Forecasting AirBnb revenues in Geneva

The context and the problem

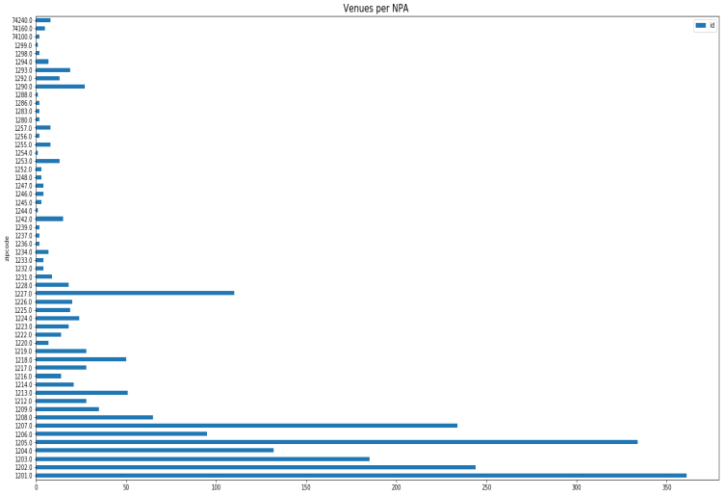
- Airbnb is a popular website to offer short time accomodation to customers
- A municipality might want to know the situaiton in its city to better understand if Airbnb is going well or not in its territory and what are the main factors affecting the revenues for such service
- Some renter might also want to have insights on how to improve the revenues from its rented properties

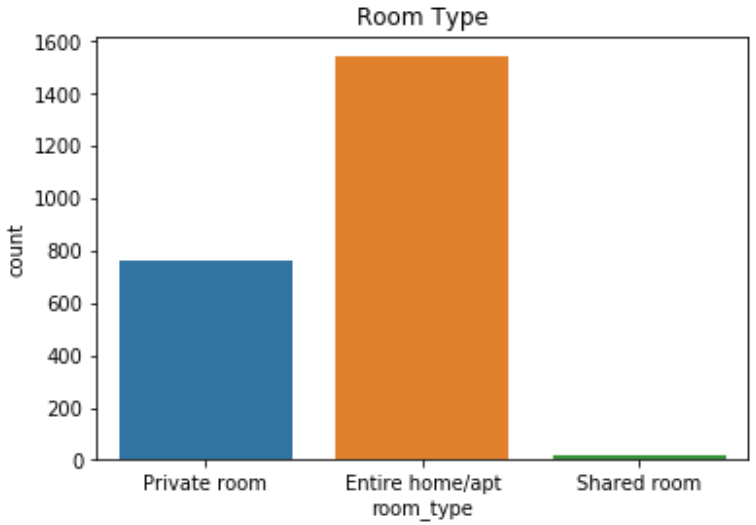
The problem:

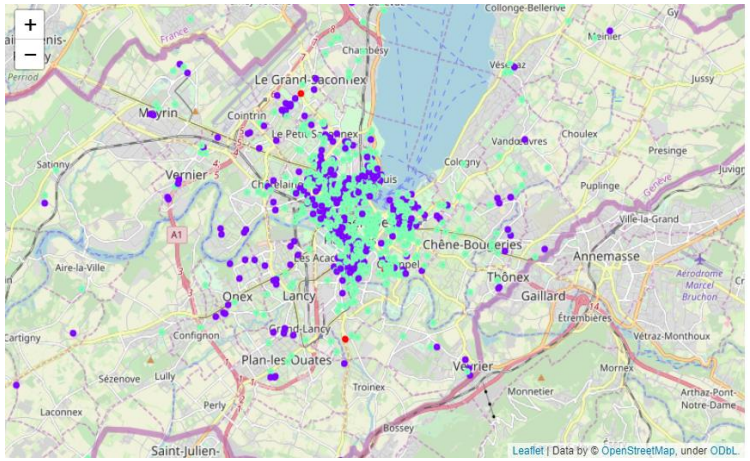
- In the project we foreacast prices and annual revenues of homes and appartments in Geneva, Switzerland, using data from
  - http://insideairbnb.com/get-the-data.html
  - Foursquare API
  - Geopy

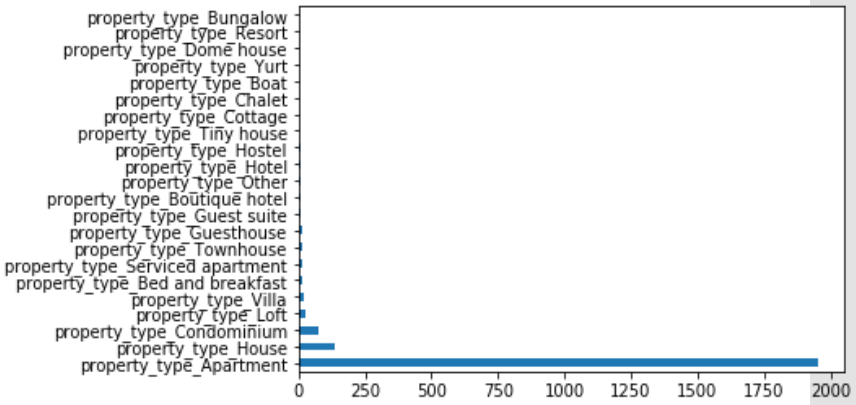
Exploratory analysis

Most venues rented ar entires homes or appartments.
Most venues are typically rented in apartments and not in houses.
The houses and apartments are more or less evenly distributed in the city center while is number is decreasing in more periferic neighborhoods.







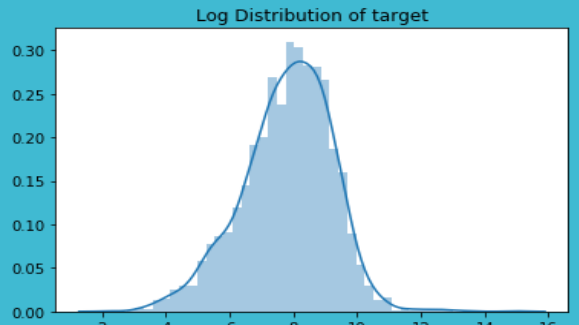


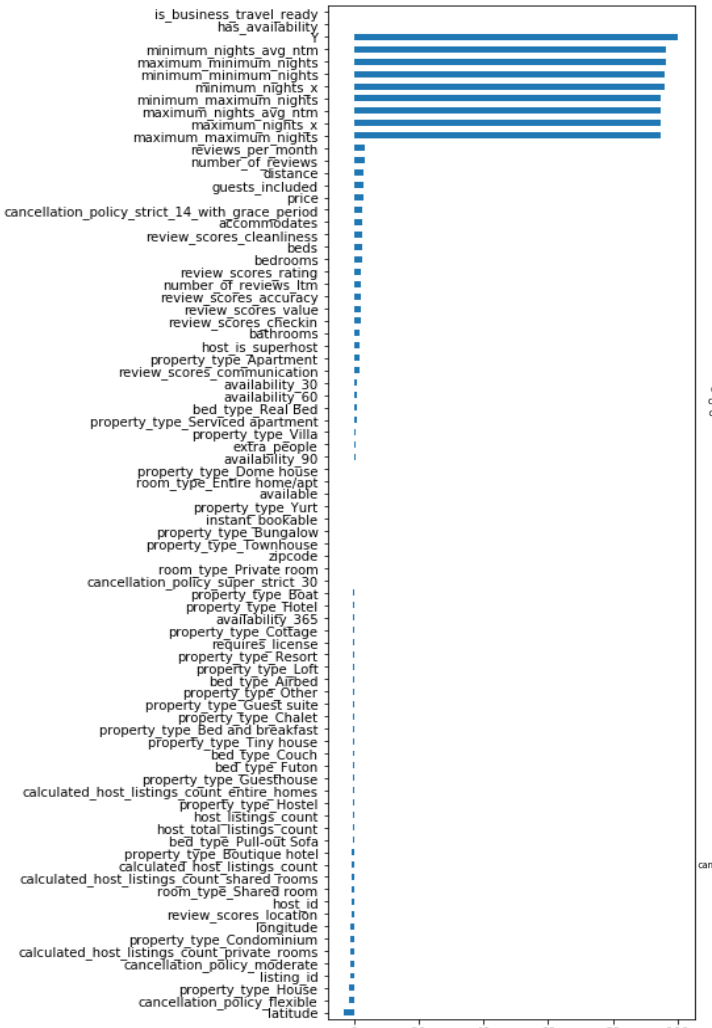
Taget Variable = Annual revenues per venue

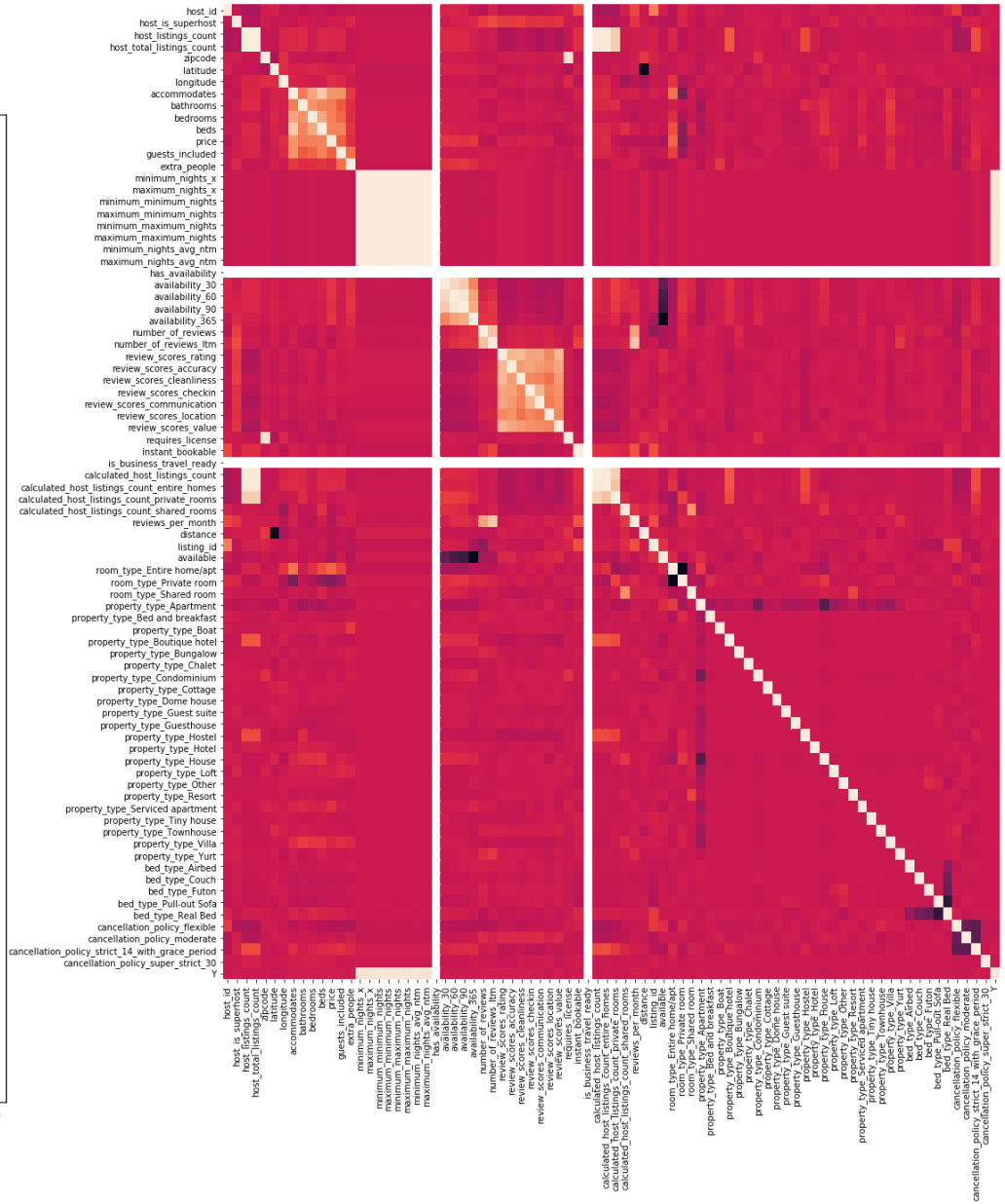
Exploratory analysis

We apply a logartimic transformation to the target variable to make its distribution less skewed towards high values

Few parameters seem to correlate heavily with the target.







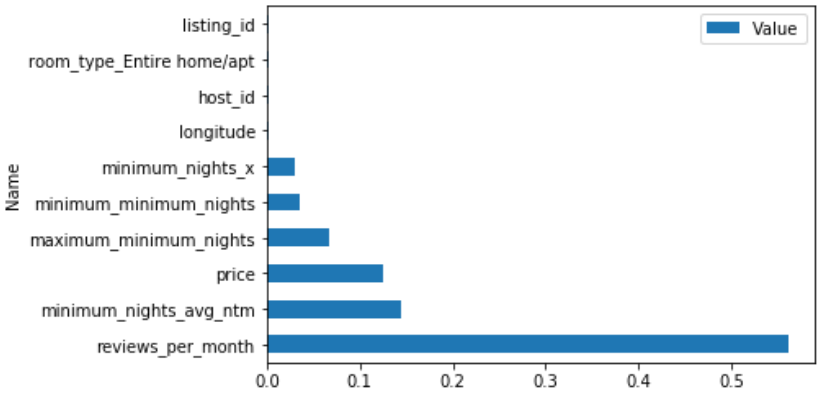
Modelling

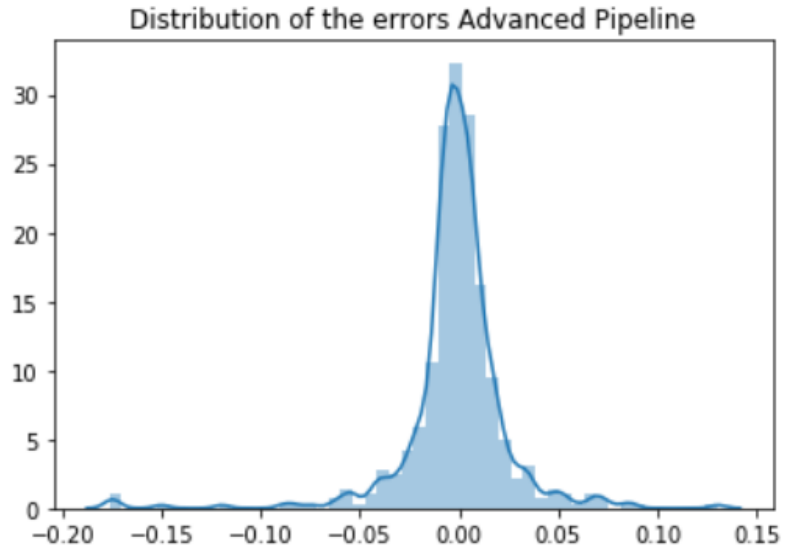
The modeling pipeline is tuned and validated on a test set.
The results are shown in the features. As suggestred by the correalation matrix the target variable, annual revenues per zyear is pretty well correlated with number of days the house is rented and the number of reviews per month. Actually this is ththe biggest result of the model.

Taget Variable = Annual revenues per venue

Model: RandomForestRegressor
CV: kFold
TestTrain: 20%/80%

Average error on annual revenues : 11.7%

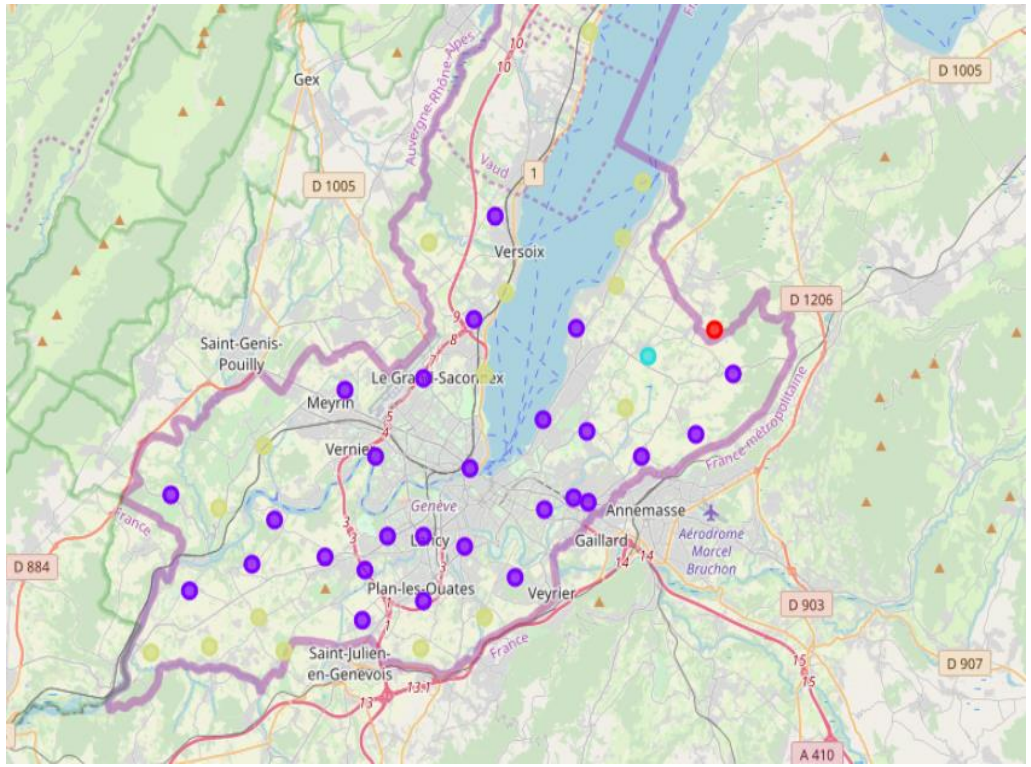


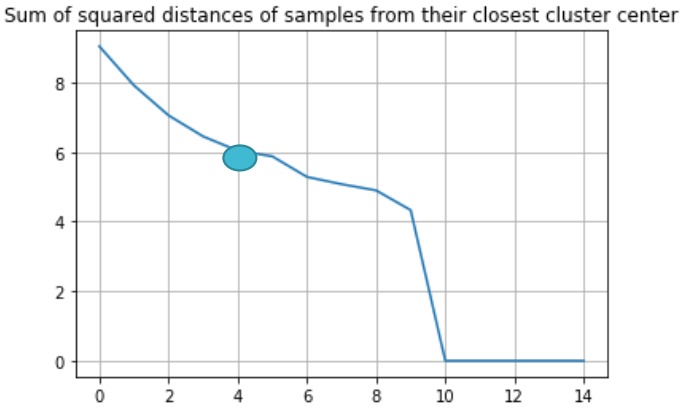


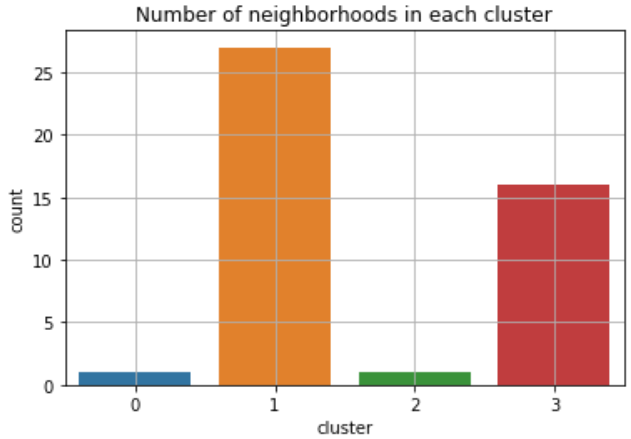
K-means clusters

To improve quality of the model the foursquare api is used to make cluster between the neighborhoods in geneva. The results are shown in the figures.
The model does not get a bette accuracy using this approach

Taget Variable = Annual revenues per venue

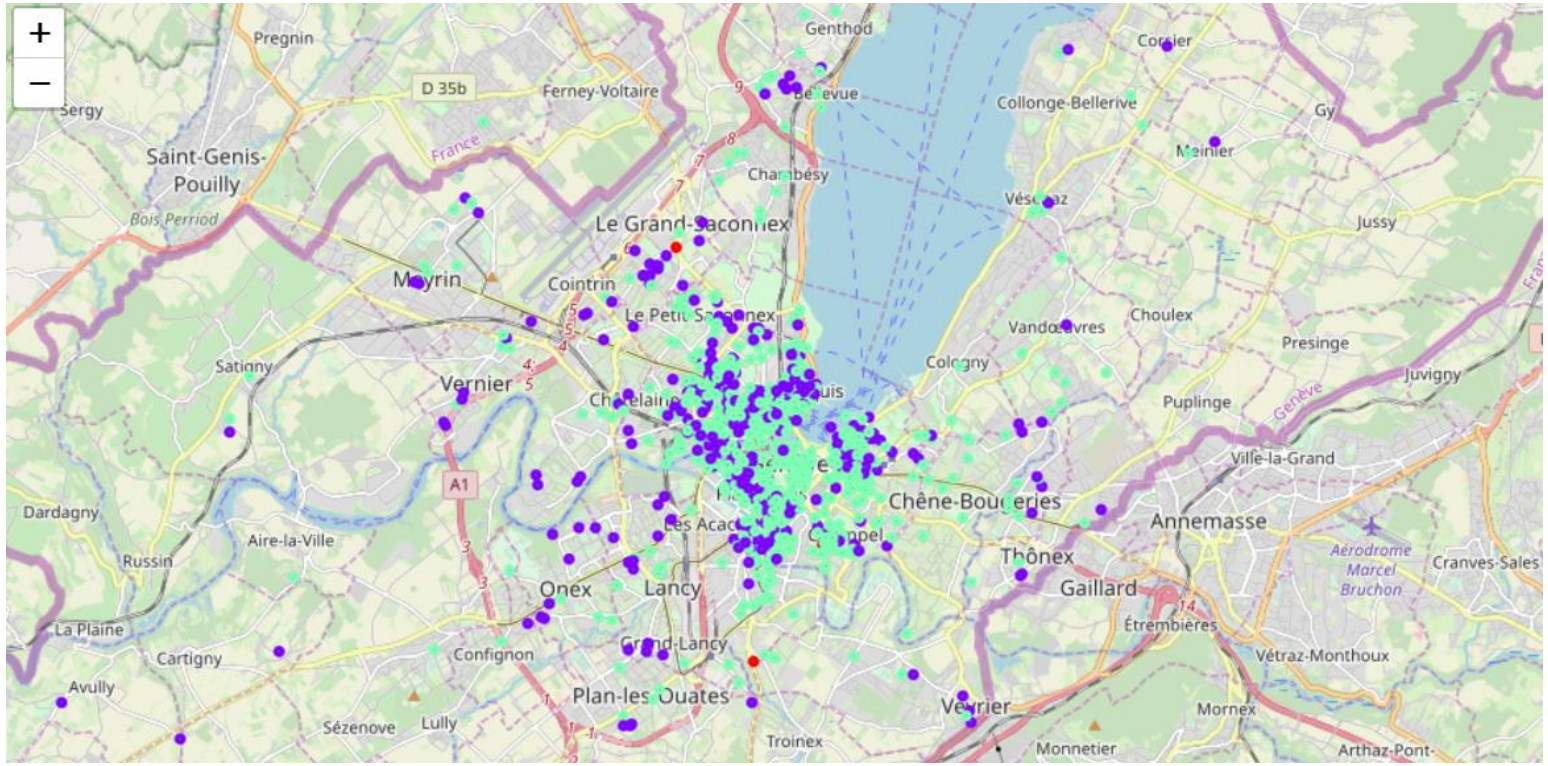






Average error on annual revenues : 12.7%

The solution



- A random Forest Regressor is traned on a sample to predict the revenues for a given house the accuracy is computed as the MSE of 11%
- The most important parameter is found to be the number of reviews an apartment has got.
- Using neighborhoods caracteristics to reinforce the lerning has no effect. This might be to the fact that Geneva is a relatively small city with lot os similarities in its territory.
- Municipalities and renters can have a better look at the scenario of AirBnb thanks to this work and might try to understand if their property can get more value from AirBnb or from other renting strategies.

Patrizio Canzi