ANALYSIS

OF

AIRBNB NEW YORK CITY

Student Name – Shreyas Vaidhyanathan

Student ID – 30305195

Tutor Name – Jeffery Liu

Lab – Thursday 6PM

INDEX

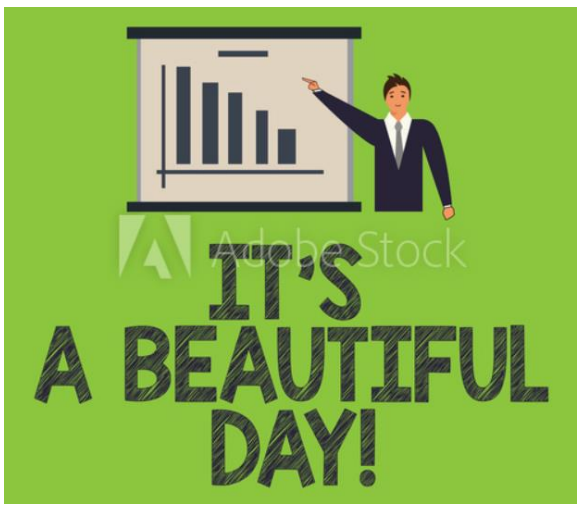
Introduction

AirBNB is an online marketplace that connects people who want to rent out their homes with people who are looking for accommodations in that locale. It is currently available in more than 81000 cities and 191 countries worldwide and is a viable option for the visitors as it helps them connect and stay with the locals at a reasonable price. This also helps them understand the local culture and their cuisines better.

In this project, we analyse the AIRBNB data and help come up with insights that help the visitors as well as the locals. The data set chosen has a few useful columns such as Name of the Listing, Description of the Listing, Neighbourhoods, Location of the Airbnb, property type, amenities provided and the price using which we help to answer questions like what are the keywords and amenities to be used and provided by the hosts in order to attract visitors, What are the closest Airbnb's to popular places like Statue of Liberty, Central Park, Empire State Building etc. We would also like to find out an overview how the Airbnb's listings vary over different regions and type of rooms that become available at different price ranges.

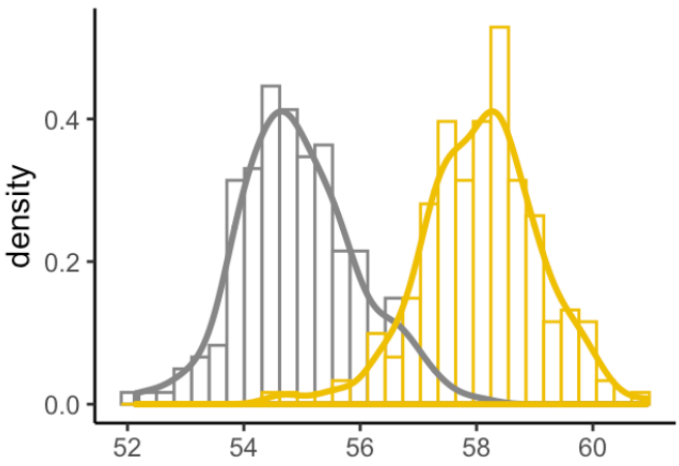
Design

In order to help the hosts figure out the keywords they should be using and the amenities to be provided, we had to reformat the data into 2 columns, a column that had the word / amenities and the other column that has a count of their occurrences. Once this is done, we need to decide on whether we would like to plot it using word clouds, bar charts, line charts, lollipop charts etc. We decide on going with the Word Cloud for the keywords as keywords are words that are suppose to describe the property and we wouldn't want to be using the top rated word if it does not describe the looks of the property. On the other hand, We believe amenities can be provided by any host and an order becomes extremely crucial in this scenario. Therefore, we prefer a bar chart to indicate the amenities that are common and the ones that are rare.



In order to show the users how close the Airbnb's are to their favourite tourist destinations; we go ahead and locate the Airbnb on the map. This way, the tourists have a peace of mind and a state of assurance as to how close they would be to their desired place of interest.

Later, we wish to decide on plots that help us with an overview of how areas affect pricing and what kind of rooms can one look for at a given price range. In order to answer the first part, we can choose graphs similar to a bar chart having various price bins or a density plot. In this case, we go ahead with a density plot as the number of Airbnb's are not evenly spread and some neighbouring groups have more listings than others. A Density plot in this works best as it indicated a ratio rather than an actual count of Airbnb's. In order to answer the second part of the question, i.e., the type of rooms one can expect for a given price range, we can either go ahead with choosing a bar chart or a pie chart. We go ahead with a pie chart in this case as the types are rooms are limited and a pie chart would look better and easier to interpret.



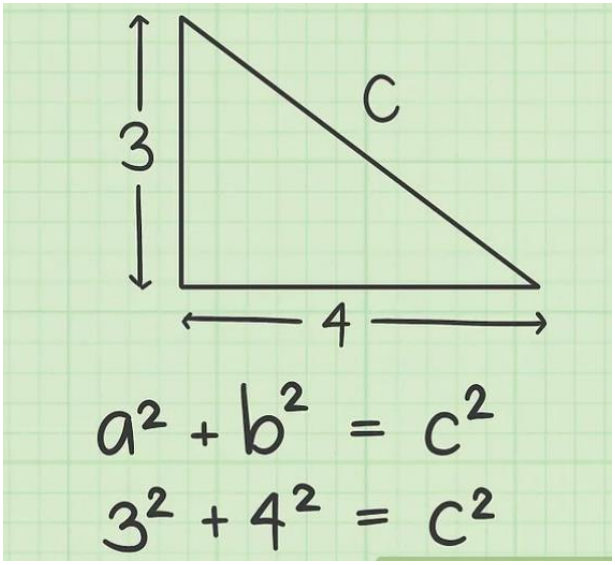
Implementation

In order to attain a working model that would help analyse all the questions above, we initially had to create 3 additional files, cheat_counts.csv and expensive_counts.csv that contains keywords used in cheap listings and their counts, and places.csv that contains list of popular places and their latitude and longitude. The first 2 data sets were created using the primary data set (cleaned.csv) and the latter has to be created from scratch with the help of Google.

The libraries we use for this package are :

```
# loading all the required packages
library(shiny)
library(leaflet)
library(ggplot2)
library(shinythemes)
library(wordcloud)
```

We then read all the required files mentioned above. We also create a few additional columns in our primary data set that holds the displacement of the Airbnb from various popular destinations. This is done using the hypotenuse formula.




```
temp_lat = places$lat[places$place == "statue_of_liberty"]
temp_lon = places$lon[places$place == "statue_of_liberty"]
clean$distance_sol = abs(abs(clean$longitude) - abs(temp_lon)) + abs(abs(clean$latitude) - abs(temp_lat))
```

We do something similar for all the popular places.

This is something I tried, and it worked wonders. Since we are only considering a very small part of the whole globe (New York city), we can consider the earth to be a flat surface. Although, this gives us a distance in degrees, it does not matter as we require is the closest 15 Airbnb's to each destination.

Once all the pre-processing is out of the way, it is time to build the UI and SERVER.

In the UI section, we initially select a theme for our website. We go with 'UNITED' as our theme because it matches with the colours of Airbnb's logo. We then mention the tabs needed. We create 3 tabs namely "Keywords and Amenities", "Hot Spots" and "Overview". Below we mention the options

provided for each tab. All this summarises the information for the sidebar Panel. Now, moving to the Main Panel, we indicate the plots and explanations that should appear in different tabs. In this application, we call them word_cloud, leaflet_map etc.

In the Server section, we plot the graphs considering all the options a user can select using conditional if else statements.

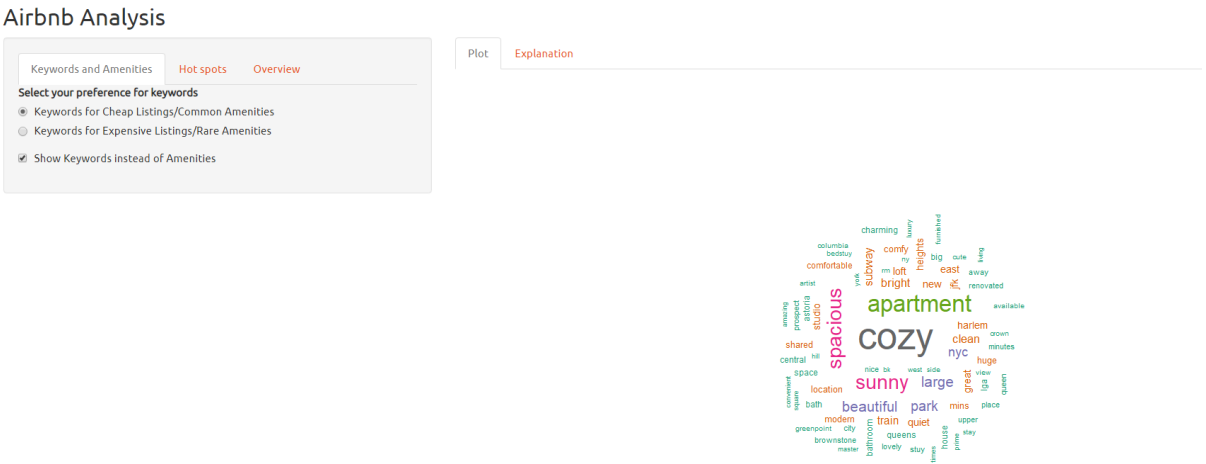
The whole app is later run using the following command.

```
# Running the application
shinyApp(ui = ui, server = server)
```

The code is saved as app.R.

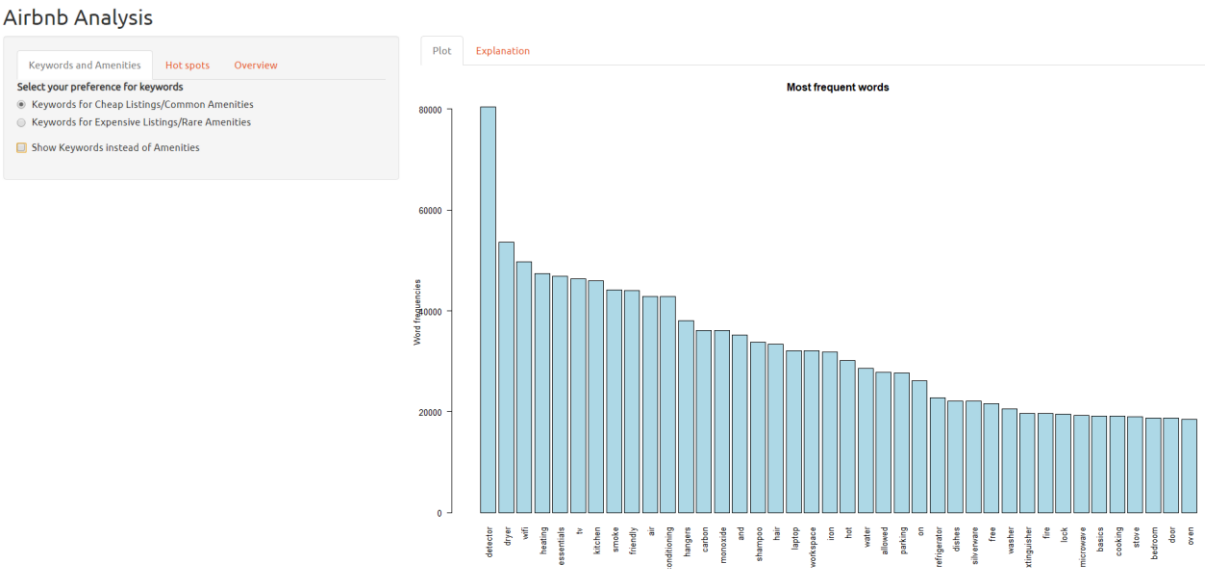
User Guide

The initial home page of the application is shown below. The default tab selected is the Keywords and Amenities tab. Initial the page shows a word cloud that indicates the keywords used in cheap listings.



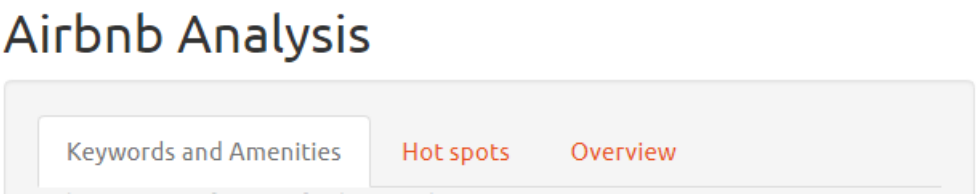
The user can then manually click on the radio buttons and check boxes to view the word clouds / bar charts as per his liking. The check box is to switch between the keywords and amenities whereas the radio buttons are used to switch between cheap and expensive listings.

The check box when ticked displays keywords using word clouds whereas when unticked, it displays a bar chart indicating the amenities.



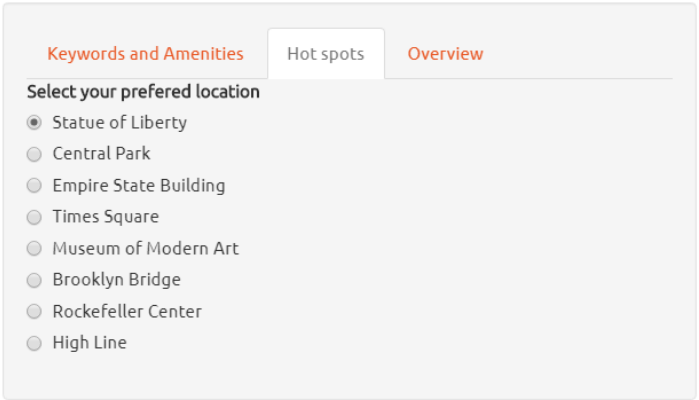
The Explanation tab beside the plots tab gives a short explanation as to the insights seen.

The user can also switch the main tabs that analyse / visualise a different aspect of the entire project.

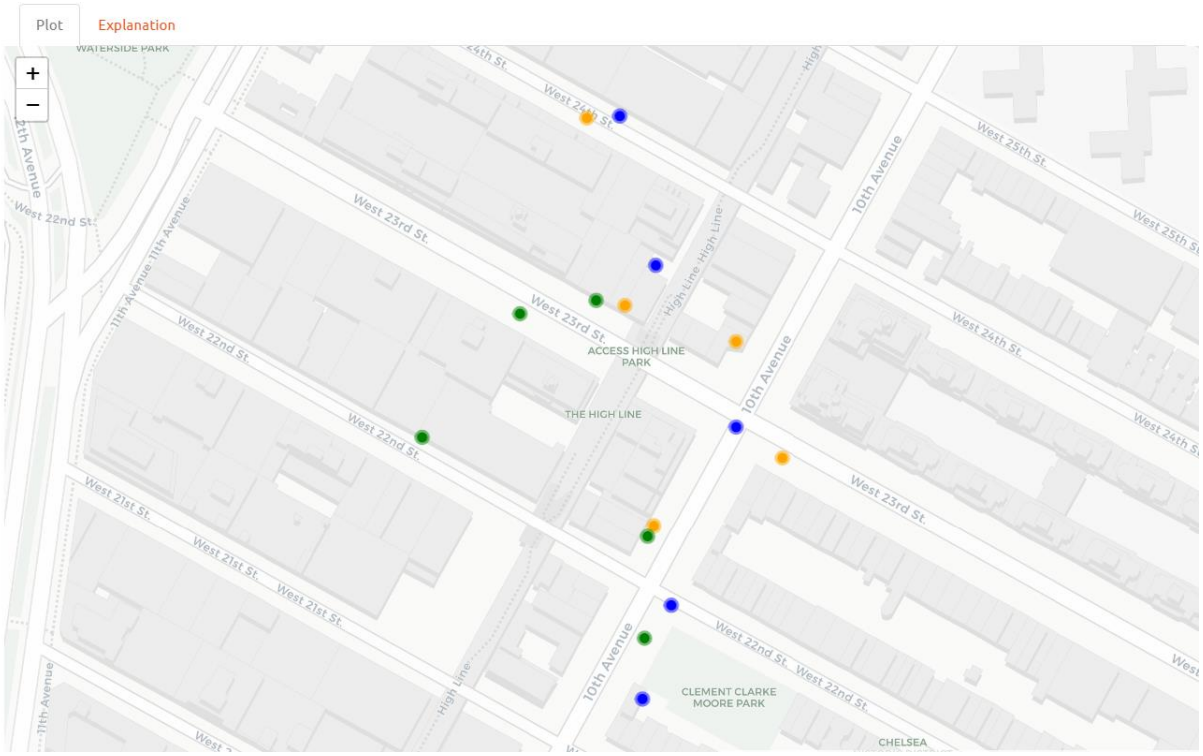


The Hot spots tab helps users visualise the closest Airbnb's to their favourite places. This is does with the help of a Leaflet Map.

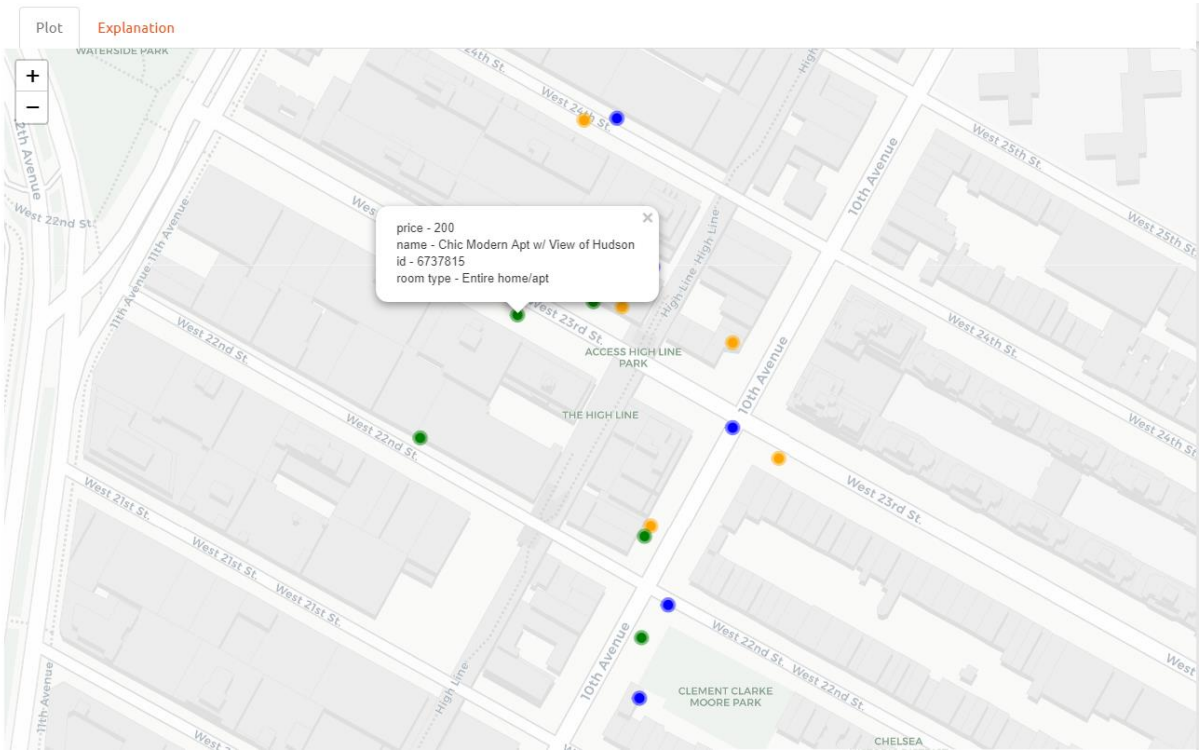
Once the user clicks on Hot Spots, the user now gets access to a new set of options to choose from. The user can now click on one of his favourite places to view Airbnb's closest to it.



The map below is generated when the user clicks on High Line.

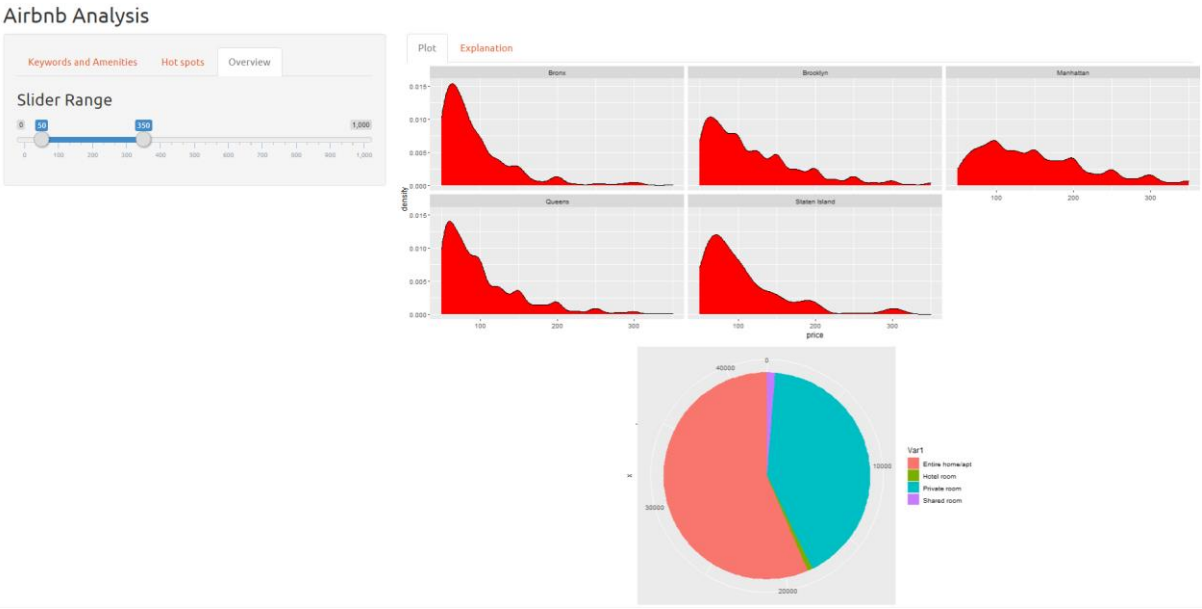


From the image, it can be seen that the user can view the Airbnb's location as well as the High Line to get a sense of how close the property is to the High Line. The user can also hover the mouse over the points to view a short description of the property and click it in order to view a detailed description of the problem.



The Explanation column also descries the Leaflet Map for the understanding of a layman.

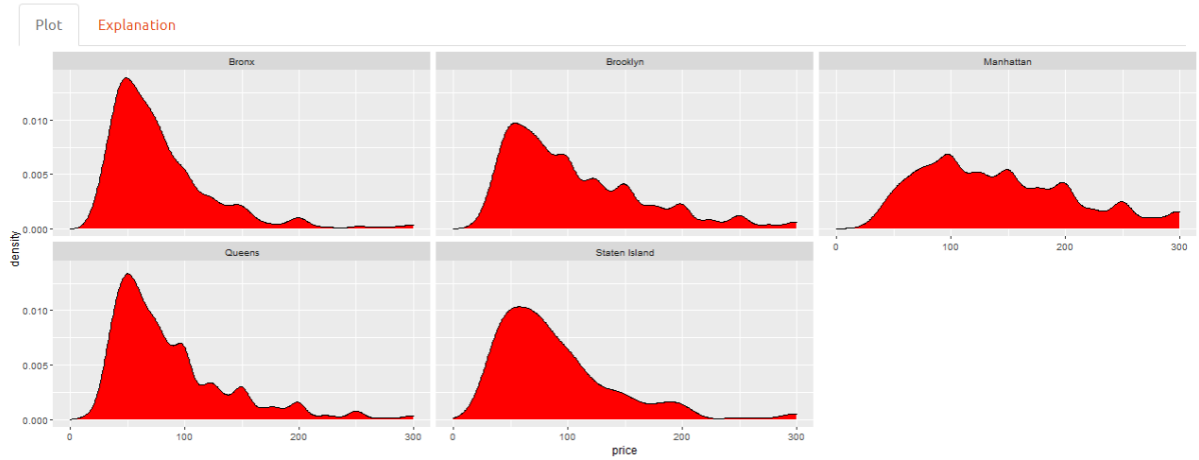
Moving on to the last tab (Overview tab), the user sees



The slider helps the user select a price range of his choice and the density plots and the pie charts refresh after every change made by the user.

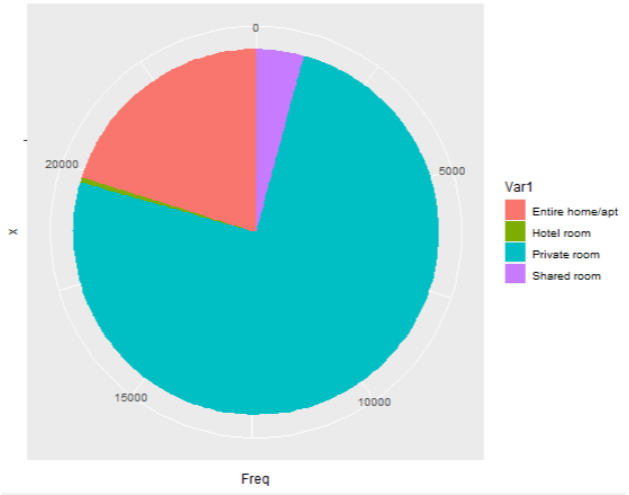
The density plots help visualise the ratio of houses at a particular neighbourhood group for a particular price range. Since this is a density plot, the Y-axis is scaled and helps in analysing how areas affect pricing. Also, the pie chart indicated the ratio of room types available at the price range selected.

The density plot for listings between 0 – 300 looks like :



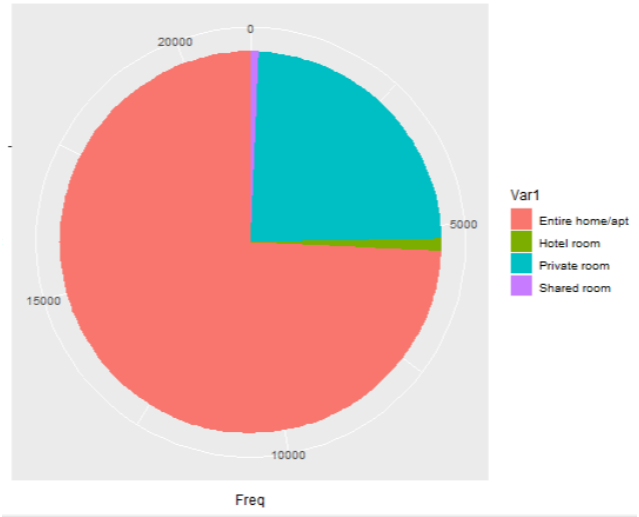
It can be seen that the ratio of houses for Staten Island, Queens and Bronx becomes negligible as we approach a price of $ 200 and more but the ratio of properties in Manhattan and Brooklyn are pretty significant indicating that they are expensive places to live in.

Also, on arranging the slider to a price of 0 – 100, the pie chart appears like –



This helps in understanding that when your budget is low, the ratio of entire house / apt becomes very less and therefore we must opt for a Private room as the options available are a lot more.

When the slider is used to select a price range of 100-200, the pie chart changes to -



This helps us understand that increasing out budget by a $ 100 helps us in opening various options for an entire house / apt.

Conclusion

In this project, we were successful in delivering the most impactful keywords along with finding out the most significant amenities provided by the various hosts. We were also able to successfully locate the closest Airbnb's to the most famous places of New York city using the hypotenuse formula, something I am really proud of in this assignment. Also, unknowingly but intuitively, we find that almost the properties close to the famous locations are expensive that the rest. We were also successful in building a user interactive web page for a user to decide on where he should stay, what type of room should they choose depending on their budget.

Overall, I feel pretty satisfied with the outcome. I would like to have linked this data set with the crimes of new York data set if we were provided additional time to try and see if there were a relationship between the crime intensive areas and the price, giving the visitors more information of places that they should avoid.

Bibliography

https://en.wikipedia.org/wiki/Airbnb

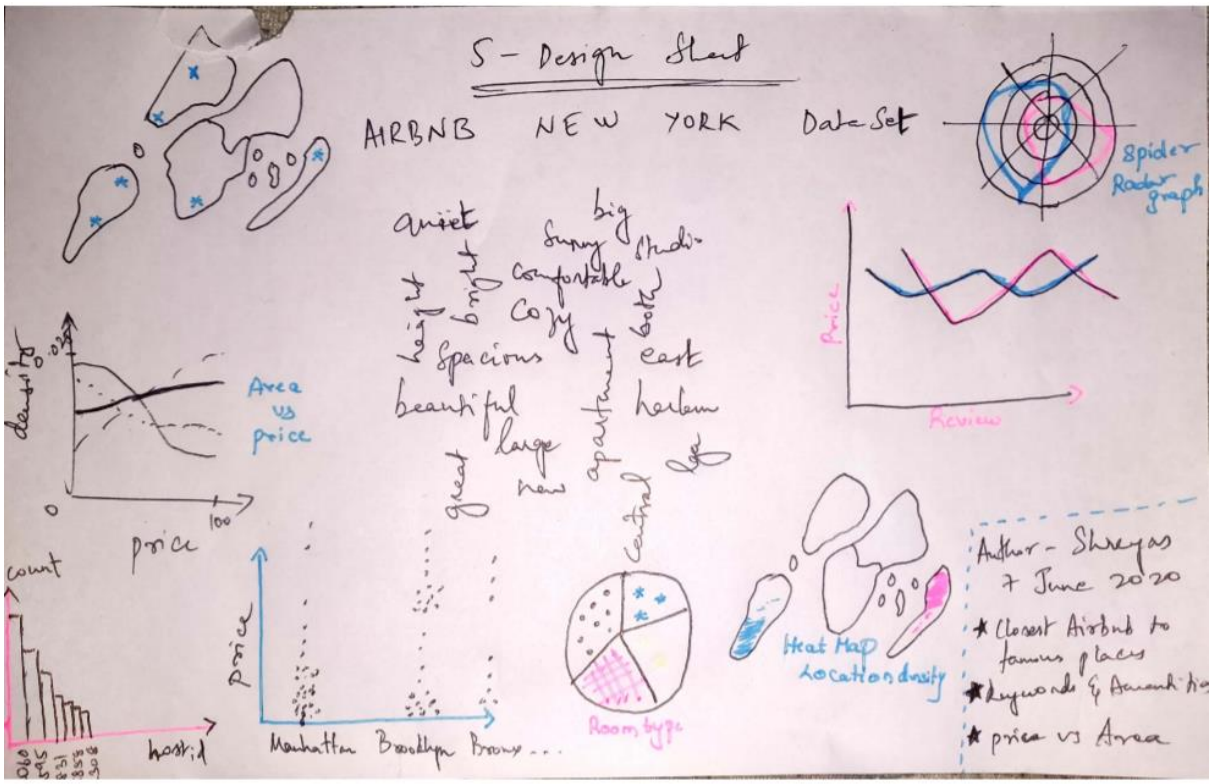
https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data

http://insideairbnb.com/

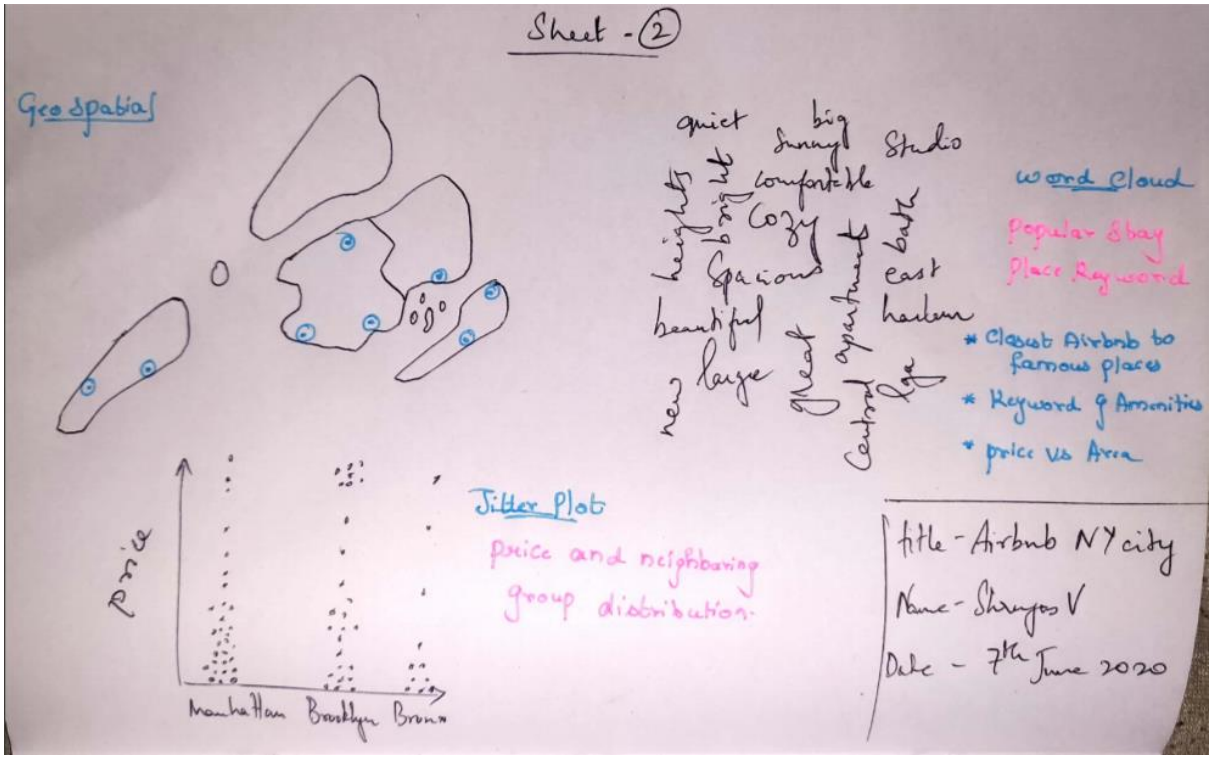
https://shiny.rstudio.com/

Appendix

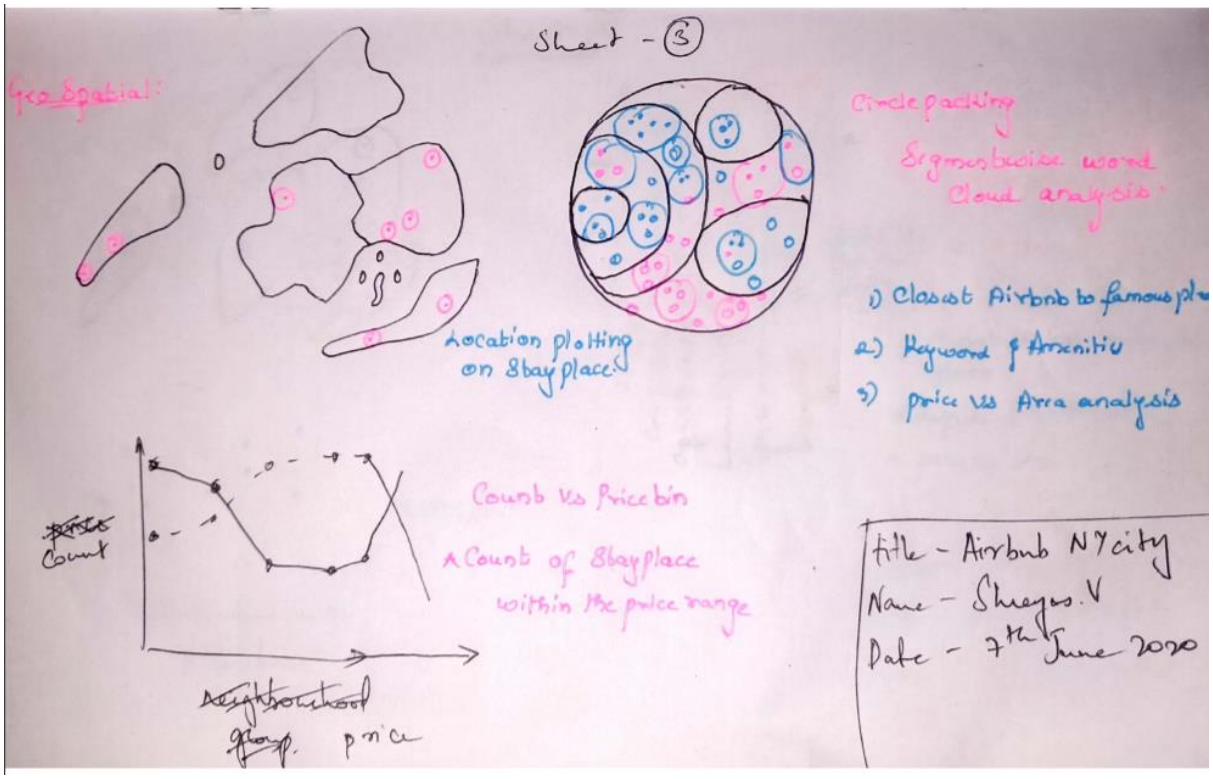
Sheet 1



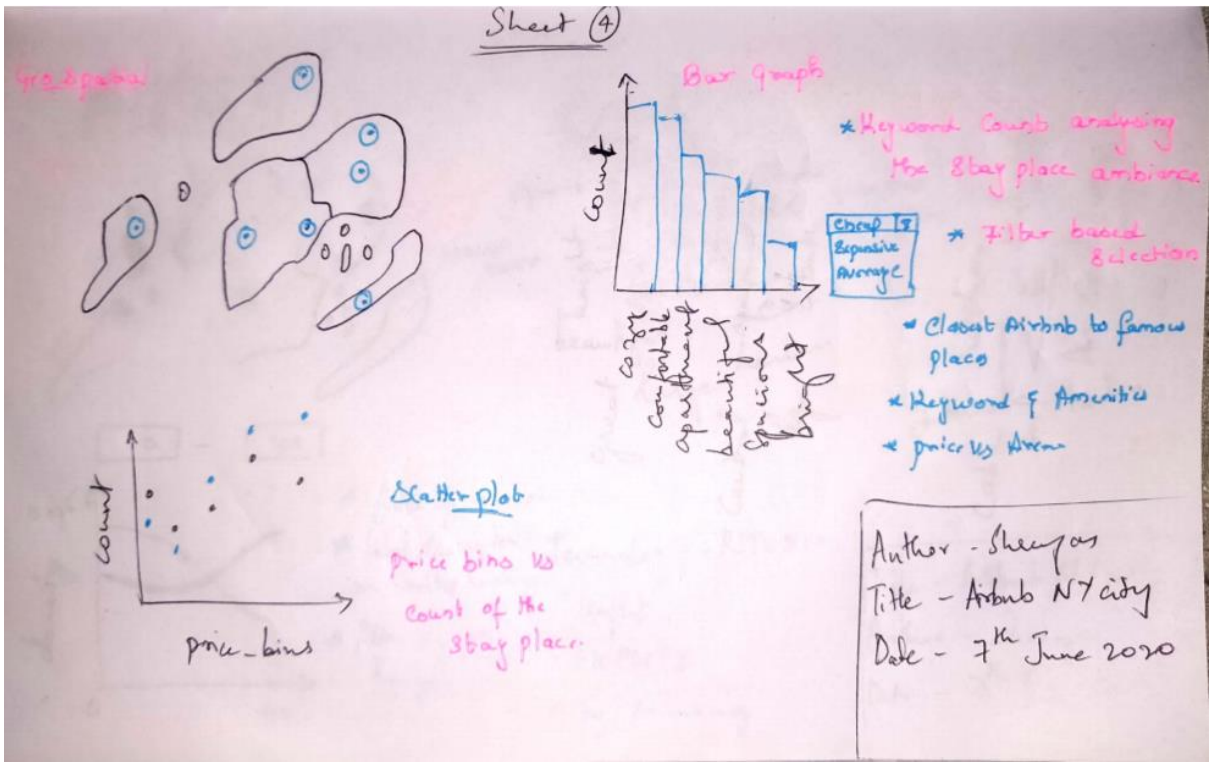
Sheet 2



Sheet 3



Sheet 4



Sheet 5