A CASE STUDY TO IDENTIFY OPPORTUNITIES TO INCREASE OVERALL REVENUE OF AIRBNB

By Sagnik Chakravarty

Agenda or objective

- Airbnb has seen a significant decline in revenue the past few months but now that the covid restrictions are slowly being lifted, people are starting to travel once again as the world slowly but surely moves back towards normalization.

- Therefore this is a prime time for us to get back into the upswing of business again and this study aims to identify opportunities for us to take advantage of in pursuit of this goal.

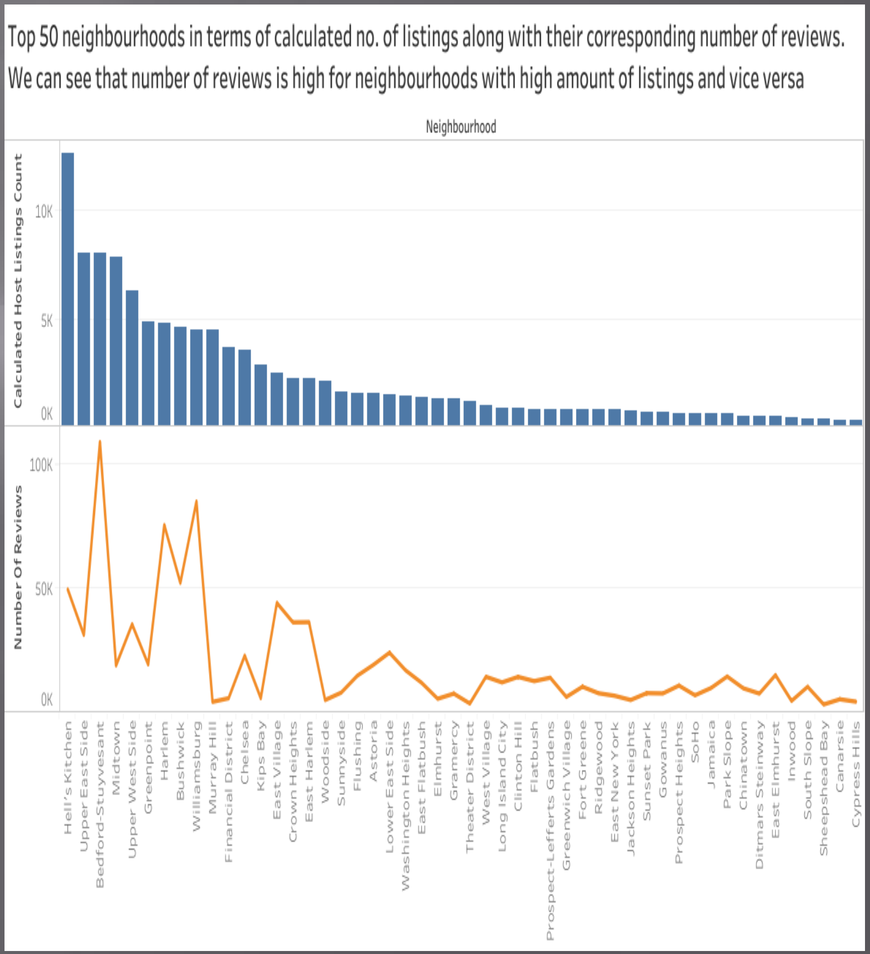
- This study aims to identify which types of hosts to acquire, the neighborhoods we need to target, price ranges preferred by our customers, most popular localities, etc so that our revenue can start it's upswing sooner rather than later.

Contents

- Choosing our target variable: Calculated host listings.
- Top 10 hosts to acquire: Host name vs calculate host listings vs availability.
- Calculated host listings is higher when availability is high and vice versa.
- Neighbourhood groups vs calculated host listings: Manhattan and Brooklyn are the dominant neighbourhood groups which customers prefer.
- Top 10 neihbourhoods with highest amount of listings.
- Calculated host listings vs price: Amount of listings per host is higher at lower prices and it decreases as price rises.
- Room type vs calculated host listings vs average price: Entire home apartments are the most popular room type.
- Conclusion.
- Methodology.

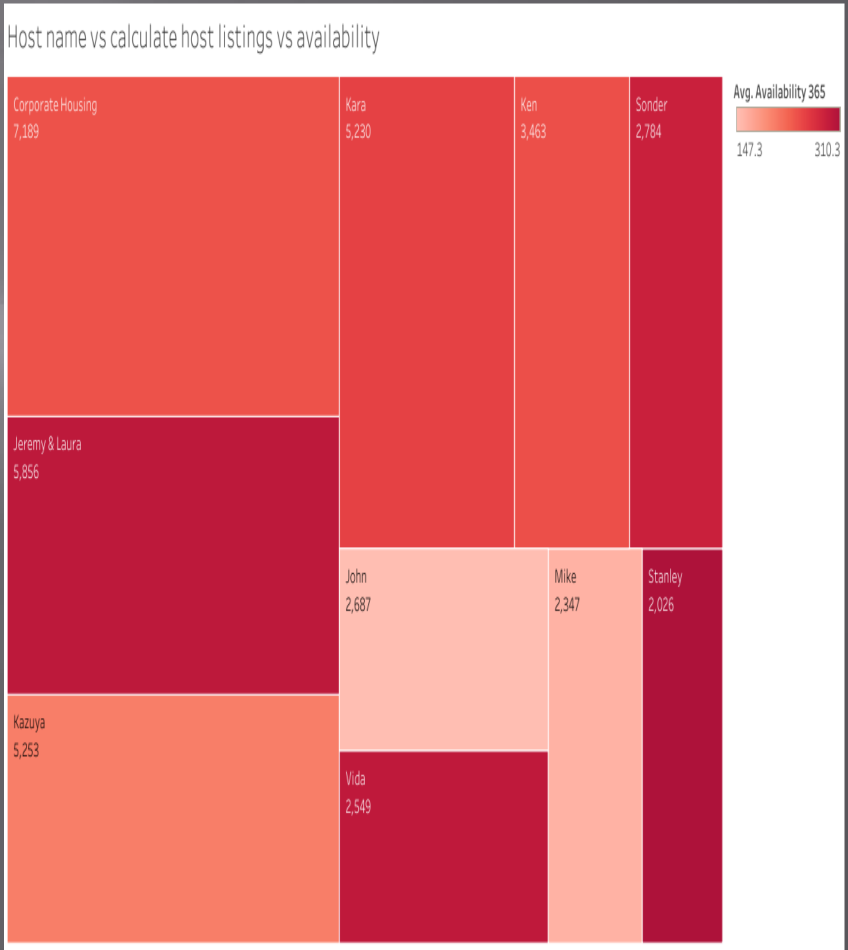
Choosing our target variable: Calculated host listings

- Calculated host listings is the amount of listings per host which is an indicator of demand for holiday homes since supply follows demand.
- We found that neighbourhoods that have higher amount of listings also get more number of reviews. As we know that reviews are a good indicator of customer engagement so we can conclude that higher amount of listings lead to higher customer engagement i.e. proving our earlier point about amount of listings being an indicator of customer demand.



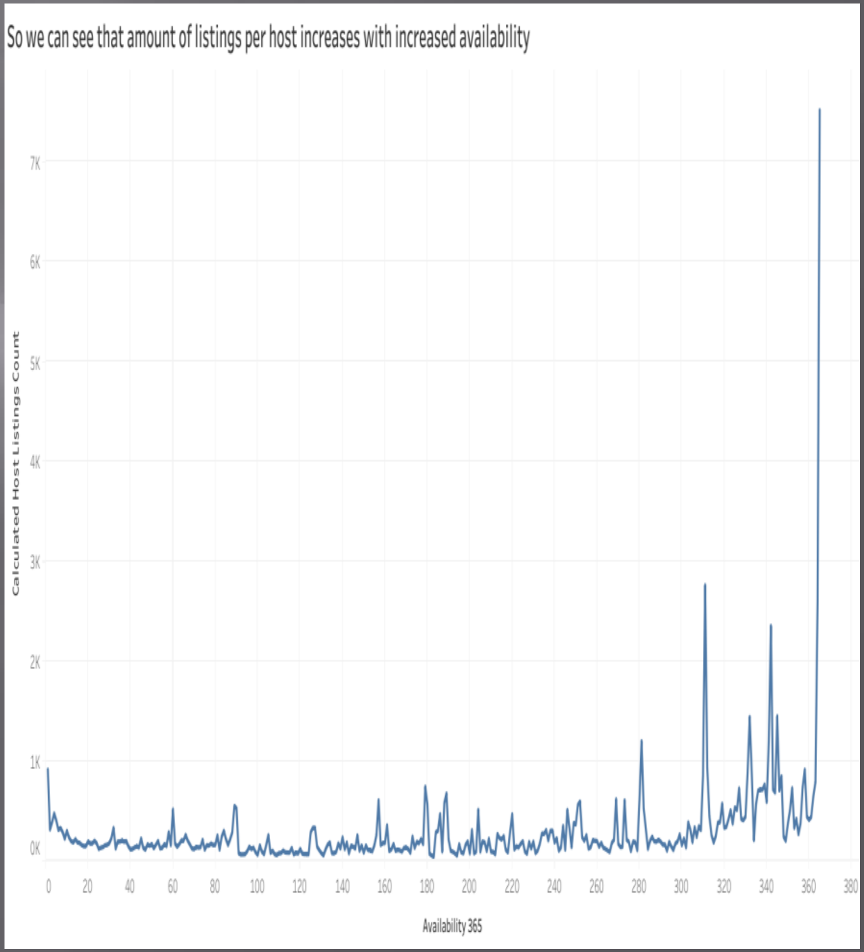
Top 10 hosts to acquire: Host name vs calculate host listings vs availability

- Corporate Housing, Jeremy and Laura and Kara are the top 3 hosts we should target for acquisition.
- We did this by creating a heat map and checked the top 10 hosts in terms of highest amount of listings per host. We assigned the size of the individual hosts in the heat map to calculated host listings i.e. greater the size of the grids higher is the host's calculated host listings measure.
- We assigned the color of the heat map to indicate the average availability of the hosts. So the more reddish hosts in the heatmap mean they have higher individual availability throughout the year and vice versa.



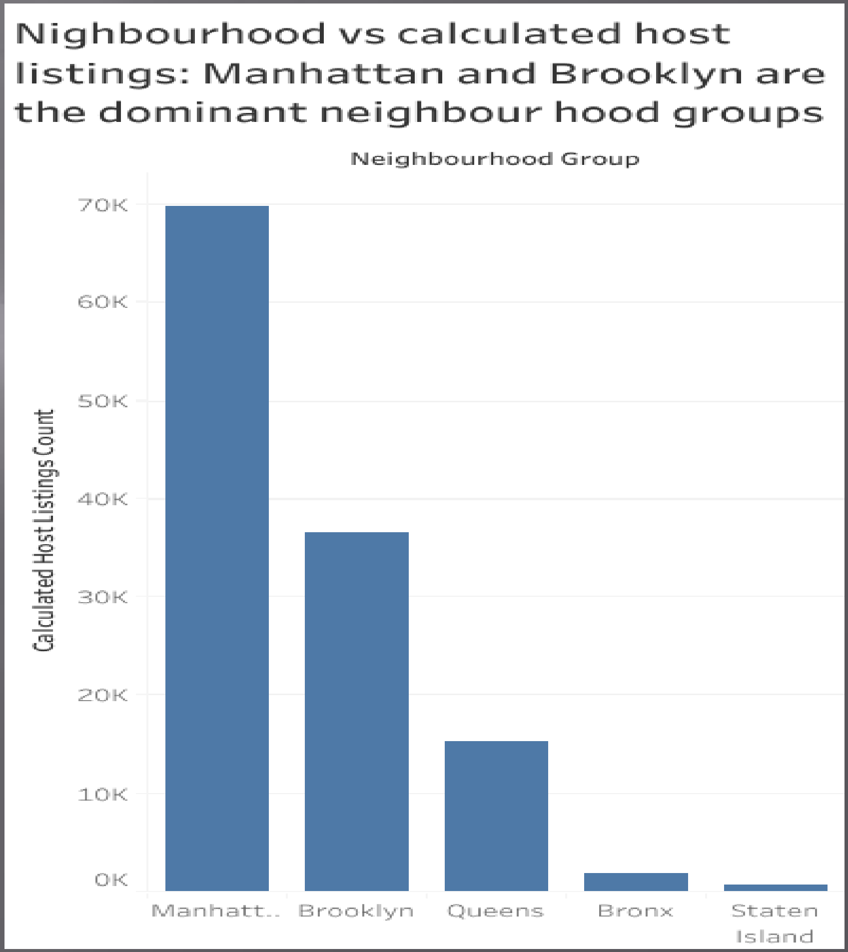
Calculated host listings is higher when availability is high and vice versa.

- We found that calculated host listings i.e. the amount of listings per host increases with increased availability.
- We plotted a line plot and saw that as availability increases the amount of listings increase.
- So we should target listings that have higher availability throughout the year.



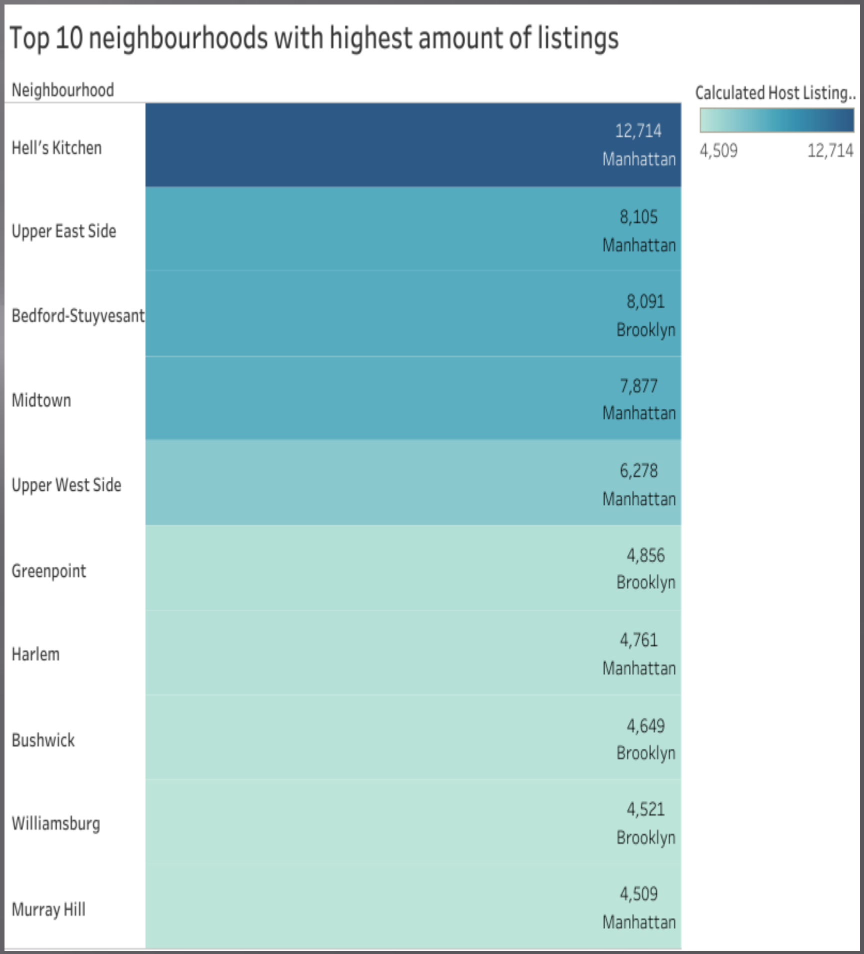
Neighbourhood groups vs calculated host listings: Manhattan and Brooklyn are the dominant neighbourhood groups which customers prefer.

- Manhattan and Brooklyn have most amount of total listings per host. As number of listings is an indicator of demand for holiday homes so we can conclude that customers prefer to travel to Manhattan and Brooklyn the most.
- We found this by plotting a bar graph as shown in the right with neighbourhoods in the x axis and calculated host listings in the y axis.



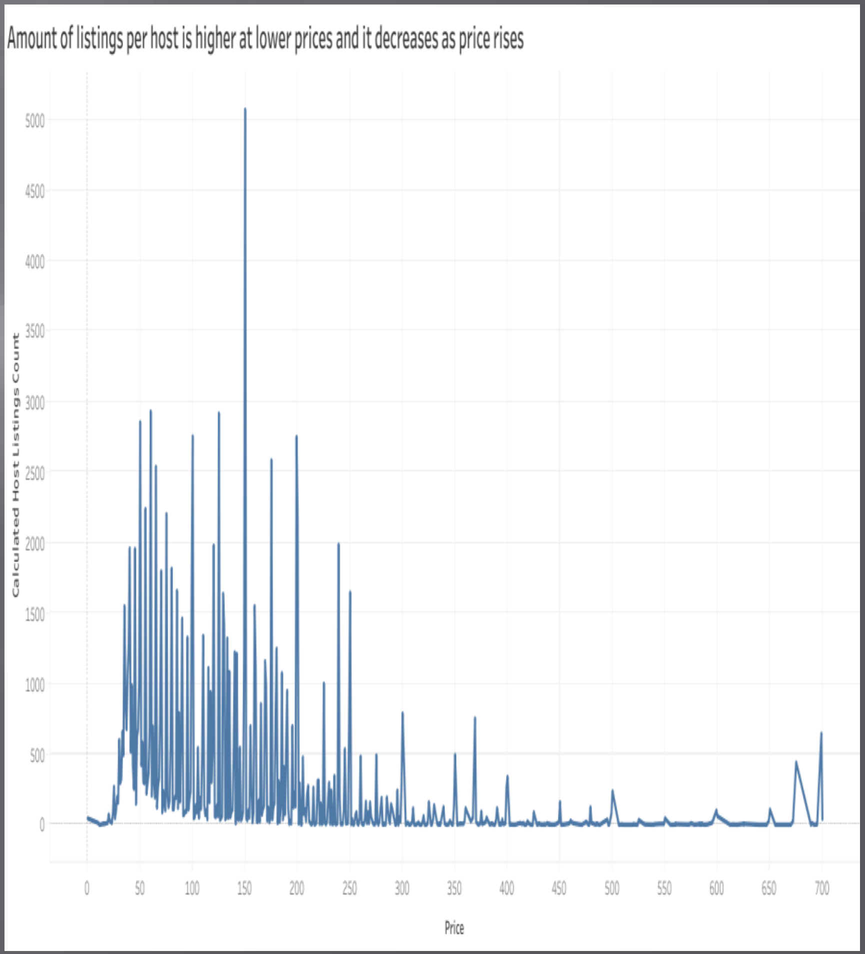
Top 10 neihbourhoods with highest amount of listings

- Hell's kitchen, Upper East side and Bedford Stuyvesant have the most amount of listings per host. Therefore we can conclude these are the top 3 neighbourhoods or localities with highest demand from our customer base.
- We created a highlight chart to find the top 10 neighbourhoods in terms of calculated host listings.
- We also added neighbourhood groups as labels to the chart.
- From the chart to the right we can see all ten of them are predominantly from Manhattan and Brooklyn thereby further proving our point that Manhattan and Brooklyn are the most popular neighbourhood types.



Calculated host listings vs price: Amount of listings per host is higher at lower prices and it decreases as price rises.

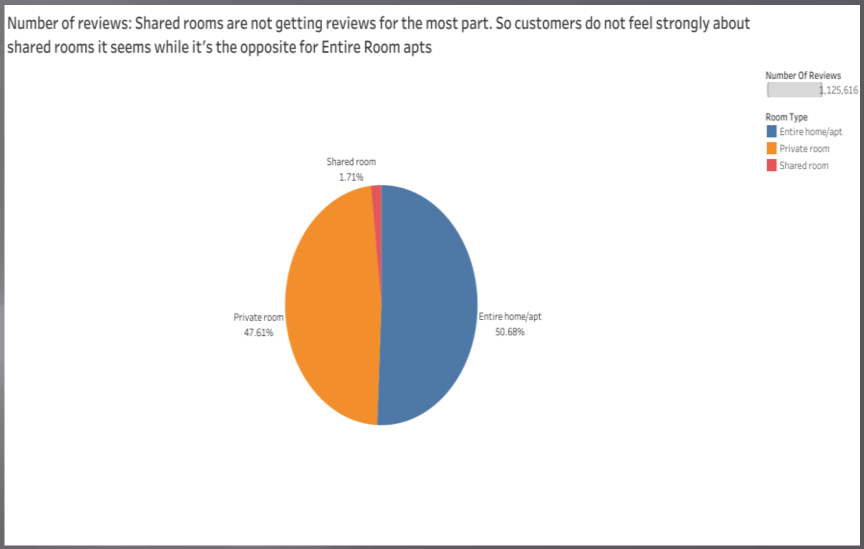
- We plotted a line graph to check the relationship between price and calculated host listings to check which price ranges are preferred by customers.
- We can see that with increase in prices the amount of listings per host decreases. We already concluded before that amount of listings is an indicator of demand thus conforming to the inverse relationship between price and demand.
- From the graph to the right we can see that the amount of listings is high for price range: $35 and $250.
- Hence we should be targeting this price range for attracting customers.



Room type vs calculated host listings vs average price: Entire home apartments are the most popular room type

- Entire home apartments have the most amount of listings with Private rooms following at second. Shared rooms have the least amount of listings per host indicating it has the least demand.
- We found this by plotting a bar graph with room type on x axis and calculated host listings on y axis.
- We also found that average price is highest for entire home apartments while shared rooms have the least average price.
- We found this by plotting a line chart with room type on x axis and average price on y axis.
- So taking these observations into consideration we can conclude that it's better to focus on entire home apartments and less on shared rooms.

- We created this pie chart to have a look at the percentage share of different room types in the total number of reviews shared by customers.
- From this pie chart we can clearly see that shared rooms are not getting reviews for the most part as it's share is miniscule compared to entire room apartments and private rooms
- So we can conclude that customers do not feel strongly about shared rooms it seems while it's the opposite for Entire Room apartments.



Conclusion

- Corporate Housing, Jeremy and Laura and Kara are the top 3 hosts we should target for acquisition.
- We found that amount of listings per host increases with increased availability. So we should target listings that have higher availability throughout the year.
- Manhattan and Brooklyn are the dominant neighbourhood groups which customers prefer.
- We also found that neighbourhoods that have higher amount of listings also get more number of reviews. As we know that reviews are a good indicator of customer engagement so we can conclude that higher amount of listings lead to higher customer engagement i.e. amount of listings is a valid indicator of customer demand.
- Hell’s kitchen and Upper East side are the most popular neighbourhoods.
- Customer prefers price ranges between $35 and $250.
- Entire home apartments have the most amount of listings with Private rooms following at second. Shared rooms have the least amount of listings per host indicating it has the least demand.
- We can also see from this pie chart that shared rooms are not getting reviews for the most part. So customers do not feel strongly about shared rooms it seems while it's the opposite for Entire Room apartments.

Methodology

- We first downloaded the AB_NYC_2019 dataset and imported into a jupyter notebook using pandas for cleaning the data.
- We imported all the necessary libraries like numpy, pandas, matplotlib and seaborn.
- We then quickly checked the dimensions and shape of the data for our understanding with info and shape functions and then moved on to the data cleaning step.
- We first checked for duplicates in the data. None were found so we moved on to checking for missing value.
- We checked for missing values in all the columns and checked percentage of the missing values for each column as well.
- We saw that columns name and host_name have 0.03% and 0.04% percentage of null values which is very miniscule. So we had no issue in dropping these rows from the dataset.
- We also see that last_review and reviews_per_month has 20.54% null values. We know that last_review is supposed to be a date column and if we replace the missing values with the mode or a 'missing' category. It won't help us with the further analysis as it won't be treated as a date column. So we are dropping these row.
- Now that the all the null values are taken care of we moved to the outlier checking process.
- We plotted box plots for the numeric variablles to check for outliers.
- We identified the columns that may have outliers from the boxplots and then conducted univariate analysis on them to further investigate if there are legitimate outliers or not.
- We checked for how big the difference in the 99th percentile value and the max value is or how big the gap between mean and median is to check for outliers. We also used business logic to determine if the outliers made sense or not.
- For e.g. in the column minimum_nights we found that there is a massive difference between the max value(1250) and 99% percentile value(31) which means there is something wrong. So we capped minimum_nights to 40.
- After dealing with outliers we finally arrived at the clean dataset which had 38003 rows and 16 columns.
- We then finally exported the clean dataset as an excel workbook so that we can import the same to tableau for exploratory analysis and generating insights.
- We then imported the clean dataset into tableau for the analysis, creating graphs and generating insights.