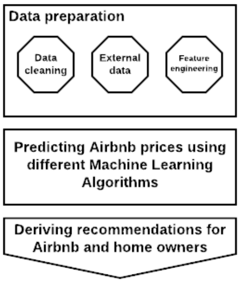
Drivers of Airbnb prices and recommendations -
Deriving pricing recommendations for landlords and Airbnb

airbnb Team HacKings



On the task:

Having been founded in 2008, Airbnb reshaped travel possibilities offering a personalized way of experiencing the world. With around 50,000 listings solely in NYC, the company excessively gathers data. **Adequately evaluating** the information with statistical techniques enables **empirically-justified business decisions**. Shedding light on the central determinants of the listings' price allows Airbnb to optimize their platform, and owners of properties to increase the perceived value of customers to increase and fully exploit their willingness to pay. The following outline describes the machine learning algorithms used to predict prices and our final recommendations.

Potential drivers of listings' prices:

In principle, **home owners are unrestricted** in setting the daily price for their apartment. Due to the high number of Airbnb listings around the world, the price is to a large extent governed by the **forces of a competitive buyer's market**. In that scenario, the price is driven by the customers willingness to pay (WTP), which is, in turn, determined by the features of the house. The provided dataset entails a wide array of information that are possibly determinants of the WTP ranging from the host name to the minimum nights per stay. To expand the information and improve the forecasts, we **added promising features** including information regarding the score of recent reviews and whether the home owner is a superhost from an **external dataset**. Lastly, we incorporated coordinates from subway stations and relevant landmarks in NYC to calculate a variable that measures the distance to the closest subway/landmark.

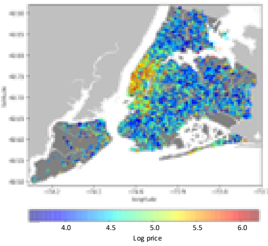
Selected features:

Longitude
Latitude
Review scores
Neighbourhood
Room type
Minimum nights
Number of reviews

Assessment of forecasting accuracy based on the RMSE:

Before feeding the data into the machine learning algorithms, which indicate the feature importance, the visualization of the data motivates to **take the logarithm of variables** for which the distribution is significantly skewed. Furthermore, running separate regressions for different room types represents a chance to reduce the regression error further. In other models, distinct room types were differentiated by dummy variables. The finalized sets for independent and dependent variables are then employed in the regression. To get a first impression for the order of magnitude of the results and to sanity check the results, we first estimated a simple **linear regression** model. Subsequently, we used **Ridge regression**, a technique for analyzing multiple regression data that suffer from multicollinearity, which we had already eliminated a priori, and **Lasso regression**, where data values are shrunk towards a central point. Additionally, we used **XGBoost** and **RandomForestRegression**, both of which are decision-tree-based algorithms, the former of which uses a gradient boosting framework, and the latter employs an ensemble of randomized regression trees. Lastly, we also ran a **Neural network**, the specifications of which are described in the attached Notebook, which yielded the smallest RMSE.

Plotting prices onto a city map of NYC:



Manhattan is the region with the most expensive listings.

Recommendations for Airbnb and home owners:

Based on the regression results, we identified several levers to positively influence the listing's price and therewith maximize returns of Airbnb and home owners. For the purpose of the challenge, the guiding principle for the selection of recommendations is **profit maximization**. It is needless to say that **other factors such as sociability** also contribute to the **welfare of home owners**. Hence, factors beyond the financial impact would need to be considered despite the analysis. Starting with home owners, the recommendation with the highest financial benefit attached to it is to **offer entire homes instead of private/shared rooms** only. For all machine learning algorithms, the home type was by far the most important determinant. Secondly, the data suggests that the superhosts has a positive impact on the price that is charged for a home. Hence, we recommend that **home owners should actively work towards meeting the superhost criteria**. Although the analysis was limited to listings with less than 31 minimum nights, this variable is negatively correlated with the price. **Reducing the minimum nights** would enable home owners to exploit the WTP of customers further. For Airbnb, we identified **review rewards as the most important measure** to increase customer engagement and indirectly through that also the willingness to pay for others customers that comes with a larger number of reviews. While the proximity to subways did not significantly increase the price, the should **introduce location-specific perks** to target home owners in expensive districts. Lastly, we recommend Airbnb to **give home owner guidance regarding a target price** that incorporates all additional factors considered in this analysis. With that, Airbnb ensures that home owners optimize price setting not only based on the location.

Algorithm	RMSE
Neural network	**0.305**
XGBoost	0.349
RandomForest	0.362
Linear regression	0.369
Ridge regression	0.375
SVG	0.417
Lasso regression	0.596

Top 3 recommendations:

for home owners:
- *rent out entire homes*
- *become a superhost*
- *lower minimum nights*

for Airbnb:
- *reward reviews*
- *location-specific perks*
- *suggest pricing*

Please consult the Jupyter Notebook for a comprehensive description of the recommendations,

Timo Debono and Sandro Bednorz