**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.

**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. These are the tables we used:

**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.

**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.

The following variables were used in our study:

Discrete Variables: listing id

Categorical variables: available

**Reviews:**

The reviews dataset contains 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:**

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

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:

* 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

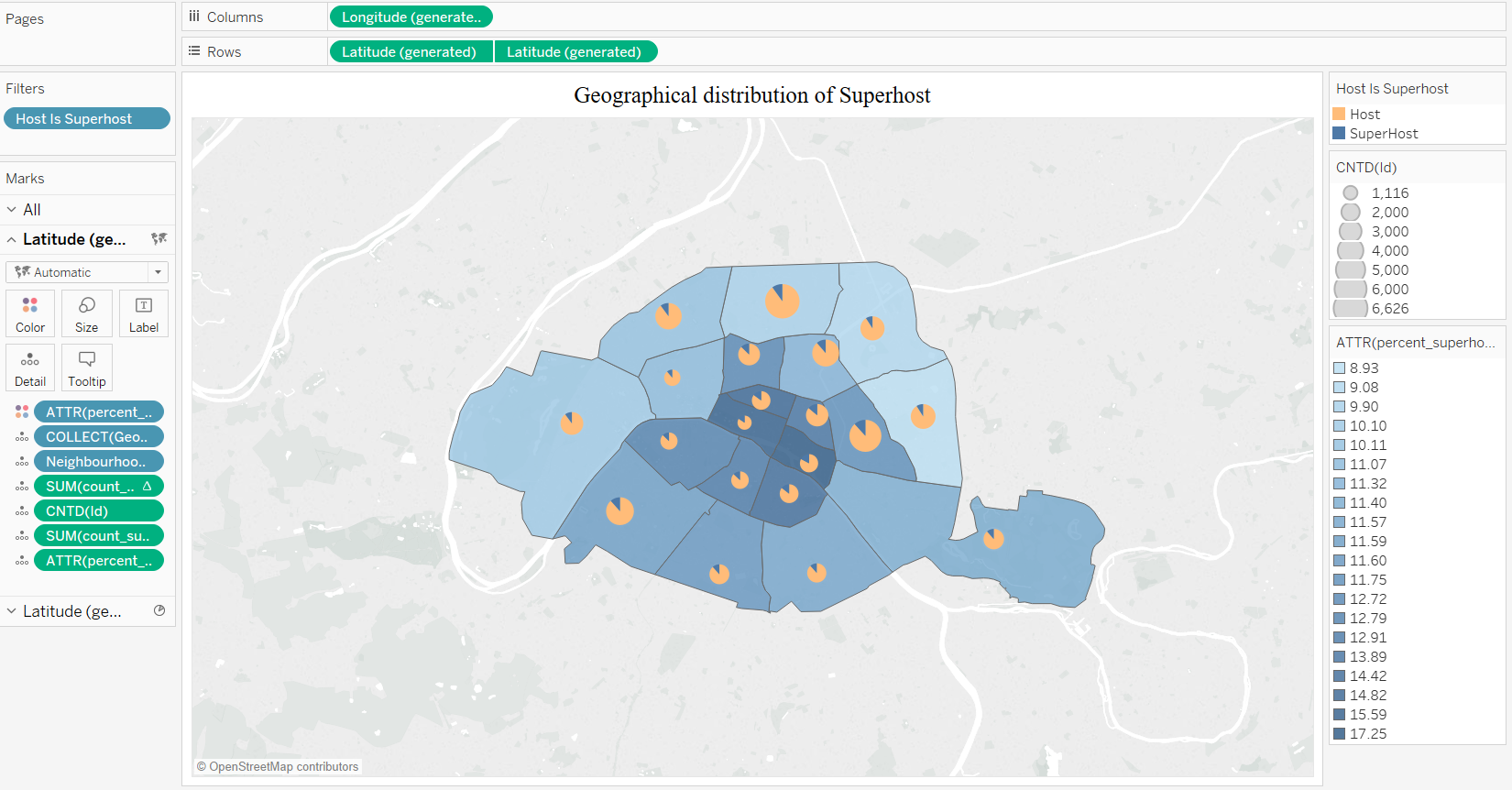
**Data Analysis and Insights:**

**1. How are the superhosts distributed geographically?**

**Data Analysis:**

We analyzed the proportion of superhosts against total hosts by geographic location. Percentage of superhosts to total hosts is varying about 1-10 percent in each neighborhood. We also found that the average price for the super hosts is higher compared to the normal hosts in all the neighborhoods.

**Graph:**

****

**Insights:**

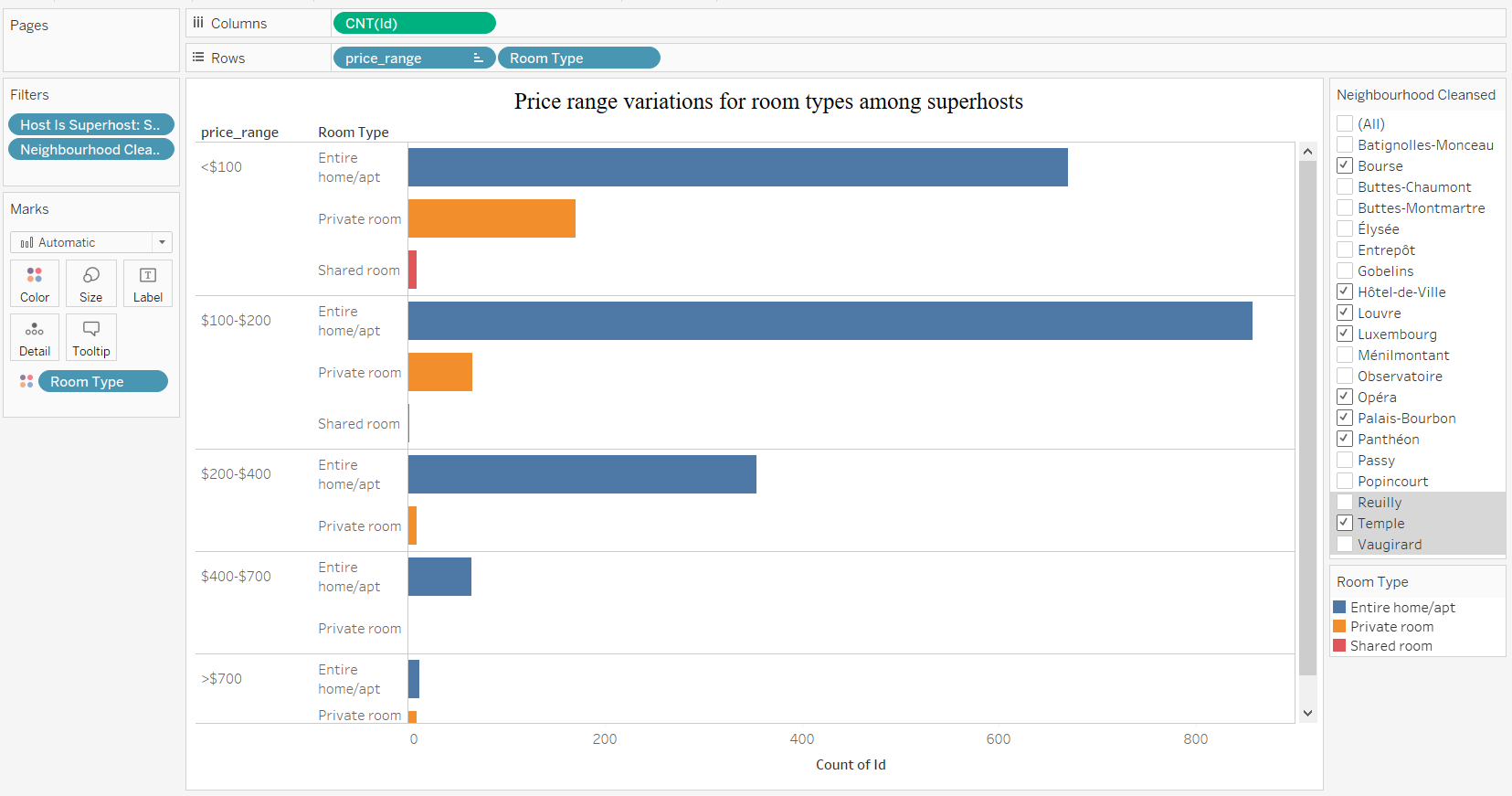
We found that the neighborhoods with the highest proportions of superhosts coincide with the top 5 tourist destinations in Paris. This indicates that Airbnb has a strong presence in the areas with the largest income in tourism.

**2. 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. 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.

**Graph:**

****

**Insights:**  
Surprisingly, entire home/apartments are listed more than private rooms and shared rooms among all price ranges. 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 behavior that when choosing among luxurious stays, the room type does not matter as much.

**3.** **Given a superhost would like to host either short term or long-term stays, how should they decide on their cancellation policy?**

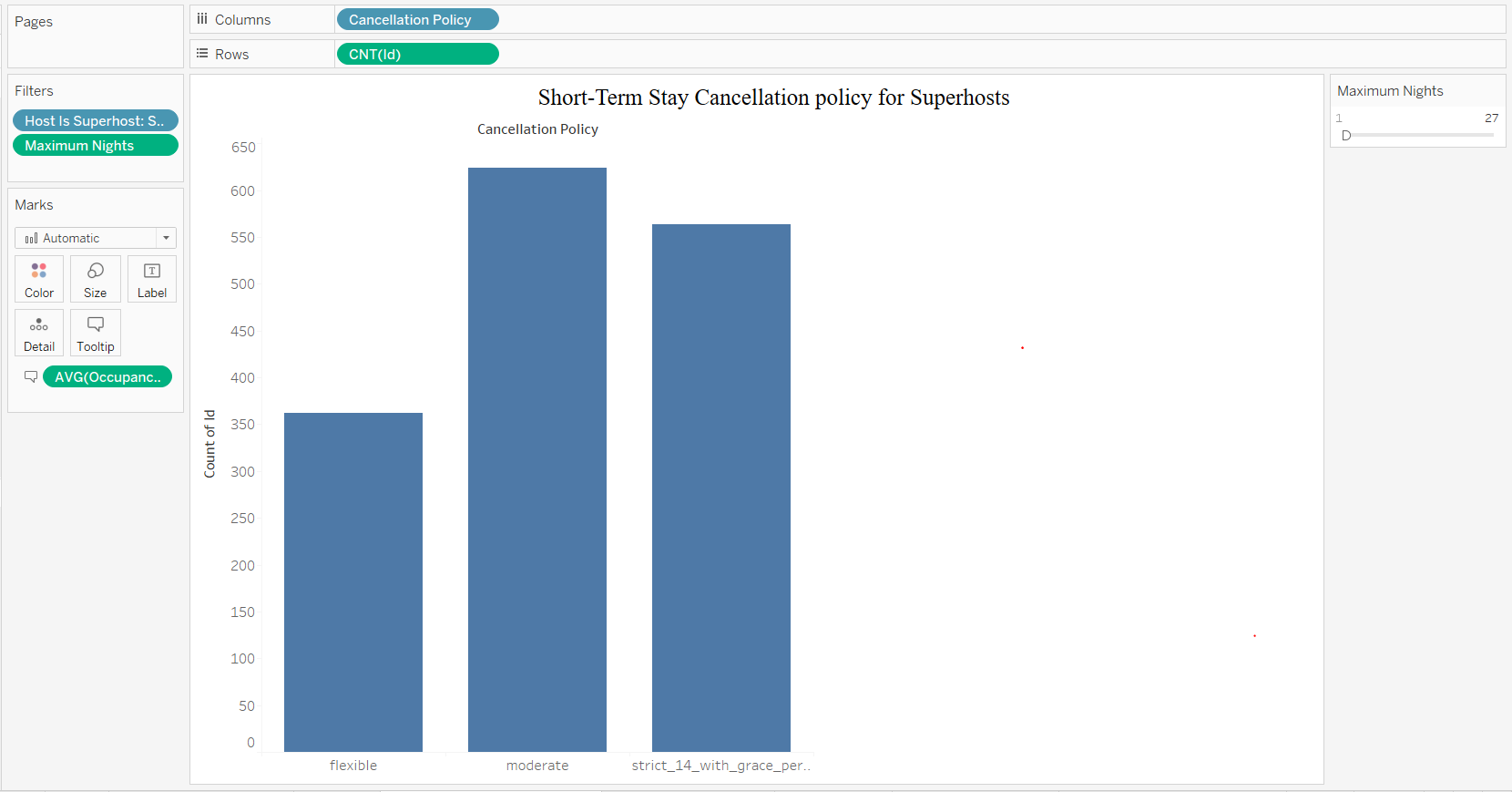
**Data Analysis:**

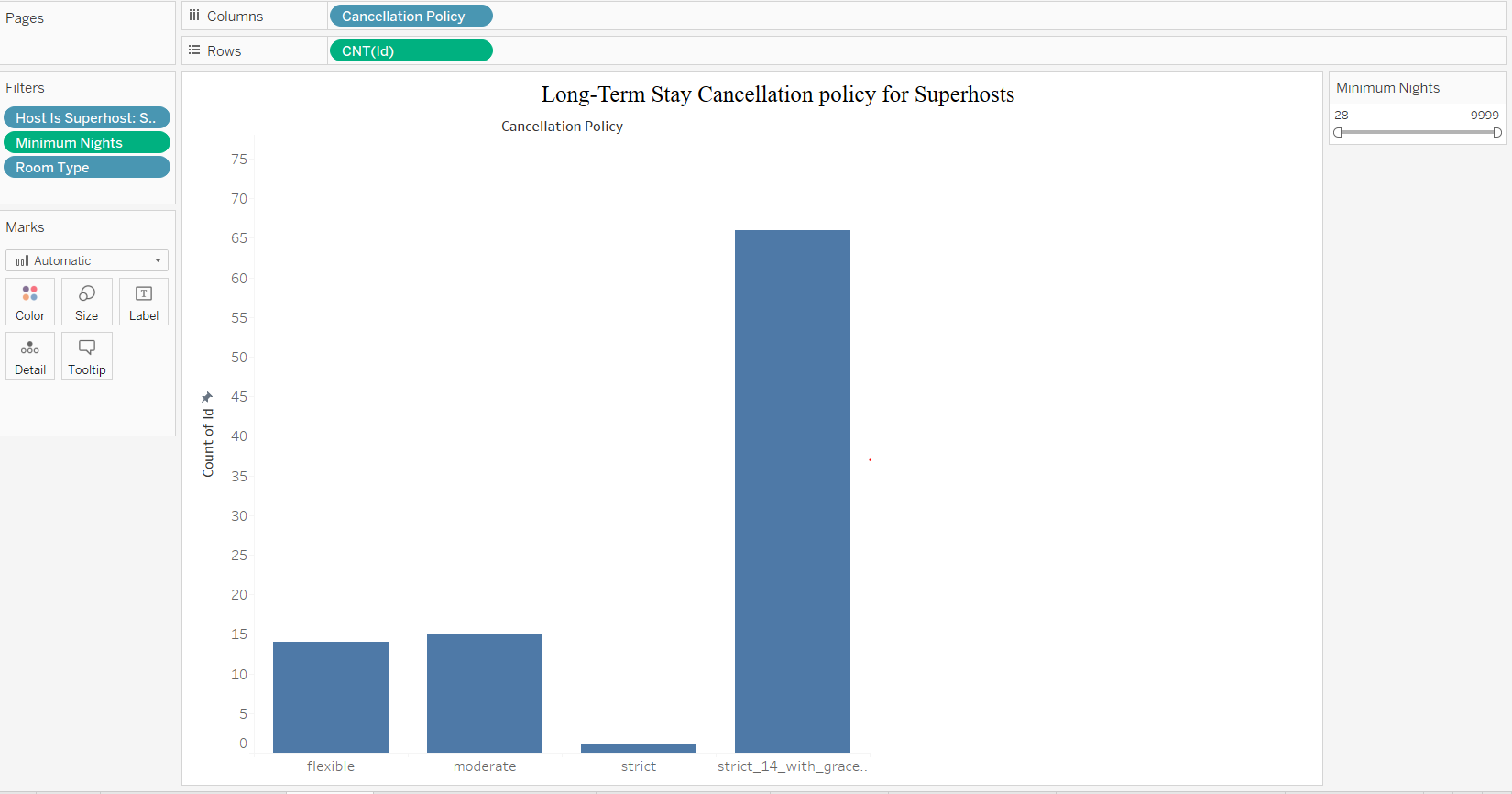
We analyzed that there is a variation in occupancy due to the cancellation policy used by superhosts. We defined the short term and long-term stays based on a 4-week period. Listings with a maximum night’s requirement of 27 were considered short term and that with a minimum night’s requirement of 28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

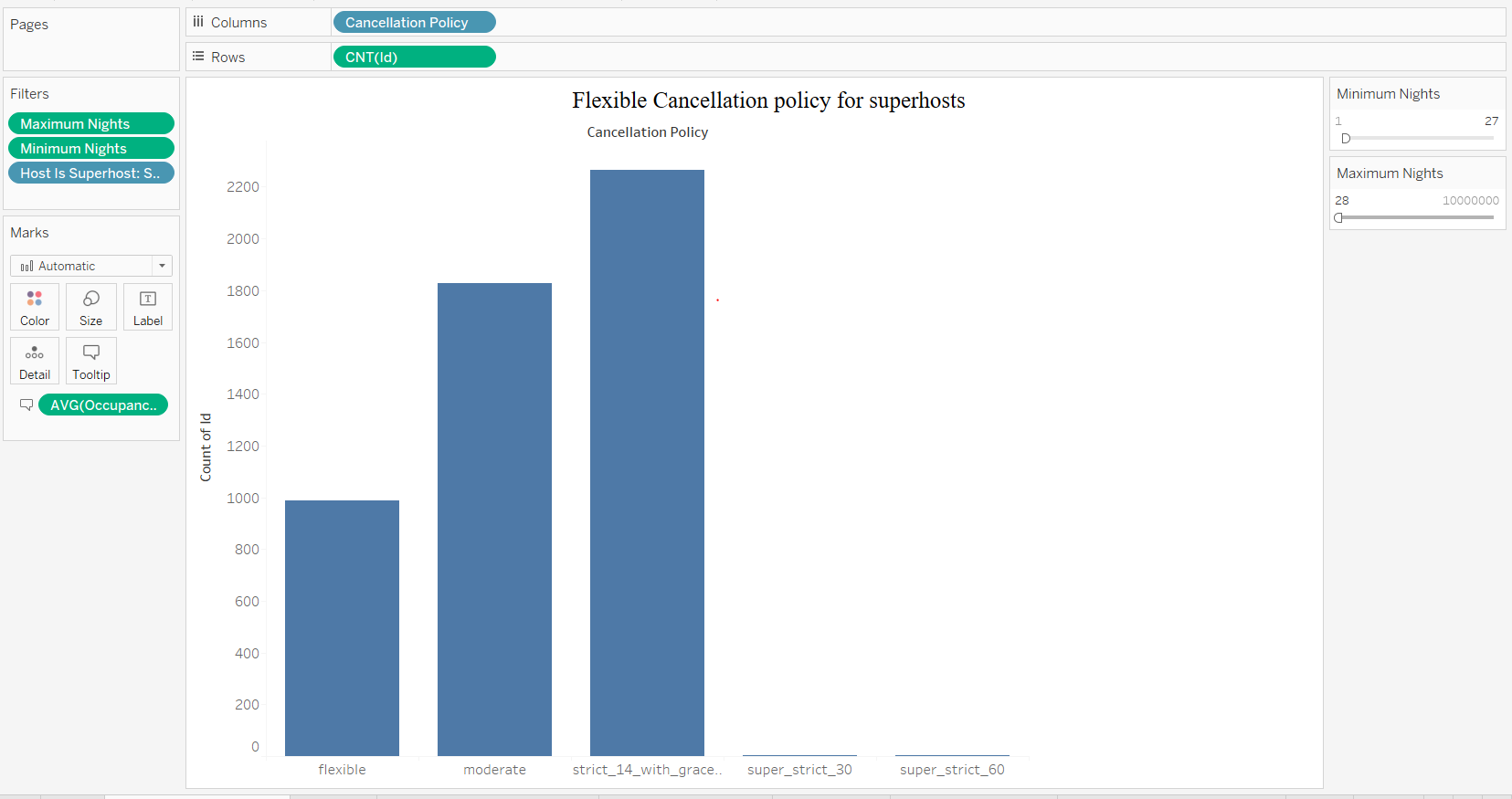
**Insights:**

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:**

****

****

**4**. **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.

**Graph:**

**Word cloud for the most praised features**

**Insights:**

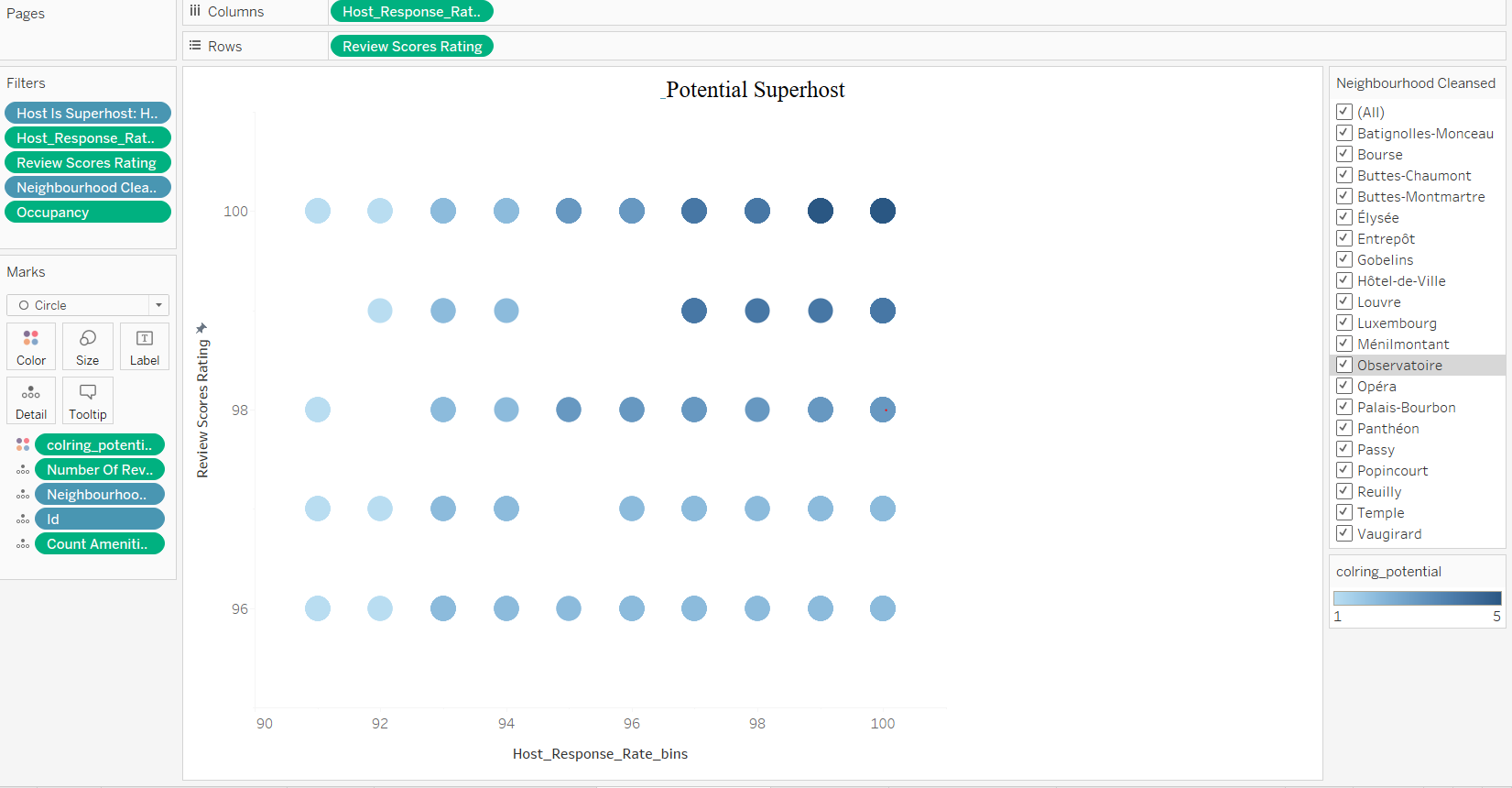
The French word “bien situé” means good location, “très bien” means great. Location and transportation are the main considerations for guests.

**5. How many normal hosts are close to becoming superhosts? What should they improve to gain this status?**

**Data Analysis:**

Prerequisites for superhosts are: 90% response rate, 4.8+ rating, 10+ stays and 0 cancellation. We inferred the number of stays from the occupancy column in calendar table, so we found the hosts that satisfies the first 3 prerequisite. This left us with the cancellations, which is unavailable from the dataset. 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.

**Graph:**

****

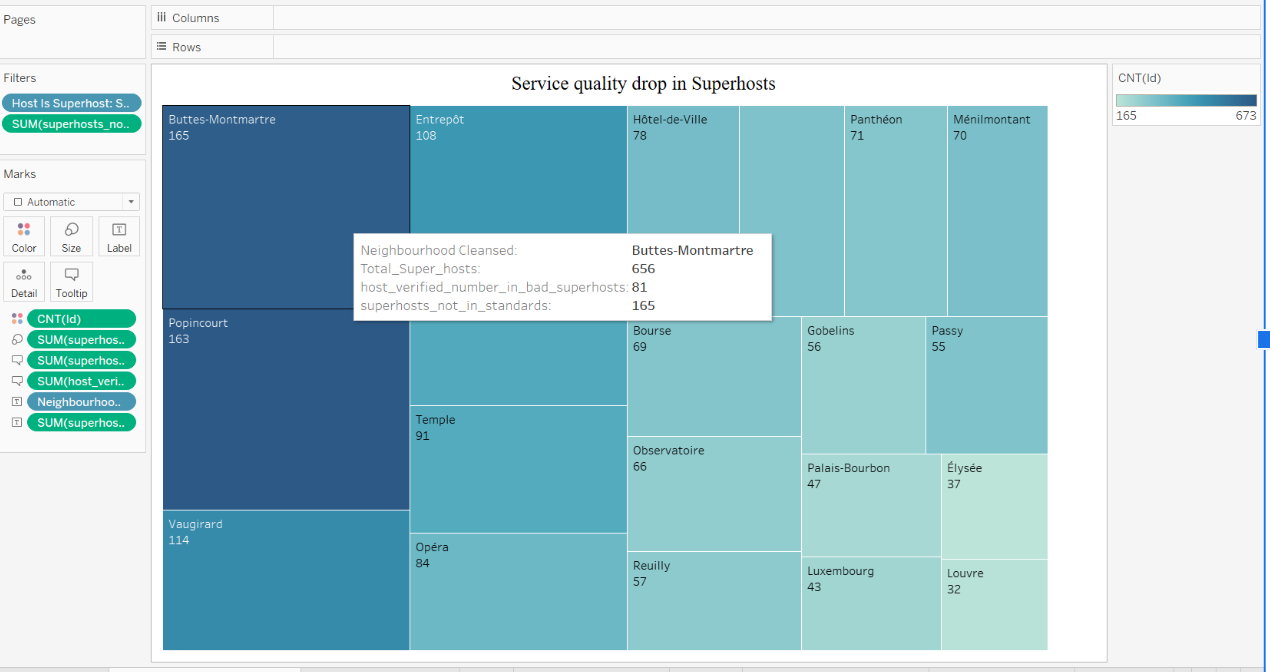
**Insights:**

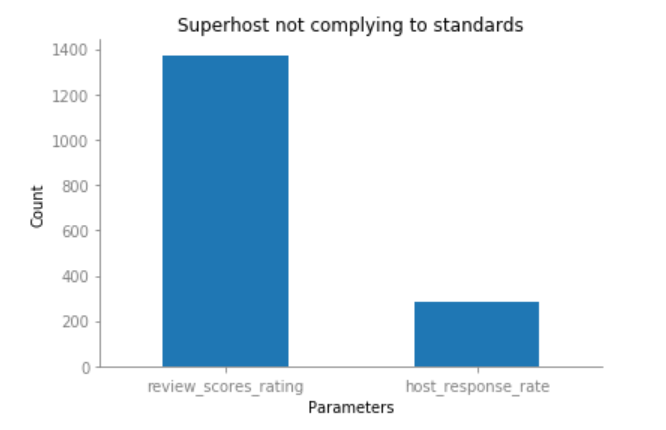
From this graph, we can tell that many hosts with good potential are suffering from this policy. Airbnb should look deeper into the cancellations and check whether there are any extenuating circumstances.

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

**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. We want to analyze them further with their neighborhoods and other driving features.

**Graph:**~~~~



**Insights:**

For the superhosts who failed to meet the criteria, we observed there were fewer of them from the top neighborhoods discovered in Question 1. Also, the top reason of their failing is due to a low review score rating. The list of superhost not meeting the prerequisites is nearly 20% of the superhosts present in that neighborhood. Further analysis showed more than half (about 53%) of superhost’s identity is not verified.

**7. How many superhosts retain their status across quarters?**

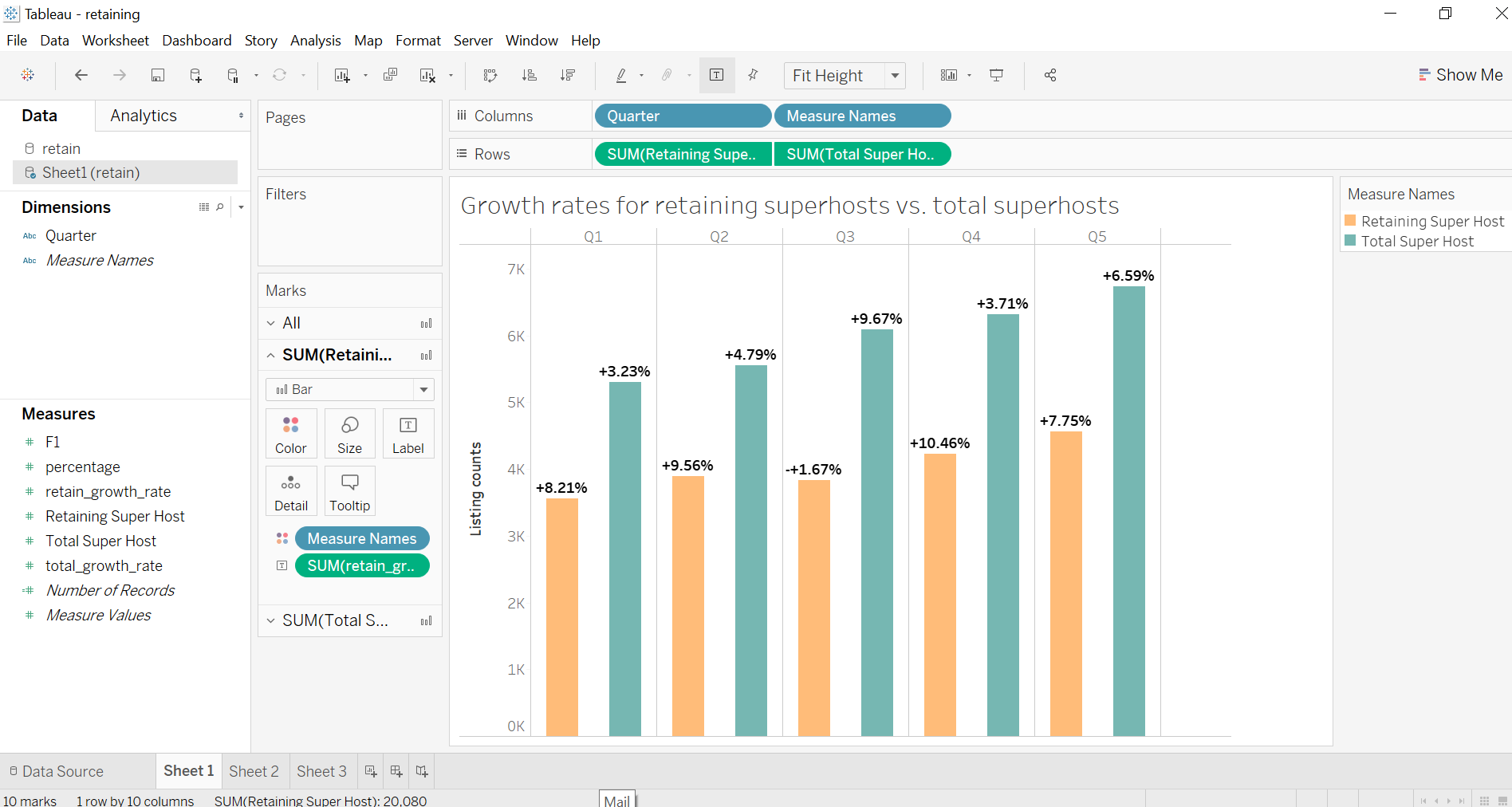
**Data Analysis:**

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 month and the status will not change within the same quarter, we chose to use the superhost numbers from February, May, August and November for 2018 and February for 2019 to avoid any duplicates. The percentages show the growth rate for total superhosts and retaining superhosts.

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

**Graph:**



**Insights:**

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 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.

**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.