

Study of Superhost with a Revenue growth perspective for Airbnb in Paris, France

Group members:

CHING-MIN HU (013726154)

Dandan Zhao (013795392)

Nandini Puppala (013752713)

Prathusha Koouri (013710658)

Qiao Liu (013802893)

Venkata Anil Kumar Thota (012525734)

Introduction and Motivation of study:

Airbnb is a popular home sharing platform where anybody with an open space can become host and offer it to the global community. Any guest with a specific need can search for accommodation on the Airbnb website with the price, amenities, reviews and location filters.

Airbnb recognizes 'experienced hosts who provides a shining example for other hosts, and extraordinary experiences for their guests [here](#)' as Superhosts. According to Scott Shatford's analysis from [airdna](#), Superhosts 'globally earn 60% more revenue per available day'. So we want to understand the distribution of Superhosts: where they are concentrated geographically and what percentage they are among all hosts in Paris. By exploring, we can find areas to expand our Superhost ratio to create more revenue.

Description of the Dataset:

We have considered the Paris dataset from [insideairbnb.com](#), which is a third-party website collecting Airbnb data from publicly available sources.

Data extracted on Feb/05/2019 and Feb/06/2019 with calendar data from February 2019 to January 2020 and updated listings information extracted in February.

Listings:

The listings dataset contains all the information related to individual listings within a month. It has 106 variables and various row numbers depending on the specific month.

Neighborhood:

The neighborhood data is a geojson file that exhibits different neighborhoods of Paris.

Calendar:

The calendar dataset contains the availability of individual listings on a daily basis from February 2019 to January 2020.

Design principles and Perceptual Properties:

The aim in our project is to transfer abstract into physical attributes of vision effectively.

Design Principles used:

- Gestalt laws of proximity and connectedness are used in Graph 1, Graph 4 and Graph 5.
- Maximized data-ink ratio by using text instead of legend in Graph 1, Graph 4 and Graph 5.
- Minimized chart junk by deleting grid lines and using white background in Graph 2 and Graph 3.
- Layered and separated regions with different colors in all the graphs.

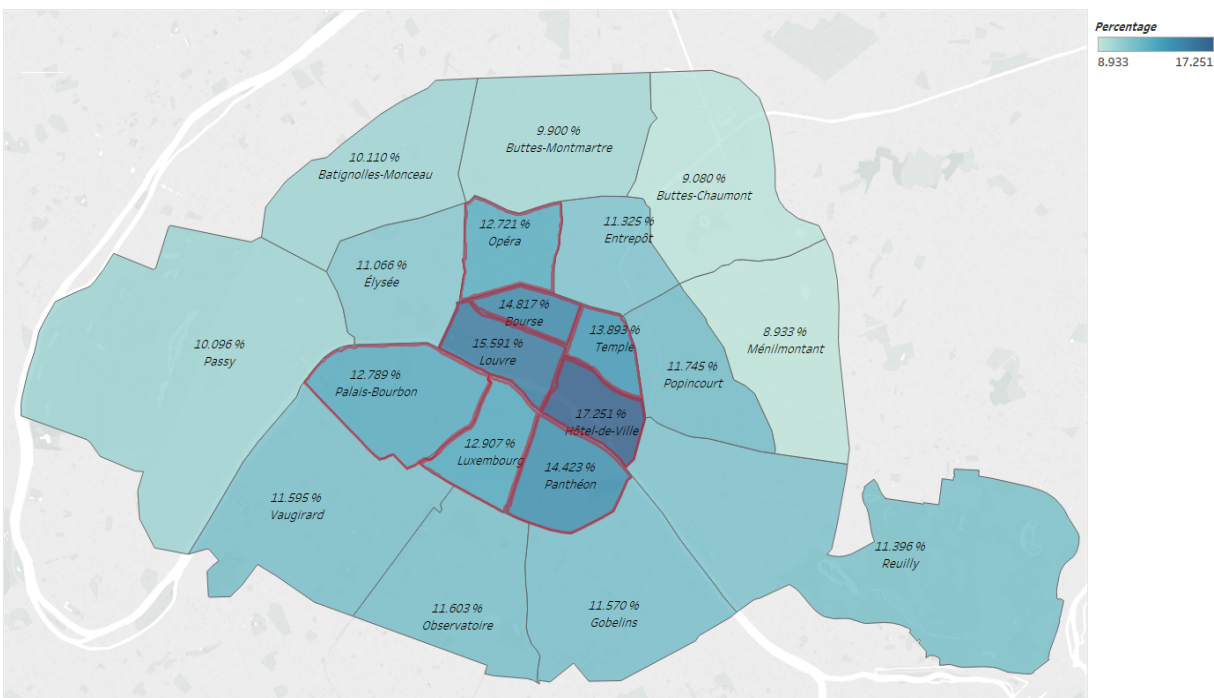
Preattentive visual properties used:

- Used shape, numbers and length that does not distract attention and provides the most information in all the Graphs.
- Color-blind friendly palettes and differences in colors are clearly visible in Graph 1, Graph 4 and Graph 5.
- Differences in colors are clearly visible and we also used hue to differentiate categories in all the graphs.

Data Analysis:

We first explored the distribution of Superhost listings from a geographic perspective. By calculating the ratio of listings that belongs to Superhosts to the total listings count, we discovered a few neighborhoods with higher percentage of superhost listings. The neighbourhoods marked in red in the below map shows the discovered **top 8 neighborhoods**. The list in descending order of percentage is: **Hôtel-de-Ville, Louvre, Bourse, Panthéon, Temple, Luxembourg, Palais-Bourbon and Opéra**.

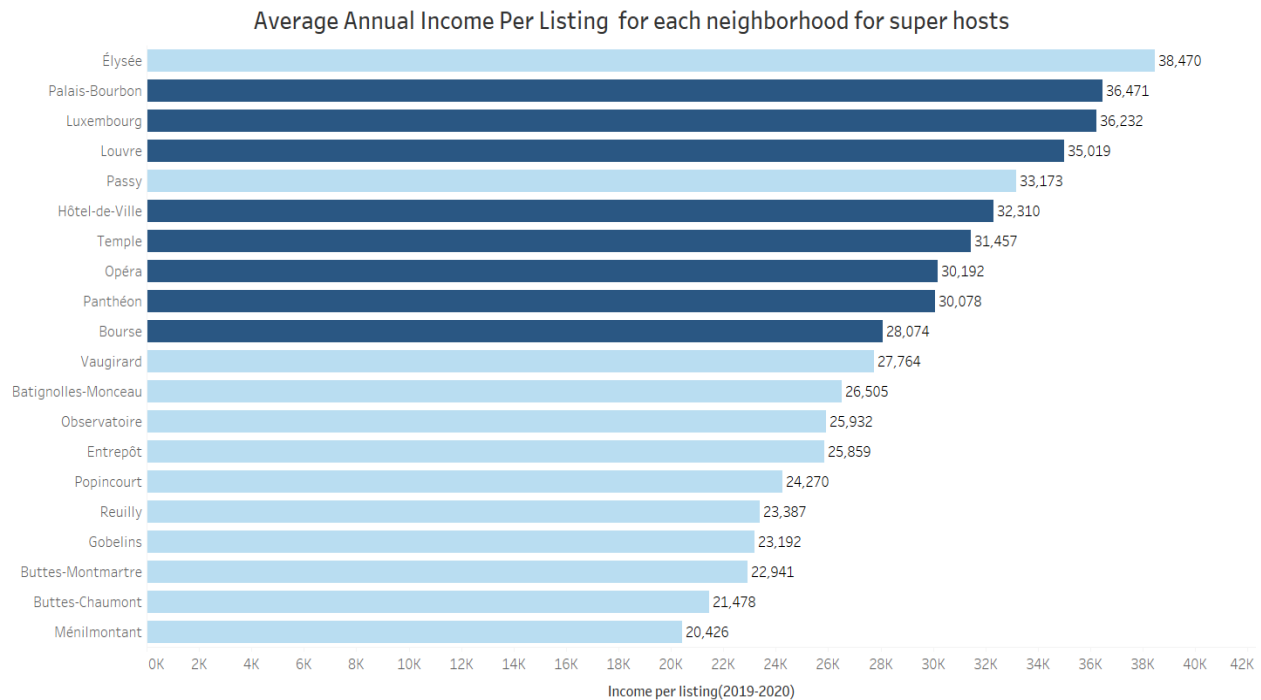
Geographical distribution of the percentage of Number of Superhost listings to total Number of Listings



Graph 1

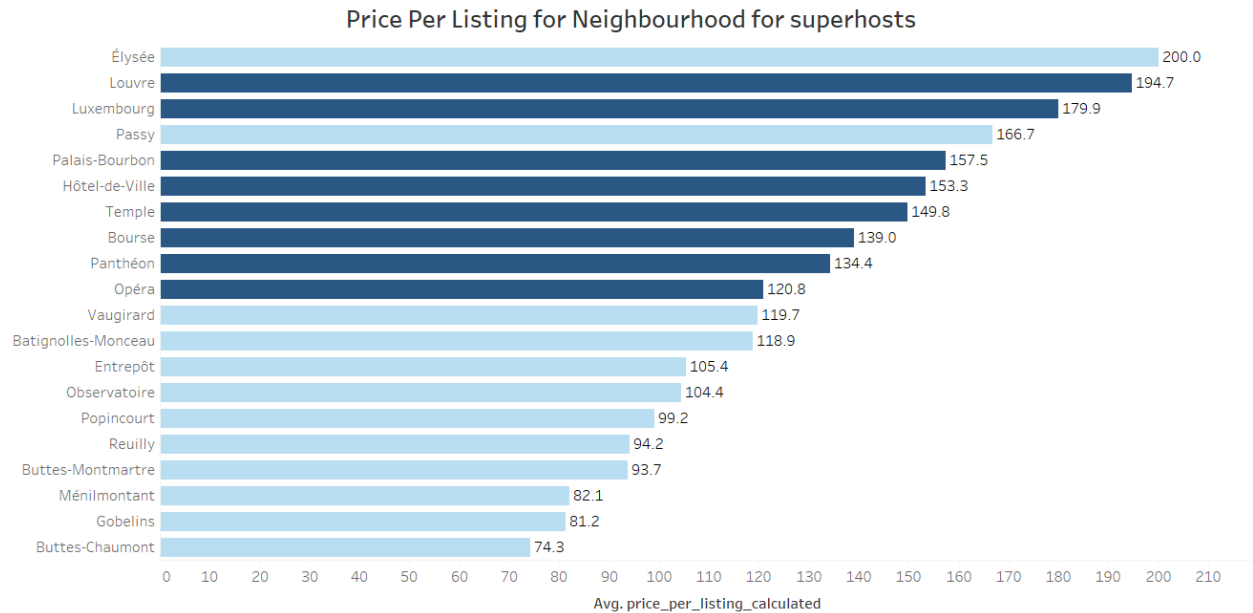
Because we were taking a revenue perspective, we want to know if these neighbourhoods with a higher percentage of Superhost listings also contributes more to our revenue. In the absence of actual booking data, we used the availability information in the calendar table to find the occupancy for each listing. If the listing shows unavailability in the calendar, we consider it as booked.

We multiplied the occupancy with the price and calculated the average annual income of the Superhosts by neighbourhood in dollars. Shown in Graph 2. We used a darker hue to distinguish the top 8 neighborhoods in the graph.



Graph 2

We want to understand whether the higher average income comes from a higher listing price or higher booking rate. Without the actual booking information, we plotted the average listing price by neighbourhood in Graph 3.



Graph 3

Graph 3 confirms that the neighbourhoods with higher percentage of Superhosts also have a higher average listing price and income.

We found 2 exceptions: Élysée and Passy. After some research, we realized that Passy is one of the wealthiest neighbourhoods in the city. From graph 3 we also found that it has a higher average listing price. The high price generates more income. We also explored more about Élysée in later paragraphs.

Because we got a similar list of neighbourhoods with both higher concentration of Superhost listings and higher average Superhost income, we wanted to explore if location is a determining factor. We considered 2 possibilities:

1. Some Superhosts made multiple listings in the same neighborhood, causing the ratio to increase.
2. Environments of these neighborhoods contribute to having more Superhosts.

For the first possibility, we found the distinct host IDs from the listings data and calculated the percentage of Superhosts to total hosts. The top 9 neighbourhoods with a higher percentage of Superhosts is marked black in the below map:

percentage

8.36 20.08

District	Percentage
Reuilly	10.84 %
Gobelins	10.79 %
Observatoire	11.31 %
Vaugirard	11.39 %
Palais-Bourbon	13.88 %
Luxembourg	14.14 %
Panthéon	13.65 %
Opéra	12.60 %
Élysée	12.79 %
Bourse	16.45 %
Louvre	16.32 %
Temple	16.53 %
Hôtel-de-Ville	20.08 %
Popincourt	11.61 %
Ménilmontant	8.36 %
Entrepôt	11.24 %
Buttes-Montmartre	9.97 %
Buttes-Chaumont	8.91 %
Batignolles-Monceau	9.36 %
Passy	10.47 %

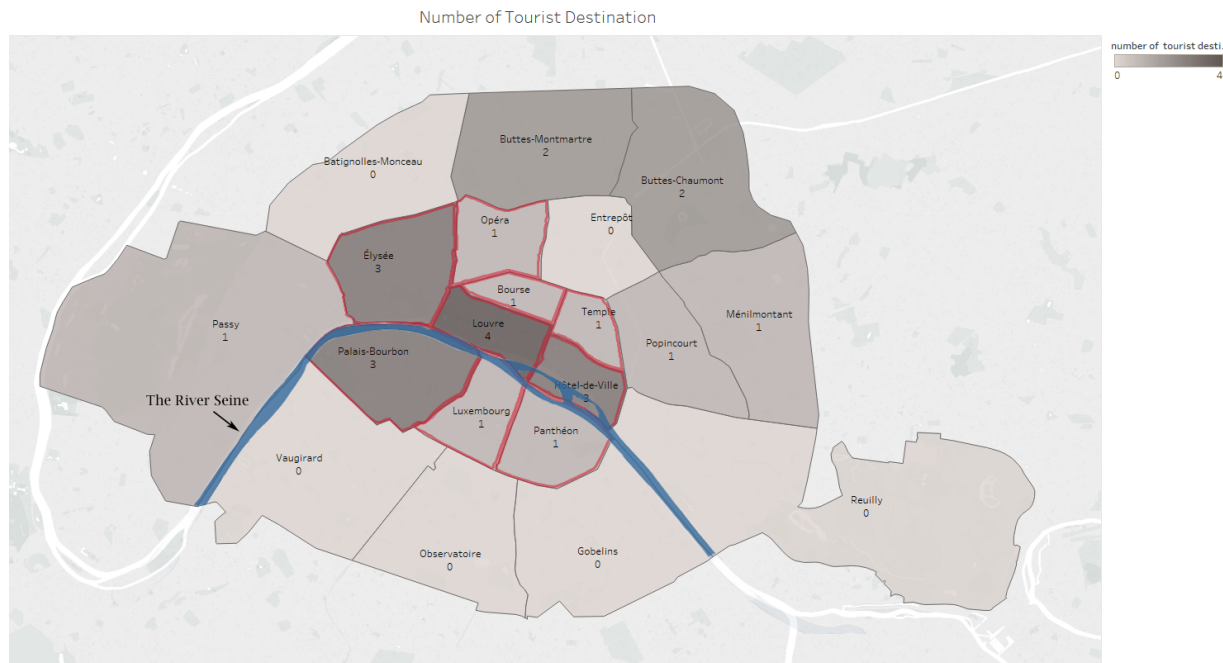
Graph 4

The top 8 neighbourhoods with the highest percentage of Superhosts are similar to the top 8 neighborhoods with the highest percentage of Superhost listings, with one exception of Élysée. This means that Élysée has a high percentage of Superhosts, just not as many multiple listings from the same host. The new finding could explain why Élysée shared the same high average income with the other neighborhoods in the list of Top 8 neighborhoods with higher percentage of Superhosts.

Because the list of neighbourhoods with higher percentage of Superhosts did not change much after we removed the multiple listings from the same Superhosts, we ruled out the possibility that the higher percentage of Superhost listings is due to multiple listings from the same hosts. Then, we continued exploring the second possibility that the environment of these neighbourhoods affects percentage of Superhosts.

We first explored whether there's better public transportation in the top 8 neighborhoods with higher ratio of Superhosts. We checked the Paris bus/metro station map and found that the Paris city has a well-developed public transportation system. All neighbourhoods have multiple bus/metro lines and there was no major difference in commute among neighbourhoods.

Next, we explored whether there were more tourist attractions in the top neighbourhoods that we were interested in. We gathered the top 27 tourist attractions in Paris from travel websites such as: [planetware](#) Trip Advisor and the [parispass](#). We plotted the number of tourist attractions by neighbourhoods and got the graph below. The top 9 neighbourhoods with a higher percentage of Superhosts are marked red:



Graph 5

Insights

- The neighbourhood Élysée has a high percentage of Superhosts and not as many multiple listings from the same host. Élysée gains a higher income than any other neighbourhood in Paris. This suggests that the company should seek increasing the number of new Superhosts instead of Superhosts lists multiple properties.
- We discovered that the top 9 neighbourhoods with a higher ratio of Superhosts also have a higher number of tourist attractions. After digging deep into the scenery of Paris, we realized that the top 9 neighbourhoods were centralized around the River Seine, which is the center of the vintage Paris view.

Key Challenge

The data was collected by a third-party site that has no connection with Airbnb and its competitors. We do not have official documentations on the variables, so we had to infer.

Conclusion:

From this study, we could find that the superhosts are concentrated more in certain neighborhoods due to 2 main reasons:

1. The location has many tourist attractions as well as nice scenery. The presence of the River Seine added to the scenic views surrounding the neighborhoods.
2. Income is higher for Superhosts in these regions.

Airbnb can increase their presence in similar neighbourhoods to increase their revenue.