

# **Melbourne Airbnb Dataset Visualization**

**Srikanth Manthattan**

## **Table of contents:**

Introduction	3
About Airbnb	3
Contents of Dataset	3
Questions	3
Description of Dataset	3
Design	4
Sheet 1	4
Sheet 2, 3, 4	4
Final selected graphs	5
Sheet 5	5
Implementation	5
Ui.r	5
Server.r	5
Global.r	6
How shiny knows repetitive ui.r, server.r, global.r?	6
Required Libraries	6
User Guide	7
Guide for Map the lisitngs	7
Guide for reviews by time	9
Guide for Number of reviews w.r.t to neighborhood	9
Conclusion	11
Reference	11
Appendix	12
Sheet 1	12
Sheet 2	13
Sheet 3	14
Sheet 4	15
Sheet 5	16
Steps for executing the application	17

## Introduction

### **About Airbnb:**

Airbnb is the online hospitality service which enables the people to lease or rent their houses for short term which includes the apartment renting, private rooms rental, dorms or homestays. Melbourne is ranked as the 6<sup>th</sup> in the top ten list globally for airbnb users. In this project I want to explore and visualize the Melbourne Airbnb dataset. An initiative took by insideairbnb provides the Airbnb datasets of many cities.

### **Contents of Dataset:**

There are mainly two datasets listings and reviews. These datasets are in csv format. The listing dataset contains 105 columns in it. The dimensions of the listings and reviews datasets are 22909 \* 105 and 526415 \* 6 respectively.

### **Question:**

By exploring this dataset I want to answer the following questions.

1. Which locations have the most Airbnb's in Melbourne
2. What kind of Airbnb are present in Melbourne
3. Which websites provide the most Airbnb's data

### **Description of Datasource:**

The datasource is collected from [melbourne\\_airbnb\\_datasource](#).

- The listing dataset has 106 different columns and 22909 entries in it. Although there are many columns we need only few columns for our exploration. These required columns are then extracted and cleaned which is explained in the data wrangling section of this document.
- My major analysis and exploration will focus on the listings dataset.
- The reviews dataset has 6 columns and 526414 rows in it. This dataset is not required for my initial questions. But this can be used to extend my project with the following questions
  1. On what days of the week does the reviews will high.

## **Design:**

For creating a final application, I need to have an overview of what exactly I need to do. What choices of graphs I need to select to answer the questions. Which graphs can show the better visualization and easily understandable. For all these, I need to create a draft version of my application, there exists a beautiful concept named as five design sheet which we can use for creating a draft of our application. Based on these five design sheets draft as core our work in visualizing will be easier as we already have put down everything that we thought to do. Below is the five sheet design for my application. Each sheet is explained below and you can see the images of five design sheet in the appendix section

### **Sheet1 (Backtracking):**

This is the first sheet that I have created, in this sheet we need to do a brain storm and get an idea which graphs can be used to answer the question. Collect all of the graphs in this sections, and we need to filter out which are the best out of all collected graphs.

For displaying Airbnb on the map I have used a leaflet open street heatmap as this has a bright map, and users can easily identify the location.

For showing the different airbnbs average pricing, reviews, rating score we can use a pie chart and a histogram.

For displaying the trend of reviews over years, a bubble chart or a trend time series can be used as these will explain the trend of any substance very clearly

A stacked bar chart or a grouped bar chart can be used to display the different count of Airbnb grouped by different roomtypes with respect to the councils

### **Sheet2, sheet 3, sheet4 (Layout):**

In the sheet2, 3, 4, we have filter out our ideas that is we need to go through each every graph, map that we brainstormed in the first sheet and select the appropriate one.

For displaying the reviews trend, I have select a trend time series graphs over a bubble chart. Because, the bubble chart is little more complex to understand the review when compared with the trend graph. Users can easily understand the trend ups and downs over the years in this Airbnb market.

Rather a simple pie chart for displaying for the average price, average rating, average review score, a histogram will do the work easily for coding. In pie chart, it takes percentage of all and display, it would be difficult for the users to calculate the percentage instead if we use a histogram which directly displays the total count users finds it more easy and likely to use.

**Final graphs selected are in bold below:**

**Pie chart & Histograms**

**Time series graphs vs bubble chart**

**Grouped bar chart vs stacked bar chart**

**Leaflet open street hot map**

**Sheet 5(Implementation):**

In this 5<sup>th</sup> or final sheet, we will decided which map should be displayed first over the other and how our application should be that is what needs to be displayed and what comes first and then after. So to answer all of my questions I have clearly selected the leaflet map for showing Airbnb should be the first interactive page that should appear and an histogram which displays the average counts. Then after a reviews trend graphs to show the history of the airbnbs, how it has evolved in Melbourne, at what time it all started, how the Airbnb markets is now. This second page will be very useful the investors who are trying to invest their money in airbnbs. And the final page in my application is going to display the number of airbnbs, what kind of airbnbs i.e., different roomtypes present in the different parts of the council. A grouped bar chart will clearly distinguish the data and display it to the user so that he/she can easily understand to see which is the most popular council for airbnbs.

## **Implementation:**

The interactive visualization for this Melbourne Airbnb has been done in shiny.

Different files that are created as part of this implementation are

- i.      Ui.r
- ii.     Server.r
- iii.    Global.r

### **Ui.r:**

This is r file in which I have represented all the User interface components like slider input, select input. Various kinds of user input accepting widgets has been placed to make it interactive so that the user can select different options in which he wants to see the data and understand clearly.

### **Server.r:**

This is the file in which all the data handling for visualization is being done. On the basis of users input with several widgets in the ui.r, I have used a reactive function to filter the data accordingly in the server.r and send the responses back to the ui.t to display.

### **Global.r:**

This is the file in which all the required libraries are loaded, all the required data files are loaded other manipulations like assigning respective color to the factorized data, created choices for the user input widget.

### **How shiny knows their repetitive ui.r,server.r, global.r?**

After creating these three files, to run a particular shiny application we need to place these three files into one folder. Any replicated file with the same names as above files should be present in that folder.

### **Required libraries:**

**Shiny :** This is the default library that should be loaded every time you create a shiny application

**Shinydashboard:** This library is used to create an interactive and good looking dashboard in the user interface. Rather than simply creating a basic application, if we use dashboard user can find it attractive and likes to use over application over other which has similar idea.

**Leaflet:** The leaflet library is used to show the list of airbnbs on the Melbourne map.

**Dplyr:** This is a data manipulation package , the functions such as filter(), summarise() provided by this package are very useful in the server.r for manipulating the data based on the user input.

**Shinythemes and ggthemes:** These are packages that are provided by R, to make the application look good. There are different themes that we can use to our application.

**Ggplot2:** This package is used to draw the bar graphs, histograms in the visualization

**Dygraphs:** This is javascript charting library, this package provides various kinds of chart visualization on the xts time series data.

**Xts:** As mentioned above the dygraphs needs the xts time series data, I have imported this package to get the required time series data.

**Rcolorbrewer:** This package is used to assign different colors to the different types of factored data.

In this visualization project I have used fluidrow to divide the page accordingly for selecting the inputs and for displaying the results. A shiny theme ‘yeti’ is used for creating a good visuals.

## User guide:

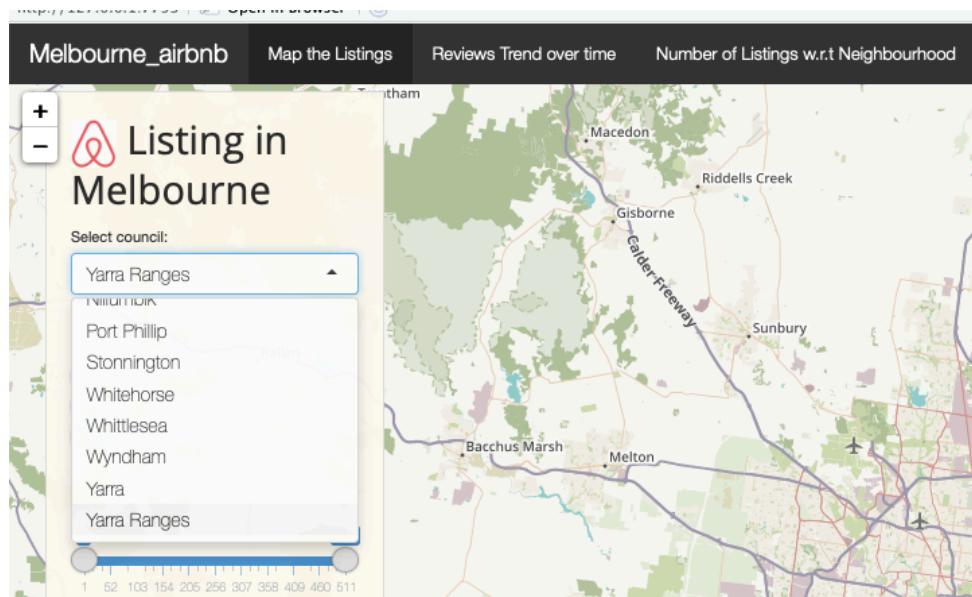
This visualization has three main tabs in it for visualizing the data. They are,  
Map the listings,  
Reviews by time,  
Listings, Neighborhood and Hosts.

### Guide for Map the listings:

This is the homepage or main page whenever this application is launched this is the first page to display.

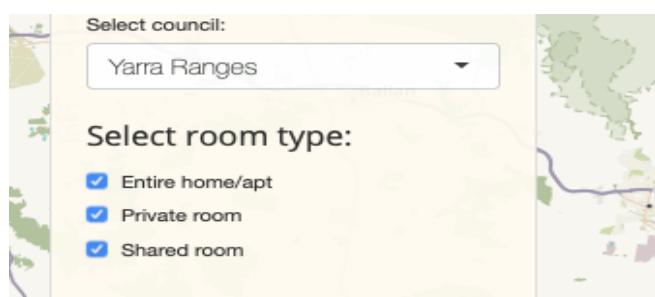
Inputs widgets:

**Dropdown:** for selecting different councils present in Melbourne



Click on the drop down and select the council which you interested. The default selected council will be Yarra Ranges

**Checkbox:** to filter the different room types



After selecting the council, click on the selected roomtype to filter out the data. By default all room type will be selected.

**SliderInputs:** for filtering the airbnbs based on the reviews, price, and rating



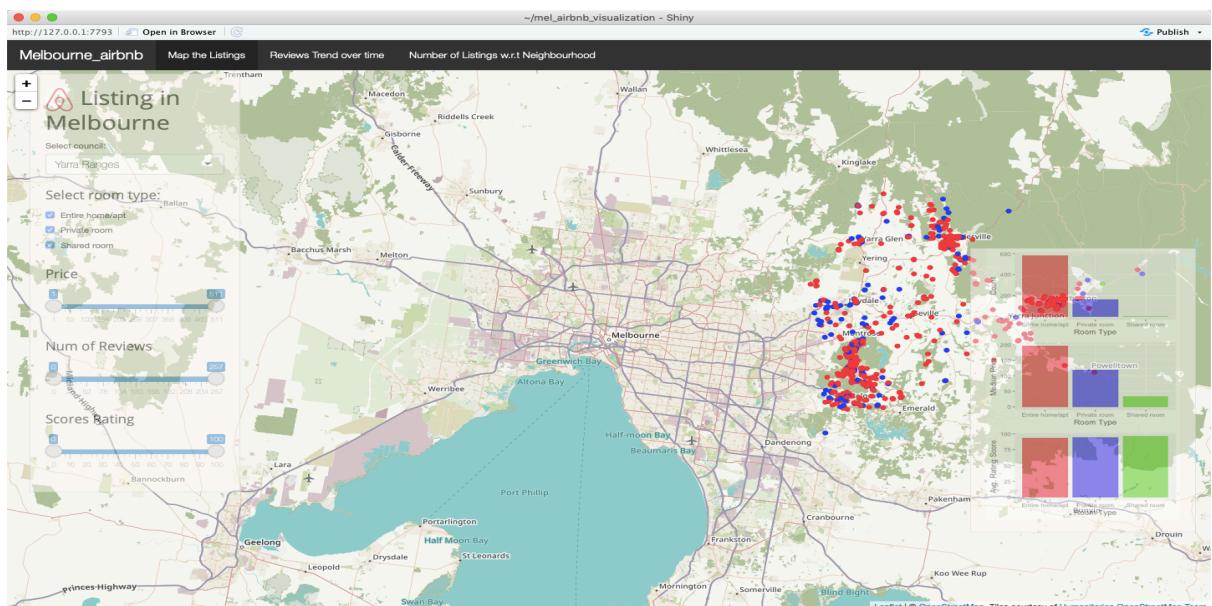
These are three different sliderinputs that we have for price, reviews and rating score for different airbnbs. Please select you required range to subset the airbnbs for clearly visualizing.

**Note:** The input selections are defined as draggable, these can be dragged from any position and can be placed at any position

Outputs:

#### Leaflet Map:

In the leaflet map, I have used ***open street hot*** map because this map has clear outline with bright outlines.

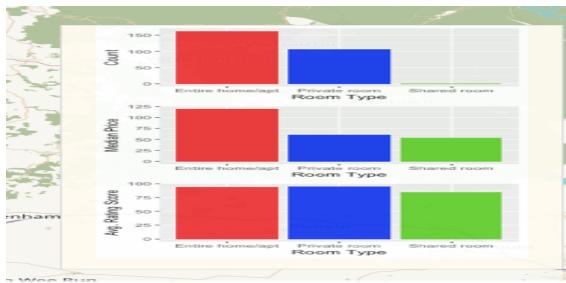


You can see various dots that are being pointed on the Melbourne map. These are different Airbnbs present and are represented with different colors.

- █ Entire room /apartment
- █ Private room
- █ Shared room

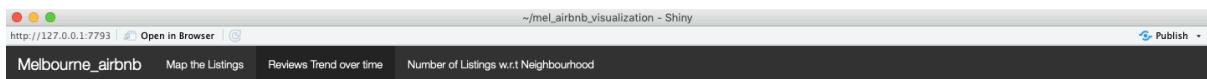
#### Statistics output:

There is also an other output that is being displayed in this page. It shows the median of the salary w.r.t to roomtypes, count of different roomtypes, Average rating price. These all are reactive and the output histogram will change based on the inputs selected.



### Guide for Reviews by time:

This page gives the user about the information of different years review trends. User can give roll number to clearly see the review trend. Hovering on the trend line will display the number of reviews, the month and the year on the top right side.



### Number of Reviews Over Time

The graph on the left side is an interactive trend graph

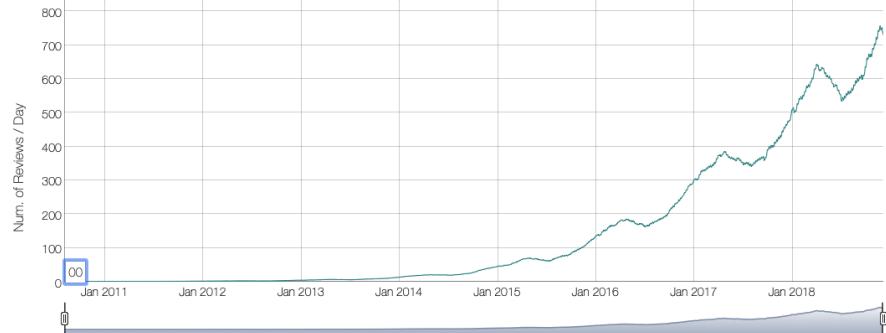
Hover your mouse pointer on the graph to see the number of reviews

The X axis provides the month and Year, Y axis represents the total count of reviews

The range selector at the bottom can be used to zoom any review and check clearly

You can also specify the roll period to see the trend clearly

Show Grid ?



### Guide for Number of listing w.r.t to Neighborhood:

In this page, the user has two kinds of inputs. One is the dropdown for selecting what kind statistic should be shown (count, percentage). The second input is the score ratings slider input.



## Listings by Neighbourhood

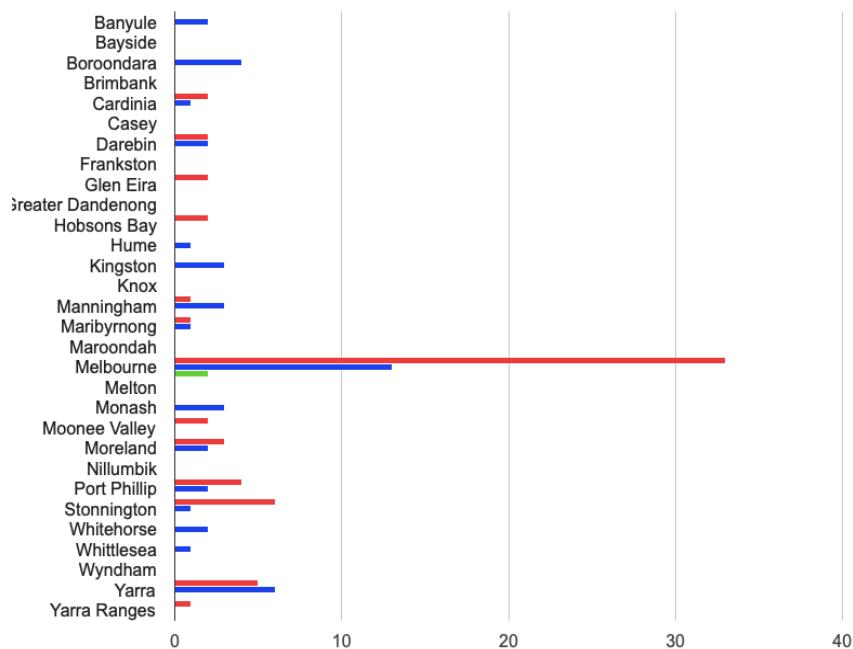
Showing as:

Count

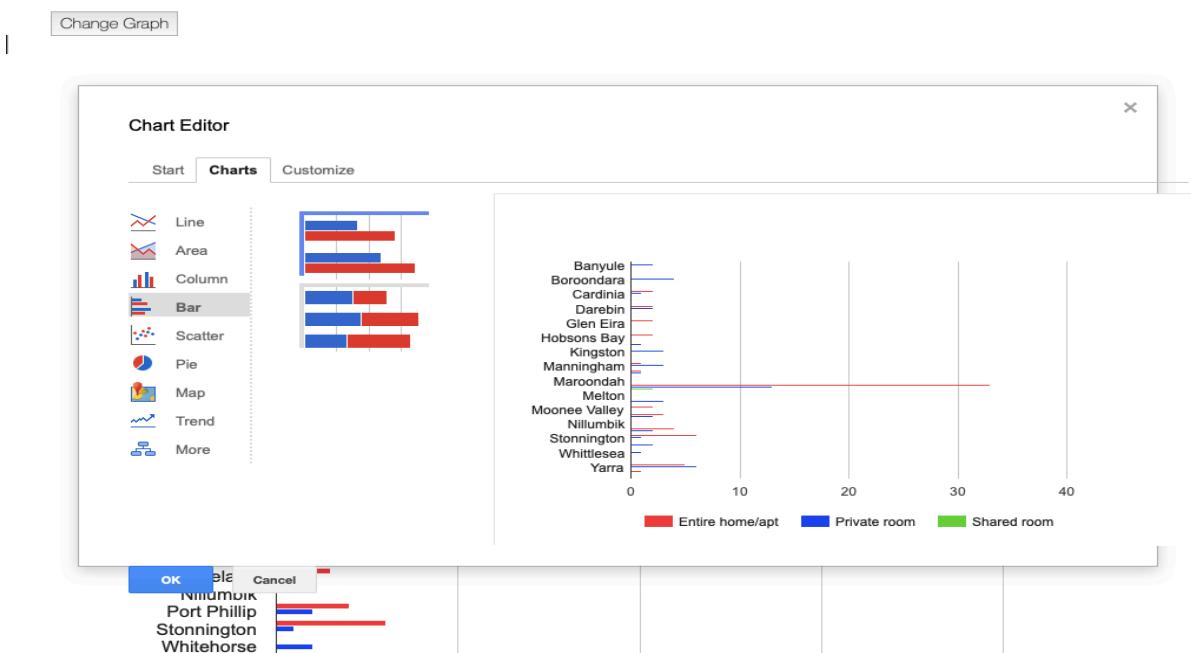
Scores Rating



The output is a bar graph which will display the different count of roomtypes in the different councils.



I have used renderGvis for this page, the specialty of this library is that it can create its own graph if the input is the time series data. The user can click on the change graph button to select different type of different graphs.



## Conclusion:

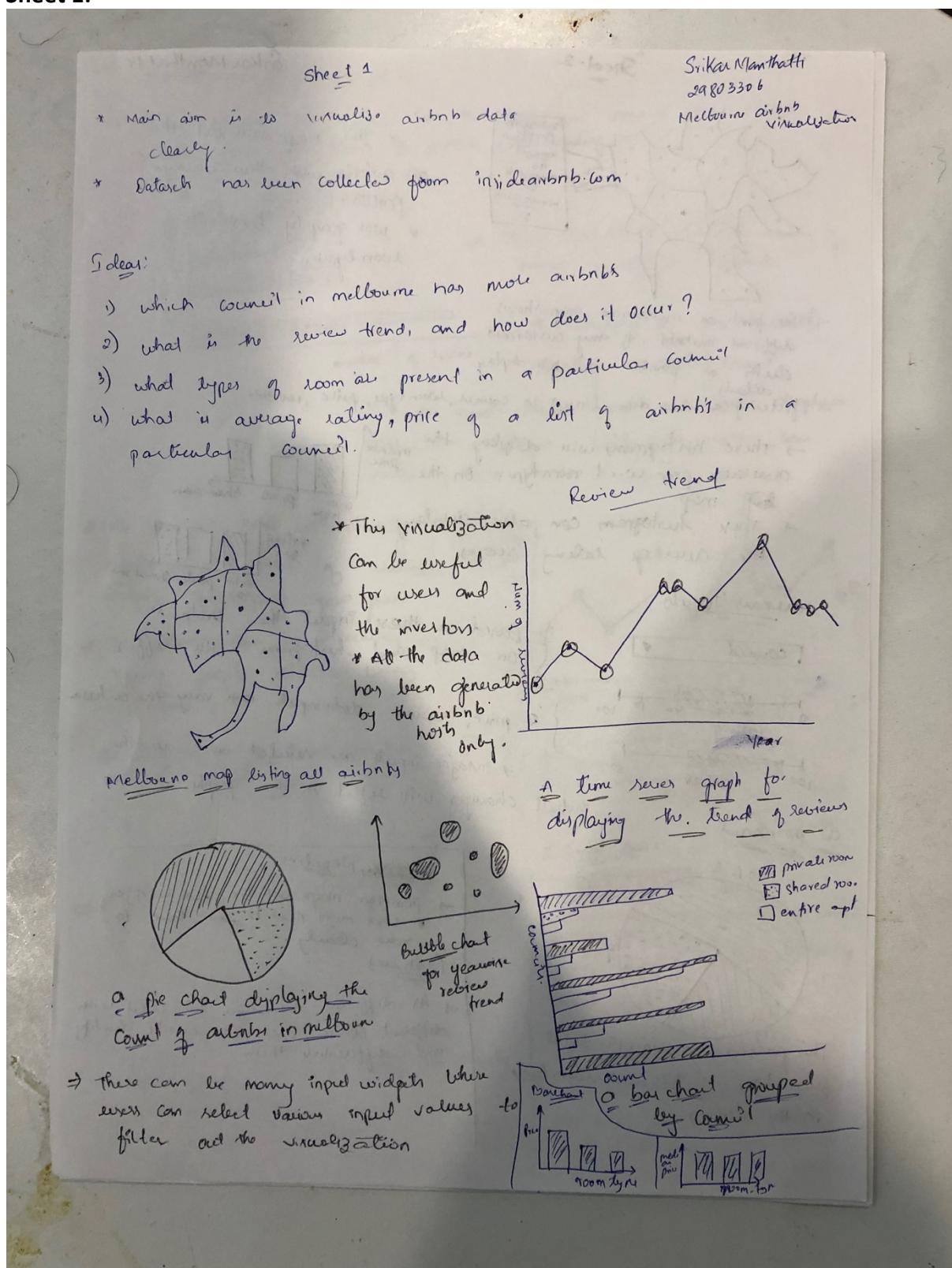
In this project for Airbnb visualization, I have learnt a lot of insights on how to analyze the data, how to clean the data, how to make interactive visualization on the given data so that a naïve user can easily understand it. This dataset has been downloaded from [insideairbnb.com](http://insideairbnb.com). The initial dataset has a lot of unused and noise in it. In this process of visualization, I have learnt on how to clean the data, after cleaning the data how to extract useful insights from that data so that any underlying trends or information can be identified.

## References:

- Basic maps with shiny <https://5147.yalongyang.com/book/module03/>
- CSS for Shiny <https://shiny.rstudio.com/articles/css.html>
- Dates and times in R <https://www.stat.berkeley.edu/~s133/dates.html>
- Empirical Cumulative distributive functions <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/ecdf.html>
- Five design sheet <http://fds.design/index.php/resources-and-publications/>
- How to choose the correct graph <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>
- Inside Airbnb <http://insideairbnb.com/get-the-data.html>
- Shiny Examples <https://shiny.rstudio.com/gallery/superzip-example.html>
- Shiny code quality <https://shiny.rstudio.com/articles/#code-quality>
- Shiny Themes <https://rstudio.github.io/shinythemes/>

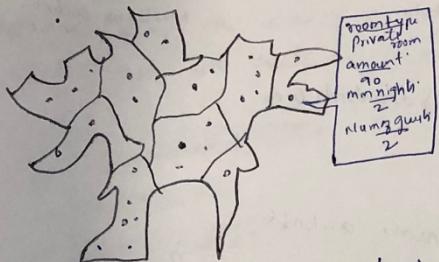
## Appendix:

### Sheet 1:



## Sheet 2:

Sheet-2



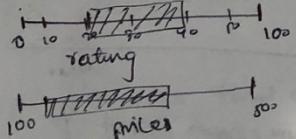
Srikan Manthattri

- \* This map will list the all airbnbs with their average miles.
- \* will group by based on room types.

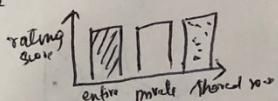
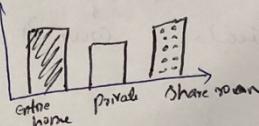
- The point on melbourne map shows different airbnbs, if any customer clicks on points it will display details.
- filtering can be done based on council, room type, price, review.
- These histograms will display the average price w.r.t room types On the ~~map~~ map
- These histograms can also display the average rating scores

Various Inputs:

[council]

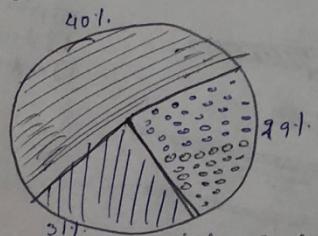


} Based on these input the Output on map and histogram will differ



- } a price, review slide inputs can vary the outcome
- \* Major inputs are selected and all the output changes will reflect on the graph.

a pie chart



This pie chart displays the total proportion of different room types in the airbnbs.

Positives Negatives

- position maps will be less in size → user might need to zoom them to view clearly.

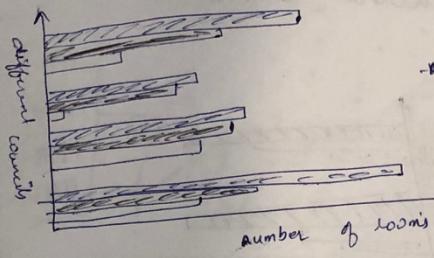
Positives

- All different room types are coded with different colors, users can easily identify and differentiate them.

### Sheet 3:

#### Sheet 3

- After displaying the airbnbs on the map we need a more clear visual that will create a precise understanding.



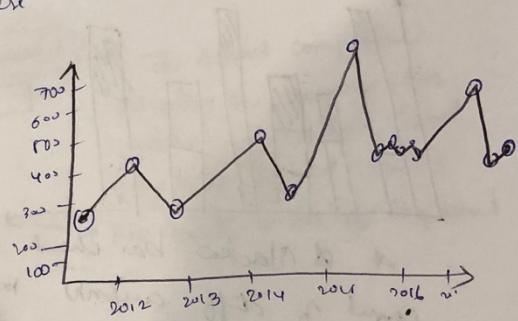
\* These bar charts are the number of houses in different councils

\* hovering on this should display the total count

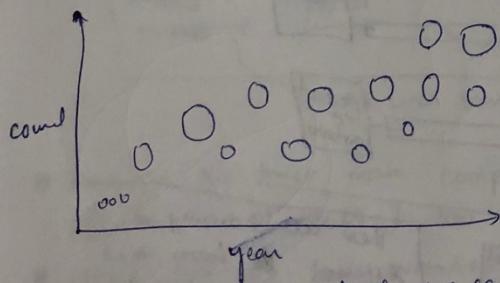
\* An input slider and a count dropdown will be displayed left side, user can select these

\* The review trend will clearly explain the trend of various airbnbs from past few years

\* This trend can be increasing as many people started using airbnbs,



An alternative for trend chart is bubble graph

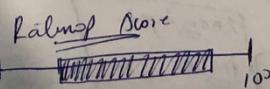


\* In bubble chart clearly we can divide the council and show the density of reviews in the no. of bubbles

\* 

Count	display
-------	---------

 based on count or percentage



Positive  
\* The review trend is clearly displayed, user can view it, investors can also clearly see the profits in airbnbs

\* the bar charts displayed will let user know, which area has more airbnbs & which room type is high.

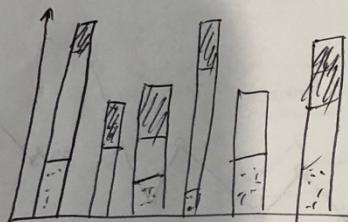
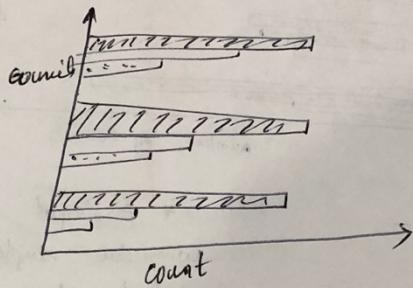
**Sheet 4:**

Sheet 4

Final month 7)

- \* As all the required metrics can be ~~seen~~ clearly explained with the before mentioned visuals.

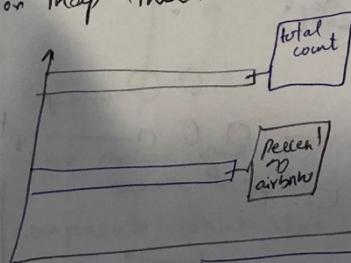
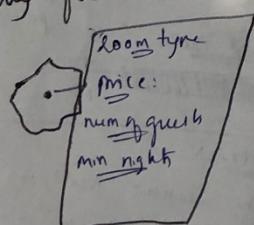
- \* Here a user can select to change the graph and display it as the pie chart or ~~selected~~ time trend



- \* A stacked bar chart can also be used to review the count of diff accounts in a given commt

Focus

- \* Any focus (or) hovering on map shows

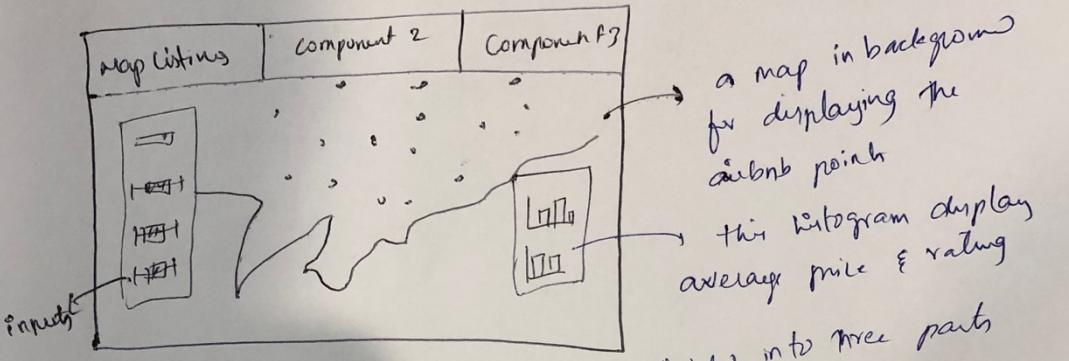


These all hovering on the different boxes that carry useful information that any customer can understand.

**Sheet 5:**

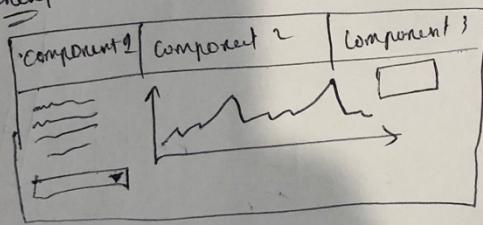
Sheet 5

Frikar Manthattip



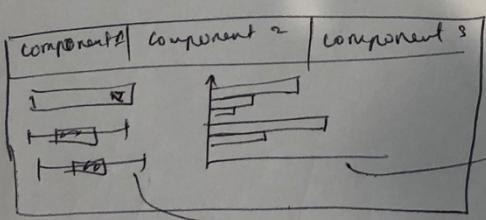
- \* The whole application has been divided into three parts
  - (i) Component 1 (shown above)

Component 2



- \* The Second component display the review over years. A simple trend graph is being used.
- \* No inputs in this component

Component 3



output bar chart

- \* There are the three main components which will entirely fit all questions that need to be answered.
- \* user can get an overall idea on how airbnb market is, and in which area it is high
- \* They can also view average rating, average price with respect to various countries.

### **Steps for executing the application:**

1. Open the required files ui.r, server.r, global.r
2. Click run app button on the server.r or global.r for the declared variables to get initialised first.
3. Do not click runapp on the ui.r as it shows error that particular variable is not defined, because all the variables are declared in global.r