



Insights from Airbnb, NYC dataset

Submitted by : Rithesh Kamath



Agenda



- Objective
- Background
- Data Preparation & Cleaning
- Data Analysis
- Recommendations
- Appendix:
 - Data Assumptions
 - Data Methodology



Objective

- 
- Understand key insights from Airbnb NYC business from the provided dataset & Improve strategies to revive the business as the restrictions have started lifting.
 - To prepare for the next best steps that Airbnb needs to take as a business



Background



- For the past few months, Airbnb has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, Airbnb wants to make sure that it is fully prepared for this change.
- The different leaders at Airbnb want to understand some important insights based on various attributes in the dataset so as to increase the revenue

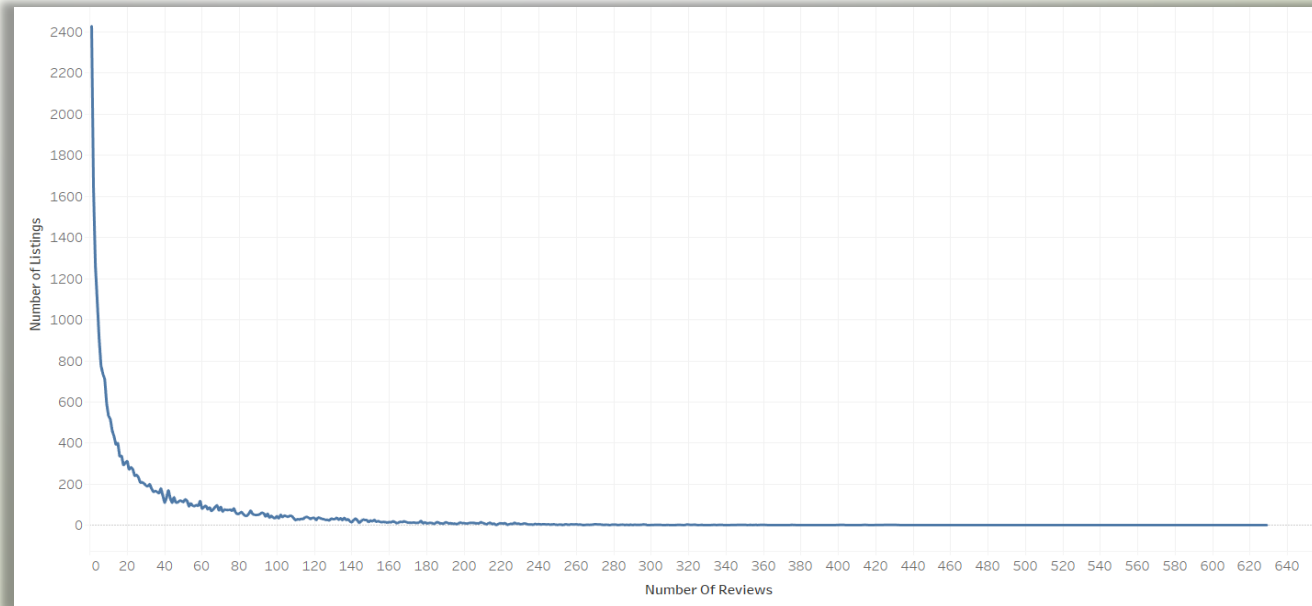


Data Preparation and Cleaning

- The data was cleaned from any missing values and outliers
- The data was analyzed using simple statistical analysis of the data
- The visualizations were derived using the Excel & Tableau tools to understand key inferences from the analysis

Preference for Booking

- Right skewed distribution of price vs listings
- Assuming that the column “Number of reviews” corresponds to the number of bookings, most of the listings in New York City have average to low number of bookings & very few listings have a very high number of bookings which is expected



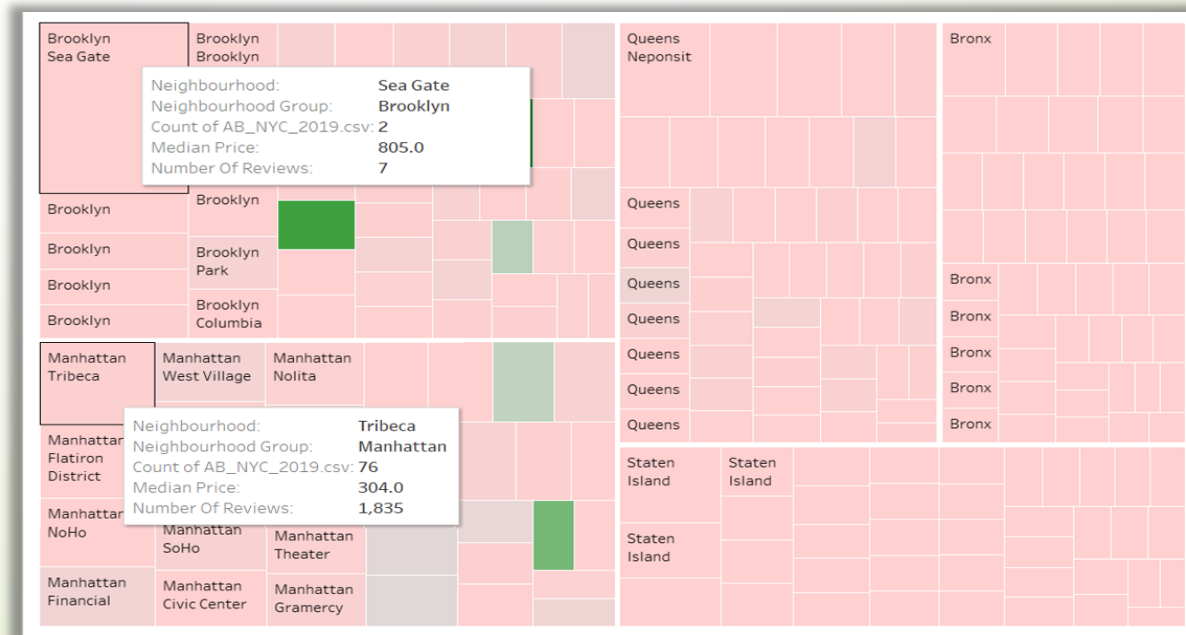
Highest listing per Host Name Vs Room Type

- There are certain hosts with a lot of listings in New York City.
- Sonder (NYC) is supposed to be a hotel brand which is a business in itself & is also taking Airbnb platform for renting out its properties as well.
- There are many such hosts with a large number of listings under them, perhaps a little investigation should be done to check if these are in line with Airbnb policies & practices.

	A	B	C	D	E
1	Highest listings per Host name & Room type				
2					
4	Host_name	Entire home/apt	Private room	Shared room	Grand Total
5	Sonder (NYC)	199	8		207
6	Michael	125	72	10	207
7	David	106	91	1	198
8	John	69	119	2	190
9	Alex	100	53	6	159
10	Maria	57	71	2	130
11	Mike	94	25		119
12	Anna	62	49	3	114
13	Melissa	26	57	19	102
14	Jason	56	43		99
15	Chris	60	33	4	97
16	Daniel	53	44		97
17	Jessica	53	41	1	95
18	Sarah	44	44	1	89
19	Mark	59	29	1	89
20	Andrew	57	29		86
21	Anthony	31	52		83
22	Laura	42	38		80
23	Nina	26	52	2	80
24	Amy	46	30	1	77
25	Rachel	35	40	1	76
26	Jennifer	37	38	1	76
27	Michelle	42	31	2	75

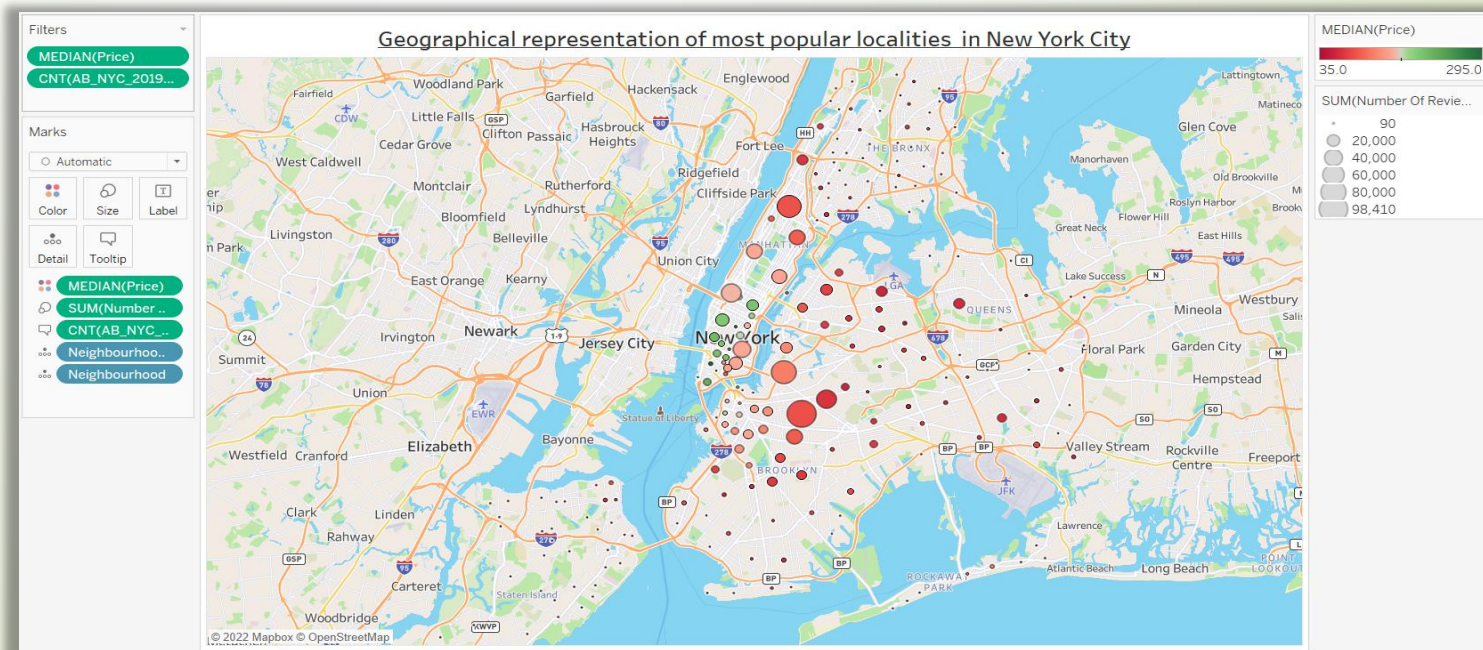
Price of properties in NYC Areas

- Even though Manhattan has high Median price, Sea Gate Neighbourhood in Brooklyn has the highest median price which is much more than the poshest neighbourhood in Manhattan.
- This is because there are only 2 listings for this neighbourhood & to calculate average doesn't make much sense here as compared to 76 listings of Manhattan's Tribeca neighbourhood



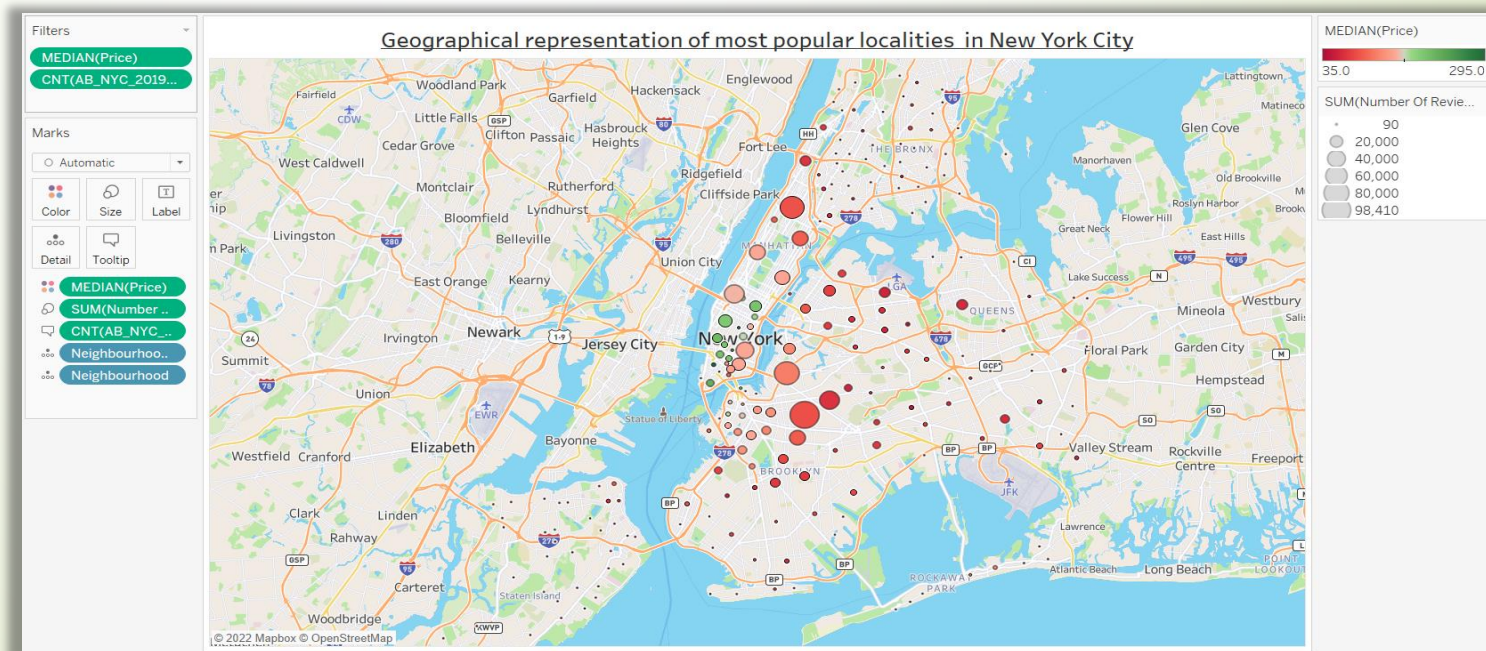
Popular localities in NYC Areas - I

- There is a huge latent demand in the Manhattan neighbourhood & so the average prices are also high here.
- Another interesting observation here is the Staten Island where there is very less number of reviews (bookings) at very low prices. So focus has to be given more on Staten Island by trying to get some more properties.



Popular localities in NYC Areas - II

- Another interesting observation would be around JFK which is one of the busiest airports in the world yet still there are not much listings around JFK. Perhaps a collaboration with airlines would be an option where the passengers can settle during their long layovers and hence acquiring some properties around JFK makes sense





Recommendations - I

- Most of the listings in NYC have average to low number of bookings & very few listings have a very high number of bookings. This is a problematic trend as this seems to be a perfect example for 80-20 Pareto Rule where 80 percent of business comes from 20 percent of the customers which is probably not very healthy. Potential platform level changes can be made so that next time when a user is trying to book a property, the platform should not recommend top properties but rather some random properties through which variability can be introduced & have smaller operators get equal share of the business.
- There are certain hosts with a lot of listings in NYC. Sonder (NYC) is a hotel brand which is a business in itself & is also using Airbnb platform for renting out its properties. There are many such hosts with a large number of listings under them, perhaps a little investigation should be done to check if these are in line with Airbnb policies & practices.



Recommendations - II

- There is a huge latent demand in the Manhattan neighbourhood & so the average prices are also high here. Another interesting observation here is the Staten Island where there is very less number of reviews (bookings) at very low prices. So focus has to be given more on Staten Island by trying to get some more properties. Perhaps by conducting tourism drives by tying up with the local ferry services from Brooklyn.
- Another interesting observation would be around JFK which is one of the busiest airports in the world yet still there are not much listings around JFK. Perhaps a collaboration with airlines would be an option where the passengers can settle during their long layovers and hence acquiring some properties around JFK makes sense.
- Lastly, Harlem is the most popular neighbourhood in Manhattan. Perhaps next time when a user is trying to book a property in this locality, the search engine could be optimized to show properties around Harlem such as “Morningside Heights” or “East Harlem” thereby having smaller operators get equal share of the business.



Appendix – Data Assumptions

- Number of bookings per listing was not provided in the dataset, so we have assumed that “number_of_reviews” columns corresponds to number of bookings. The data was analyzed using simple statistical analysis of the data
- Null Values have been assumed to have not much of an impact on our analysis, hence null values were dropped from the columns - last_review, number_of_reviews and availability_365



Appendix – Data Methodology

- A combination of excel & tableau has been used to get insights for our data analysis & visualization.
- Since there were some extreme values in column “Price”, we have used Median (average) for analysis purpose & not mean
- Please find the detailed Methodology Documents by clicking [here](#).



Thank You!