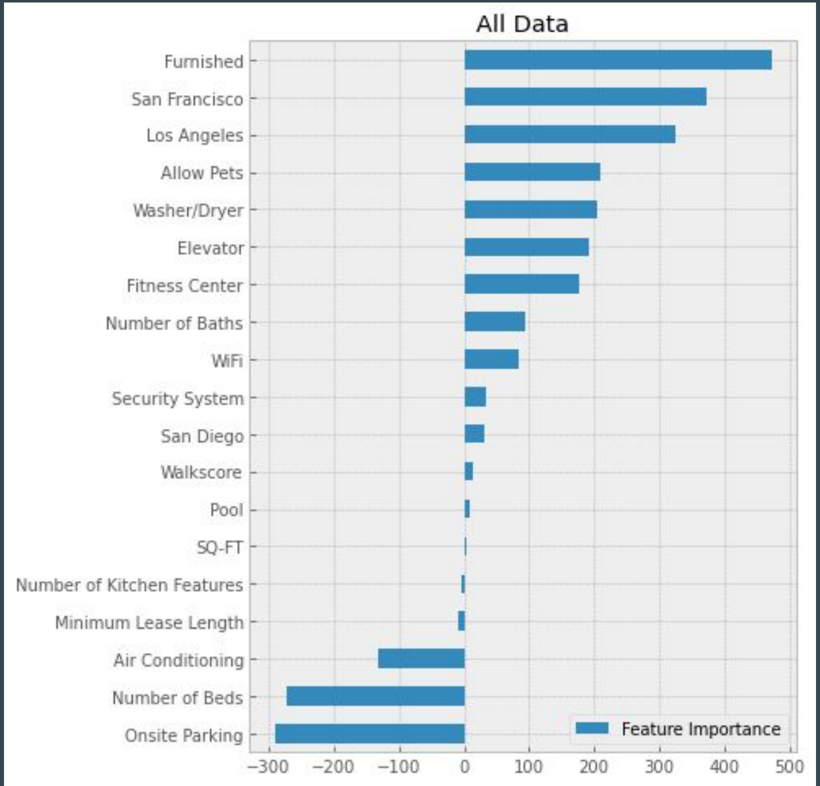


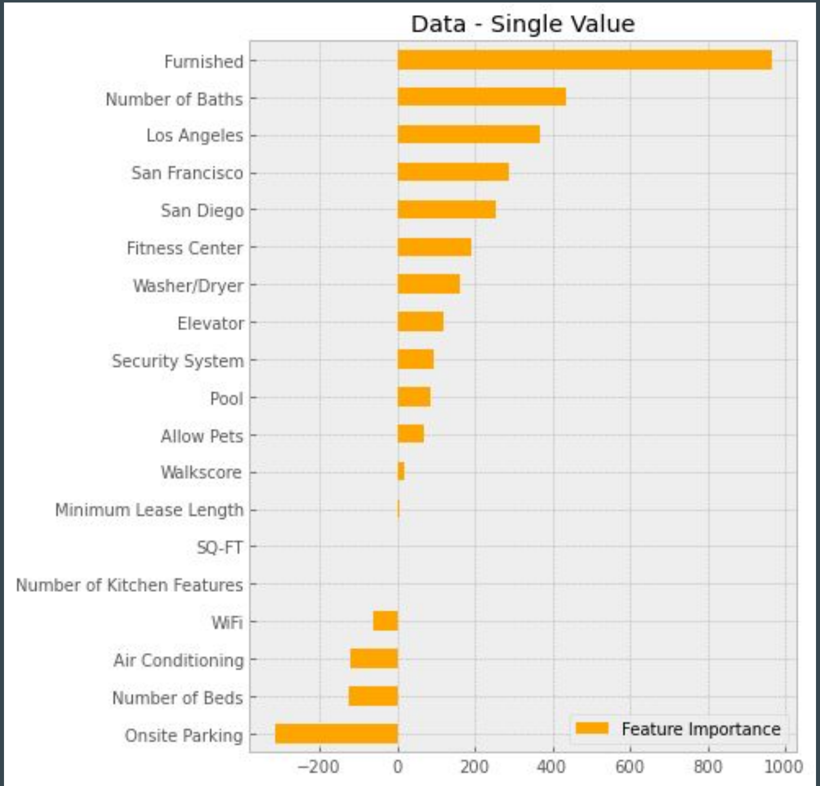
Appendix 6: Feature Importance - Random Forest



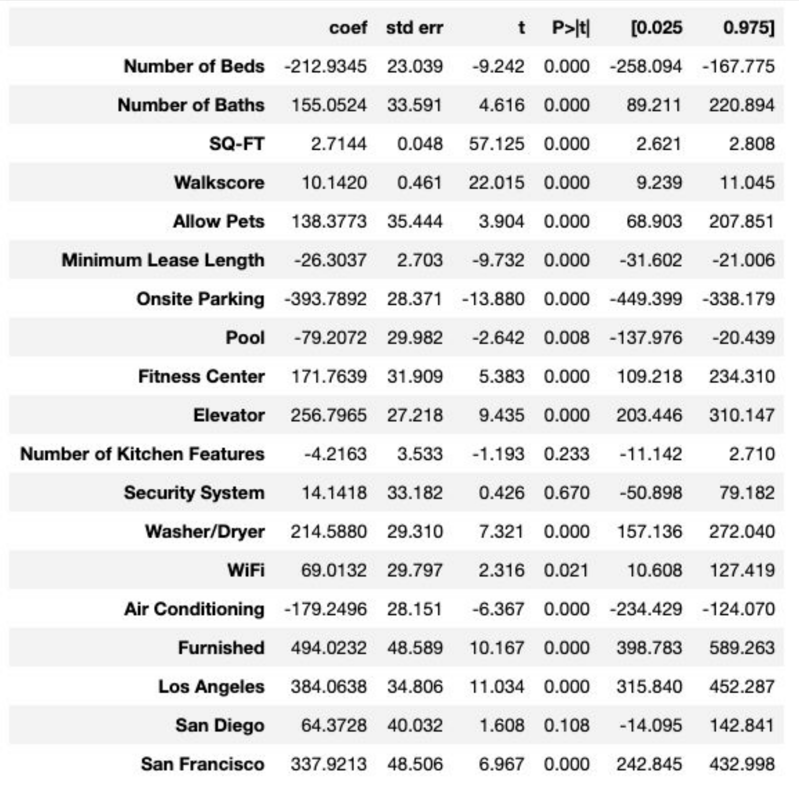


Appendix 7: Feature Importance - Linear Regression



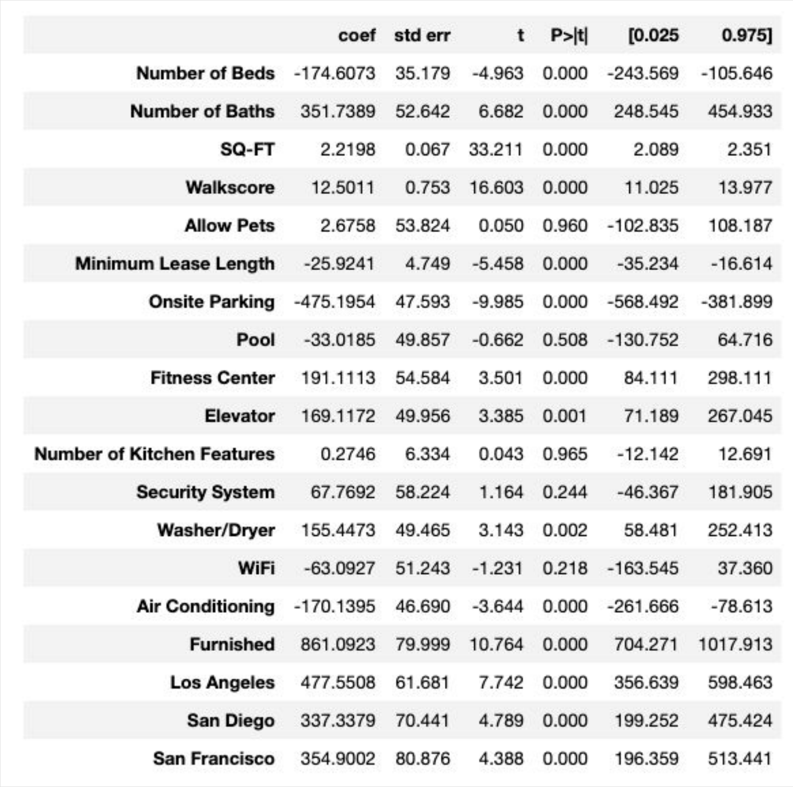


Appendix 8: Linear Regression Summary - All Data



	coef	std err	t	P>|t|	[0.025	0.975]
Number of Beds	-212.9345	23.039	-9.242	0.000	-258.094	-167.775
Number of Baths	155.0524	33.591	4.616	0.000	89.211	220.894
SQ-FT	2.7144	0.048	57.125	0.000	2.621	2.808
Walkscore	10.1420	0.461	22.015	0.000	9.239	11.045
Allow Pets	138.3773	35.444	3.904	0.000	68.903	207.851
Minimum Lease Length	-26.3037	2.703	-9.732	0.000	-31.602	-21.006
Onsite Parking	-393.7892	28.371	-13.880	0.000	-449.399	-338.179
Pool	-79.2072	29.982	-2.642	0.008	-137.976	-20.439
Fitness Center	171.7639	31.909	5.383	0.000	109.218	234.310
Elevator	256.7965	27.218	9.435	0.000	203.446	310.147
Number of Kitchen Features	-4.2163	3.533	-1.193	0.233	-11.142	2.710
Security System	14.1418	33.182	0.426	0.670	-50.898	79.182
Washer/Dryer	214.5880	29.310	7.321	0.000	157.136	272.040
WiFi	69.0132	29.797	2.316	0.021	10.608	127.419
Air Conditioning	-179.2496	28.151	-6.367	0.000	-234.429	-124.070
Furnished	494.0232	48.589	10.167	0.000	398.783	589.263
Los Angeles	384.0638	34.806	11.034	0.000	315.840	452.287
San Diego	64.3728	40.032	1.608	0.108	-14.095	142.841
San Francisco	337.9213	48.506	6.967	0.000	242.845	432.998

Appendix 9: Linear Regression Summary - Data No Range



	coef	std err	t	P>|t|	[0.025	0.975]
Number of Beds	-174.6073	35.179	-4.963	0.000	-243.569	-105.646
Number of Baths	351.7389	52.642	6.682	0.000	248.545	454.933
SQ-FT	2.2198	0.067	33.211	0.000	2.089	2.351
Walkscore	12.5011	0.753	16.603	0.000	11.025	13.977
Allow Pets	2.6758	53.824	0.050	0.960	-102.835	108.187
Minimum Lease Length	-25.9241	4.749	-5.458	0.000	-35.234	-16.614
Onsite Parking	-475.1954	47.593	-9.985	0.000	-568.492	-381.899
Pool	-33.0185	49.857	-0.662	0.508	-130.752	64.716
Fitness Center	191.1113	54.584	3.501	0.000	84.111	298.111
Elevator	169.1172	49.956	3.385	0.001	71.189	267.045
Number of Kitchen Features	0.2746	6.334	0.043	0.965	-12.142	12.691
Security System	67.7692	58.224	1.164	0.244	-46.367	181.905
Washer/Dryer	155.4473	49.465	3.143	0.002	58.481	252.413
WiFi	-63.0927	51.243	-1.231	0.218	-163.545	37.360
Air Conditioning	-170.1395	46.690	-3.644	0.000	-261.666	-78.613
Furnished	861.0923	79.999	10.764	0.000	704.271	1017.913
Los Angeles	477.5508	61.681	7.742	0.000	356.639	598.463
San Diego	337.3379	70.441	4.789	0.000	199.252	475.424
San Francisco	354.9002	80.876	4.388	0.000	196.359	513.441