Inn the Neighborhood

Accessing the key to rental spaces

Report prepared by Ryan McLean,
Data Scientist

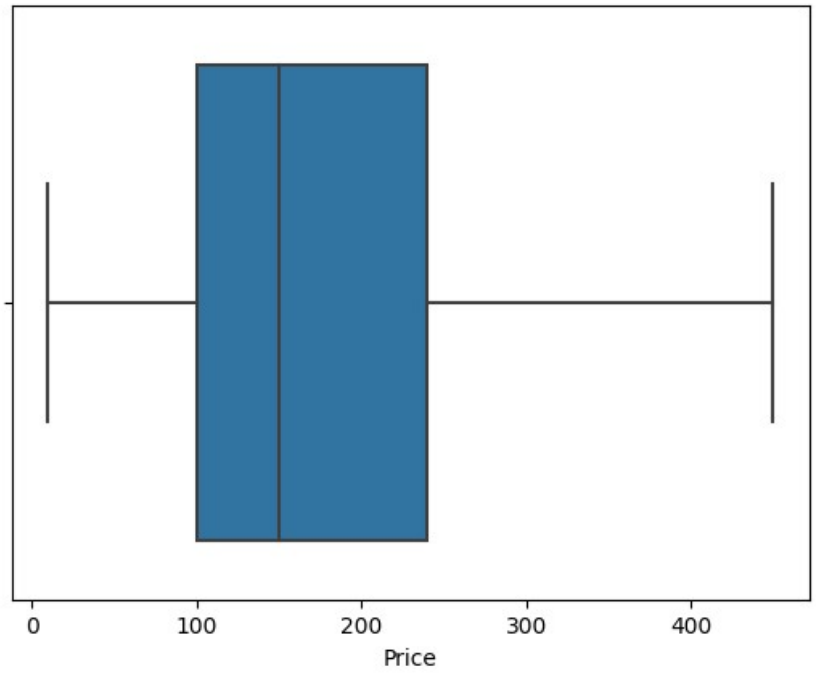
Inn the Neighborhood strives to create a premium online platform for short-term rentals

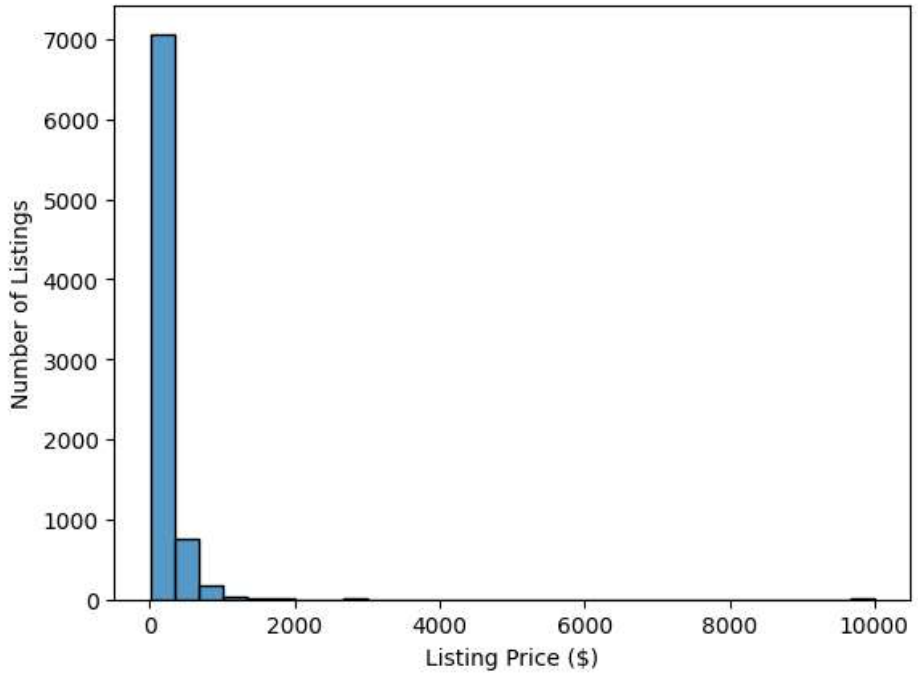
In the last decade, short-term house rentals have increased dramatically

A variety of people in San Francisco visit our site to list their apartment, home, etc.

Only 2% of site visitors actually list their home for rental

Although 50% of potential customers list their home between $100 and $240 per night, the variation in the data suggests that customers don't have a good sense for the amount they could be making by renting





Our plan forward

The data and analytics team will use machine learning to accurately predict how much a user could earn; increasing customer confidence, conversion, and retention

Final list of home features that predict price

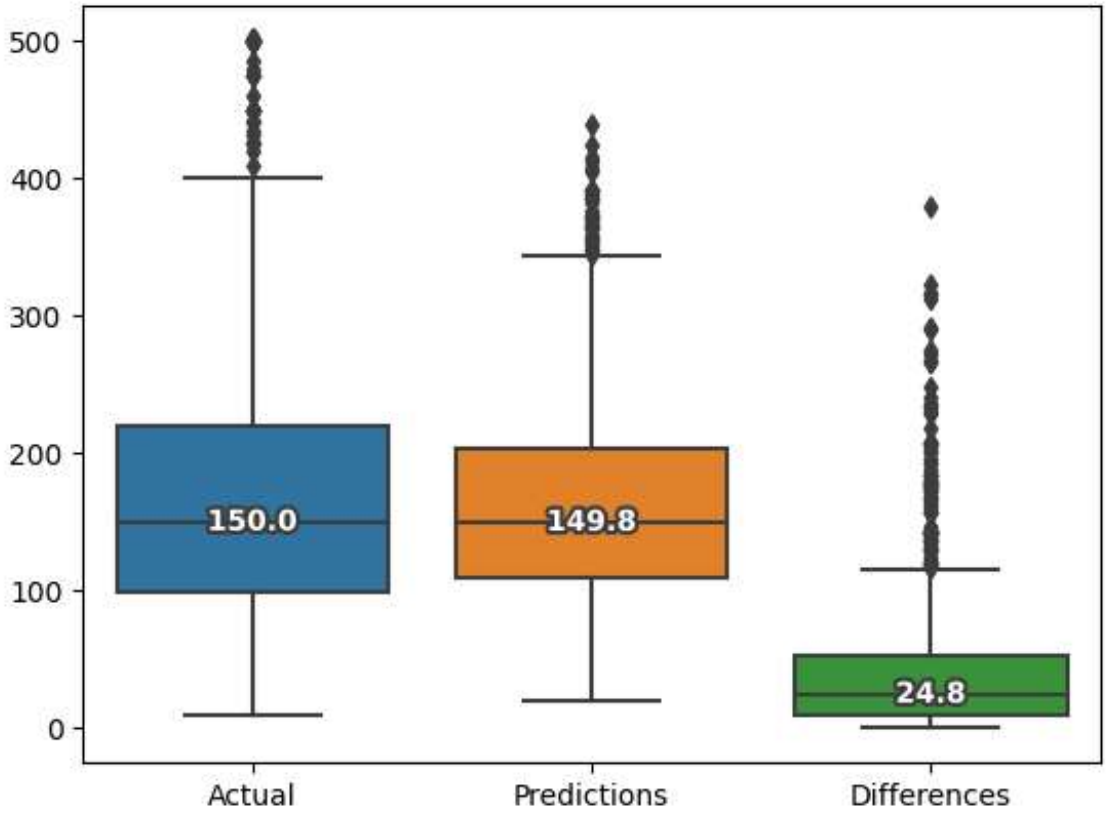
Model trained on 7,251 homes entered on our site

Data Processing Steps

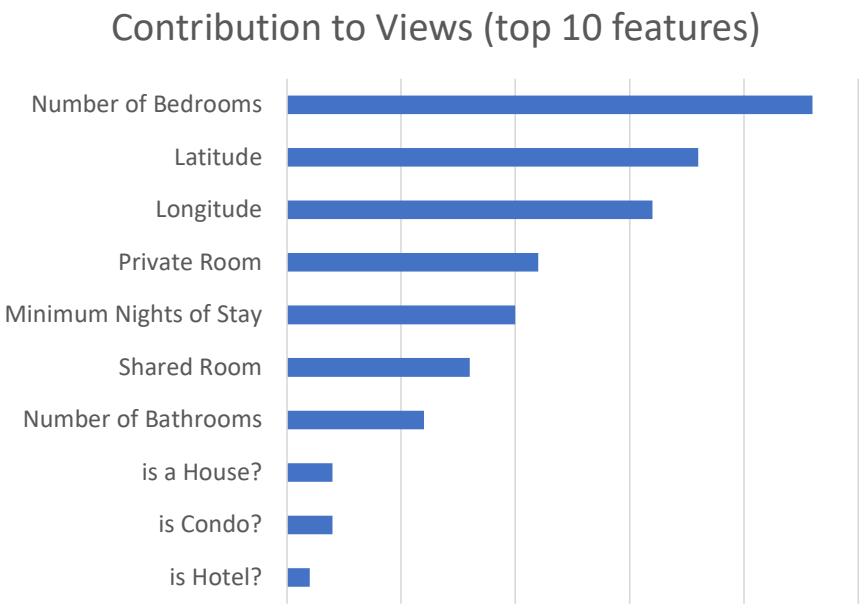
- Outliers: Taking out the most extreme listings to get an accurate model (house size, nights per stay, price). Removed about 10% of total sample
- Missing values: Many customers did not complete the form with # of beds and bathrooms, so these were filled in with the average
- Price changes: Since there was such a large dispersion of listing choices, the data were logarithmically transformed to strengthen the model after removing top 5% of the sample
- Property Type: Over 95% of listings fell into 5 categories so the 26 categories were condensed for clarity
- Bathrooms: Outliers removed and missing values imputed
- Bedrooms: Outliers removed and missing values imputed
- Minimum Nights: Outliers removed
- Machine Learning Preprocessing: I performed dummy encoding for categorical features and variable scaling to ensure that the information was scaled and interpreted correctly within the model

After evaluating multiple algorithms, the Random Forest algorithm was chosen for best performance

- The main KPI of the model was to estimate price within $25 of the actual price.
- The trained model had a median difference of $25 between estimated and actual price per night
- 25% of all estimations were within $10
- 75% of all estimations were within $53



Price is driven by size, location, and level of privacy



Listing price was most strongly predicted by the number of bedrooms and the location, followed by the minimum nights of the stay and whether the room was shared or private.

Using this algorithm, the median difference between estimated and actual listing price was $25

With this information, we can provide customers with confidence about what they can earn by listing their property with us, increasing customer conversion and retention

Conclusion

Data Science will help us unlock the key to the rental landscape