



A Tool for Setting the Price of Your Airbnb

Why?

- ❖ For a host, it can be difficult to know if you have set your Airbnb at the best price.
- ❖ Setting the price just **\$20 too low** can cost thousands each year.

The Data



- 20K listings in Amsterdam
- 106 Features

Data Wrangling/Cleaning

- Target variable - outliers removed
- Columns dropped for a variety of reasons
 - Null values
 - Textual data
 - Unique identifiers
 - Not generalizable to other cities
 - Correlated with other features
 - Categorical columns without enough diversity
- Useful categorical columns set aside for later encoding
 - Neighborhood
 - Property type
 - Room type
 - Cancellation policy
- Ensure remaining columns are numeric type

Feature Engineering

Features were added in iterations, and model performance was tested at each stage.

- Iteration 1
 - “Distance to city center” feature created
- Iteration 2
 - Categorical columns encoded, and added to model
- Iteration 3
 - “Top” amenities created as features

Choosing a Model

- Models tested:
 - Linear Regression
 - Decision Tree
 - Random Forest
- Scoring methods
 - MAE
 - RMSE
 - R-squared

Model Performance

Iteration 1

	Linear Regression	Decision Tree	Random Forest
MAE	37.02	36.29	33.94
RMSE	49.46	48.45	45.40
R-squared	0.38	0.40	0.48
Time to Train (s)	0.16	19.75	699.41

Iteration 2

	Linear Regression	Decision Tree	Random Forest
MAE	34.38	36.03	32.89
RMSE	46.40	48.15	44.42
R-squared	0.45	0.41	0.50
Time to Train (s)	0.24	26.49	1047.57

Iteration 3

	Linear Regression	Decision Tree	Random Forest
MAE	33.757	35.930	32.723
RMSE	45.684	48.120	44.258
R-squared	0.470	0.411	0.502
Time to Train (s)	0.180	28.120	1136.190

Individual Prediction Performance

- 60% of predictions within \$30 of actual price
- More feature engineering needed before model can be put into production

Feature Importances

Random Forest

	feature	importance
5	accommodates	0.255132
29	distance_to_city_center	0.112987
0	host_since	0.055777
62	Entire home/apt	0.049190
20	availability_90	0.045402
22	number_of_reviews	0.038520
9	extra_people	0.035805
21	availability_365	0.032950
7	bedrooms	0.031303
23	number_of_reviews_ltm	0.029197

Linear Regression

	feature	coefficient
62	Entire home/apt	76.310771
63	Private room	44.397271
39	Gaasperdam - Driemond	37.360949
41	IJburg - Zeeburgereiland	30.590906
67	indoor_fireplace	19.299585
59	Houseboat	18.516507
31	Bijlmer-Oost	15.596112
5	accommodates	15.507844
7	bedrooms	14.891070
35	Centrum-West	14.615807

	feature	coefficient
29	distance_to_city_center	-19.520960
48	Slotervaart	-17.192927
53	Apartment	-15.274404
32	Bos en Lommer	-15.022482
56	Condominium	-14.006873
38	De Pijp - Rivierenbuurt	-12.788246
37	De Baarsjes - Oud-West	-12.624574
75	beachfront	-12.322881
33	Buitenveldert - Zuidas	-11.742051
47	Oud-Oost	-10.777648

Future Optimizations

- Incorporate host ratings
 - Test dropping rows vs imputing nulls
- Impute nulls
 - For columns with significant missing data
- “Square footage”
 - Not collected with Amsterdam dataset, but likely found in other Airbnb datasets
- NLP with text columns