**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.

For this study, we want to gain insights on how to increase revenue by improving the multiple aspects of host services in Paris, France. We consider the three stages of being an Airbnb host: normal host, almost superhost and superhost.

Superhost is a status that’s recognized by Airbnb as ‘experienced hosts who provide a shining example for other hosts, and extraordinary experiences for their guests(<https://www.airbnb.com/help/article/828/what-is-a-superhost>)’. According to Scott Shatford’s analysis (<https://www.airdna.co/blog/airbnb_superhost_status>) from AirDNA, Superhosts ‘globally earn 60% more revenue per available day’. By analyzing the multiple aspects of superhosts, we want to build a growth path for normal hosts to improve their services and become superhosts, hence increasing  Airbnb’s revenue.

The path to become superhosts can take effort. We want to provide striving normal hosts with the experience offered by the superhosts as an example of services. While exploring the dataset, we also found a large portion of normal hosts who have been hosting before 2019 that satisfy the review score ratings, host response rate and number of stays requirements for superhosts, but were not rewarded the status. We consider this group as potential superhosts and want to study more about them.

By studying these three groups of hosts, we want to gain insights into Airbnb’s superhost program and identify the driving factors that would improve the business revenue.

**Data :**

We have considered the Paris dataset from insideairbnb.com, which is a third party website collecting Airbnb data from publicly available sources. Location information for listings are anonymized by Airbnb, which is not in the exact coordinate of the actual building. It consists of the following tables:

**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. The following variables were used in our study:

Discrete Variables: Id

Categorical variables:amenities,  room\_type property\_type,host\_is\_superhost,host\_identity\_verified

Continuous variables: price, latitude,longitude,neighborhood\_cleansed, review\_scores\_rating, ~~review\_scores \_accuracy, review\_scores\_communcation, review\_scores\_location, review\_scores\_value, review\_socres\_cleanliness, review\_scores\_checkin,~~ host response rate, number\_of\_reviews.

**Neighbourhood:**

The neighbourhood data is a geojson file that exhibits different neighbourhoods of Paris.

**Calendar:**

The calendar dataset contains the availability of individual listings on a daily basis from February 2019 to January 2020. The following variables were used in our study:

Discrete Variables: listing id

Categorical variables: available

**Reviews:**

The reviews dataset contain the reviewers and their comments for each listing.

The following variables were used in our study:

Discrete Variables :listing\_id

Text Variables: comments

**Design principles** **and** **Perceptual Properties:**

(Data Visualization is a graphical display of abstract information. The aim in our project is to transfer abstract into physical attributes of vision effectively. Following  principles and perceptual properties are used in our graph):

Design Principles used:

* Gestalt laws: closure, symmetry, proximity and connectedness
* maximize data-ink ratio by using tooltip instead of legend
* minimize chart junk by deleting grid lines and using white background for the graphs
* maximize data density by using stacked bars
* layering and separation with different colors
* utilized multi-functioning elements

Preattentive visual properties:

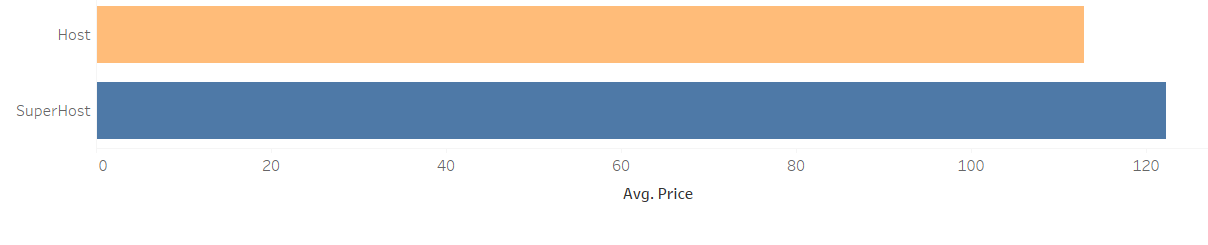
* Used shape, texture, length and numbers that does not distract attention and provides the most information
* colors with low saturation to avoid strong visual interference
* color-blind friendly palettes
* hues to differentiate categories
* differences in colors are clearly visible
* blue for large regions

--------------------------------------------------------------------------------------------------------------------------------------------

**Data Analysis and insights:**

**1. Is there a difference in price between normal hosts and superhosts?**

**Graph:**



**Data Analysis:**

We used the price for each listing to calculate the average price and grouped this by host status from the host\_is\_superhost column. Null values are excluded. We did not consider listings that have price = 0.By comparing the average price between superhosts and normal hosts, we found that the average listing price for superhosts is 8.4% higher than that of the normal hosts.

**Insights:**

The super hosts can have a higher listing price compared to normal hosts.

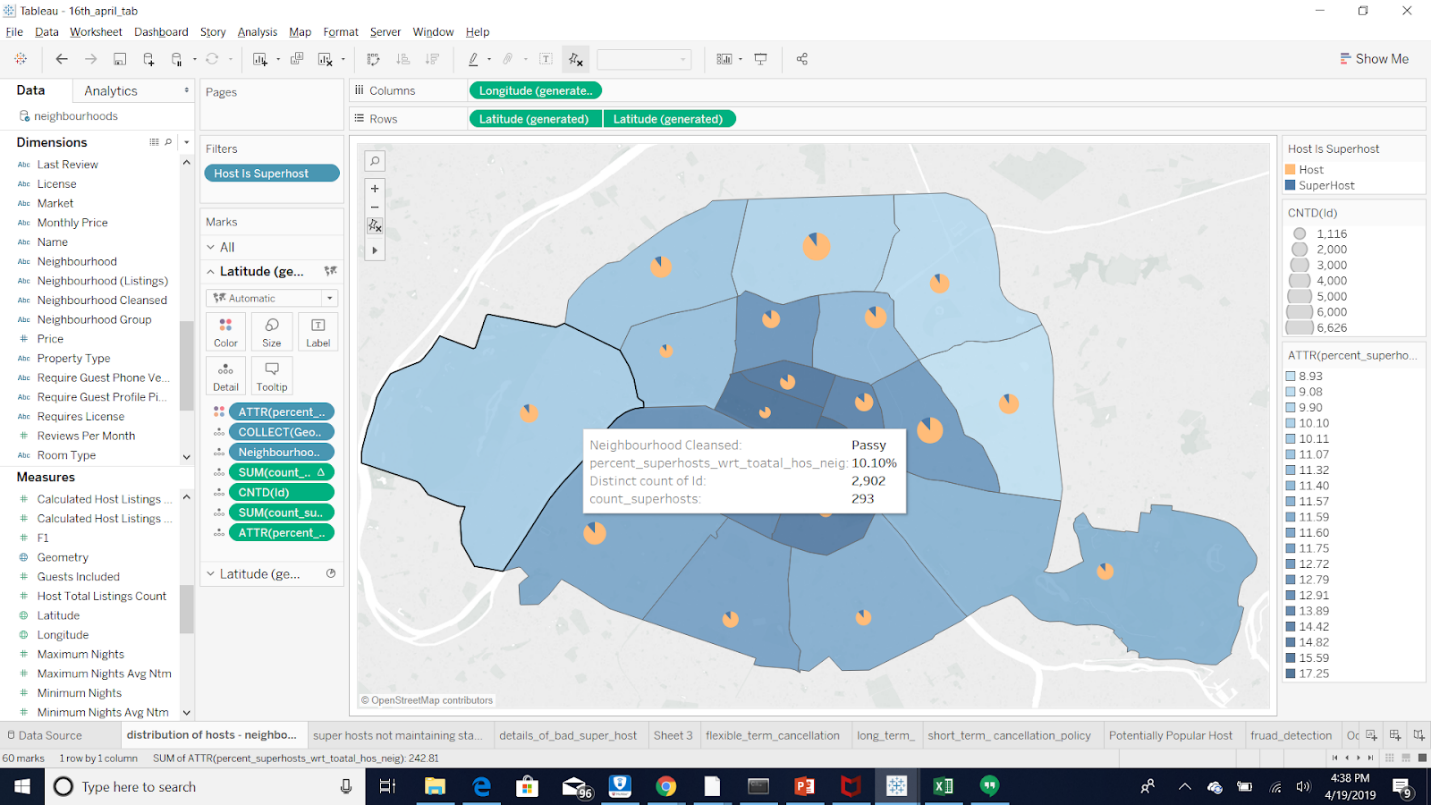
**2. How are the superhosts distributed geographically? Are there any neighborhoods with higher proportions of superhosts?**

**Data Analysis:**

We could find that there is a variation in the number of listings among the neighborhoods. We  analyzed the proportion of superhosts against total hosts for all the neighborhoods. We also found that there is a variation in the average price and the number of amenities listed between superhosts and normal hosts. We explored this variation in individual neighborhoods.

Percentage of superhosts to total hosts is varying in each neighbourhood, there is about 1 to 10 percent variation amongst the neighborhoods. And the average price for the super hosts is higher compared to the normal hosts in all the neighborhoods

**Graph:**

****

**Insights:**

We found that the neighborhoods with the highest proportions of superhosts coincide with the top 5 tourist destinations in Paris. ~~(~~This indicates that Airbnb has a strong presence in the areas with the largest income in tourism. )

(include 1 more)

Hotel-de-Ville has highest proportion of superhosts.

Menilmontant has the least ratio of hosts converted to superhosts

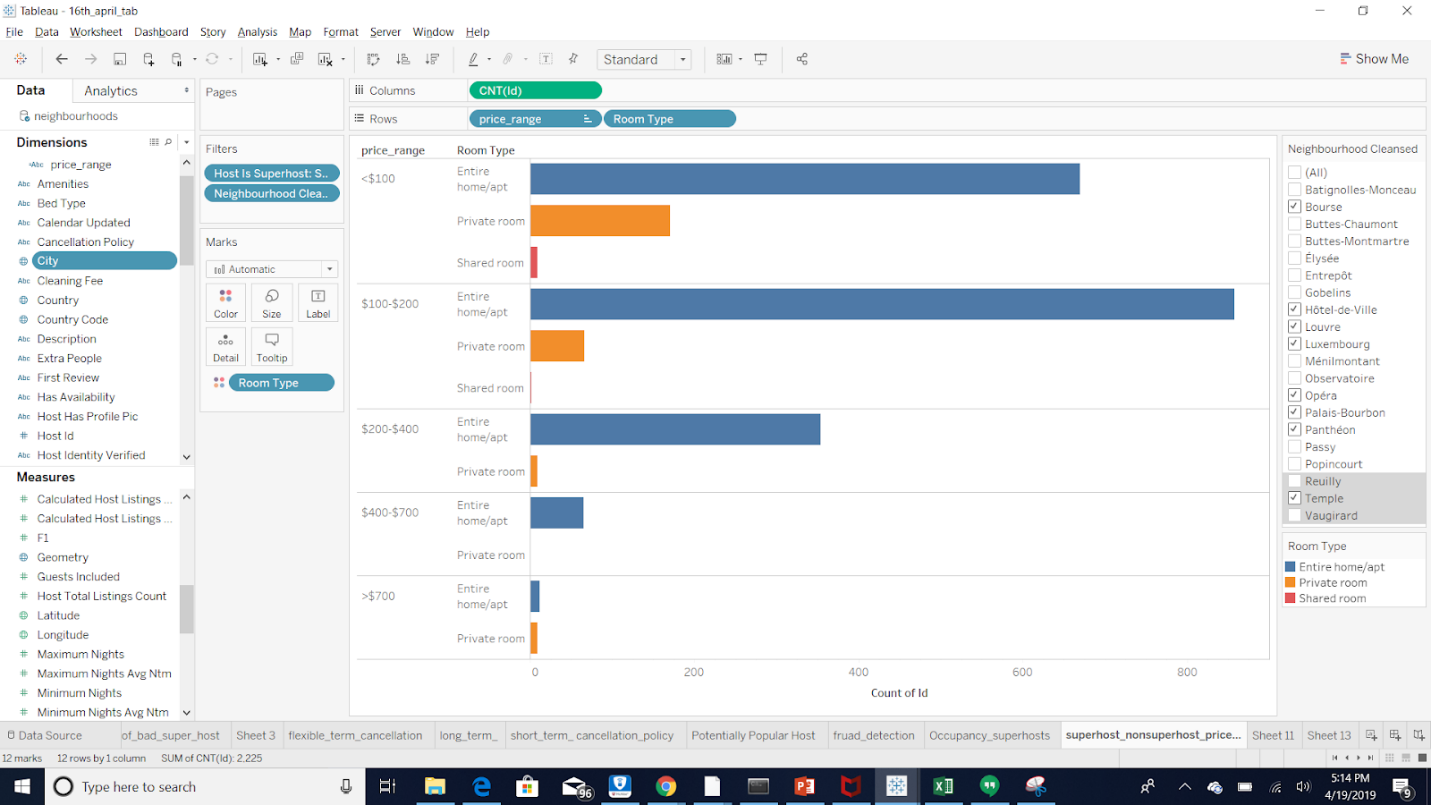
As the area of neighbourhood increases less number of hosts are converted to superhosts.

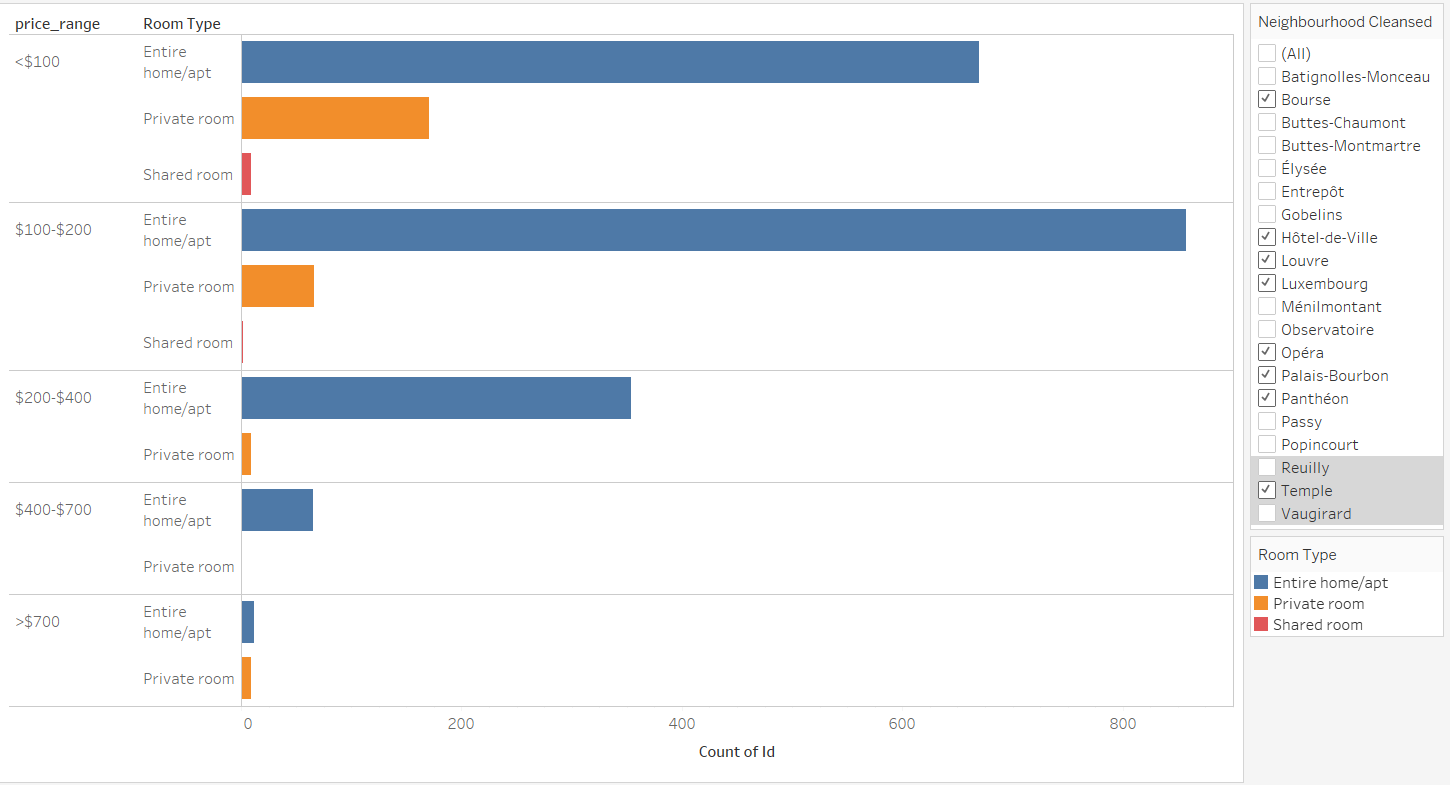
Most superhost ratio we can see in the interior regions of Paris

3. Given a superhost in a specific neighborhood and with a specific room type, can we come up with a common price range as reference?

**Data Analysis:** Having found that the room types play a major role in the listing price, we explored the price ranges for different room types followed by the super hosts.

**Graph:**

****

****

**Insights:**  
Surprisingly, entire home/apartments are listed more than private rooms and shared rooms among all price ranges. While the number of listings for private rooms and shared room decline naturally with increase in prices before reaching $700, entire home/apt reached its peak at the $100-$200 bucket.

We also found that private rooms reached the lowest in the $400-$700 bucket, but bounced back when the price is over $700. (This may be an indication of guest behaviour that when choosing among luxurious stays, the room type does not matter as much. )

4. Given a superhost would like to host either short term or long term stays, how should they decide on their cancellation policy? Can we give them some suggestions?

**Data Analysis:**

We analyzed that there is a variation in occupancy due to the cancellation policy used by superhosts. We defined the short term and long term stays based on a 4 week period. Listings with a maximum nights requirement of 27 were considered short term and that with a minimum nights requirement of 28 were considered long term . Flexible term refers to the listings that can host for both short term and long term stays.

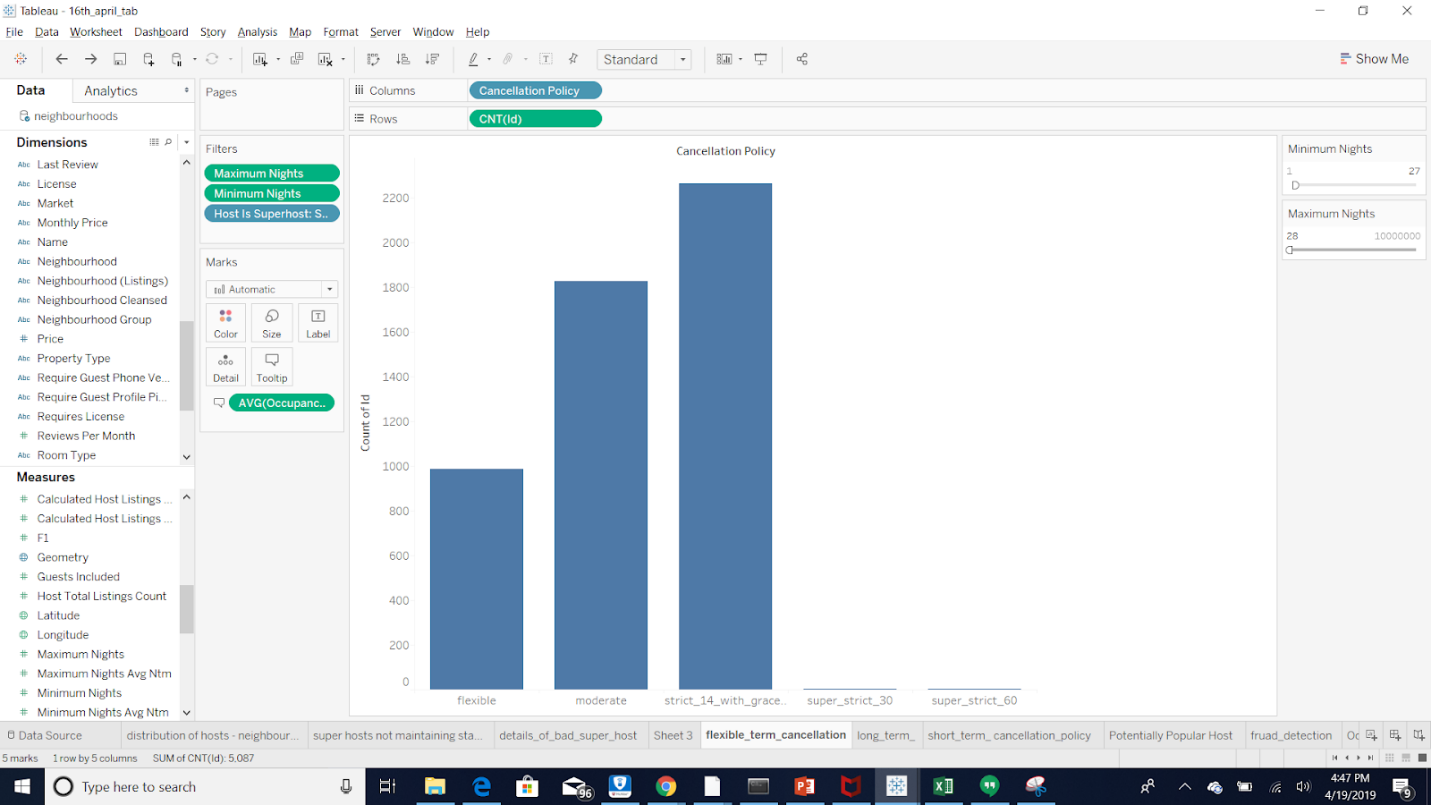
**Insights:**

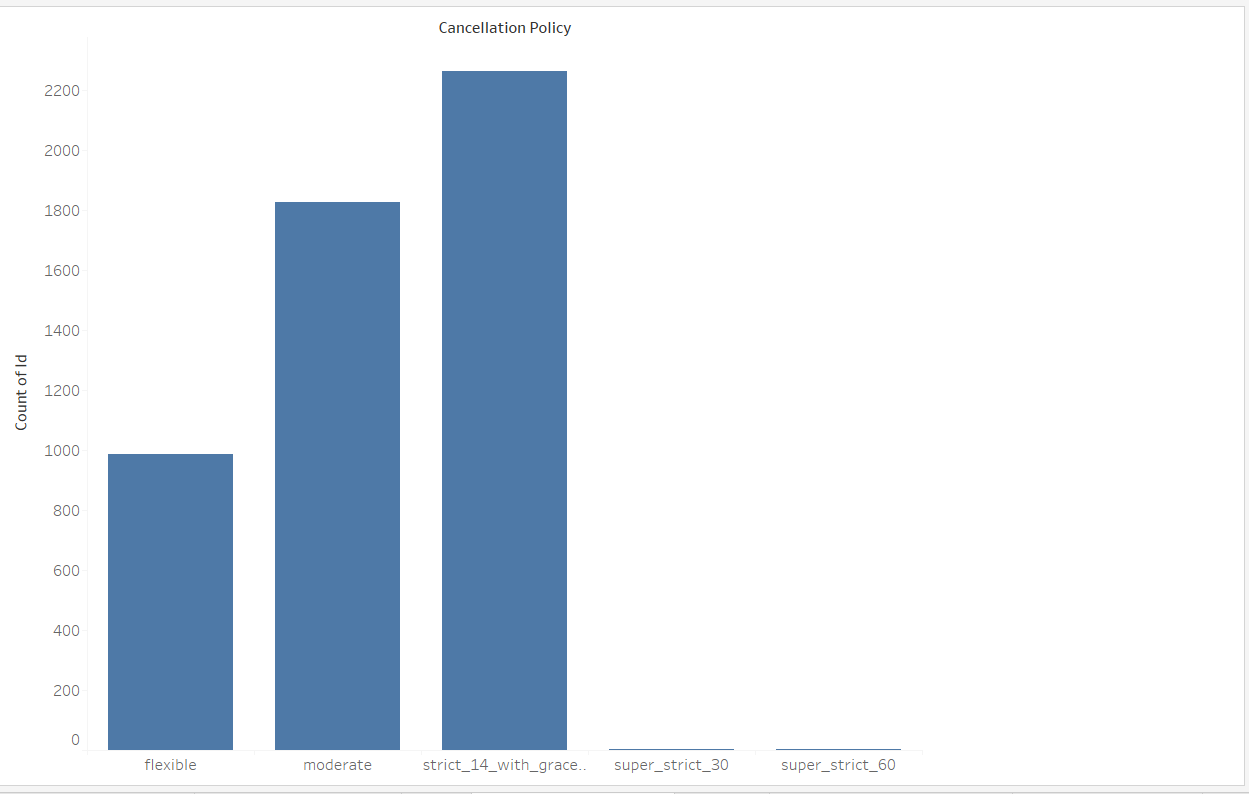
~~We have observed that superhosts with short term stay listings have majorly used moderate cancellation policy , so we can suggest the normal hosts or to change their existing policy as per superhost or new potential host to use the cancellation policy as per the superhost.   Long Term cancellation policy is more~~

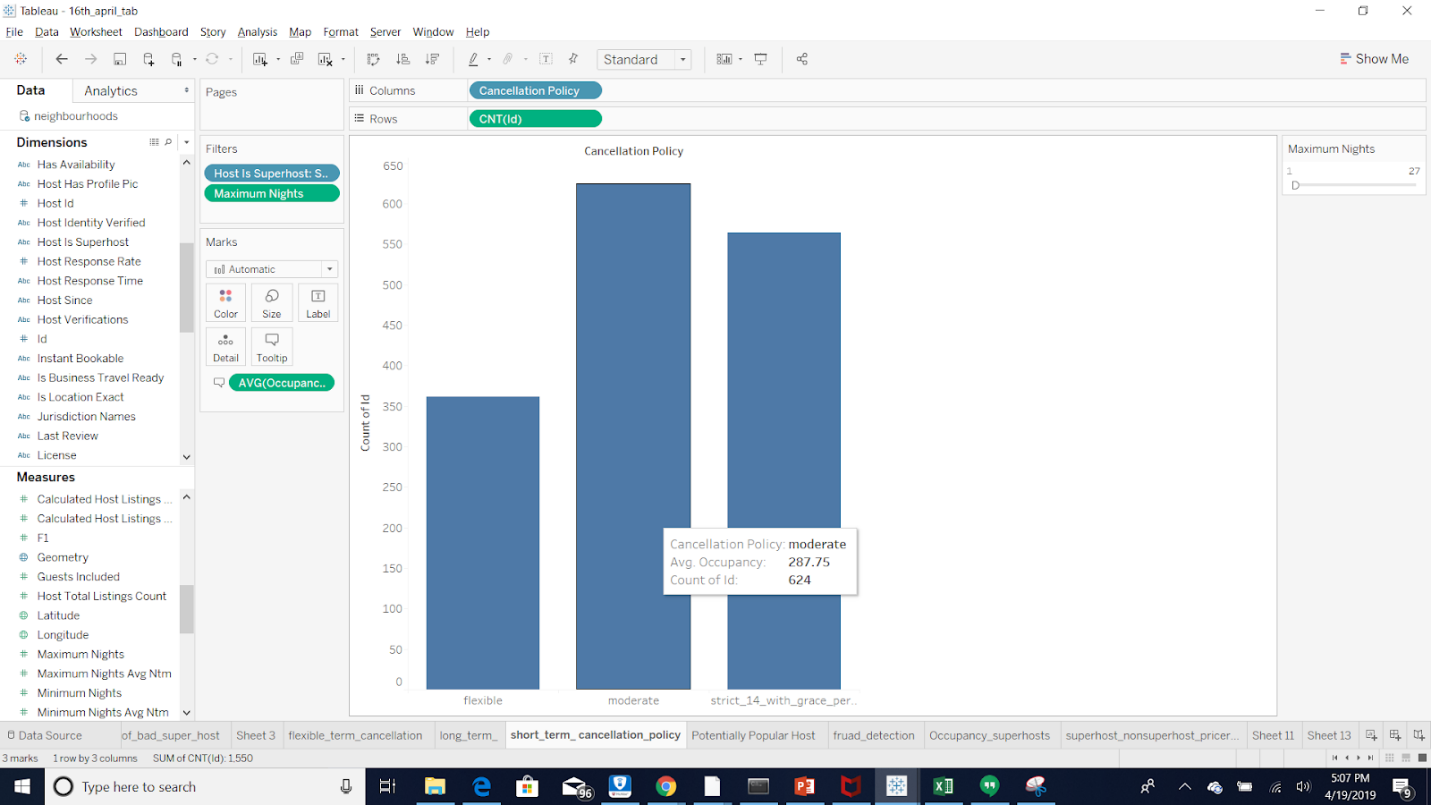
This exhibits different business behaviors for the 2 types of hosts. For the superhosts with a long term stay listings, they prefer well planned schedule with minimum last minute changes. And for the ones with short term stay listings, they are more flexible and adaptive to changes.

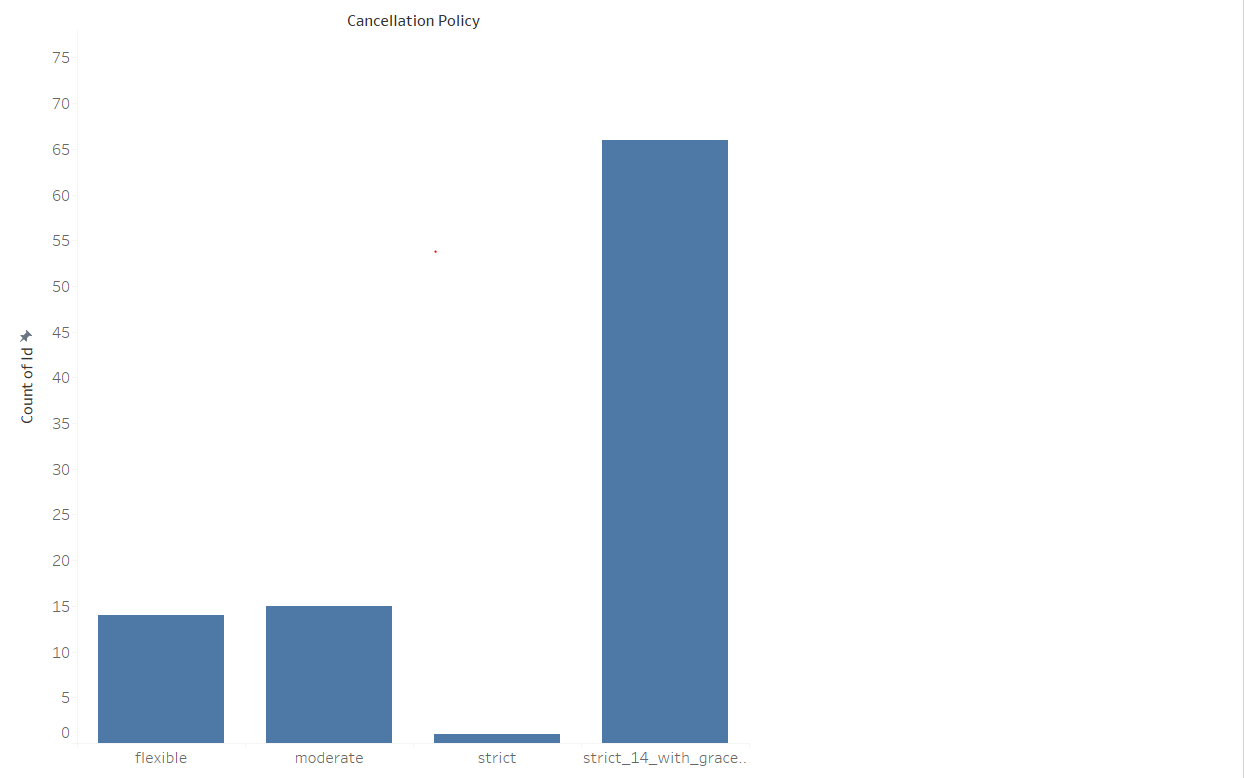
Also the flexible term has the highest number of listings.

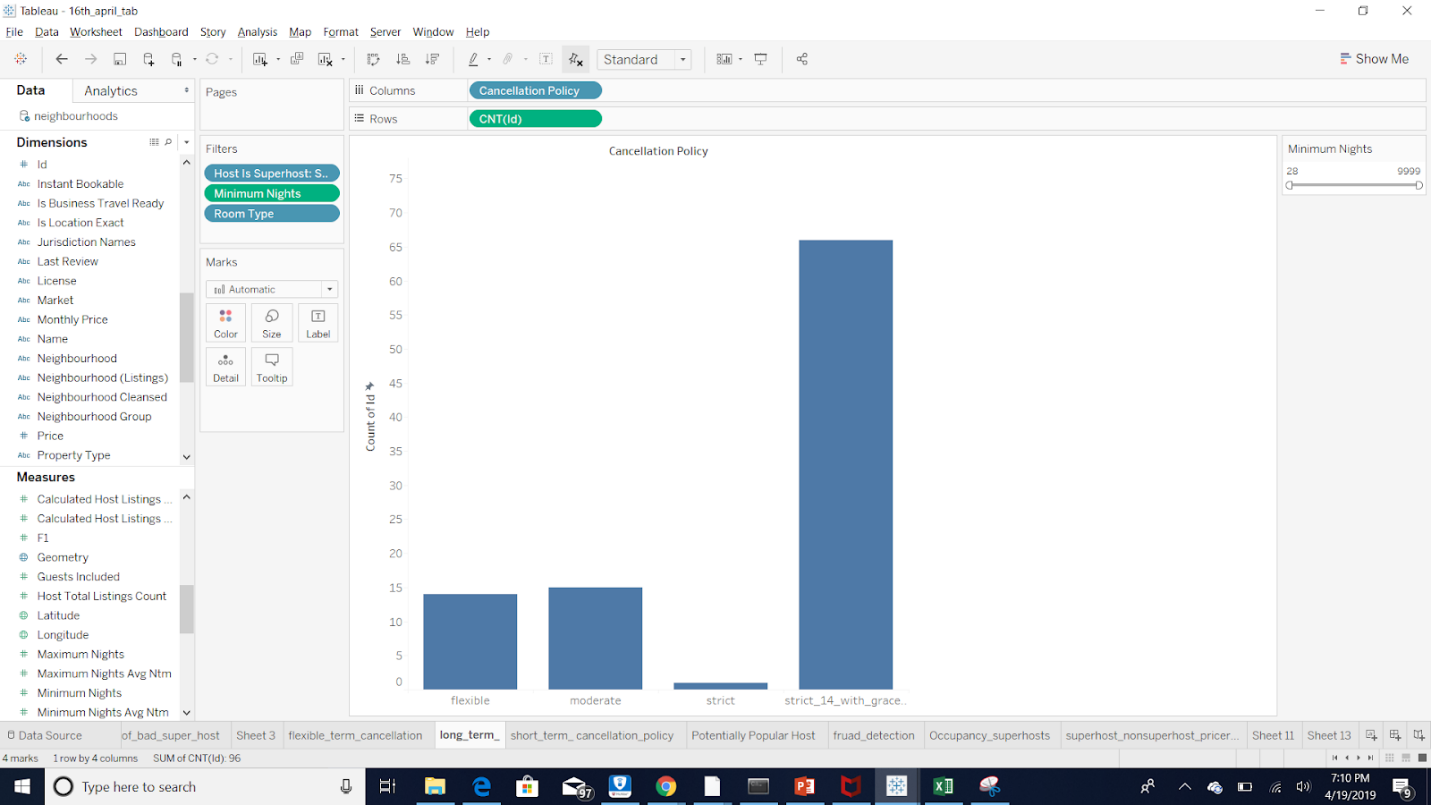
**Graph:**

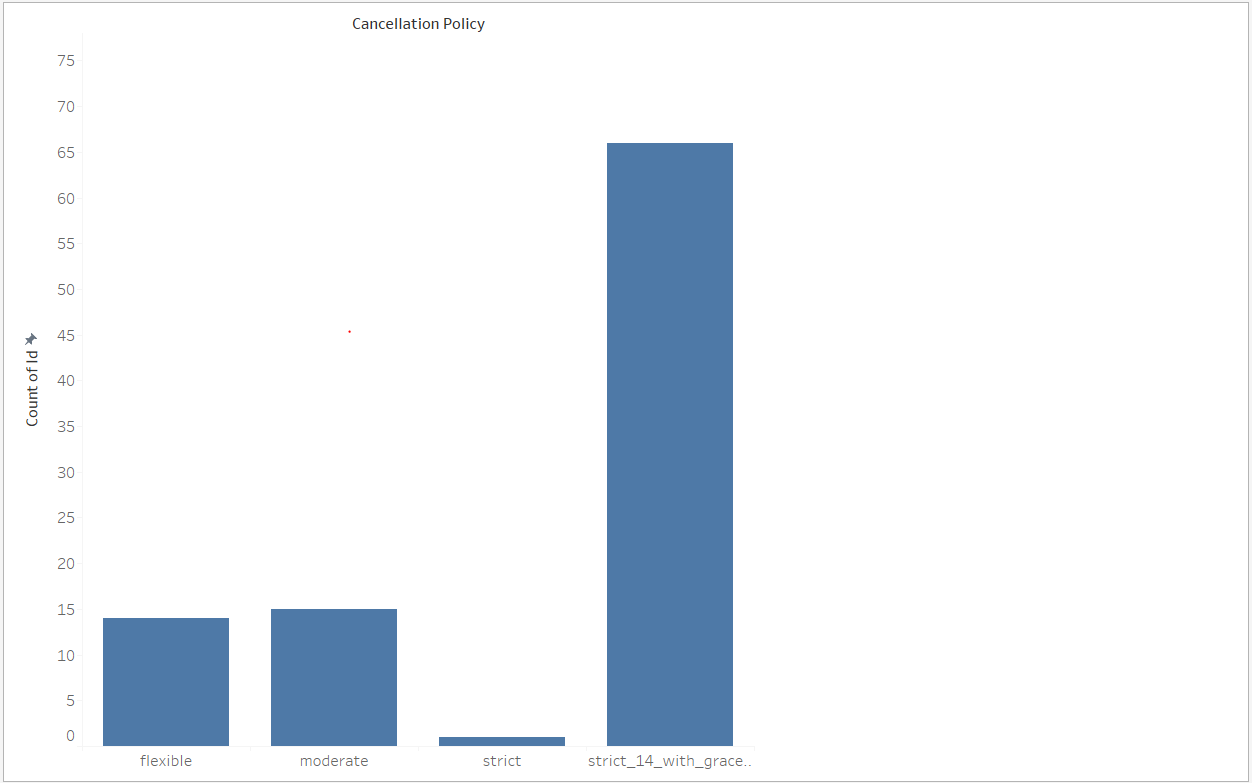
****

****

****

****

****

****

5. What are the most praised features of superhosts?

**Data Analysis:**

Considering listings of past one year, we filtered records where review\_score\_rating is 100 and host is superhost, generated the word cloud for their reviews to show the most used words, from which we can infer what makes guests more satisfactory about their stay.

**Graph:**



**Insights:**

Through the graph, we immediately notice that a lot of French words were frequently mentioned, it is not surprising that most French people rent the room in Paris. Words “bien situé” means good location, “très bien” means great. Obviously, the word “great location” mentioned most frequently. Hence, the location of property is likely one of the most important considerations of Guest.

6. How many normal hosts are close to become superhosts? What should they improve to gain this status?

**Data Analysis:**

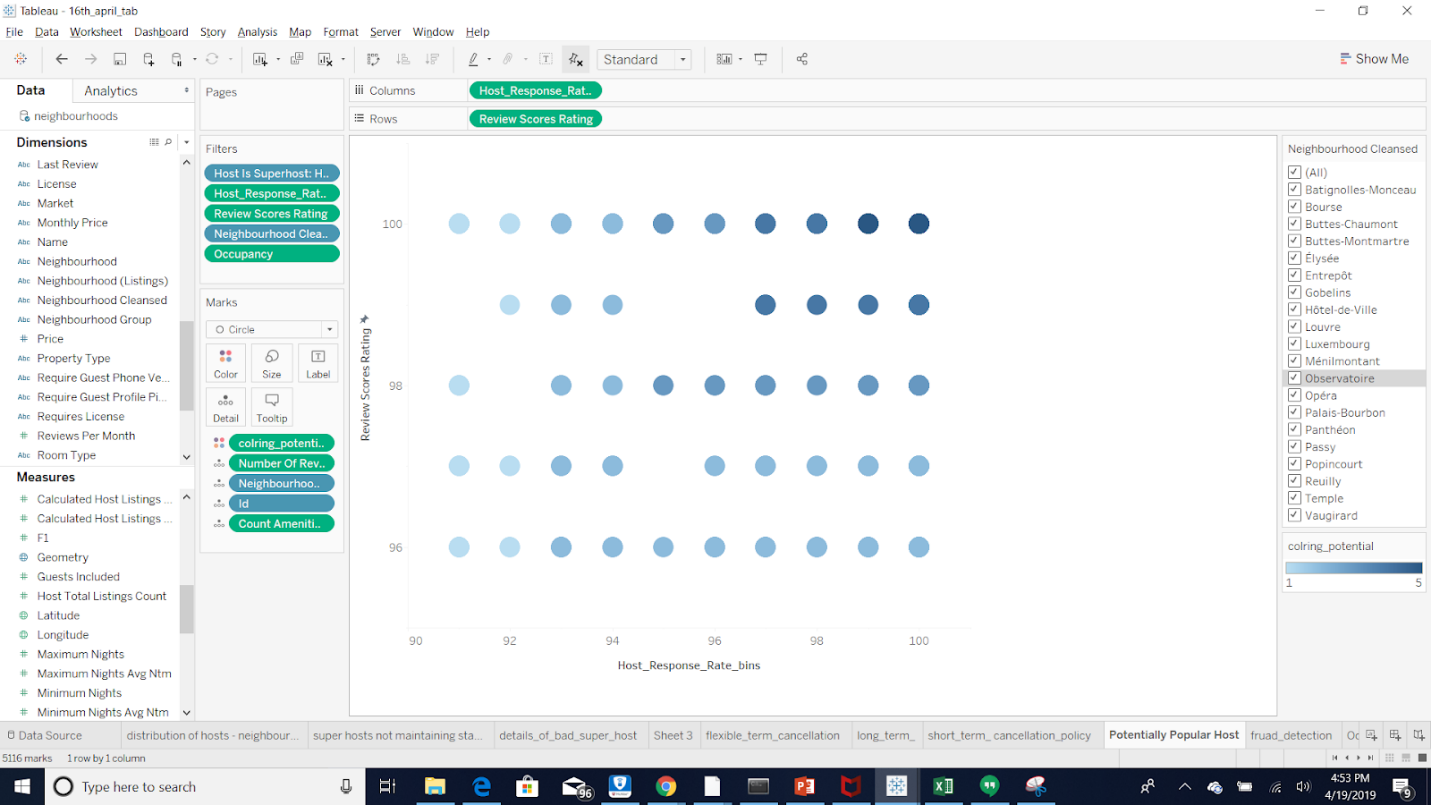
One of the prerequisites for superhosts is a 90% response rate. Which means that they have to respond to 90% of new messages within 24 hours. The second prerequisite is the overall rating based on reviews should be greater than 4.8.  The third prerequisite is having over 10 stays. We inferred the number of stays from the occupancy column in calendar table. This left us with the cancellations, which is unavailable from the dataset. So we assumed that these hosts with qualified ratings and number of stays were not recognized as superhosts because they have some cancellations. To avoid counting the new hosts who simply haven’t been checked for the status, we limited our study to hosts who have been hosting before the end of 2018.

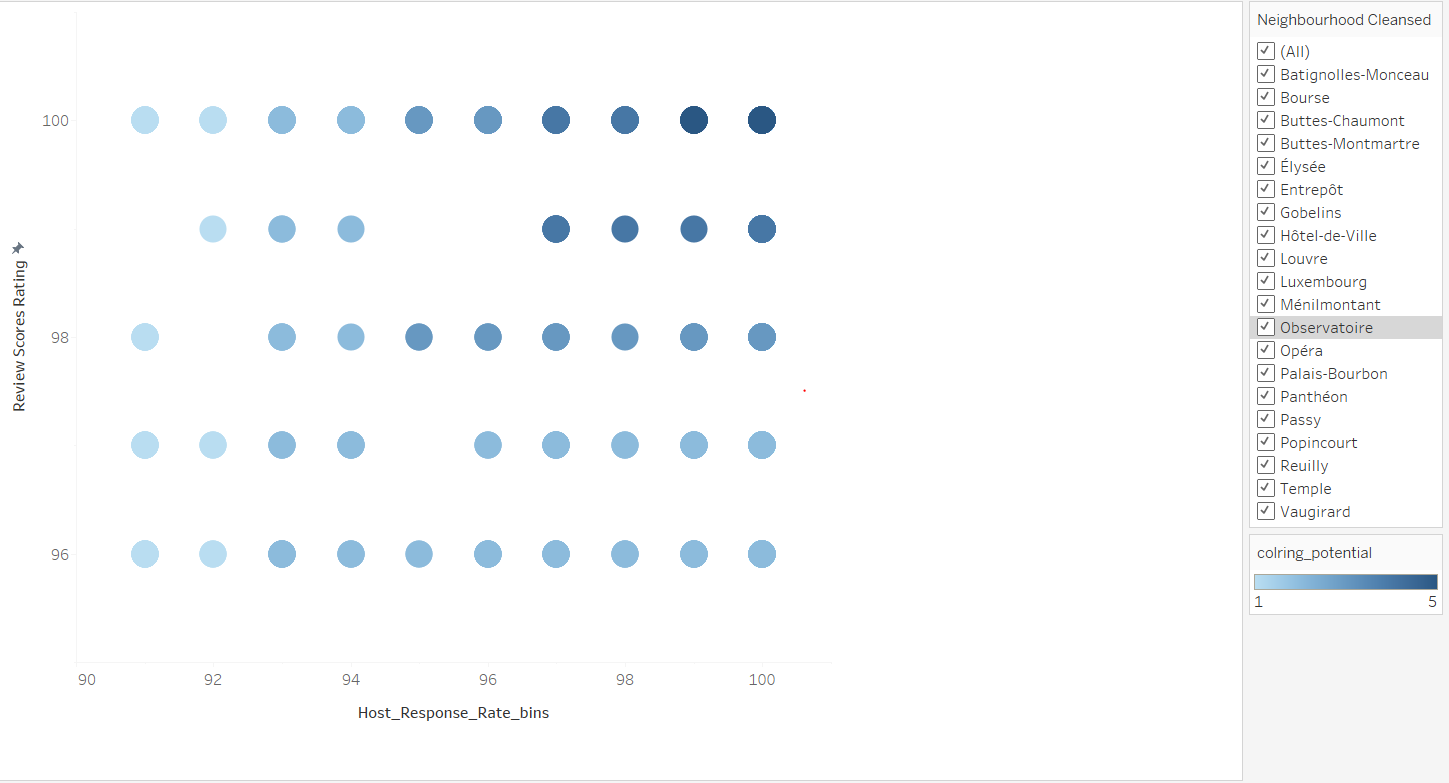
**Insights:**

~~From this graph, we can tell that many hosts with good potential are suffering from this policy. Airbnb should look deeper into the cancellations and check whether there are any extenuating circumstances. Implementing this type of strategy will increase the superhost count, which increases the revenue.~~

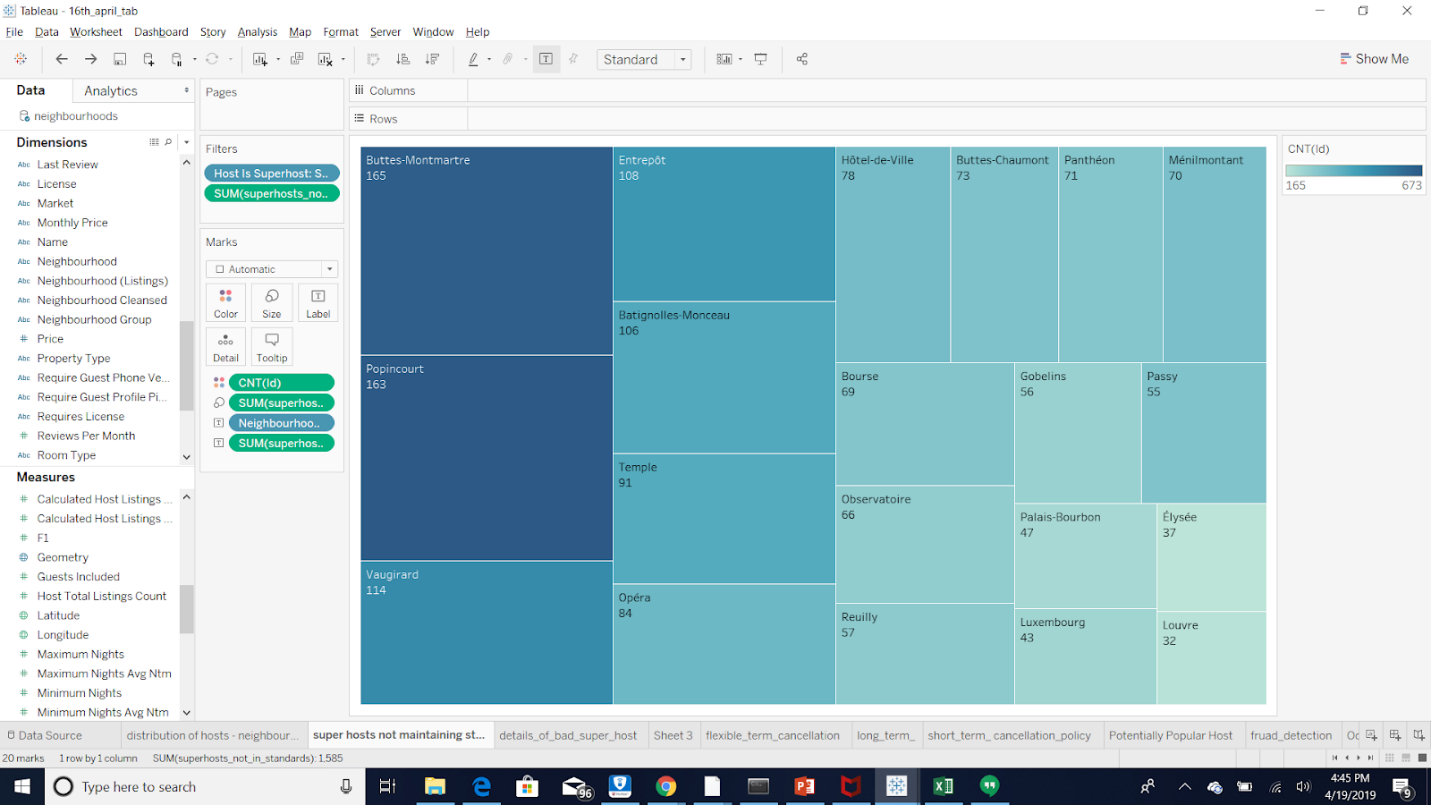
This tells us the company should inform their superhost on taking cautions when cancelling a booking. We could suggest adding an alert when a host clicks on the cancellation button.

**Graph:**

****

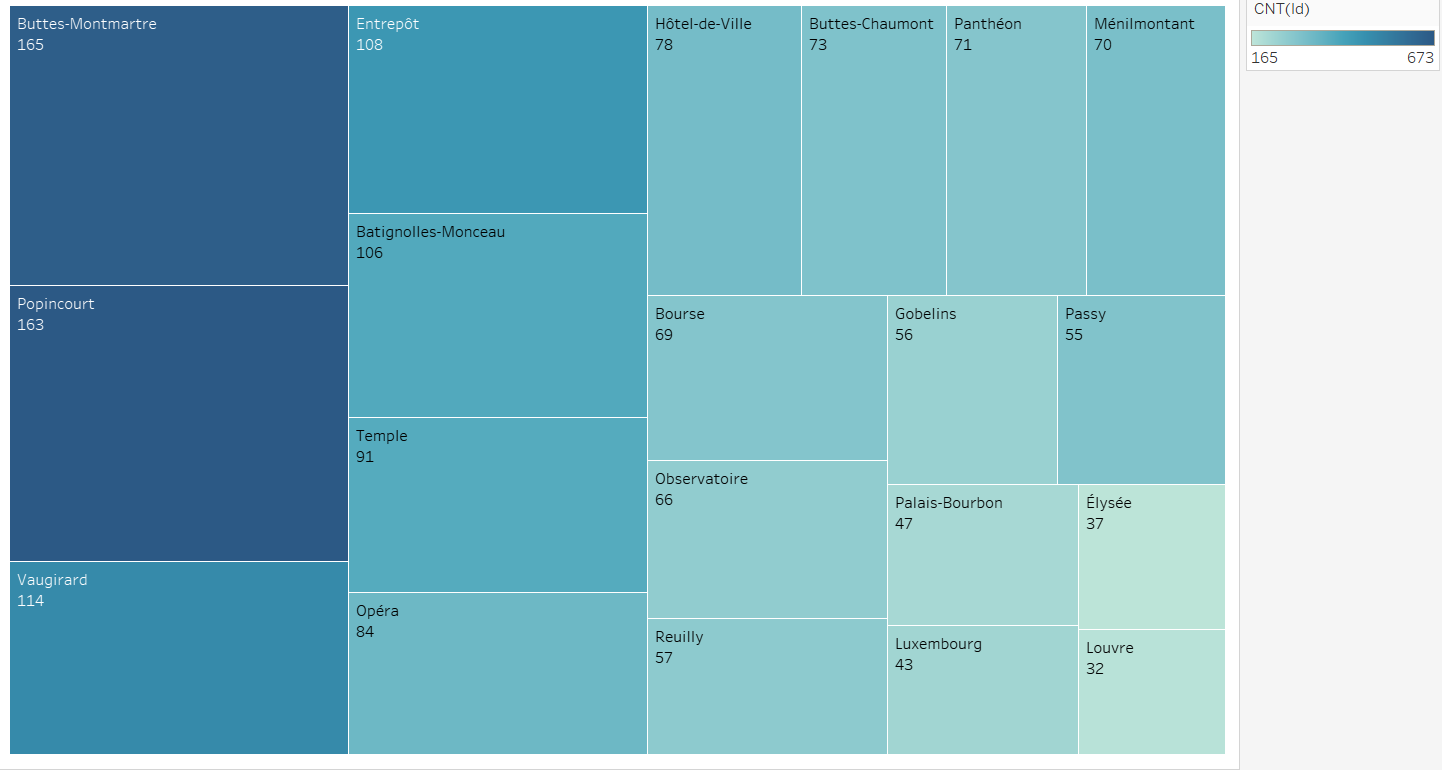
****

7. Is there anything we should be concerned about the superhost community?



Analysis:

As per the prerequisites specified in the Analysis-6 for SuperHost, we have observed that many superhosts are losing their status in between the term of verification and becoming burden to Airbnb. We want to analyze them further with their neighbourhoods and other driving features.



Insights:

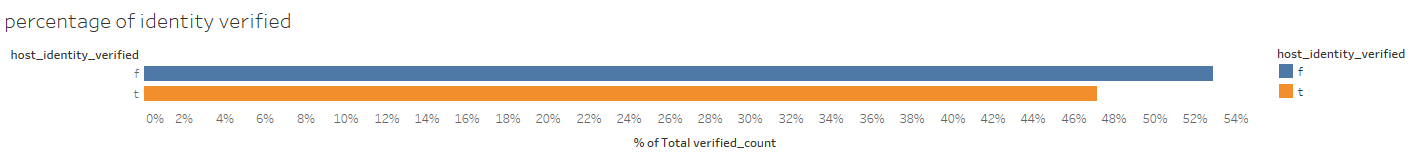
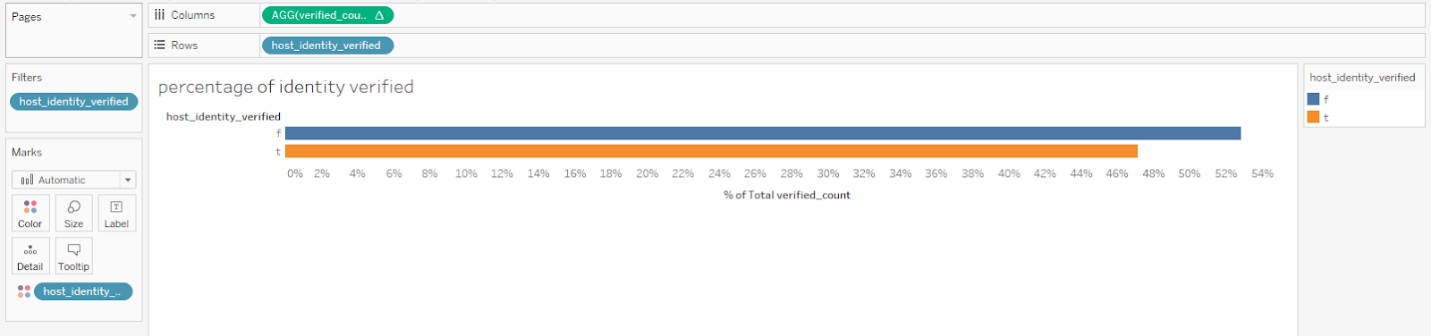
We can observe that the top neighbourhoods as we see in the graph 2 which have high superhost proportion do have very less number of them not meeting the criteria.

The list of superhost not meeting the prerequisites is very large nearly 20% of the superhosts present in that neighbourhood.

**Data Analysis:**

Airbnb is a company built on trust between hosts and guests, so we deem it important for the hosts, especially the superhosts, to be trustworthy. We considered 2 types of behaviors to be of concern for the superhosts: 1. The host does not maintain the requirements for superhosts after they’ve gained the status. 2. The superhost’s identity is not verified. The first situation could potentially damage the reputation of Airbnb’s most prolific community, and the second situation would damage the company’s image irretrievably should an incident occur.

**Graph:**



**Insights:**

Airbnb verified the identity of host to know the host is real person or not. Identity verification is a critical process, it can gain the confidence for the guests and also prevent the fraud. In Paris, more than half (about 53%) of superhost’s identity doesn’t verified.

**8.  How many superhosts retain their status across quarters? What is the growth rate for retaining superhosts compared with total superhosts?** (the superhost policy has a strong hold on its superhosts)

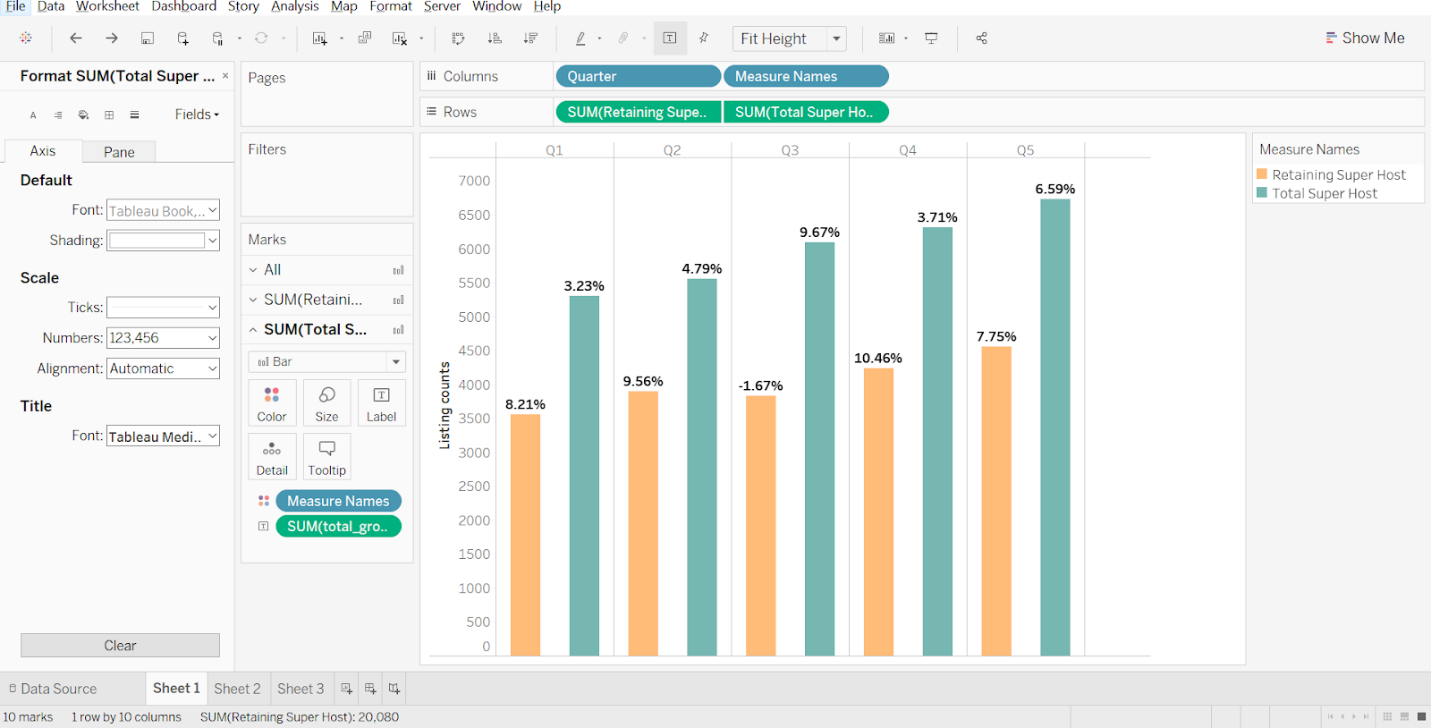
**Data Analysis:**

In this graph, we compared the total number of superhosts to the number of superhosts that has obtained their status for the past 5 quarters to understand the retention  rate for superhosts.

The superhost status is checked and evaluated on the 1st day of January, April, July and October for each year.

Because the data we have were not collected at the 1st day for each months and the status will not change within the same quarter, we chose to use the superhost numbers from February, May, August and November to avoid any duplicates. The numbers show the growth rate for total superhosts and retaining superhosts.

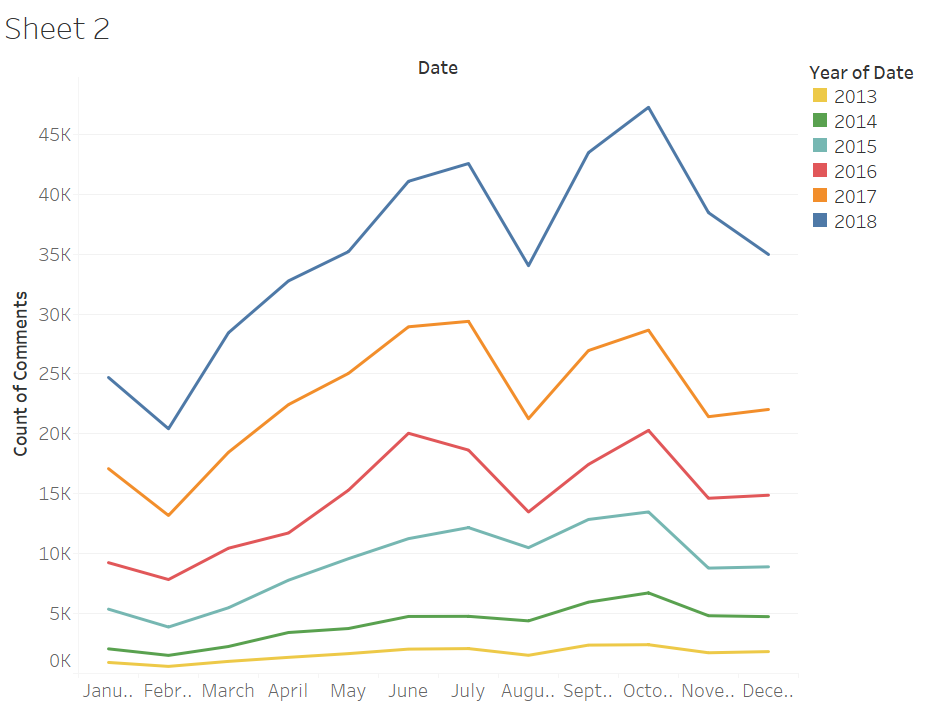
Q1: January to March, 2018, Q2: April to June, 2018, Q3: July to September, 2018, Q4:October to December 2018, Q5: January to February, 2019.

We discovered that although the total number of superhosts is growing steadily for the past five quarters, the retention rate experienced a drop at the third quarter in 2018.

**Insights:**

With further exploration, we realized that the drop in Q3 may be related to the seasonal fluctuation in August shown below:

In this graph, we used the number of reviews as an indication of booking rate. We can see there’s a constant drop in august from 2013 to 2018, indicating a steady behavior pattern.



Other from the third quarter, the retention rate for superhosts exhibits a larger growth rate than total superhosts. This indicates that Airbnb’s policy has a strong hold to its high quality hosts.

**Key Challenge**:

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

**Conclusions:**

From this study, we learned the multiple aspects of the superhost community for Airbnb in Paris,  France. We came up with a portfolio of superhosts which can be of an example for the other hosts.  By analyzing the data and combining it with researches on the business, we obtained several insights on the behavior and business patterns for the driving force of Airbnb’s revenue. Not limiting to the good side of it, we also studied the potential problems in the management of it by Airbnb.

(From the first 4 questions, we analyzed the behavioral patterns for superhost in Paris, France.)

Presentation Video Link: <https://www.youtube.com/watch?v=S8KmqU-B1l8&feature=youtu.be>