Process Book

- Motivation and background:
  In 2020, Airbnb made an initial public offer to the investors. This is a revolutionary decision not only because this time period is under the COVID-19 influence, but also the changing of the overlooked picture of traveling by the name of Airbnb. Instead of staying in a hotel, travelers now have an option to select a local host place where local people can list their house for travelers to stay and live like a local. As the business analyst students who live in New York, we are curious about what Airbnb has brought to the New York area, specifically the traveler perspective of staying within the Airbnb.
- Research questions:Here is the list of our research questions:

What is the lowest price in each New York area neighborhood (from 08/07/2018-04/07/2021)?

What is the highest number of reviews in each neighborhood (from 08/07/2018-04/07/2021)

What are the popular rooms? (from 08/07/2018- 04/07/2021)

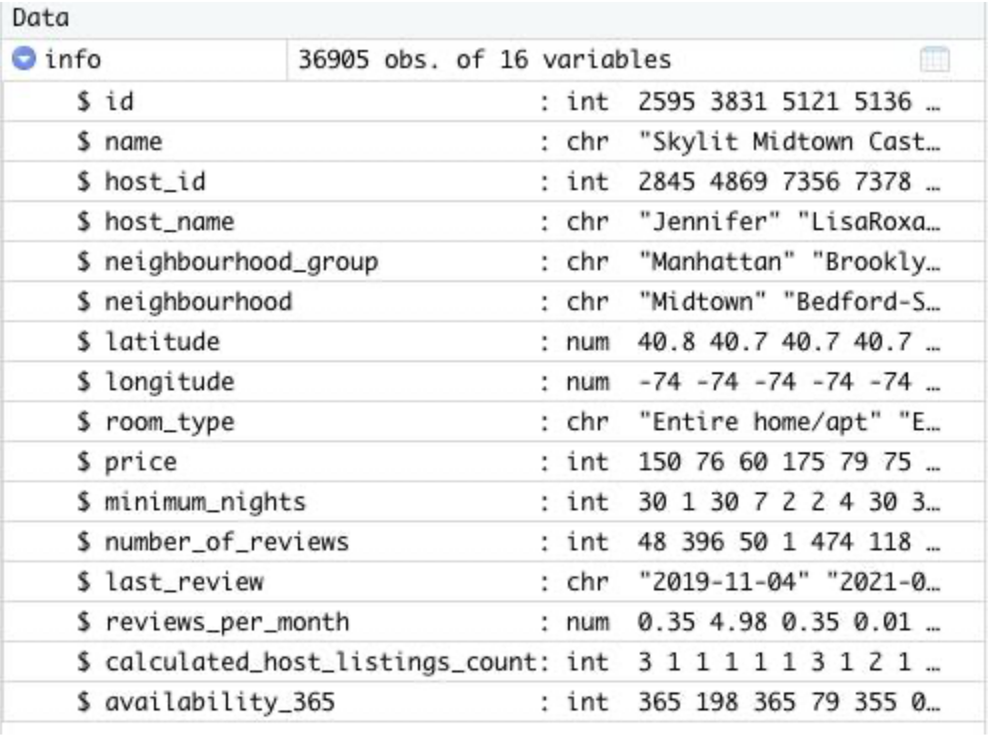
Which neighborhood has increased the number of rooms listed on Airbnb during the pandemic (from 08/07/2018- 04/07/2021)?

Why has this neighborhood increased the number of rooms listed (from 08/07/2018-04/07/2021)?

What are the most popular room types within that neighborhood (from 08/07/2018-04/07/2021)?

- Data: We found our dataset by searching on google with keyword airbnb dataset. The first website that pops up : http://insideairbnb.com/get-the-data.html. This is a website that collects public data from Airbnb. Once on this website we did another keyword search with New York and chose the dataset under this section name as “listing.csv” because the description mentioned that this is a dataset that is good for data visualization.

Furthermore, we use R with str() to show what we have from the dataset.



```
Data
info        36905 obs. of 16 variables
  $ id                            : int  2595 3831 5121 5136 …
  $ name                          : chr  "Skylit Midtown Cast…
  $ host_id                       : int  2845 4869 7356 7378 …
  $ host_name                     : chr  "Jennifer" "LisaRoxa…
  $ neighbourhood_group           : chr  "Manhattan" "Brookly…
  $ neighbourhood                 : chr  "Midtown" "Bedford-S…
  $ latitude                      : num  40.8 40.7 40.7 40.7 …
  $ longitude                     : num  -74 -74 -74 -74 -74 …
  $ room_type                     : chr  "Entire home/apt" "E…
  $ price                         : int  150 76 60 175 79 75 …
  $ minimum_nights                : int  30 1 30 7 2 2 4 30 3…
  $ number_of_reviews             : int  48 396 50 1 474 118 …
  $ last_review                   : chr  "2019-11-04" "2021-0…
  $ reviews_per_month             : num  0.35 4.98 0.35 0.01 …
  $ calculated_host_listings_count: int  3 1 1 1 1 1 3 1 2 1 …
  $ availability_365              : int  365 198 365 79 355 0…
```

- ETL:
  Extract: We Used R to open the .csv file. This file is from the local address and name as “listing” which was downloaded from insideairbnb.com.
  Transform: Once the file was opened on the R platform(read.csv), we applied str() function to check variables and observations. Furthermore, we use summary to detect any missing values(na). Then, we deleted all the na by using na.omit(). After that we use the library lubridate and dplyr to sort the dataset following the order of last_review. Then we made the type of last_review as Date. Then, we use library data.table to select the range we want to use for our project which is from 2018-08-07 to 2021-04-07. Finally we set this new variable to dataframe and write it as a new .csv file as our cleaned(transformed) dataset.
  Load: We load the transformed dataset to Tableau.
- Design of visualization: We considered bar, bubble, line and map type visualization. Moreover, we applied design principles such as don’t use too many colors and sort by certain order to help the audience easier understand the insight from our graphs.
- Implementation: We create filters to make our visualization looks dynamic. The filter that chooses a specific timeline for the user to select to do further analysis. (More)
- Results and conclusions:
  1.The most popular ranking category we created for travelers to use to show diversity choices while staying within the New York area.
  2. The neighborhood name as Manhattan has more houses listed on Airbnb platform. One of the reasons behind it based on our analysis is that the price within that neighborhood

became cheaper compared to the period before the pandemic. Another reason could be the increased number of hotel rooms in Manhattan.