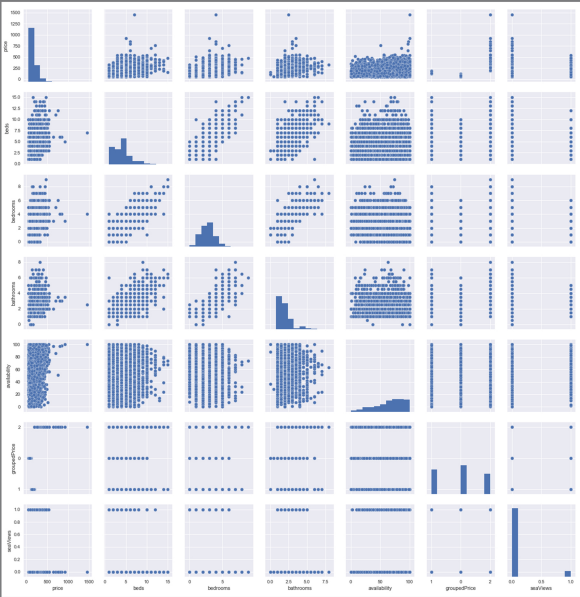


MALLORCA SHORT TIME RENTALS CLASSIFIER

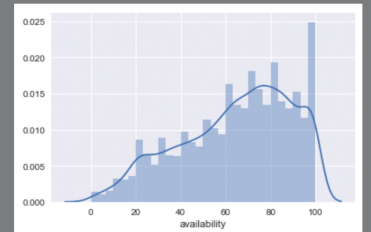
ABOUT THE DATA

- Data describes Airbnb's listing from Mallorca.
- Dataset includes more than 11019 samples.
- Clean data has 13 features.
- Label was assigned regarding the 0-33-67-100 percentiles of the logarithmically normalized data, that is to say, split into three categories: lowCost, medium and premium.
- This mosaic view compares seven fields of the dataset.

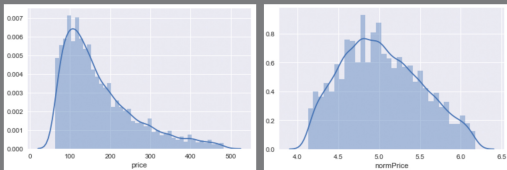


USEFULL FACTS

- Most of the villas do not have sea views (3% do, versus the 6% average).
- 11% of Communities and Houses have swimmingpool, while for others is <9%.
- Neighborhood matters: maximum average price per zipcode is \$450 per night, while minimum is \$70.
- This graph shows the percentage of available nights of every property.

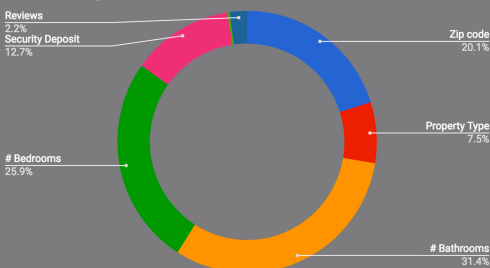


ABOUT THE TRAINING



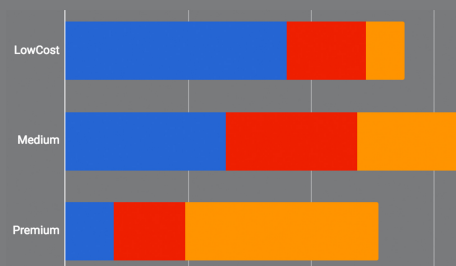
- The model was tried out both with unnormalized and normalized price, being the latter the final choice.
- 80% of the data was used for training. No cross-validation.
- The first implemented model was linear regression running with the TensorFlow library.
- The final model runs on catBoost, a gradient boosting on decision trees specialized on working with categories - the main type here.
- To improve the model new features would be needed, like the sqft.
- Half of the features were ruled out by the model, as seen bellow:

Feature weights



ABOUT THE TESTING

- 20% of the data was used for testing, and random classification guess is 33% for this problem.
- The linear regression model performed with a 41% of accuracy.
- The catBoost implementation had a 51% of accuracy.
- Misclassification does not only occur between adjacent categories, but also between premium and lowCost, which is a bad sign:



- If trained by property type, it performs worse for all of them. Also when the category-prices were calculated per capita, and when all the features were normalized.
- The most important feature to determine the final result was the number of bathrooms, followed by the number of bedrooms.
- Some features like sea views or having a balcony do not influence the classification.

TRIVIA FACTS

- There were more than 10 properties exclusively listed in German.
- One host offers an igloo on the island if it ever snows enough.
- Almost half of the properties have balcony.
- The highest number of bathrooms in a property is eight (for 11 people!).

SOME CONTEXT

- Data was gathered by insideairbnb in March'17.
- Dataset covers from March'17 to March'18.
- Analysis done by Maria Benavente for the course Knowledge Technologies and Data Mining 2.
- Credits also to Adrian Spataru and Roman Kern for the useful advice.