

Income Optimizer for Airbnb Hosts

Annie Wang | Feb 2022

Goals

Goals

- Help Airbnb hosts maximize the income by optimizing the listing price and vacancy
- Provide a tool to calculate the income in different scenarios

Business questions

- What would be a reasonable price for the unit?
- How would price affect demand, and then affect income?
- What leverages can be pulled to improve price and demand?



Data

Airbnb listings for San Francisco, San Mateo, and Santa Clara counties, collected between late October to early December 2021, from [Inside Airbnb](#).

- Listing price
- Availability in the next 30, 60, 90 days
- Location and size of the unit
- Amenities
- Review scores and review counts
- Availability of more information about the unit and the host

Method

Building models

Two-stage modeling

- Predict price
- Predict demand

Models

- Baseline
- Random forest
- XBGooest

Model selection

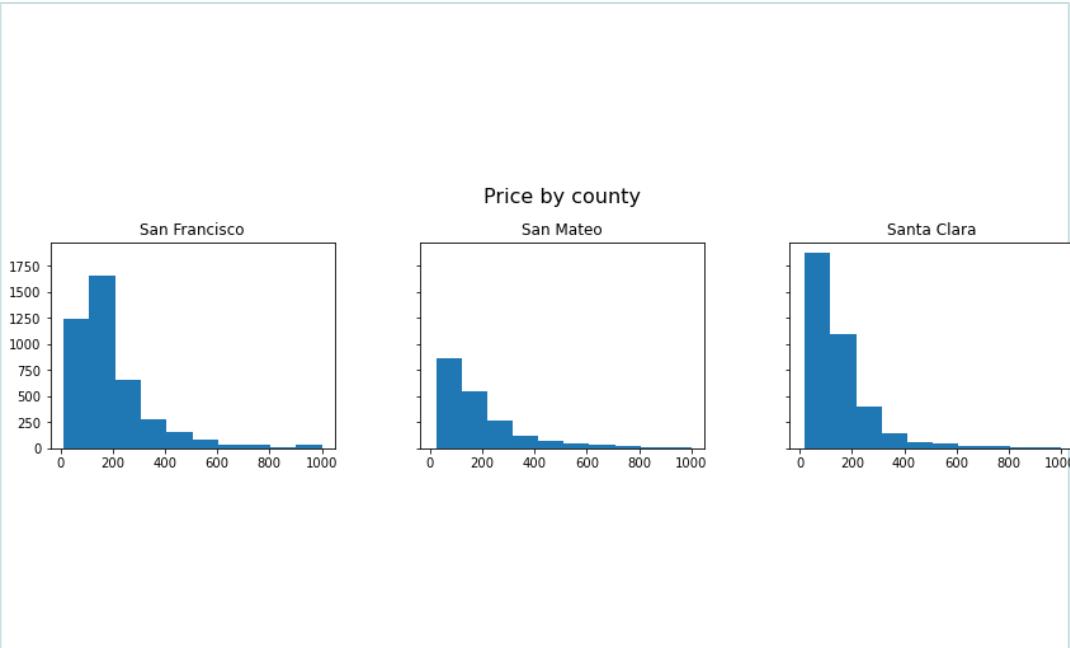
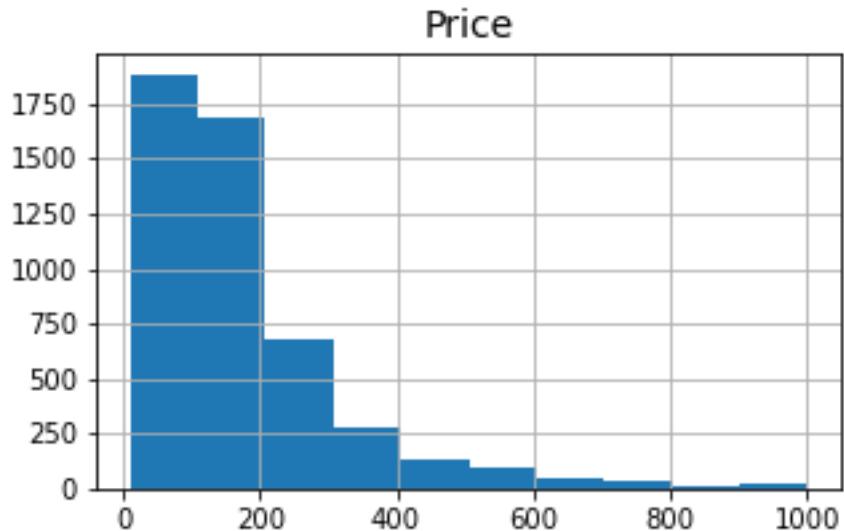
- Metrics: R2 (primary), MSE
- Grid Search CV

Improving models

Reducing overfitting

- Drop small categories in categorical features
 - Neighborhood, property type
- Tune hyperparameters
 - Depth, leaf size, etc.
 - Max number of features
 - Learning rate
 - Lambda for L2 regularization
- Drop more features based on feature importance
 - The top 50 out of the 100

Listing price



The overall average price is \$178. Most listings are below \$300.

The average price is highest in San Francisco and lowest in Santa Clara.

Correlation with price

Price is most correlated with **size** related features

It is correlated with some **amenities**: indoor fireplace, private entrance, BBQ grill and patio or balcony, etc.

Review scores on overall rating, cleanliness and location are weakly correlated with price.

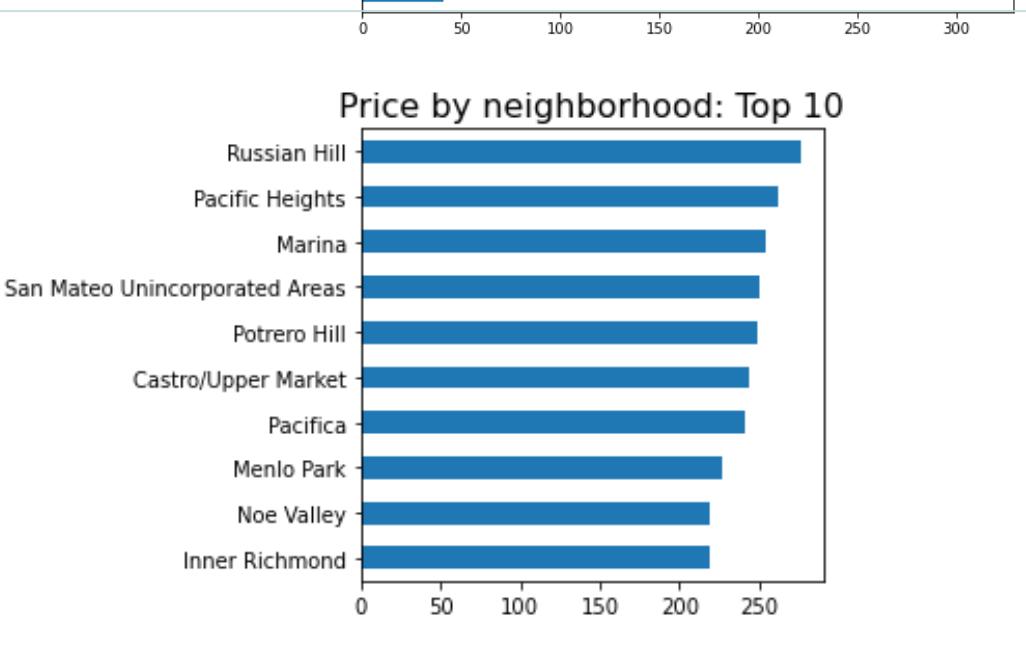


Listing price

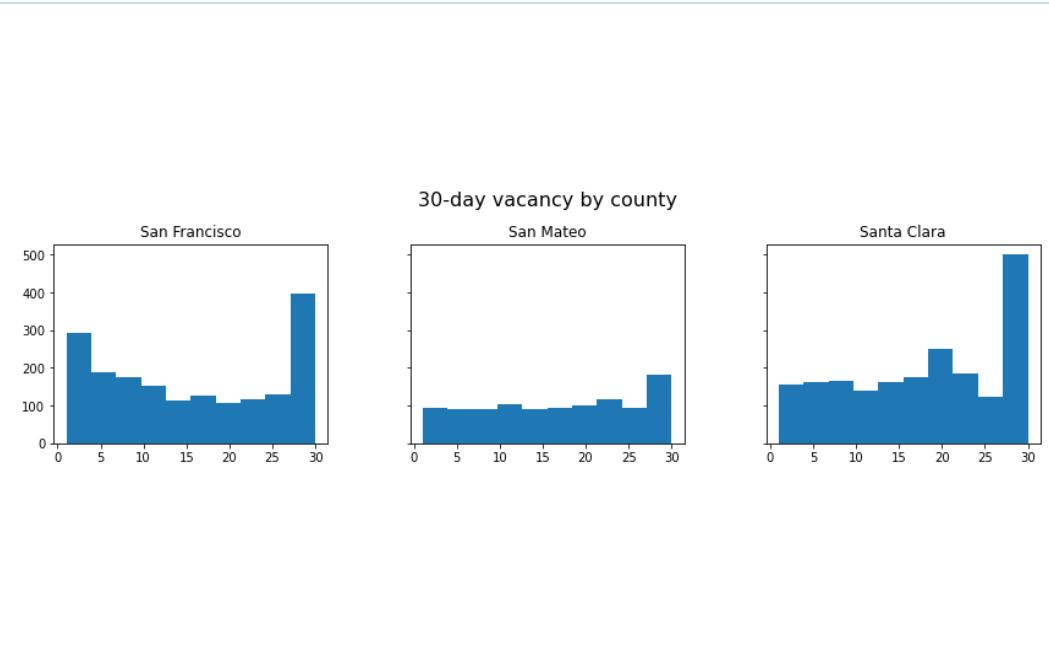
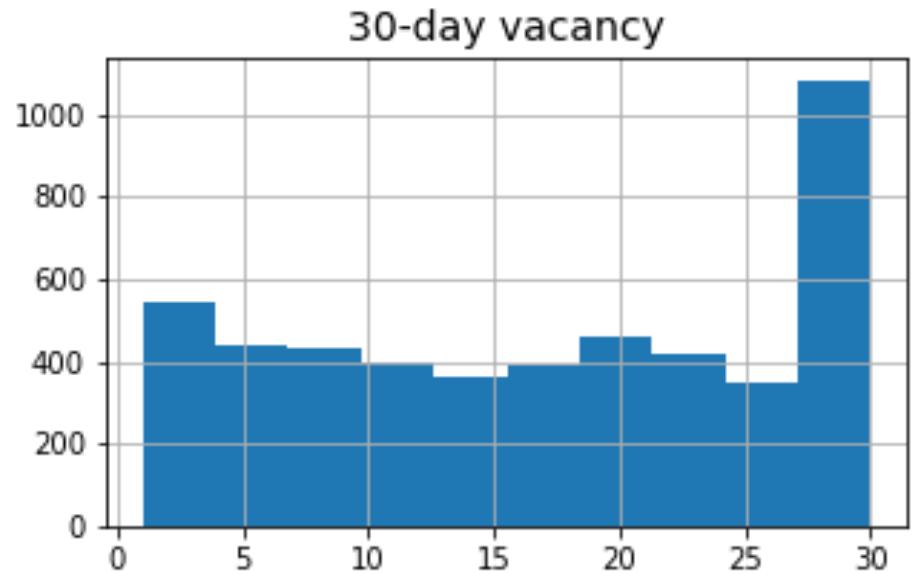
An entire unit is more expensive than private rooms in a unit.

- Hotel rooms ranked between these two.

The most expensive neighborhoods are Russian Hill, Pacific Heights and Marina, all in San Francisco.



30-Day vacancy



The average 30-day
vacancy is 17 days.

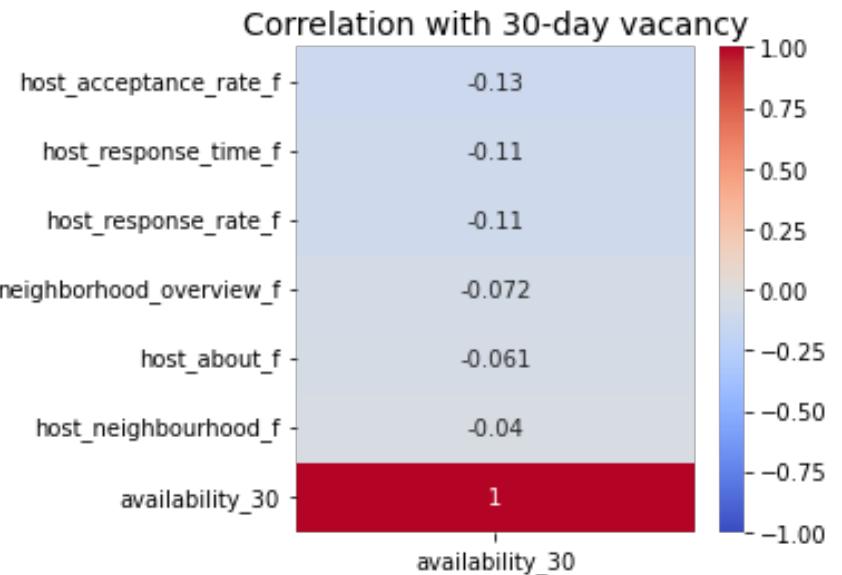
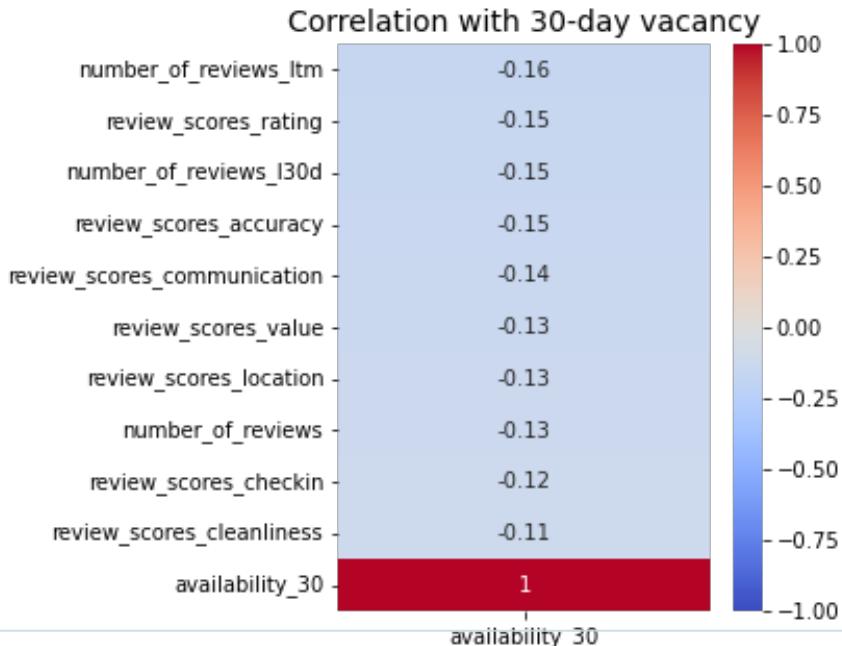
Santa Clara's 30-day
vacancy is the highest
(18.1) and that of San
Francisco (15.6) is the
lowest.

Correlations with 30-Day vacancy

30-day vacancy is negatively related to all review scores and counts.

- Correlation with overall rating is the highest

Listings that provide host acceptance rate and response time have lower 30-day vacancy.



Model summary

At both stages, the XGBoost model with grid search CV and the trimmed feature list were selected as the final model.

Predict price

Model	R2 train	R2 test	MSE train	MSE test
Baseline			21698	23040
Random Forest	0.9497	0.6574	1090	7894
Random Forest + Grid Search + all features	0.8152	0.6465	4010	8145
Random Forest + Grid Search + top 50 features	0.8126	0.646	4066	8156
XGBoost	0.9448	0.6569	1197	7905
XGBoost + Grid Search + all features	0.8449	0.6907	3364	7125
XGBoost + Grid Search + top 50 features	0.8001	0.6849	4336	7259

Predict 30-day vacancy

Model	R2 train	R2 test	MSE train	MSE test
Baseline			94.7495	94.14
Random Forest	0.8934	0.2142	10.1046	73.98
Random Forest + Grid Search + all features	0.5866	0.2212	39.1741	73.31
Random Forest + Grid Search + top 20 features	0.5954	0.1943	38.336	75.85
XGBoost	0.8724	0.1896	12.0858	76.29
XGBoost + Grid Search + all features	0.5992	0.2286	37.9753	72.61
XGBoost + Grid Search + top 50 features	0.4759	0.2184	49.6594	73.58

Feature Importance

Price

- * Size
- * Property type
- * WIFI, indoors fireplace, BBQ grill

Vacancy

- * Number of reviews
- * Host acceptance rate
- * Predicted price
- * Price
- * Scores on accuracy, value, location
- * Backyard, outdoor furniture, private entrance

- * Location
- * BA type
- * Min nights
- * Patio & balcony, pool

Predict your 30-day income

Please provide some information on your Airbnb listing

Property type

Private room in residential home

county	Number of beds	minimum_nights
San Francisco	1	1
Neighborhood	Number of bathrooms	maximum_nights
Noe Valley	0	1
Total number of people	Bathroom type	Total number of reviews
1	Private	1
Number of bedrooms	Number of reviews last 12 months	Number of reviews last 30 days
1	1	1

Please check the amenities available in your listing unit.

- Wifi BBQ_grill Patio/balcony
- Dedicated workspace Backyard Outdoor furniture
- Indoor fireplace Pool Private entrance

What are your review scores?

Overall rating	Cleanliness	Communication	Value
1.00	1.00	1.00	1.00
Accuracy	Checkin	Location	
1.00	1.00	1.00	

Is the following information provided on your Airbnb page?

- Neighborhood overview Host response time Host neighbourhood
- About the host Host response rate Host acceptance rate

Your expected listing price

How much do you want to charge per night?



Income Prediction

Predicted listing price is **186.23** dollars.

Predicted 30-day vacancy is **.13** days, and predicted booking days are **.17** days.

At your desired listing price (\$ **0**), your predicted total income for the next 30 days is **0** dollars.

A tool in
Streamlit

Conclusion and next steps

Conclusion

- To maximize income, hosts should optimize price and demand at the same time and understand their interdependence and their shared and unique drivers.
- This income optimizer helps hosts maximize their income by trying different scenarios.

Next steps

- Further reduce overfitting
 - Create higher-level categorical features (neighborhood)
 - Feature engineering: PCA
- Improve 30-day vacancy prediction
 - Deeper understanding of availability columns: voluntary vacancy
 - Use more data, more features, feature engineering