ORIE 4171

Learning with Big Messy Data

Project Midterm Report

The Prediction of Airbnb Price in NYC: the Effect of COVID-19 in the Airbnb Market

November 2020

Linna Yue (ly389)

Sophie Zhao (sbz24)

Nicolás Oriol Guerra (no227

1. Introduction

The Covid-19 pandemic gave the global economy a huge shock. The travel industry was clobbered heavily. Many trips were cancelled or held due to Covid-19 which has a devastating effect on the hotels, Airbnb and other short-term rental properties. Especially after March 22, 2020, the NYC government closed the non-essential economy and international travel restrictions enforced on February 2 and March 11. According to Forbes[1], 47% of Airbnb hosts don't feel safe renting to guests and 70% of guests are fearful to stay at Airbnb apartments. The revenue from June to August was expected a 44% decrease for the Airbnb apartments. 47% of the hosts provided month-long rental, and 29% reduced the listing prices.

According to the above facts, we make the hypothesis that the Airbnb prices, number of rentals will decrease compared to the prices before Covid-19. And because more people are practicing social distancing, it is also likely that there will be more rentals for private Airbnbs and less rentals shared Airbnbs. Different machine learning models will be tested to predict the listing price before and after the pandemic, and cross-validation will be used to choose the most accurate model. We will see whether the best model is changed before and after the Covid-19. To minimize the seasonal influence and study both the immediate and medium to long term pandemic impact, the listing data used are rentals in June and September (2019 and 2020). This way, data from right after restrictions were lifted and more recent data will be compared.

2. Initial data treatment[2]

Datasets obtained from Airbnb contain information uploaded within the first week of the following month in our study (e.g. September 2020 data was uploaded October 5th). These datasets each contain almost 50k observations with different pieces of information. They have been cleaned and missing data substituted with relevant values to make further studying easier. For example, some airbnb places had no reviews and some host names and ids were missing (less than 1%). Furthemore columns containing information about last reviews were often missing (24%) in the cases where no previous reviews had been posted about the place.

Another step in the data treatment process was to encode some of the variables such as neighbourhood (223 unique values), neighbourhood group (5 unique values) and room type (4 unique values). One hot encoding was in this conversion.

---

[1] LANE, L. *How Bad Are Covid-19 Pandemic On Airbnb Guests, Hosts?.* Forbes (Link)
[2] Code and process detailed in Data Mining and Cleaning.ipynb

3. Airbnb listings on the map

In order to compare the effect of Covid-19 on the Airbnb market, the listing with the last reviews in Sep 2019 and 2020 were selected. There were 12,086 listings being rented in Sep 2019, however, there were only 5,721 listings being rented in Sep 2020. As shown in the two maps. The listings with the price under 200 dollars, 200 - 500 dollars, 500 - 1,000 dollars, and above 1,000 dollars are represented by grey, steelblue, blue and red respectively. It's obvious that rental in Sep 2020 decreased dramatically compared with that in 2019, especially the proportion of lower price listings (under 500 dollars). One interesting phenomenon is that in Long Island, the number of rentals with 500 - 1,000 dollars price increased after the lock down. Because the social distancing makes non-crowded places more preferable than those in commercial areas.

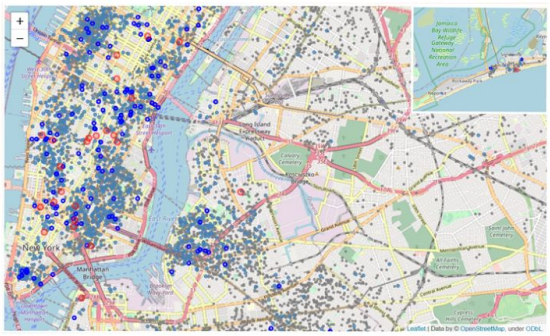

Figure 1. Listings in Sep. 2019.

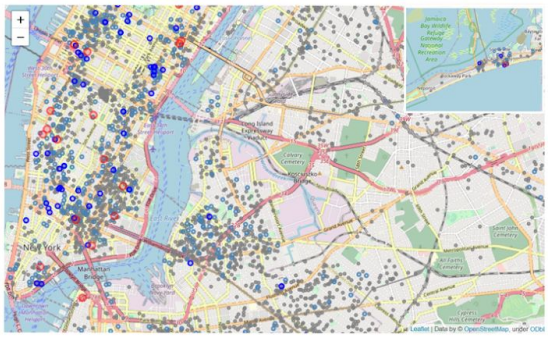

Figure 2. Listings in Sep. 2020.

4. Exploratory data analysis

Price listings were further analyzed in order to determine what factors would affect Airbnb prices after Covid 19. The months September 2019 and 2020 were selected again for this analysis. First, the changes in Airbnb prices per region in New York City were observed. Based on the histograms shown in Figures 3 and 4, Covid-19 did not seem to make a difference in the Airbnb prices among each of the neighborhood groups. However, after analyzing the prices per room

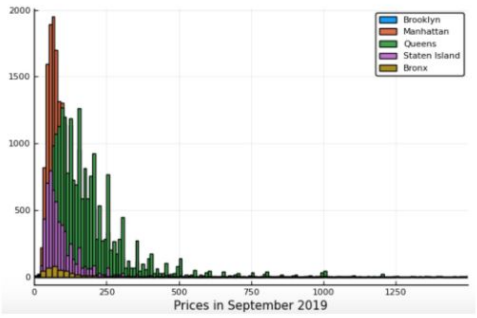

Figure 3: Prices in Sep. 2019 per neighborhood group

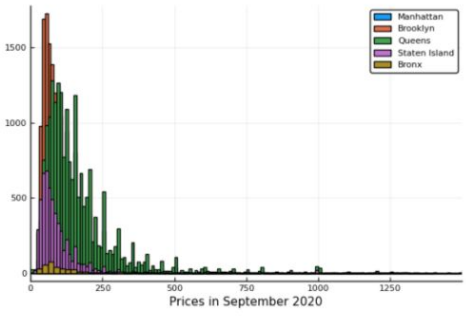

Figure 4: Prices in Sep 2020 per neighborhood group

type (entire home, private room, shared rooms, and hotel rooms) as shown in Figures 5 and 6, prices for private rooms increased in September 2020 compared to September 2019. But there was not a significant difference in price change for the other room types. The reason for this trend is because of the increased demand for renting private rooms since more people have been practicing social distancing since the beginning of Covid-19.

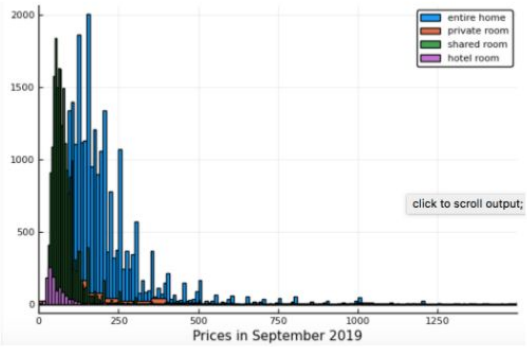

Figure 5: Prices in Sep. 2019 per room type

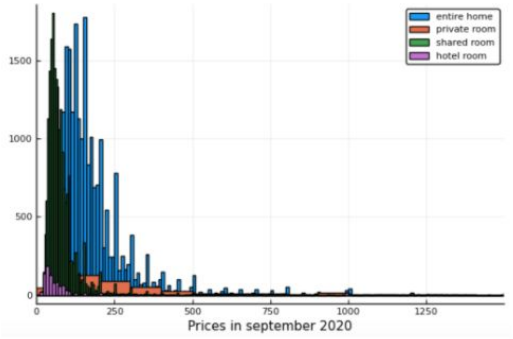

Figure 6: Prices in Sep 2020 per room type

5. Methodology and future steps

After the initial analysis and variable exploration, the following steps are to compare how closely related the different datasets are to get a measure of how the Airbnb situation has changed. The proposed idea is to fit an individual prediction model for each of the selected time intervals and gauge how well it fits the data in different periods.

For example, the model fitted with data from 2019, when used in 2020's information will indicate how drastic the change was. Moreover, the difference (if any) between June and September will be useful to infer if the situation is improving or staying mostly constant.

One possible danger of this method is that there might not be enough information or the period between both observations may be too short to come to any meaningful conclusion. In this case, the answer may be that with our method of study, there is not enough evidence to suggest that the pandemic has affected COVID prices.