AirBnB has become a very convenient and popular platform for travellers to find a short-term accommodation or renters to lease.

In this article I would like to take a close look at the data from the AirBnB service in Seattle in 2016.

In the dataset there are 3818 flats listed in AirBnB with availability provided from January to December 2016.

1. *How is Distribution of Lodgings in Seattle?*

The map below shows the distribution of lodgings in Seattle and their listing price range per night per guest (red: over USD50, orange: between USD30 and USD50, and yellow: below USD30). It suggests a radical distribution centred at the Central Business District (around Belltown) where both the density of lodgings and number of top listing prices are the highest. Convenience is one of the major factors driving the lodging business, so the infrastructure of this area may explain why lodgings concentrate there － availability of terminals where main sea routes converge, grid-shape road networks and facilities for serving high populations like town hall, museums, large shopping malls and cinemas .

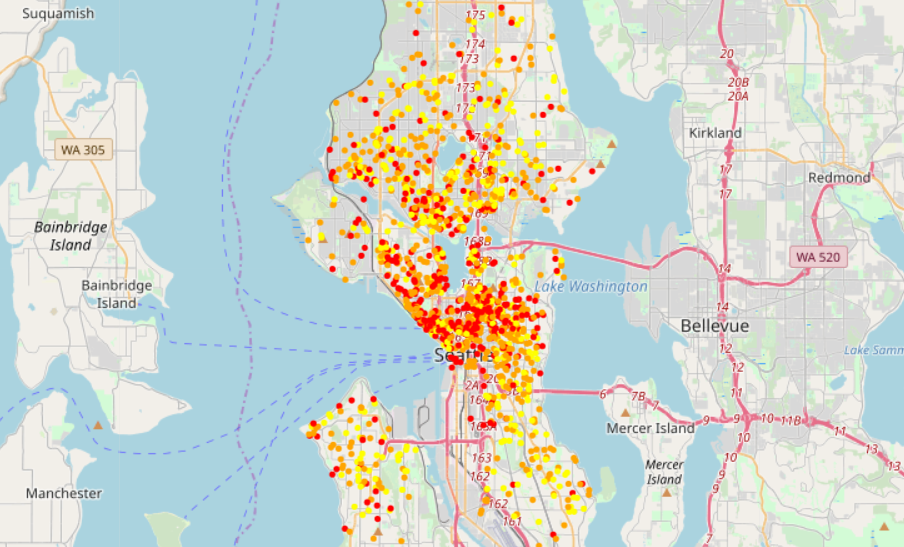


Fig 1.1) Radical distribution

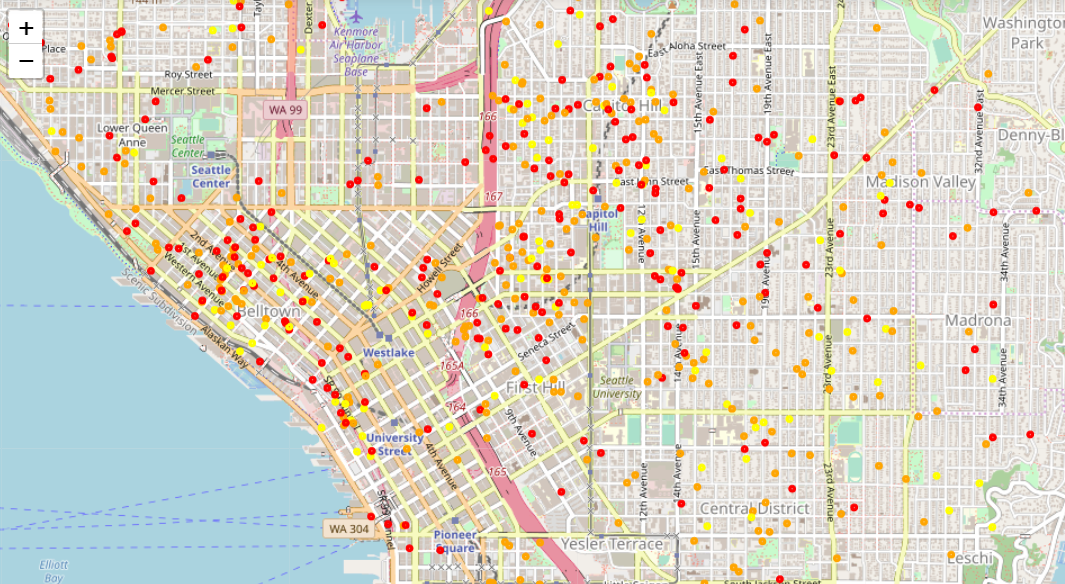


Fig 1.2) Presence of terminals where sea routes converge, and grid-shape road networks

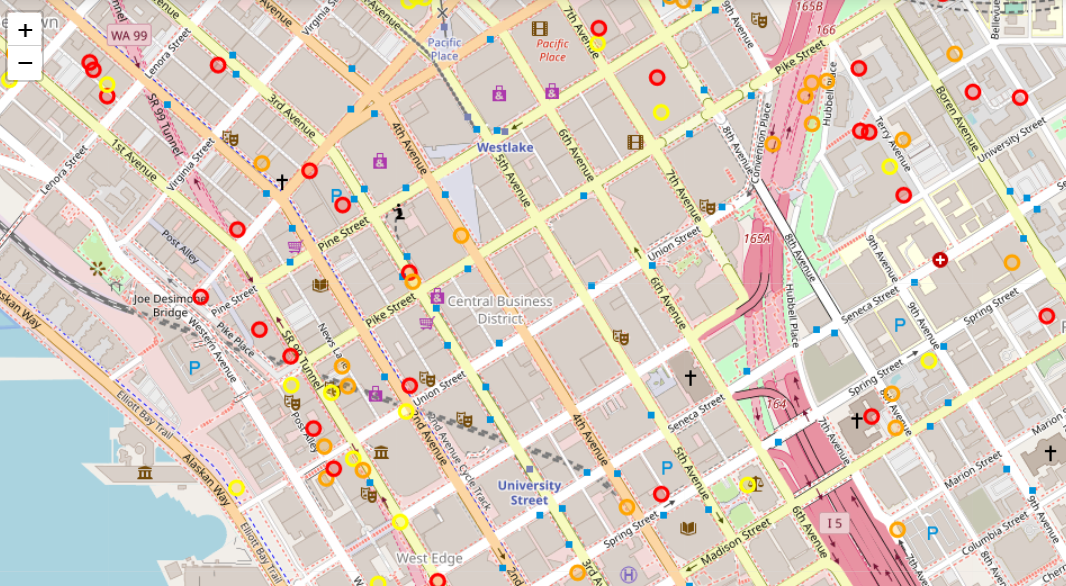


Fig 1.3) Living convenience by means of facilities

List of neighborhood with top ten highest number of lodgings:

neighbourhood

Capitol Hill 351

Ballard 213

Belltown 204

Minor 192

Queen Anne 187

Fremont 148

Wallingford 143

University District 107

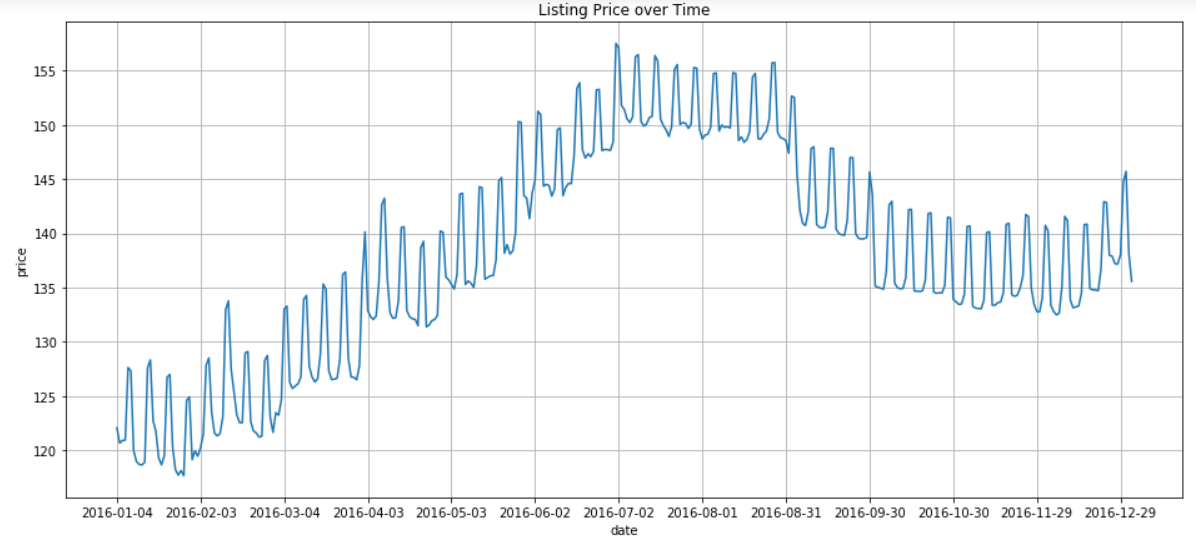
First Hill 105

Stevens 98

(Belltown ranked the third but its area size is less than half of Capitol Hill or Ballard)

*2) How is seasonal change in listing price?*

We try to understand the trend by plotting the daily average price:



The graph above shows the seasonal trend in the change of seasonal prices:

1) The average listing prices increase gradually from around $120 to $150 from January through the summer, and the prices go down after 31 Aug and maintain at around $135 till the end of the year.

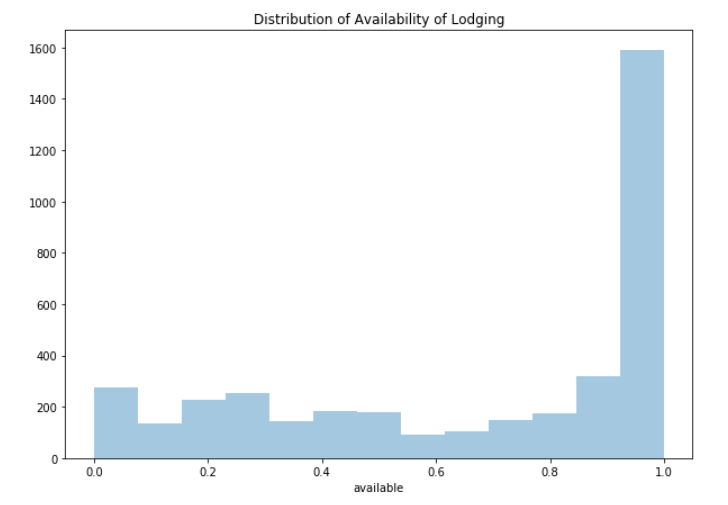
2) The average price by the end of the year is obviously higher than the beginning of the year. Since it is not very likely for a sharp drop of listing prices for more than 10% in the early quarter of the following year, it seems that there is a boom in this industry over the observed period.

3) Sharp peaks reported around 4-5 times a month are due to higher listing price in the weekends.

1. *OK I have some ideas on the prices, but is it easy to lease a lodging if I start a business in Seattle?*

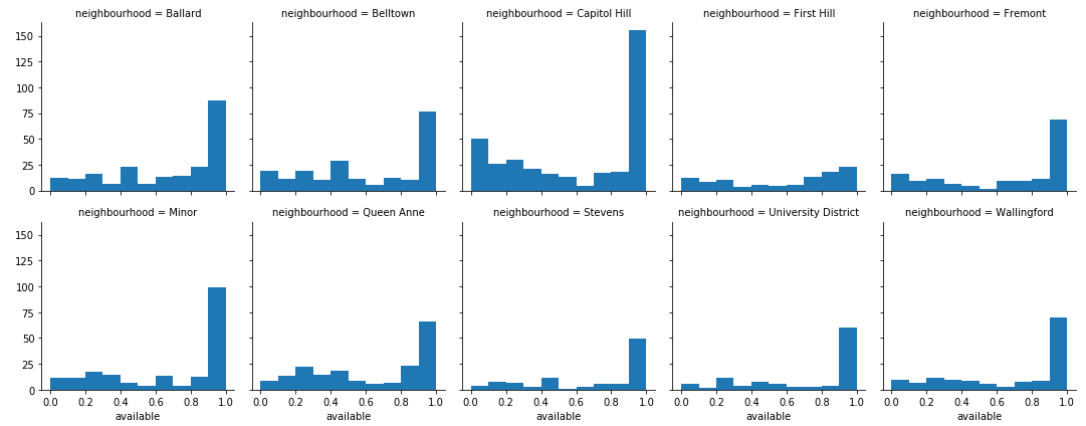
Here we take a look at the distribution of the overall availability of lodgings (number of available days / number of listing days).

It seems not convincing to start a lodging business in Seattle since there are nearly 1,600 out of 3,818 lodges available all the time.



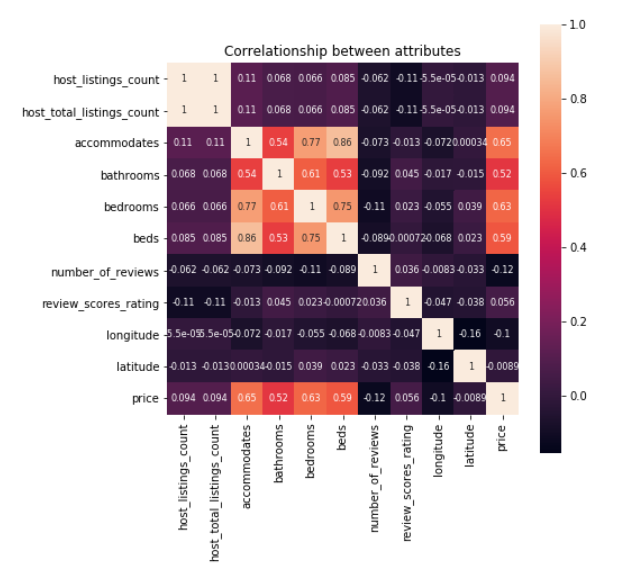
It seems that the results are not convincing enough for a lodging business to sustain in Seattle, but is it a general phenomenon of the lodging business or is there any neighbourhood easier to lease?

We can look at the following histograms from other neighbourhoods. Although the ratio of lodgings with high availability is also high, the distribution of availability in First Hill (adjacent and more inland neighbourhood of Belltown) is more even. In Capitol Hill, there is a moderately high ratio of lodgings unavailable all the time.



*4) Is it possible to predict listing price?*

First we look at the correlation coefficients of the continual parameters to see if there are obvious relationships among any of them:



1. It makes sense that the number of guests, bathrooms, bedrooms and beds, which define the size of the flat, is mostly relevant to the change of prices.
2. It is understandable that the number of reviews may affect a flat’s popularity thus posing a slight influence on the prices.
3. It is interesting that longitude has a greater effect on the listing prices than latitude, which can be explained by the geographical fact that Seattle is a spreading city along longitude with lodgings concentrating at the central part of the city.
4. Review score and rating is not relevant to the listing prices because the range of rating is very narrow (50% of the users give 96 marks according to the original dataset)

Finally, I tried to predict the listing prices by means of machine learning.

Distribution of price:

count 3153.000000

mean 126.902632

std deviation 90.205348

min 22.000000

25% 75.000000

50% 100.000000

75% 150.000000

max 1000.000000

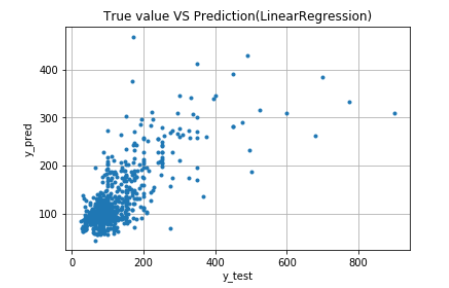
Linear Regression

RMSE train: 64.4248

RMSE test: 65.1412

R^2 train: 0.4759

R^2 test: 0.5281



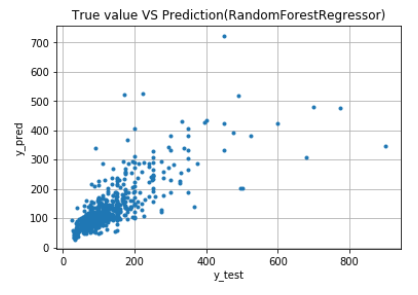
RandomForestRegressor

RMSE train: 21.5256

RMSE test: 61.2714

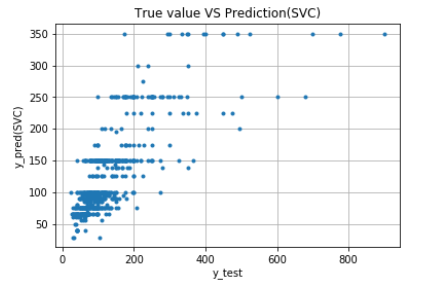
R^2 train: 0.9415

R^2 test: 0.5825



Support Vector Machine

RMSE train: 64.4248   
RMSE test: 62.3423   
R^2 train: 0.4759   
R^2 test: 0.5678



It seems that the performances of the three algorithms are more or less the same (except that overfitting occurs in RandomForestRegressor but it does not significantly affect the result)