Mitigate impact of COVID on Airbnb

AGENDA

Objective
Background
Dataset and variables
Technical Analysis
Analysis findings
Appendix –Data assumptions

OBJECTIVE

Understand the dataset having Airbnb neighbourhood information.

Looking on the techniques and functions used for analysis in python notebook

Understand customer preferences post-COVID.

Provide recommendations to increase revenue.

BACKGROUND

Major decline in revenue resulting from travel restrictions during the COVID pandemic.

Insights for revenue increase in New York region after return to normalcy.

Dataset and variables

Airbnb, Inc. is an American vacation rental online marketplace company based in San Francisco, California, United States. Airbnb offers arrangement for lodging, primarily homestays, or tourism experiences.

We have the dataset for mitigating the business and increasing the profit margin

Content: This data file includes all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Variables/Columns:

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Exploring data

Basic analysis of data

After importing the data set into python notebook we used functions like shape ,info ,describe to look at the structure and variance of the data

Findings:

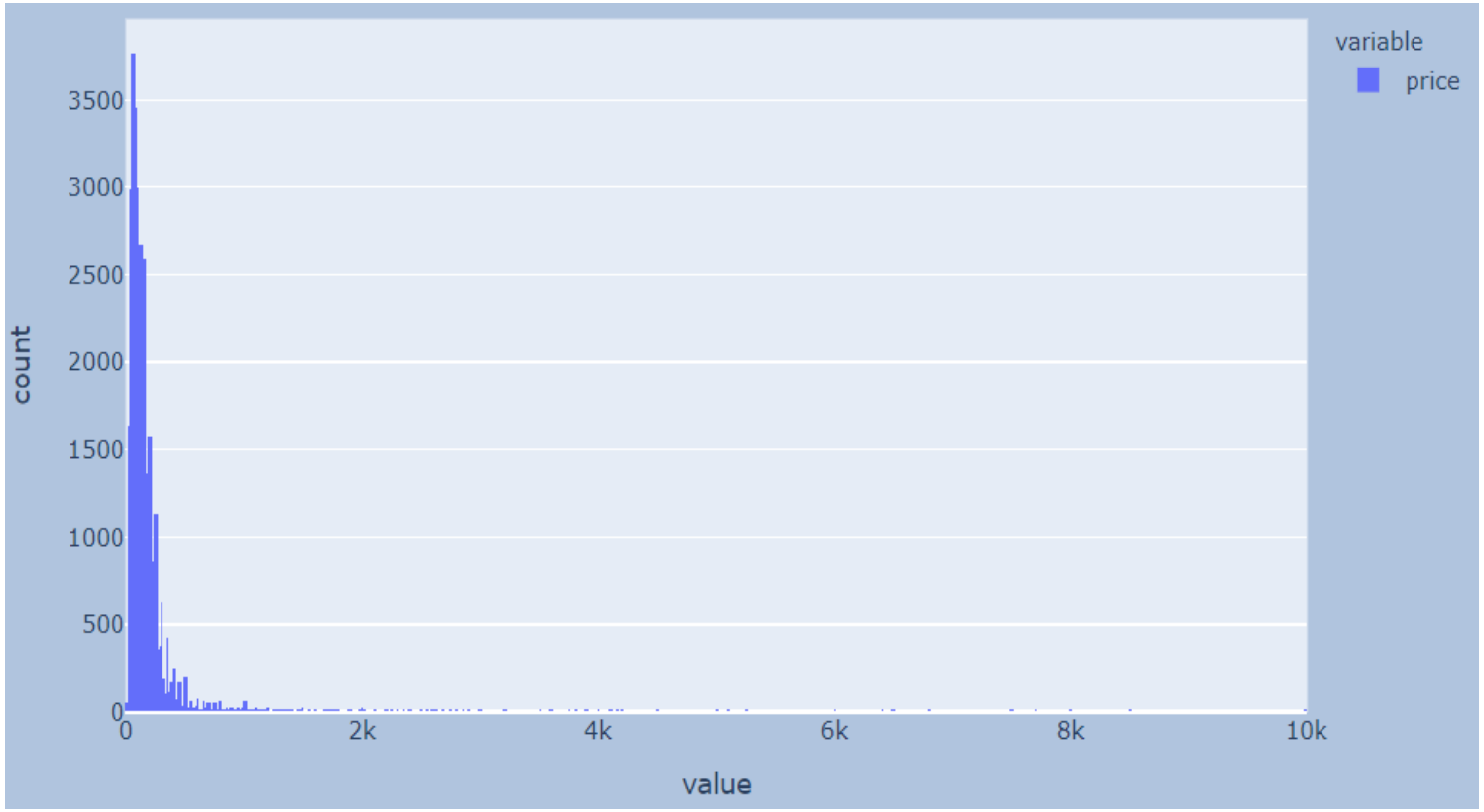
- 1. price variable mean is much higher that median, which means that distribution is right skewed.
- 2. minimum_nights too has mean much higher than median, so this is also right skewed distribution. There are possible outliers in this variable.
- 3. number_of_reviews and reviews_per_month are also heavily right skewed, since the difference between the 90th percentile and maximum value is huge.
- 4. calculated_host_listings_count is also heavily right skewed.

Checking the price variable distribution

We can see the price distribution is heavily right skewed

Most of the prices are in the range $50 to $1k

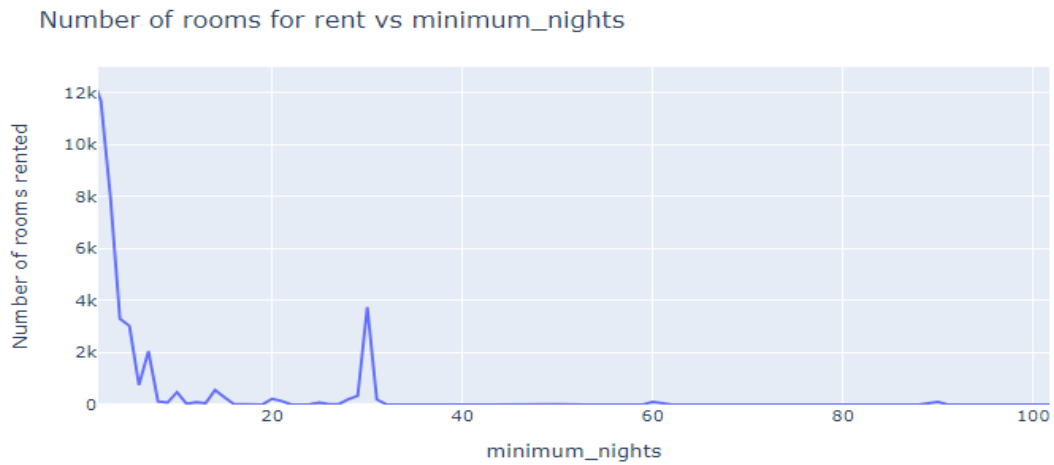
But may vary to as high as $10k



Analysis of neighbourhood_group by minimum_nights

Used groupby and count on minimum nights and plotted line graph using plotly express

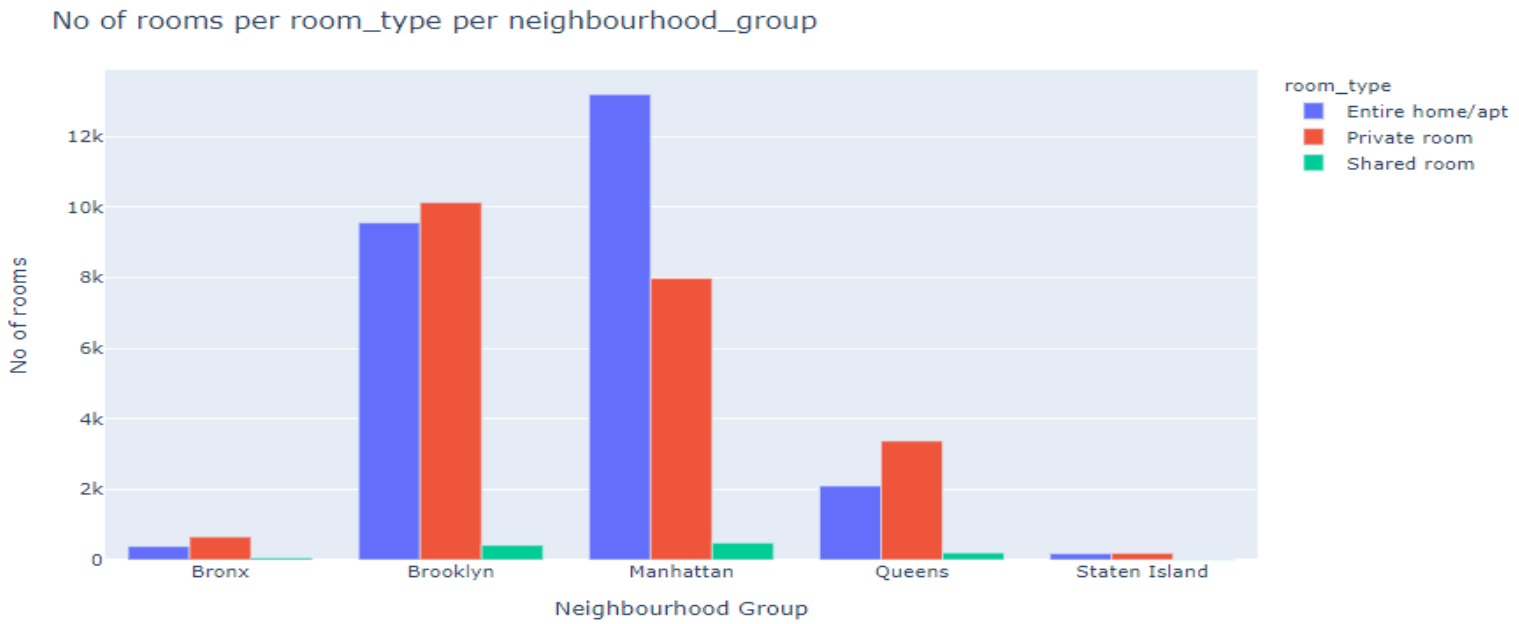
- Most people prefer to rent airbnb rooms for one-three night.
- Consequent smaller spikes are at 7, 10, 14- these rooms also seem to be very popular.
- Many people also prefer to rent for a month- maybe these are those who come for business purposes.



Different neighbourhoods vs Room types

Created clustered bar charts to show no of rooms for each neighbourhood groups for different room types available

- Brooklyn and Manhattan are the most popular neighbourhood groups for rooms. In Brooklyn the most popular room type is Private room(50.4%) by a small margin over Entire home/apt(47.5%), and in Manhattan it is Entire home/apt(61%) by a huge margin over private room(37%).
- Shared rooms are not very popular among travellers. The maximum % they constitute in any region of the total no of rooms for rent is 5% in Bronx.



Availability 365 vs number of rooms for rent
Availability 365 vs Reviews

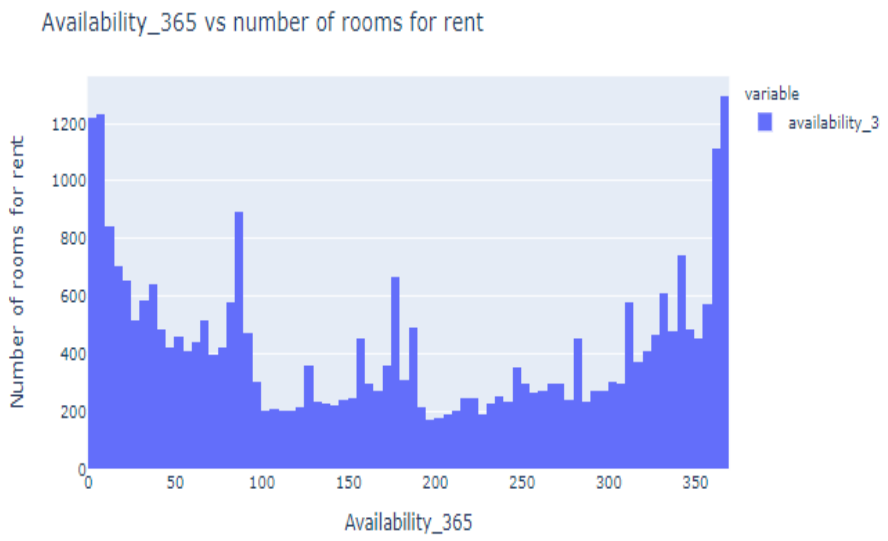
Plotted histogram for avalibility vs rooms for rent and used bar plot for availability vs reviews ,for the later plot we used a dataframe filtering availability less than 200 and non zero values

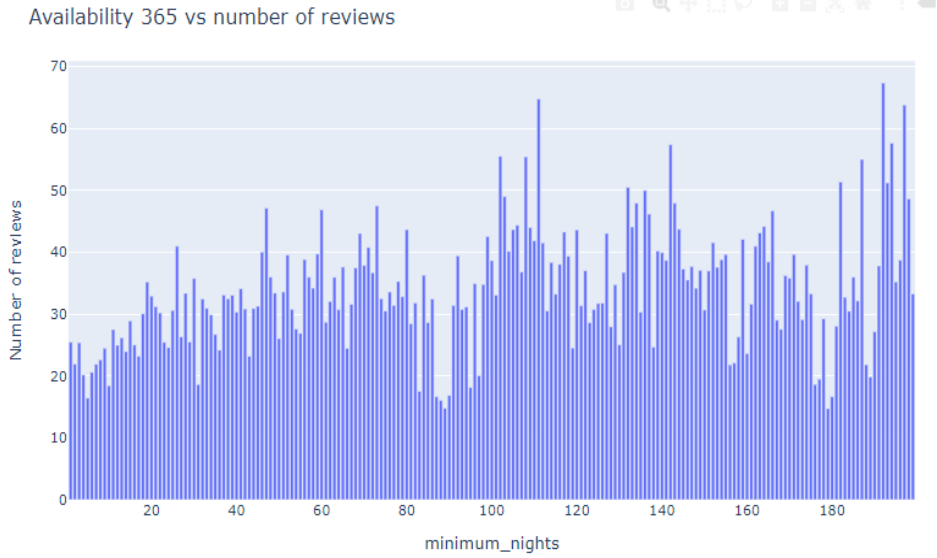
Code snippet:LimitedAvailability_365 = airbnb.loc[(airbnb['availability_365'] < 200) & (airbnb['availability_365'] != 0)]

Findings and suggestion for business:

2447 rooms with availability 1-10 days per year. This could be because these owners don't want to invest in their properties for the rest of the year, or they are using it for some other speficic reasons.

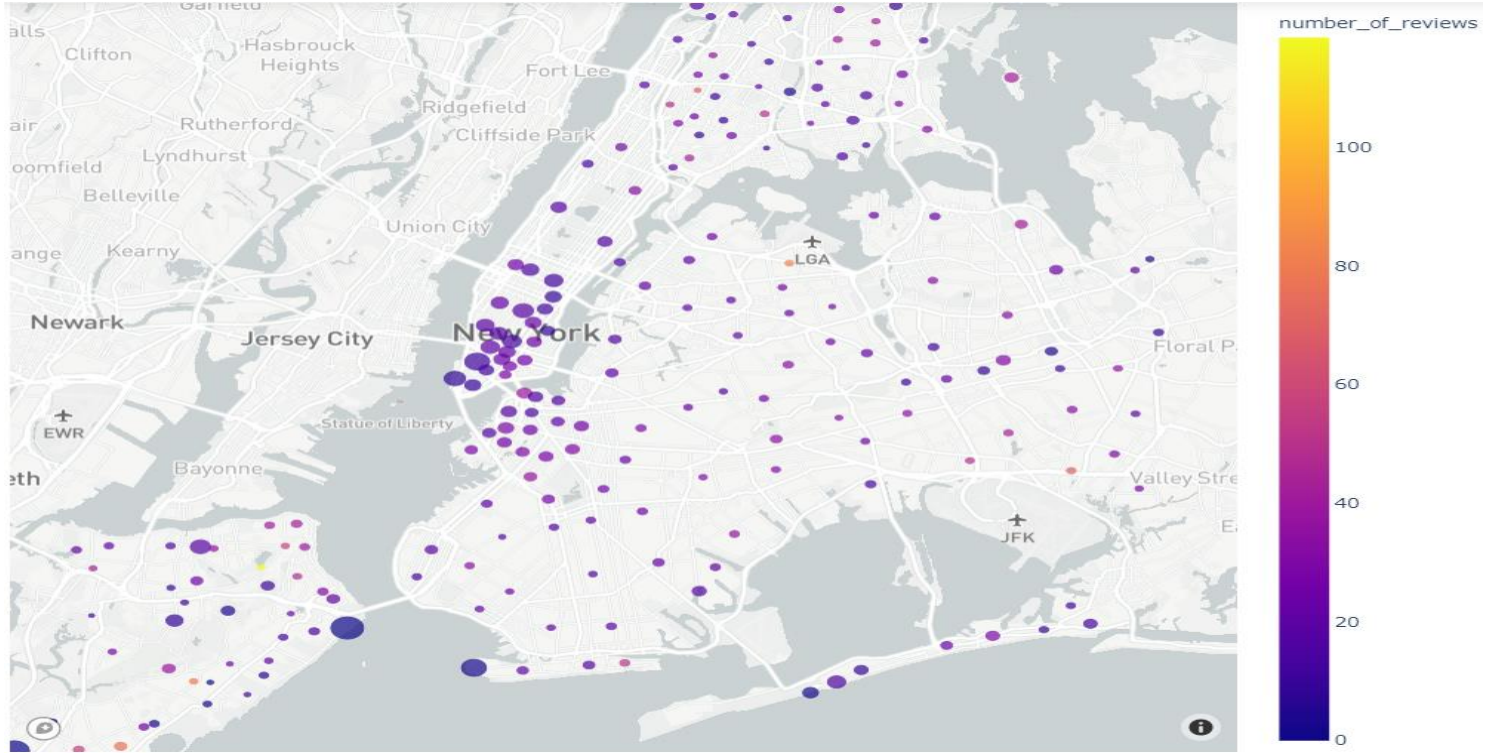
Discourage airbnb owners to keep airbnb's available for 88-90 days and 176-181 days ans people prefer renting for either 3months or 6months, and these figures are just 2-3 days less than 3months/6months.





Airbnb price for different neighbourhoods

- Two neighbourhoods in Staten Island command very high prices- Fort Wadsworth and Woodrow, even though Staten Island does not have many airbnb's. These two neighbourhoods also do not have any reviews on airbnb's, so insight is to decrease prices or remove them from airbnb.
- Silver Lake neighbourhood in Staten Island neighbourhood_group is very popular(has many reviews) but price is low. Insight is to increase prices here.
- Two neighburhoods in Queens also command high prices- Belle Harbor and Breezy Point.
- Almost all of Manhtan commands high prices as expected.
- Brooklyn, inspite of having high densities of airbnb's, does not command high prices except in some northern neighbourhoods like Vinegar Hill and Navy Yard, and even that could be due to their proximity to Manhattan.



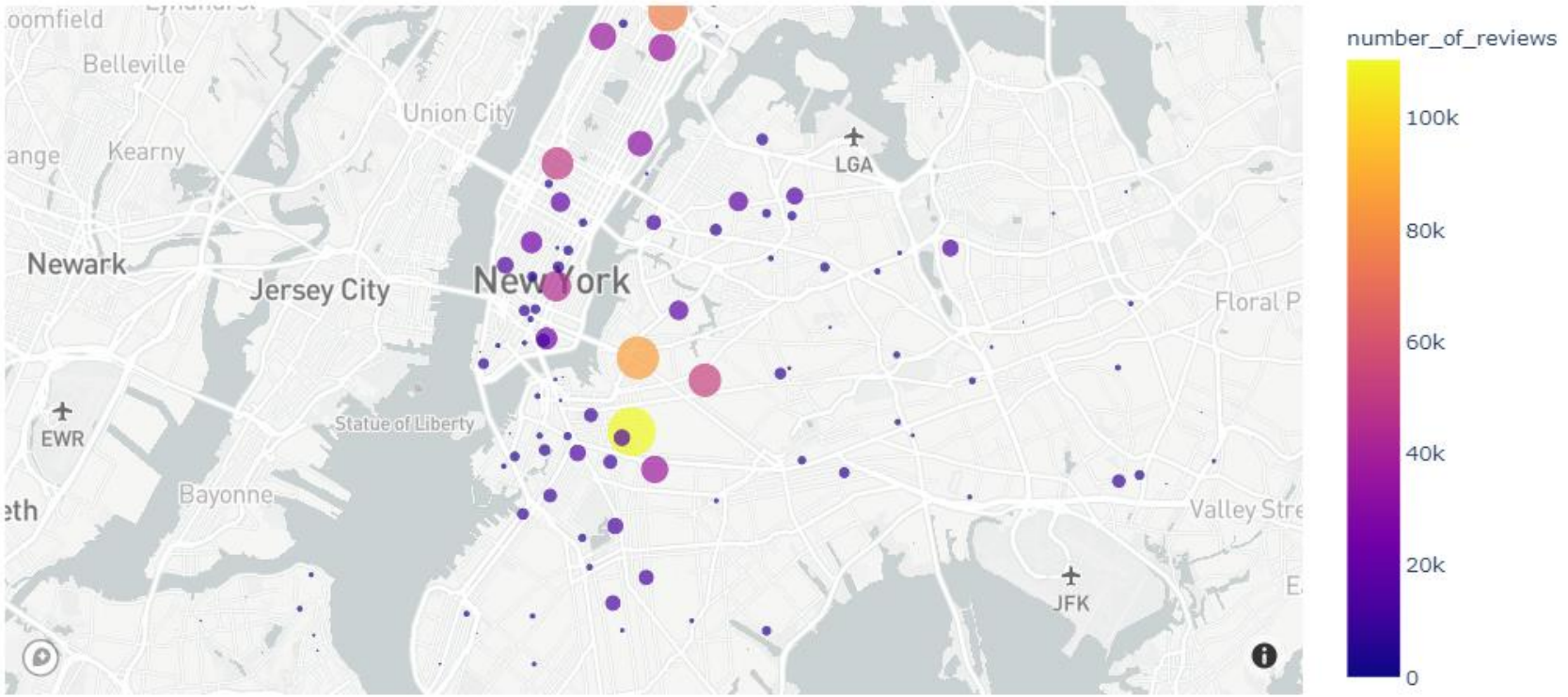
Number of reviews by neighbourhood

Used number of reviews as sum aggregation and plotted map using plotly express

Findings:

Bedford-Stuyvesant in Brooklyn has the highest number of reviews- seems to be popular.

Harlem in Manhattan and Williamsburg in Brooklyn also seem to be quite popular- they too have many reviews



APPENDIX - DATA ASSUMPTIONS

- The only two column that have significant number of nulls(~20%) are last_review and reviews_per_month. We will let those nulls remain. The rest of the dataset looks fine and does not need to be cleaned.
- Used Availibility_365 variable column using filters for with zero values and non zero values for different scenarios