Predicting AirBnB rental price

SpringBoard Data Science Intensive Capstone Project, 2021

Hien Quang

The Problem

- Peer - to - peer system
- Set your own price
- Risk of undervalued or overvalued your properties

Data

Source:
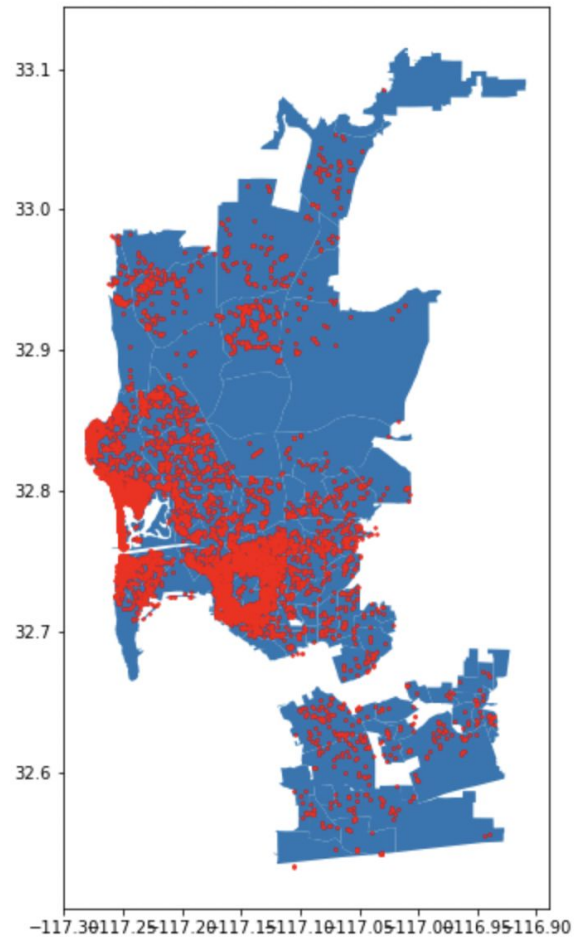
- http://insideairbnb.com/get-the-data.html)
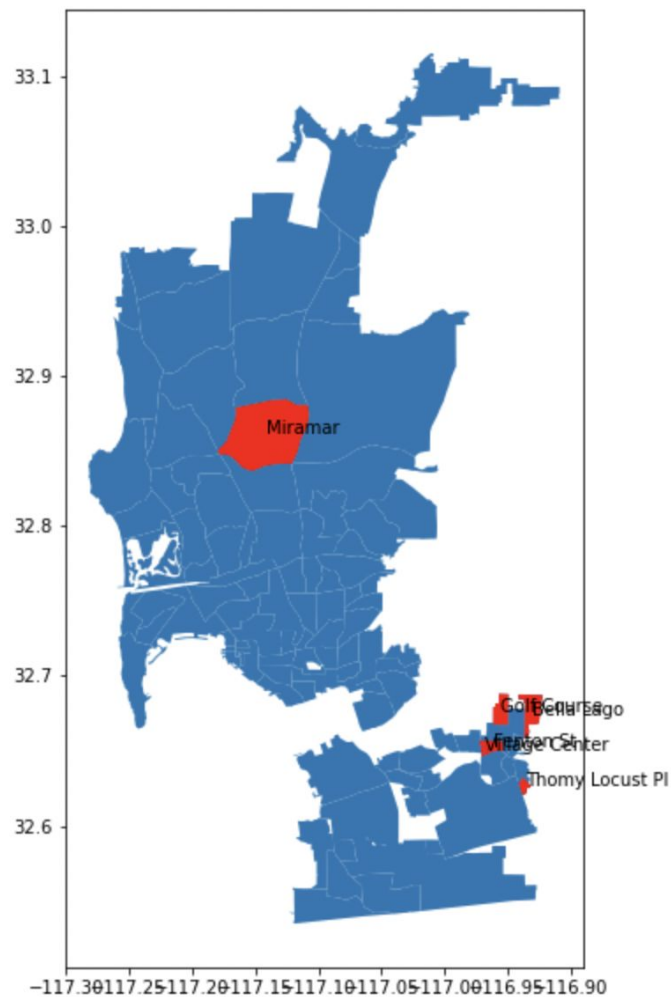- listings.csv : detailed listings data

Cleaning:

- Minimal cleaning done as needed
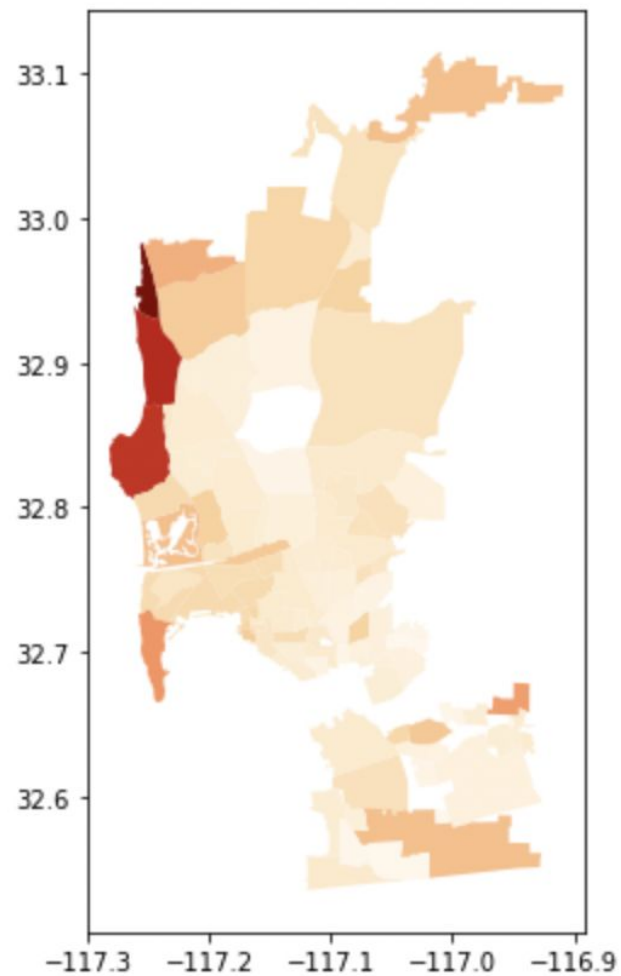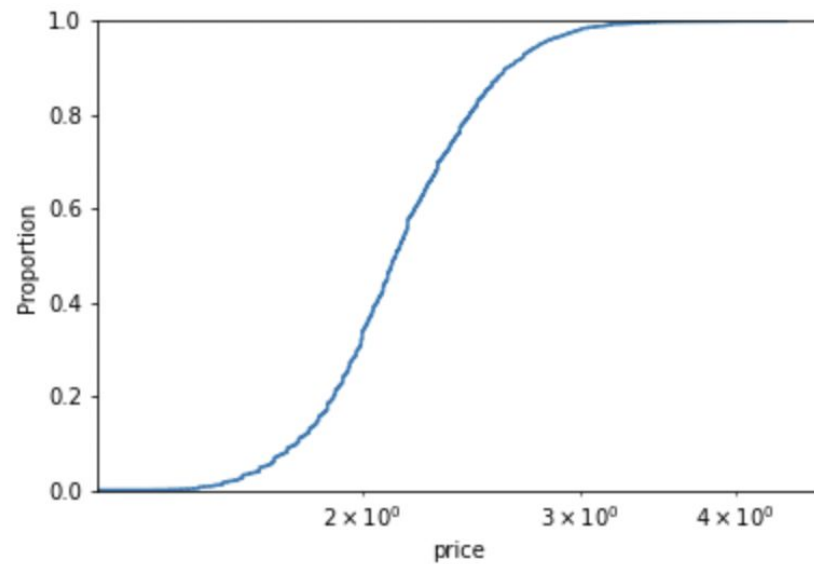- Dropped rows and columns containing no data

Listings density
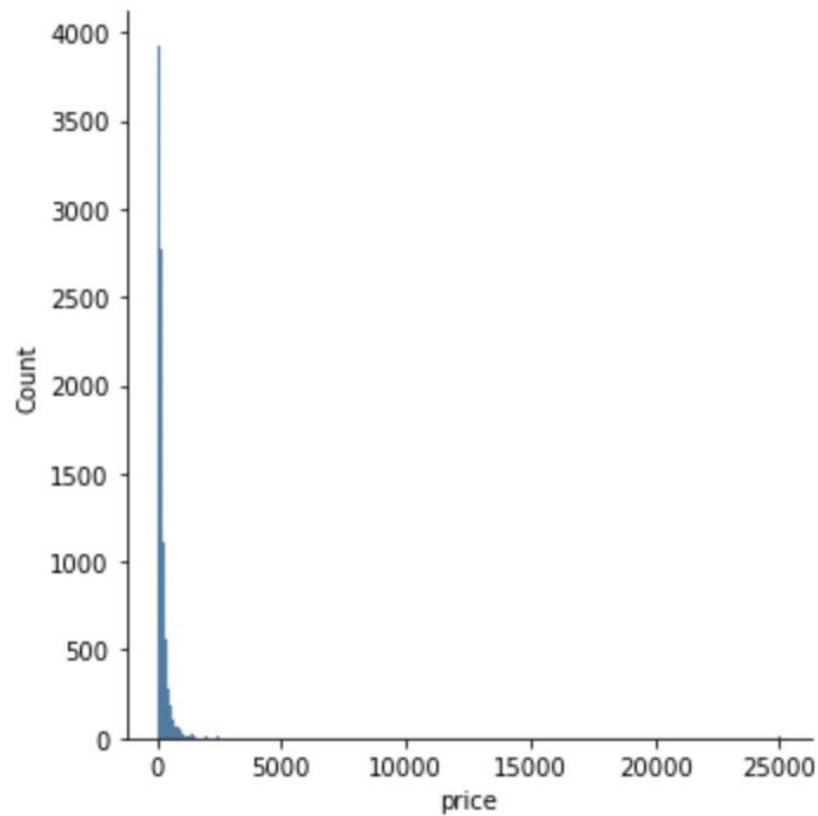
No listings counties

Average price by county

Rental pricing is highly skewed

Modeling

Baseline models

Using Pycaret - compare_models()

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
lightgbm	Light Gradient Boosting Machine	8.898580e+01	8.843717e+04	2.774652e+02	3.772000e-01	0.5194	4.671000e-01	0.313
rf	Random Forest Regressor	7.827350e+01	8.712803e+04	2.770872e+02	3.397000e-01	0.4354	3.716000e-01	19.591
br	Bayesian Ridge	1.016227e+02	9.707396e+04	2.893947e+02	3.317000e-01	0.7094	6.132000e-01	2.096
lasso	Lasso Regression	1.003021e+02	9.755420e+04	2.896762e+02	3.314000e-01	0.6503	5.829000e-01	0.121
ridge	Ridge Regression	1.132242e+02	9.936143e+04	2.949442e+02	2.980000e-01	0.8317	7.363000e-01	0.126
omp	Orthogonal Matching Pursuit	1.112825e+02	1.001828e+05	2.961379e+02	2.923000e-01	0.8166	7.017000e-01	0.131
en	Elastic Net	1.009071e+02	1.017875e+05	2.974381e+02	2.915000e-01	0.6211	6.069000e-01	0.082
gbr	Gradient Boosting Regressor	8.872650e+01	9.868847e+04	2.951061e+02	2.840000e-01	0.5111	4.819000e-01	5.442
et	Extra Trees Regressor	7.600620e+01	9.027132e+04	2.770742e+02	2.594000e-01	0.4218	3.282000e-01	27.431
huber	Huber Regressor	9.161110e+01	1.080697e+05	3.075620e+02	2.389000e-01	0.5355	3.997000e-01	4.111

Tuned Models

- Light Gradient Boosting Machine
- Optimize for MAE
- 10 - fold

	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	58.8910	11201.8570	105.8388	0.6325	0.3948	0.3374
1	58.1656	11524.3602	107.3516	0.6043	0.3969	0.2995
2	54.7202	8371.6596	91.4968	0.6962	0.3774	0.3253
3	53.1397	7868.7942	88.7062	0.7066	0.3794	0.3220
4	54.6592	7874.4361	88.7380	0.7212	0.3776	0.3215
5	55.3407	9830.4166	99.1485	0.6236	0.3934	0.3323
6	53.4761	8747.9405	93.5304	0.5713	0.3887	0.3176
7	51.2321	9609.2713	98.0269	0.5647	0.3844	0.2974
8	54.9721	8178.0841	90.4328	0.7119	0.4056	0.3375
9	55.9118	8386.0005	91.5751	0.7422	0.4068	0.3456
Mean	55.0509	9159.2820	95.4845	0.6575	0.3905	0.3236
SD	2.1508	1268.2694	6.4801	0.0622	0.0103	0.0150

Feature Importance

- Number of bedrooms
  - Positively correlated
- Number of bathrooms
  - Positively correlated
- Reviews per month
  - Negatively correlated
- Longitude
  - Higher price if closer to the beach
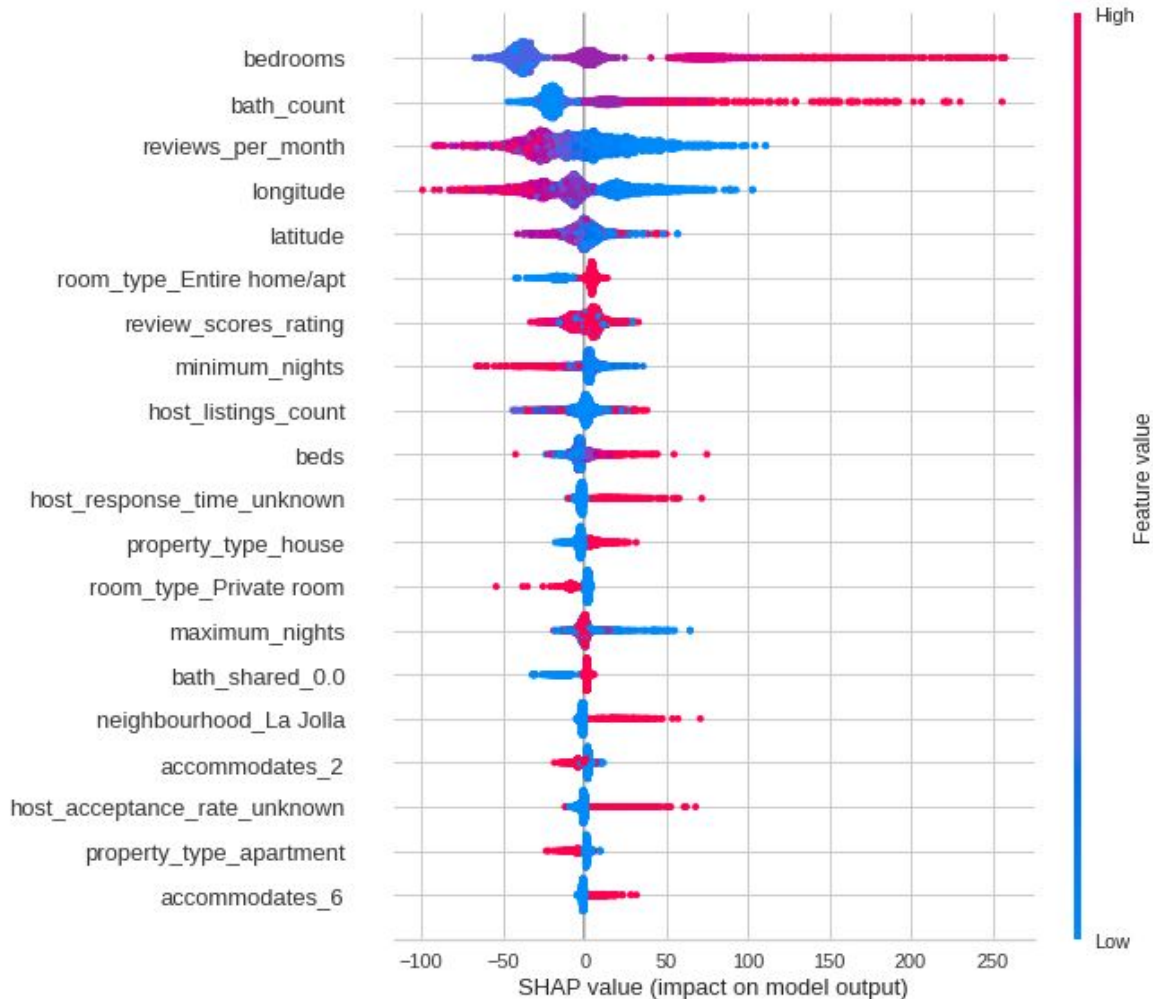
Assumptions, Limitations and Disclaimers

- Price are set fairly and at market level
- All listings were booked equally
- Rental price stays somewhat constant

Future Improvements

- Listings clustering with K-means
- Sentiment Analysis with NLP
- Deep dive into amenities list

Thank you!

Hien Quang
Email: quangthuchien@gmail.com
https://www.linkedin.com/in/hienquang/
https://github.com/Hienquang/
Project report: https://github.com/Hienquang/AirbnbCapstone/blob/main/README.md