Airbnb Rental Market Analysis 2017-2020

By Group #5:
- *Fabienne Zumbühl,*
- *Lizeth Orozco,*
- *Jairo Humberto Leon Cardoso,*
- *Stefan Blagojevic*

1. Project Overview

Goal and Motivation

The goal of our group was to take a closer look at the Airbnb Market in New York City and find out whether hosts becoming a superhost increase their prices from one year to another, as a result of their newly attained status.
Our motivation for this choice is based on the fact that all of the group members have been exposed to Airbnb and some even rent out their own apartment in Airbnb. For the group, it was interesting to look at the superhost status because this status comes with more duties than just being a regular host. The criteria as the following:

- Completed at least 10 trips OR completed 3 reservations that total at least 100 nights
- Maintained a 90% response rate or higher
- Maintained a 1% percent cancellation rate (1 cancellation per 100 reservations) or lower, with exceptions made for those that fall under our Extenuating Circumstances policy
- Maintained a 4.8 overall rating (this rating looks at the past 365 days of reviews, based on the date the guest left a review, not the date the guest checked out)

Ultimately NYC is the biggest market for Airbnb and thus an interesting object of study. Our group also expected to get most data within this market.

Main Questions

Apart from general questions our group has on the NYC Airbnb Market between 2017 to 2020, including price changes and listings per borough we wanted to find out whether hosts of listings that attained a superhost status charge more per night than previously as host. This led to our main research question which states as follows:

> From 2017 to 2020, did Airbnb superhosts charge more per night than when being a regular host the previous year?

Other questions of interest to us where following:

- What is the price discrepancy per night between hosts and superhosts?
- Which borough in NYC has most listings?
- How do the prices generally change over the years studied?
- In which boroughs in NYC do the superhosts live?
- In which borough does the median superhost charge the most?
- Which borough in NYC would be ideal to own an Airbnb based on our findings?

This exploratory data analysis was carried out from ideation to data collection and processing and data analysis over the course of 2 weeks.

2. Data Acquisition and Processing

We gathered data using Inside Airbnb csv's (http://insideairbnb.com/get-the-data.html).

Due to the size of the files from Inside Airbnb, files starting in March 2017 monthly to March 2020 were imported into the Jupyter Notebook. Each file provided a snapshot of a point in time when Airbnb pulled the listing data. In total there were 37 csv's that were cleaned based on criteria of interest. This criteria included filtering by bed type, bedrooms and listing that had complete data (dropna). Due to time limitations the decision was made to reduce the data from monthly review to yearly which narrowed the csv's in scope down to 4.

In order to find the inflection point at which the listing converted from host to super hosts the 4 data frames were combined. Using the four data frames we combined the listings in March of 2017 which started as hosts and with those same listings that then became superhosts in March 2020. This gave us the starting point at which we could review pricing per year based on host status.

3. Data Analysis

Our main research question guided us through the data analysis. Based on our expectation on the answer to that research questions, we formulated our hypothesis (and null-hypothesis), which state as follows:

Null Hypothesis: hosts increase their base price per night (not incl. "cleaning fee" and extra fees) after they achieve superhost status.

Alt. Hypothesis: hosts don't increase their price per night (not incl. "cleaning fee" and extra fees) after they achieve superhost status.

Before focusing our attention to the main research question, we also wanted to answer the other questions we had on the NYC Airbnb market stated above. We filtered our dataframe with no superhosts in 2017 to all superhosts in 2020.

We did a preliminary analysis on the price changes in the market from one year to another and found that in 2017 to 2018, 2018 to 2019 and 2019 to 2020, between 50% and 70% of listings did not increase their price (regardless of change from host to superhost or remaining a host). We further analysed the market share of listings per borough and found that throughout the four years studied, most listings are located in Manhattan and Brooklyn. A look on that with the absolut listings verified this finding. We also analyzed that the total amount of listings throughout the four years and found a decrease in absolute numbers of listings from 2018 onwards. Using the Coordinates (Latitude and Longitude) we also analysed where each of the - throughout the four years - 943 upgraded superhosts live (color coding then differently depending on the object type they rent out). We found that most superhosts live in Manhattan (East Village, Lower East Side, Harlem) and in Brooklyn. Most apartments are rented out in Manhattan, while most houses are rented out in Brooklyn.

We also analysed in which borough the median superhost listings charges the most per night and found that the median superhost charges most in Manhattan (ca. 140$ per night) and Brooklyn (ca. 120$ per night), followed by Queens (ca. 80$ per night) and the Bronx (ca- 60$ per night).

From one year to another, we wanted to see how many hosts attained superhost status and how many remained regular host. For the once that changed from year n to year n+1 from host to superhost, we calculated the price difference for each listings from year n to n+1. For year 2018 (hosts that became superhosts) we detected 468 upgrades, for year 2019 (hosts that became superhosts) we detected 216 upgrades and for year 2020 (hosts that became superhosts) we detected 468 upgrades. From those inflection points, we calculated the price differences. From those price differences we took the median for each year per borough and found that the median price change after becoming a superhost is mostly 0. We did the same analysis for hosts that became superhosts in 2018 and stayed superhosts until 2020. This analysis also led to a similar result.

Ultimately, we used a paired t-test to determine the p-value associated with our host-to-superhost sample before-and-after their change in status occurred.

A paired t-test was used because our two data sets from March 2017& March 2020 had the same number of observations (host-to-superhost inflection points). Therefore, the two samples can be paired with one another. Before-and-after observations of the same subject in two points in time.

After we isolated the inflection points which represented the change in status from host-to-superhost, we compared price discrepancies with their shift(s) in status ov.

To ensure our findings would be accurate, we approached the problem statement from three different perspectives: (1) calculating price discrepancies per listing year-over-year specifically regarding hosts becoming superhosts over the same period. (2) paired t-test of 2020 rental price to 2017 rental price for the 943 hosts that became superhosts (3) median price by borough in reference to the changes in status (H→ SH |&| SH→ H→ SH) year-over-year.

Based on our analysis, we rejected our null hypothesis that hosts increase their rental listing base price per night after they become superhosts in the NYC Airbnb market

→ p-value=0.0317 < α(o.o5) ⇒ reject the null (H0) ⇒ (H1) hosts don't raise their price after becoming superhosts!

4. Discussion of Findings

As already mentioned above, looking at the NYC Airbnb rental market from 2017 to 2020, we found that throughout those four year, 943 hosts attained a superhost status. We also found that generally speaking, most listings (regardless of being a superhost or host listing) did not change their price from one year to the following year (between 57.3% from 2017-2018, to 69.0% from 2018-2019, to 74.7% from 2019-2020. As for the market share of listing by borough, we found that most listings are located in either Manhattan and Brooklyn.

This holds true for all the four years studied. Looking at the absolut numbers of listings throughout the four years we could identify a listings decrease from 2018 onwards. Those findings are indicate that most listings don't change their price throughout the years and stay at the price with which they entered the market.

We mentioned above that based on our research approach, our finding is that the Median Superhost Rental is not increasing it's price from year n (regular Host) to n+1 (Superhost). This finding is surprising to us, since we expected the opposite to be true. Nevertheless, we can reject our Null Hypothesis.
We assume that the incentives of becoming a Superhost is not a higher price paid per customer, but attracting more customers and making more money like that, since Airbnb has the option to filter by Superhost only and thus put those listings on top of the search list. We further assume that the competitiveness in the NYC rental market is so high that Superhost might lose customers by increasing their prices due to price inelasticity of the demand. It would be interesting for future data analysis to test those assumptions.

We found four main limitations to our findings. First, our dataset was limited to 943 listings after cleaning and setting the superhost count to "false" in 2017 and to "true" in 2020". Second, our findings are not universally applicable for the NYC market, but only for the years studied.Third, our findings are also not universally applicable for the whole Airbnb market. If there was more time, would have liked to explore trends from month to month.