Liligo Test 2: Paris, France Airbnb

Robert Herczeg

2017-05-14

Data source

Data was downloaded from Inside Airbnb. 3 files were downloaded:

- listings.csv.gz
- reviews.csv.gz
- calendar.csv.gz

Listings contains detailed data about the accommodations such as host name, price, different kinds of score etc.

Reviews data includes the review, the date, the reviewer name etc.

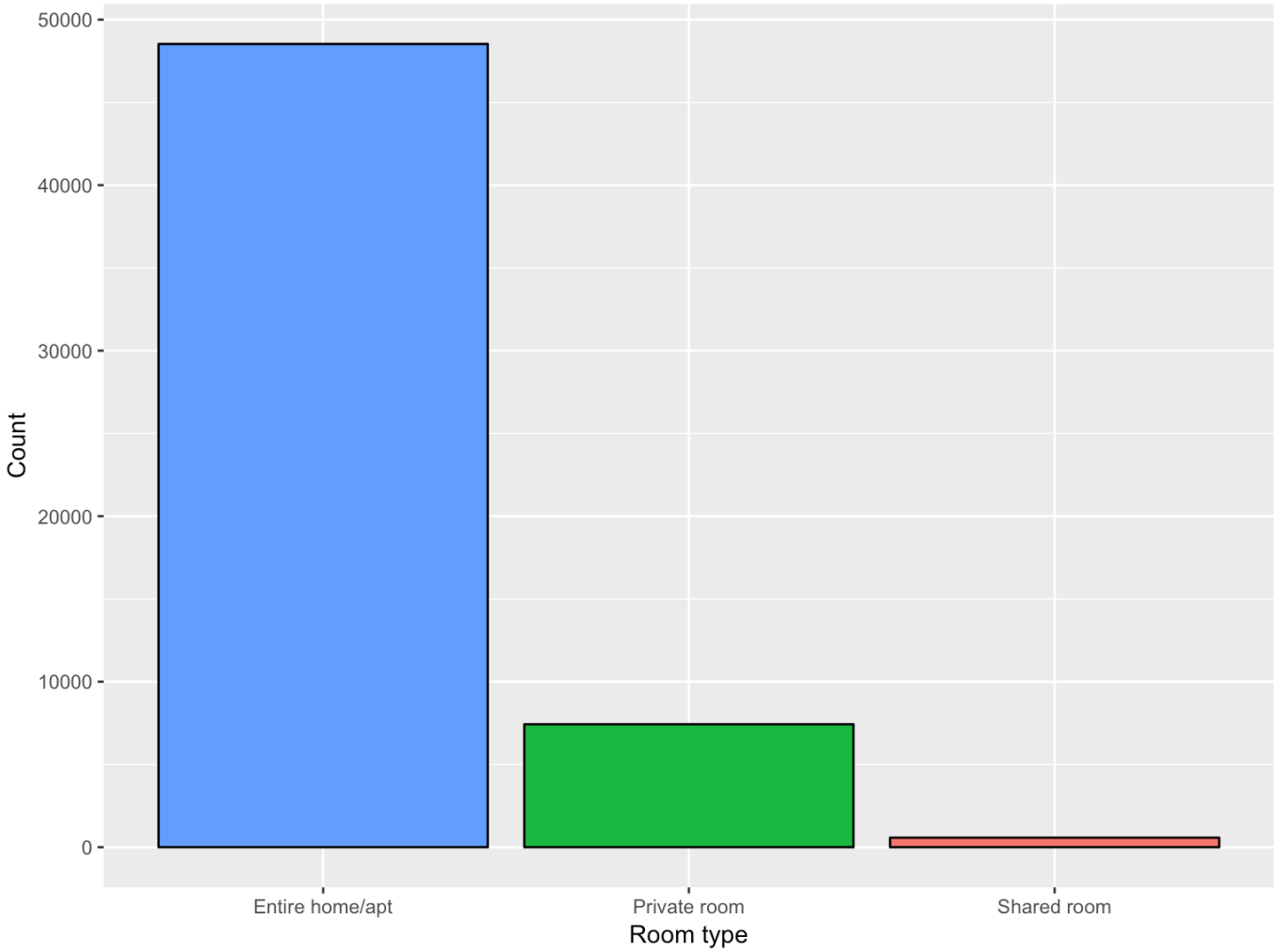
In the calendar data set, we can check the availabilty of given accommodation in a given day.

Dataset - basic info

- **48,306** hosts
- **53,920** locations
- **844,397** reviews
- There are hosts since **2008** august.
- Prices varies between 0 - 7790€. On average, an accommodation costs ~ **95€** (median - 759€).
- Reviews varies between 0 - 5510. Usually ~ **340 character long** reviews are written (median - 260).

Room type

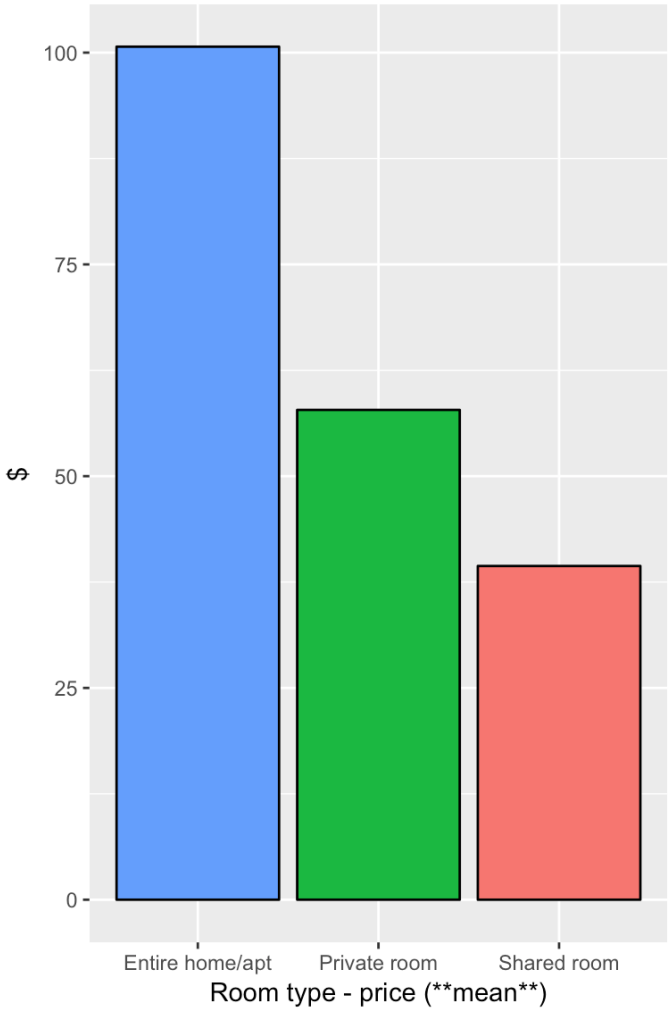
There are **56,535** locations and ** 85.84% ** of them are entire homes/apartments.

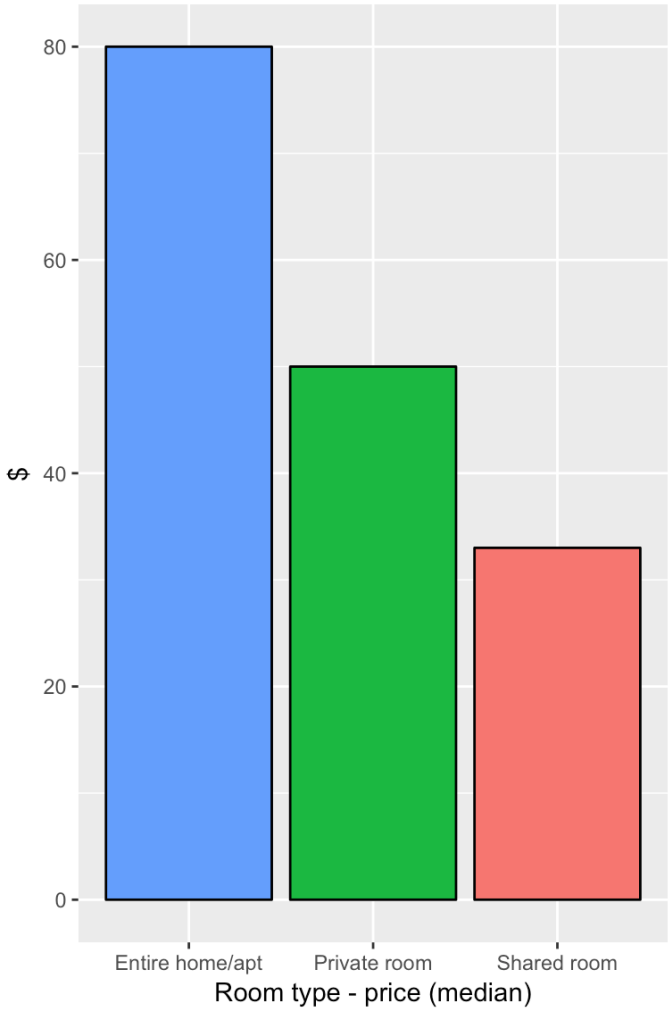


13.14% are Private rooms, while 1.02% are Shared rooms.

Room type - price (mean, median)

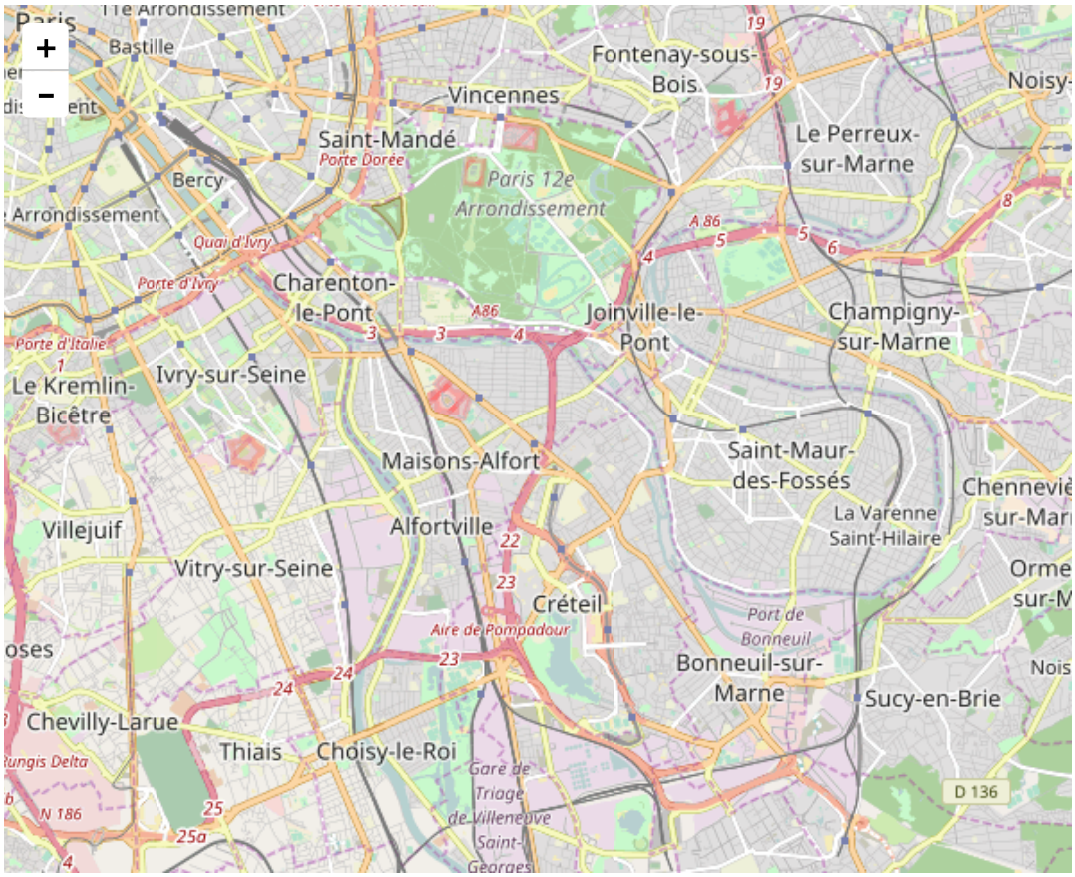
Entire home/apt are the **most expensive** compared to the other two types.





Distribution of bedrooms

Most of the accommodations have **less than 2 bedrooms**.



Leaflet | © OpenStreetMap contributors, CC-BY-SA

red >= 3; blue = 2; yellow < 2

WordCloud

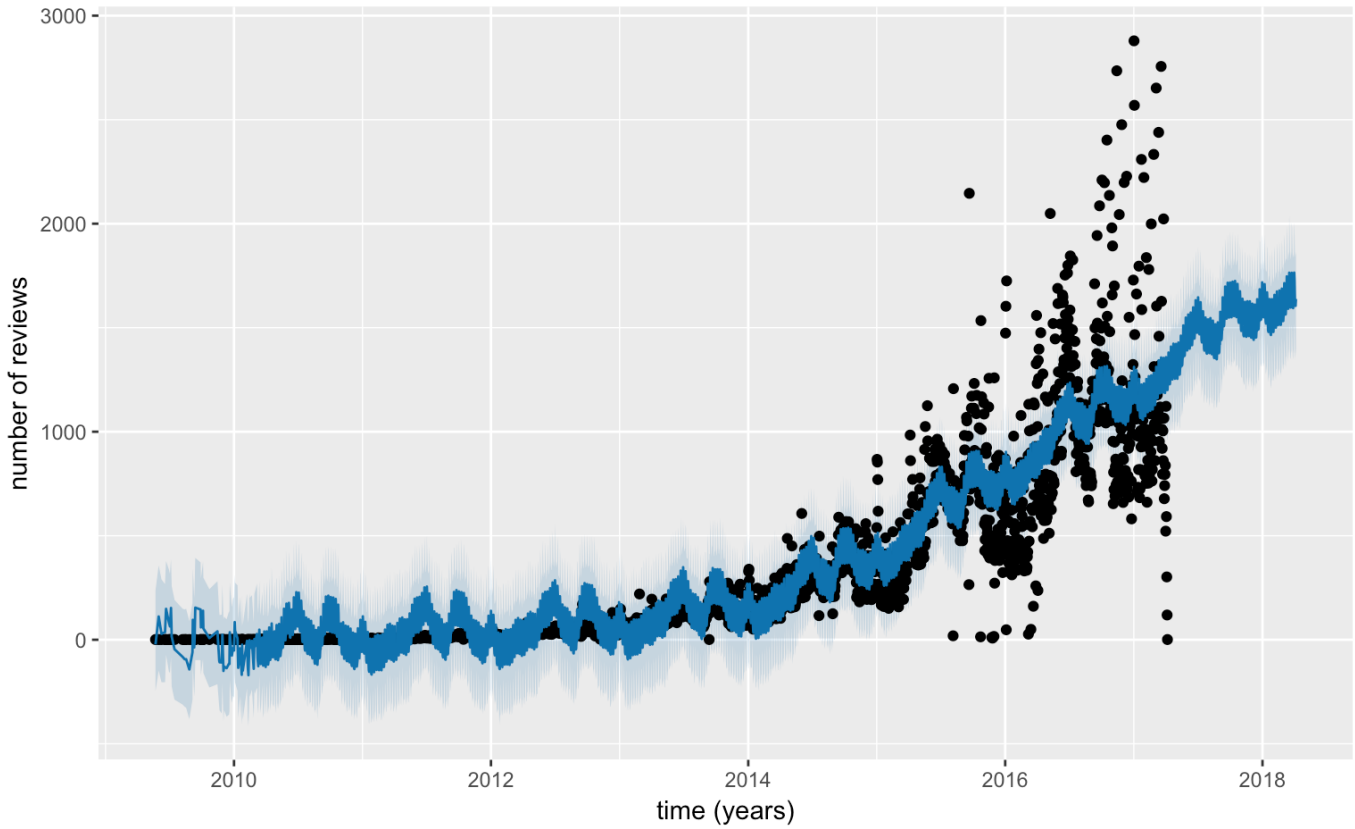
Usually, the most frequent words are **positive**. Based on 3000 random comments from reviews.

Review tendency over time

The number of reviews is continuously **increasing** since the beginning.

Review prediction - I

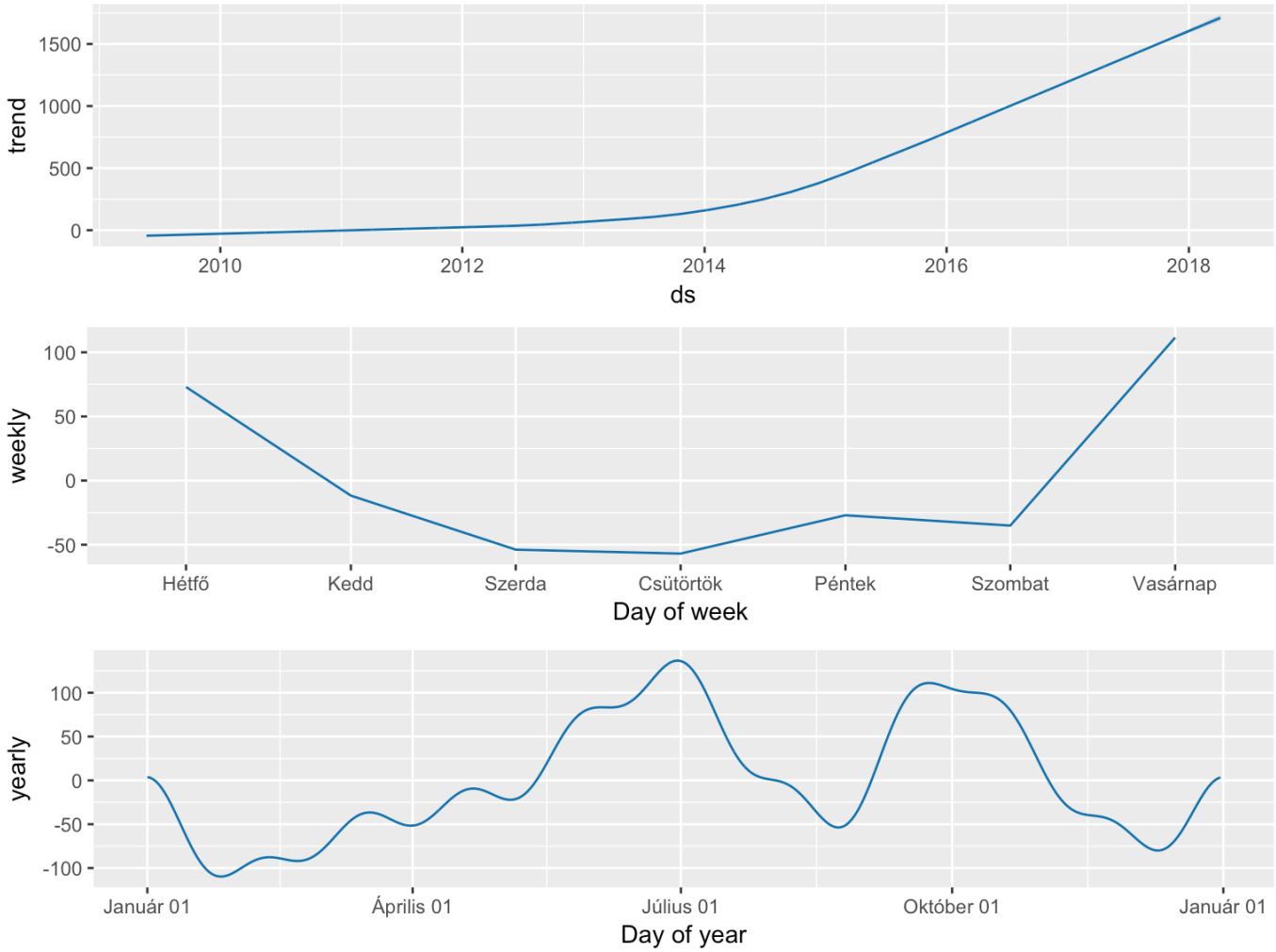
The prediction shows a **growing tendency** which continues in 2018.



Based on Facebook's prophet package.

Review prediction - 2

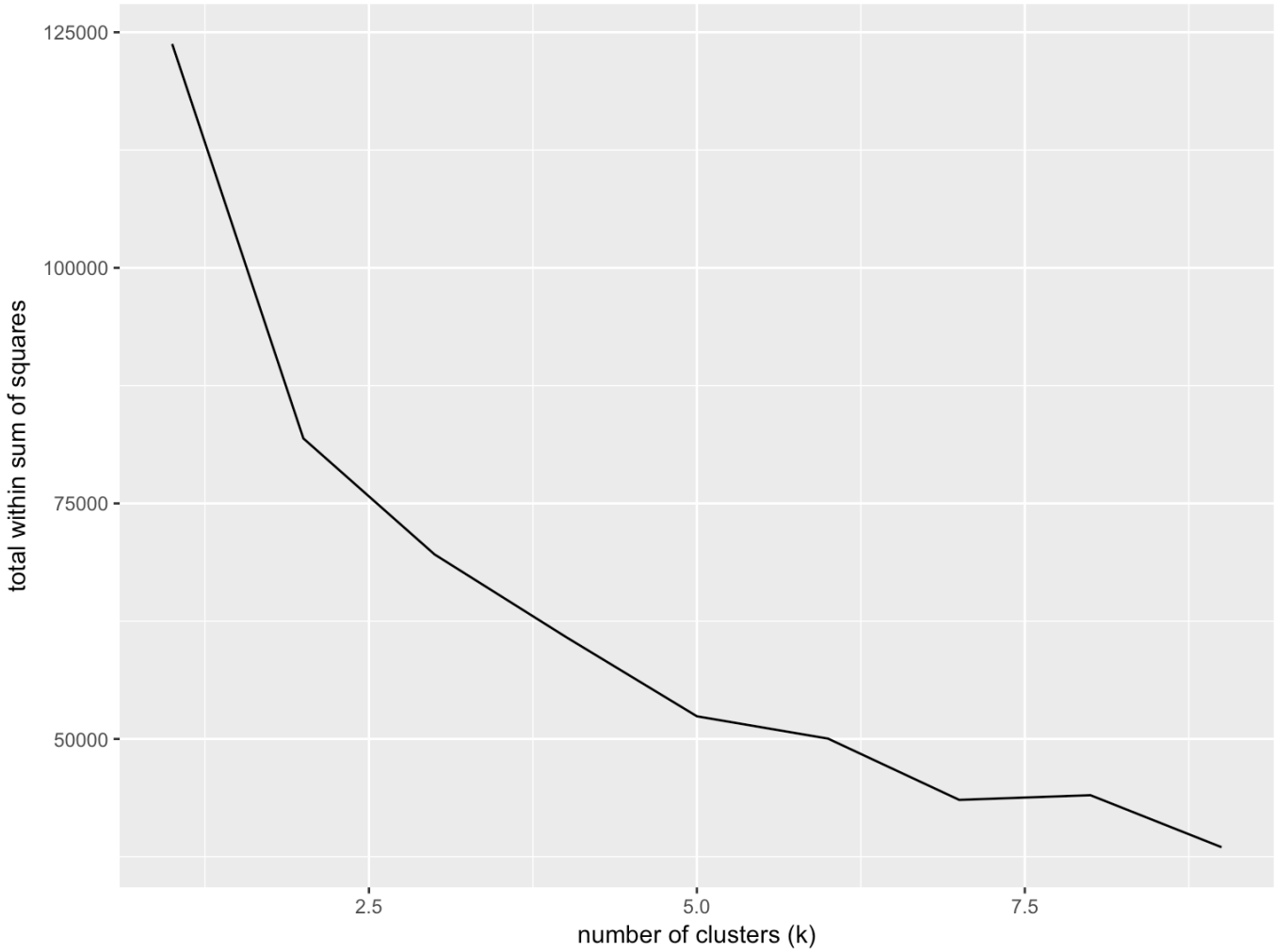
Simplified trends with three differents resolutions.



Monday and **Sunday** are the most active days to write a review. In the yearly graph, there are two peaks one in July and one in October. These clearly indicates that people usually visits Paris in the middle of **summer** and at the begining of **autumn**. Based on Facebook's prophet package.

K-mean clustering

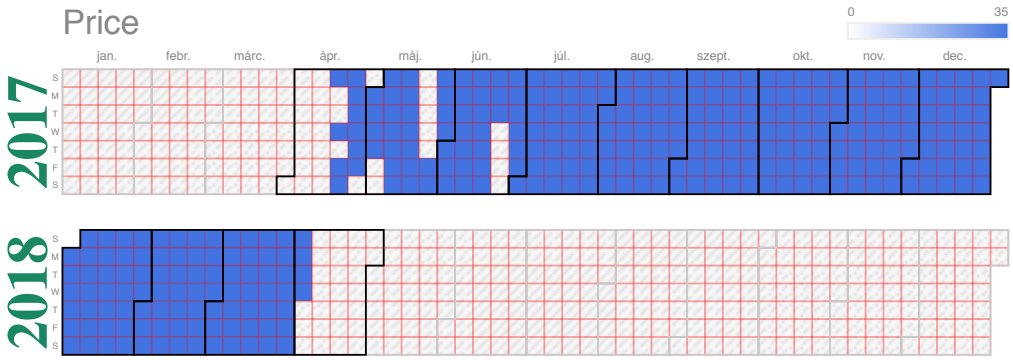
6 or **7** clusters were differentiated. Therefore, we can create at least 6 different segments.

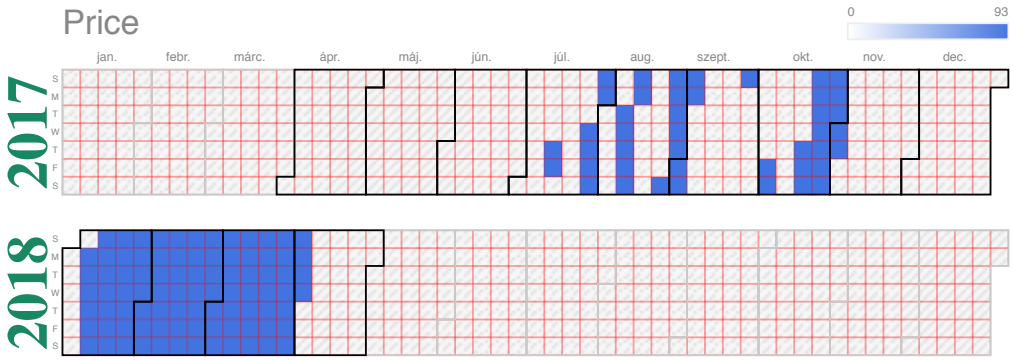


4 metrics were used: cleanliness, communication, location, value

Calendaer

Two randomly choosed hosts from calendar data set and the visualization of their availability until May, 2018.





Thank you for your attention!