FIT5147-Data Exploration and Visualisation

Submitted by: Nikhil Taralkar

Student ID: 28980433

Table of Contents

1. **INTRODUCTION**

The Melbourne property market has grown with an impressive rate of 57% over the last five years. It has moved from fifth gear into second gear, but it's not going into reverse any time soon," says Michael Yardney, CEO of Metropole Property Strategists. "In fact, Melbourne is likely to once again be amongst the best-performing property markets in 2018."

Melbourne Housing project focuses on finding the best private occupied dwellings across different suburbs in Melbourne. There is a huge shortage of housings that are affordable to low and moderate-income households in Melbourne. This visualisation helps in examining and estimating the cost and availability of private occupied dwellings that are easily affordable and accessible to those households with 40 per cent of lowest income.

Melbourne home prices are showing a substantial rise and are expected to increase by 15.9% over few years. Central Business District (CBD) of greater Melbourne metropolitan region is the major financial centre of Australia and Asia-Pacific regions. With increasing popularity of CBD, socially and financially, house prices in CBD and nearby suburbs are increasing rapidly. However, as one moves away from CBD, dwelling prices reduces substantially.

This report has a vast scope and covers a wide range of people from diverse professions and backgrounds with different interests. The application can be helpful for students as well, since it is evident Melbourne has numerous number of International students looking for accommodation all around the year.

The interactive visualization provides an insight on the following points:

- Finding a house, Unit apartment or town house in Melbourne suburbs based on several factors such as Bedrooms, Bathrooms, Car parking and Distance from CBD.
- Relation between Price and Suburb
- Distance from CBD, as one moves away from CBD, price of house decreases and vice versa.

2. DESIGN

The design is kept user friendly and is interactive. The user can drill down data according to the need. Moreover, direct search options are given so that data can be accessed easily. There are shiny widgets such as sliderInput, selectInput, radioButton and actionButton used to traverse through data and gain information based on user needs.

The visualization for Melbourne housing is created using shiny app (R). The visualization gives an overview of the private occupied dwellings in Melbourne based on several factors such as Suburb, Type of house (House, cottage, villa, Unit apartment, town house), Bedrooms, Living Rooms, Bathrooms, Distance from CBD.

A shiny Dashboard is created which comprises of five tabs, each explaining its significance though housing analysis. Five-sheet design is the base and most important step before designing any application. It gives the basic overview of the application and is easy to follow.

Note: five-sheet design is only drawn for Suburbs, Overview and Housing tab in the shiny dashboard as suggested by Prof. Kimbhal.

Following are the five sheet designs for each tab in the dashboard explaining its features:

Five sheet Design 1

The above five-sheet design helps to understand the flow of the visualisation. Layout is a map with latitude and longitude coordinates. Zoom in/out feature is added to the map. The selectInput widgets from shiny package are used for user input. This visualisation is user friendly and interactive in nature. Detailed analysis can be done using this visualisation based on the user requirement.

Five sheet Design 2
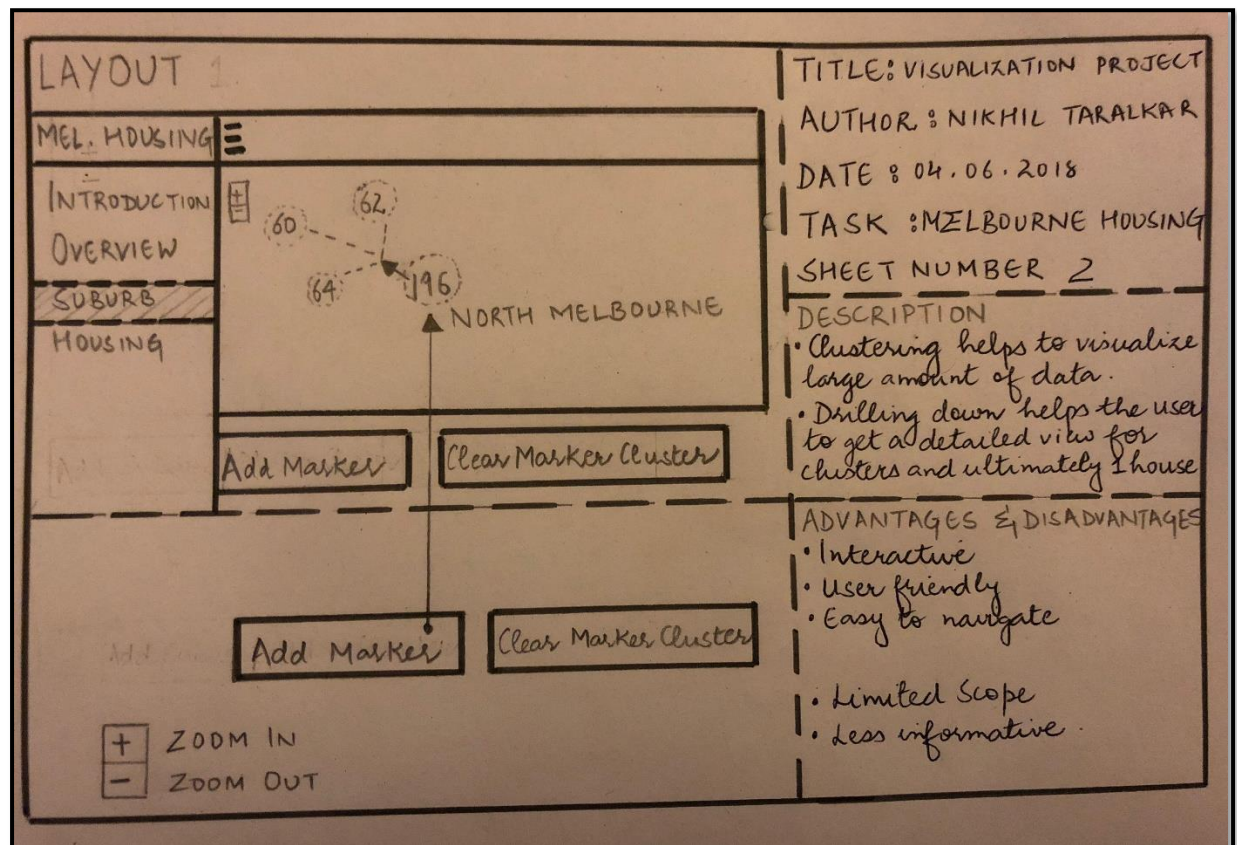
The above five-sheet design gives an overview of selection of suburb to live in. In this design, a clustering approach is used. Two actionButtons are used. The map is interactive and popup displays the price, Suburb and Address of house when clicked on marker. Zoom in/out feature is added to the map.

Five sheet Design 3
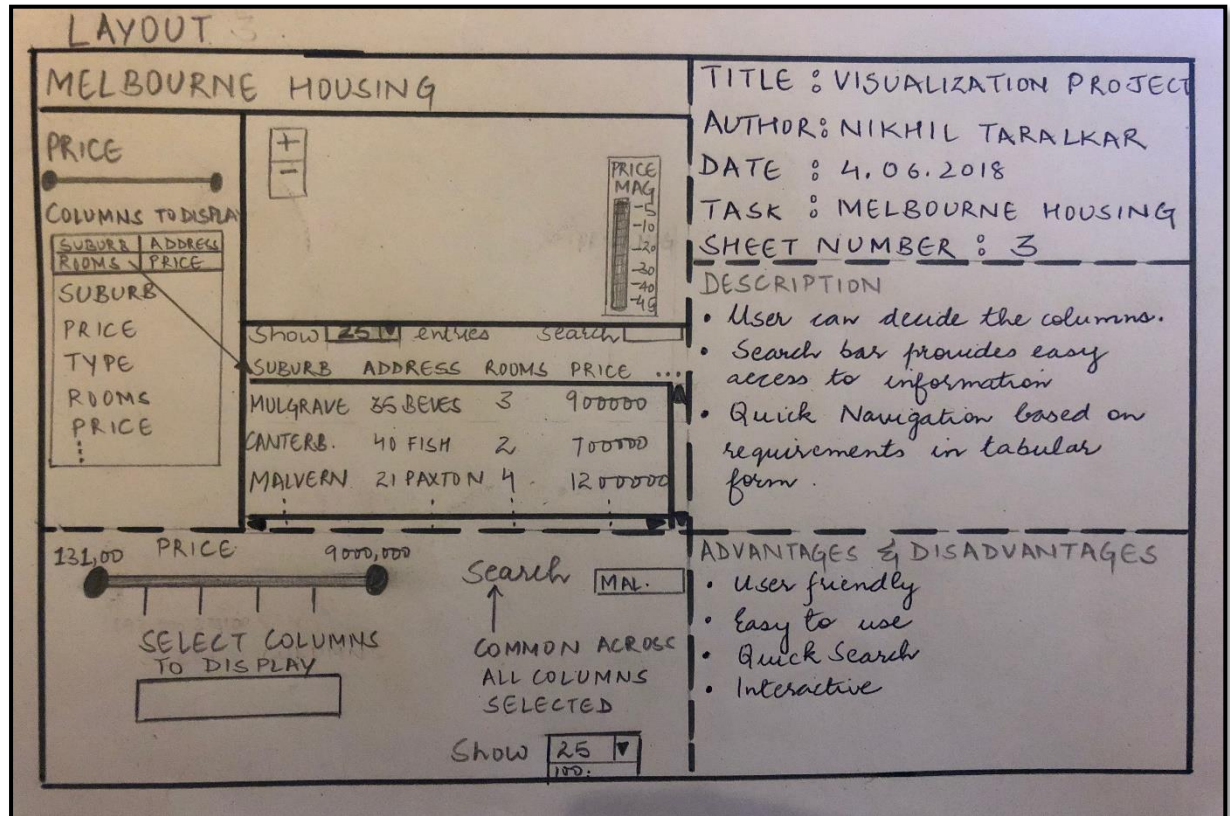
The above five-sheet design depicts and user friendly interactive visualisation. In this design, a slider is used for price variation. A legend is displayed on the right side of the map, which indicates the price magnitude. A data table is shown below the map. On clicking the marker in the map, the row in the table is highlighted and vice versa. Search box is available at the top right corner of the table. Any values from the table can be searched in the search box. Again Zoom in/out feature is added to this map

3. IMPLEMENTATION

The interactive web-based visualisation is implemented in R using shiny. Shiny is an R package used to built interactive web application in more simpler way. Since my dataset is about Melbourne housing which is more location oriented, mostly I have used interactive maps for visualisation. Following interactive visualisations were implemented using different R packages:

3.1 Interactive Leaflet map

Problem Statement: Finding a best suburb and then the best house was difficult to find from a huge dataset.

A leaflet maps are one of the best R package to plot maps based on latitude and longitude coordinates. This visualisation shows clustering of suburbs in different colours. Since, it is difficult to plot and understand huge size of data on map, clustering is used.

Clustering of suburbs uses a drill down approach. The number of clusters are formed based on the size of the data, and how close are coordinates to each other.

Packages used: leaflet and dplyr package was used for plotting map and filtering dataset respectively. Leaflet package is one of the most popular open-source JavaScript library used for interactive maps.

Leaflet has following incredible features that satisfies my requirements of dataset:

- Interactive Zoom In/ Zoom out
- Easily render spatial objects from the sp or sf packages, or data frames with latitude/longitude columns
- Use map bounds and mouse events to drive Shiny logic
- Compose maps using different combinations of:
    - **Map Tiles**
    - **Markers**
    - **Polygons**
    - **Lines**
    - **Popups**
    - **GeoJSON**

dplyr: dplyr package is easy for making data manipulation. dplyr is a powerful R-package to transform and summarize tabular data with rows and columns (data set filter)

Reason for using Interactive leaflet map:

This visualisation is basically used to help the user to see the availability of houses in a suburb and the nearby locations. The basic idea behind using leaflet map for clustering of suburbs was to show suburbs distribution across Melbourne. For any unknown individual who is new to Melbourne, it will be easy to drill down through this map and find the best suburb to live in.

Advantages:

- Before finding a house, finding a suburb is easier through this interactive visualisation.
- Wide options can be found for finding a suburb or house to live which is close to work or University.

3.2 Types of Houses

Problem statement: To find the relation between the types of houses and Price.

This visualization provides more detail into the various housing options in Melbourne. The filters allow you to see the housing options available including houses/cottages/villa/terrace, development sites or a town house.

By Navigating the visualization and looking at general trends of the graph, we can see that the most common housing option is a house/cottage/Villa, also it is easiest to get a 3-room apartment anywhere in Melbourne.

If someone is planning to get an apartment with 2 or 3 rooms/bedrooms, there is a very high probability of getting one!

3.3 Distance from CBD

Problem statement: House prices in Melbourne are very costly. With increase in population, they are expected to increase drastically. Moreover, CBD is a central business district of Melbourne. Many people travel CBD for work and for their social life. Therefore, private occupied dwellings are very expensive in CBD compared to other suburbs.

Packages Used: To implement this leaflet package was used along with dplyr. Visualisation through map is easy when it comes to finding a location or house to live. Plotting map using leaflet package is much easier than plotly, ggmap etc.

Reason for Distance from CBD:

In this simple visualization the available housing is highlighted on the map relative to their distance from the CBD. This visualization shows all the houses sold in the last six months, it indicates the distance from the CBD using colour, where areas in green are closest to the CBD, areas with red are farthest from the CBD, areas with yellow are in between.

We can see that more houses were sold in the area close to CBD and the areas in between than the areas far from it. It is safe to conclude that these areas are in demand but more interesting is the fact that there were a lot of available houses in those locations

Navigating through the visualization and looking at general trends, we can see that the distribution of price is relative to their distance from the CBD. For example, the houses that are close to the CBD have a price range from 160k to 6.5M while the houses farthest from the CBD have a price range from 410k to 970k

3.4 Interactive leaflet map with multiple factors

Problem Statement: After selecting the best suburb, finding the best house based on the different requirements was difficult to find. Also, it was difficult to know the availability of house in suburbs with user requirements. For e.g. 3 Bedrooms, 2 Rooms and 2 Bathrooms apartments are very few in Southbank, however, they are available in plenty in Canterbury and Middle park.

Another leaflet map was used that shows overall distribution of houses in Melbourne suburbs based on several factors such as Distance, Price, Number of Bedrooms, Number of Bathrooms, Number of rooms and Type of house. Leaflet map is the best way to locate a suburb and find different house based upon the requirement.

Packages Used: To implement this leaflet package was used along with dplyr. Visualisation through map is easy when it comes to finding a location or house to live. Plotting map using leaflet package is much easier than plotly, ggmap etc.

Reason for using Interactive leaflet map with multiple factors:

Although, the first interactive visualisation was helpful in determining the best location to live in, there are several other factors that plays a major role in deciding which suburb or house to live in. For example, Price, Type of house, Distance from CBD, Bedrooms, Bathrooms etc.

Any web application that is designed for buying or renting a house, comprises of below inputs of from user.

- ➢ Suburb
- ➢ Price
- ➢ Number of rooms
- ➢ Number of Bathrooms
- ➢ Number of Bedrooms
- ➢ Type of house (House, cottage, villa/Unit apartment, Duplex/Town house)
- ➢ Car parking

Considering this, I have implemented an interactive web page that considers the above mentioned minimum requirements from user. From this visualisation map, User can make a short list of the houses which are affordable and accessible to him/her as per his requirements.

3.5 Data Table with map (Housing)

Problem Statement: From several newspaper and articles, it was found that some old age people and young students find it difficult to read maps. Especially in a under developed countries with lack of education. House hunting is a tedious task for such people.

After selecting the best suburb and other factors, the next and final visualisation helps in selecting the best house from the chosen ones. Data table along with map is the best combination to analyse and make decisions.

Packages used: DT package for data table, leaflet for map and dplyr for filtering the dataset.

DT: The main function in this package is DataTables(). It creates an HTML widget to display R data objects with **DataTables**.

Reasons for using Data Table with map:

To shed more light on a data through graphically and table reading. I have used this approach because it helps in drilling down on the huge dataset and finds a precise answer to the problem. I have incorporated data table along with the map, so that even a person who is not so familiar with maps and its usage can understand by reading tabular data.

4. USER GUIDE

The visualisations are very easy to understand and is user friendly.

➢ **Interactive leaflet map for Suburbs**

The visualisation gives clustered data of houses available in suburbs. The clustering is based on area therefore, the houses close to one another form a cluster. The data would be beneficial for users who are unaware of the suburbs and are new to the town. This visualisation would give an insight into availability of houses in nearby suburbs, streets or lanes therefore, users can easily select an appropriate suburb in proximity to their office or study location.

Add Marker Cluster and Clear Marker Cluster are the two options available to plot an interactive map showing different clusters of suburbs throughout the Melbourne.
Add Marker Cluster will add clusters to the map. Once you click on one of the cluster, it expands into other clusters and so on until we get a marker. Further, if you click on markers, it will display the Price and Suburbs values as text popup.
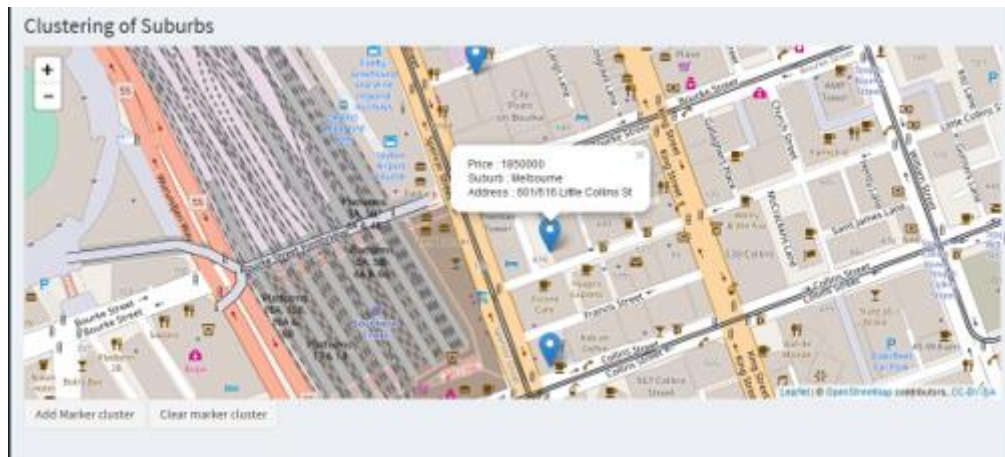

Figure 1. Before clustering

Figure 2. After Clustering

- The Latitude and Longitude from the Melbourne Housing dataset has been used for plotting map leaflet package.
- The feature of zoom-in and zoom out has also been added using the leaflet library.
- Since there is huge amount of data therefore plotting individual data points would lead to overlapping of markers creating confusion therefore, the appropriate choice was to use cluster approach which was implemented using actionButton.
- When one hovers over a cluster it highlights the area covered under that cluster using leaflet package.
- Once you click over that cluster there is drilled down view provided segregating the cluster into small individual clusters making the research more refined
- The last level of grain is the house. The user would reach to a house by drilling down. One he/ she clicks on the location marker they would get the address of the house along with the suburb and its price respectively.

➢ **Interactive leaflet map for Type of Housing:**

- A radio button is used for selection between the types of houses. Based on the selection, the map will show houses across Melbourne for each house type.
- A Zoom in/out feature is added to the map.
- On clicking the marker, a popup will appear which will display a text which contains Bathrooms, Cars, Type, Bedrooms and Rooms.

Housetype Category

House type:
- ◉ HOUSE, COTTAGE, VILLA, TERRACE
- ○ UNIT, DUPLEX
- ○ TOWN HOUSE

Figure 3. Type pf Houses

➢ **Interactive leaflet map for Distance from CBD:**

The visualisation is based on the distance from CBD. This is an important consideration for overall analysis of this project. In this simple visualization, the available housing is highlighted on the map relative to their distance from the CBD.

A slider is used to vary the distance between the suburbs from CBD. Zoom in/out feature is added to the map. This visualisation is very simple and user friendly.

A popup feature is also added which display text data of Price, Suburb and Distance from CBD.
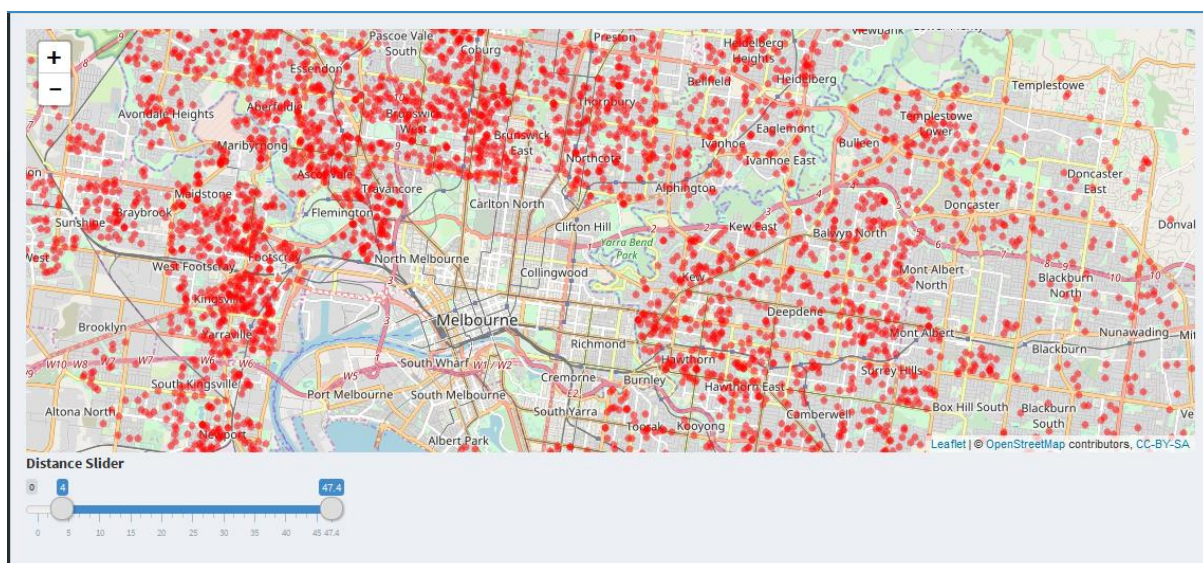

Figure 4. Distance From CBD

➢ **Interactive leaflet map for multiple factors: (Overview)**

The visualisation gives a clear picture of output data based in the user requirements by taking input from user. This visualisation is very important to user friendly. Following are the steps required to understand this visualisation

- Using the latitude and longitude coordinates, map was plot using leaflet package.
- There are two sliders used for price and distance from CBD. The user can slide and select an appropriate value based on the need. These were implemented using the shiny package.
- Another feature on the page is of selectInput widget. The user can select from a list of drop down the number of Bedrooms, the number of other rooms (study, living etc.) and the number of bathrooms.
- Moreover, there are radio buttons used which have been implemented again using shiny package. These radio buttons help the user to select the house type.
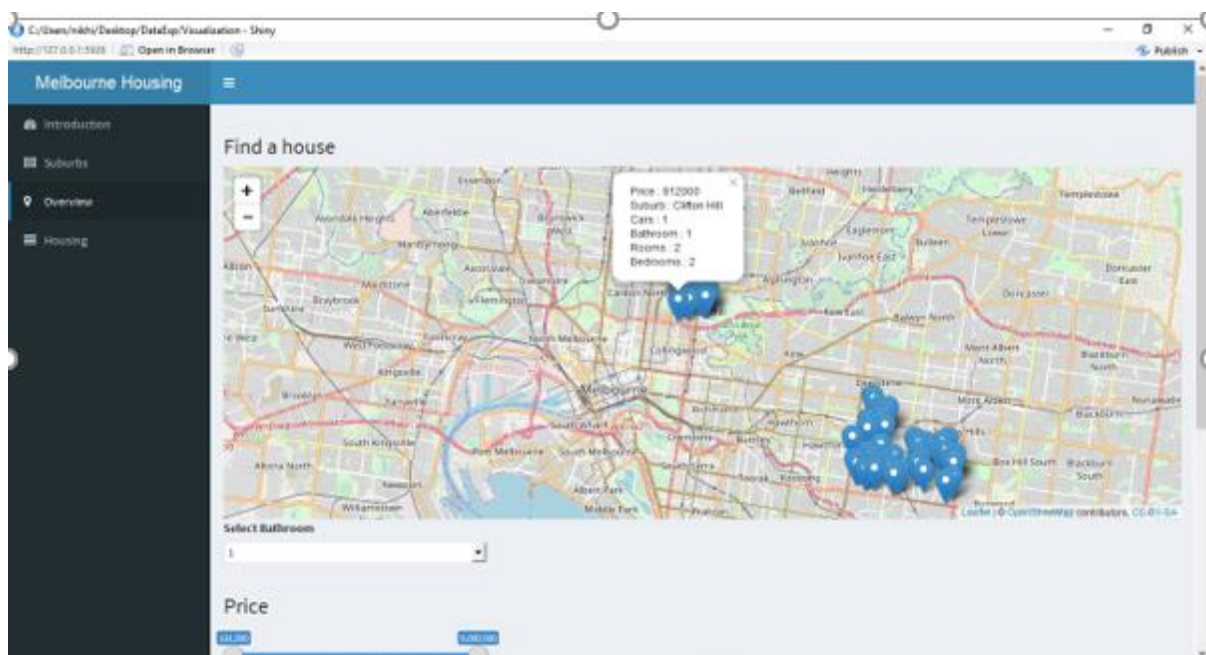
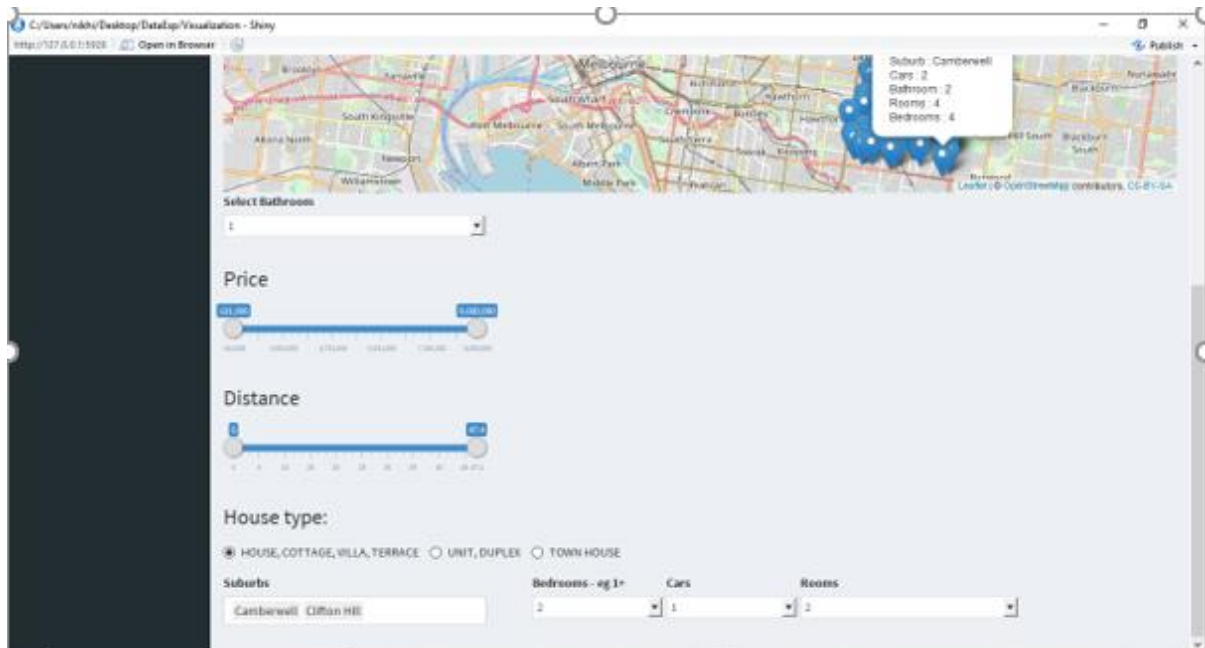

Figure 5. Interactive Visualisation for multiple factors

Figure 4. Interactive Visualisation for multiple factors

➢ **Data Table with map (Housing)**

This visualisation is very user friendly and is accessible to everyone including users who do not have prior knowledge of reading maps.

- When we click on the marker the row related to it gets highlighted in the table. Similarly, if a row is selected on the table, the respective marker gets highlighted on the map. From the map it is easy to find a location, based on the surroundings. However, it is more easy and convenient to just click on the marker in the map and look for the corresponding data in the table.
- The price magnitude shows the variation of price based on colour scheme.
- The price slider helps to select a price range.
- The search input widget above the table can be used to search the entire table based on the input in terms of alphabets, words, numbers etc.
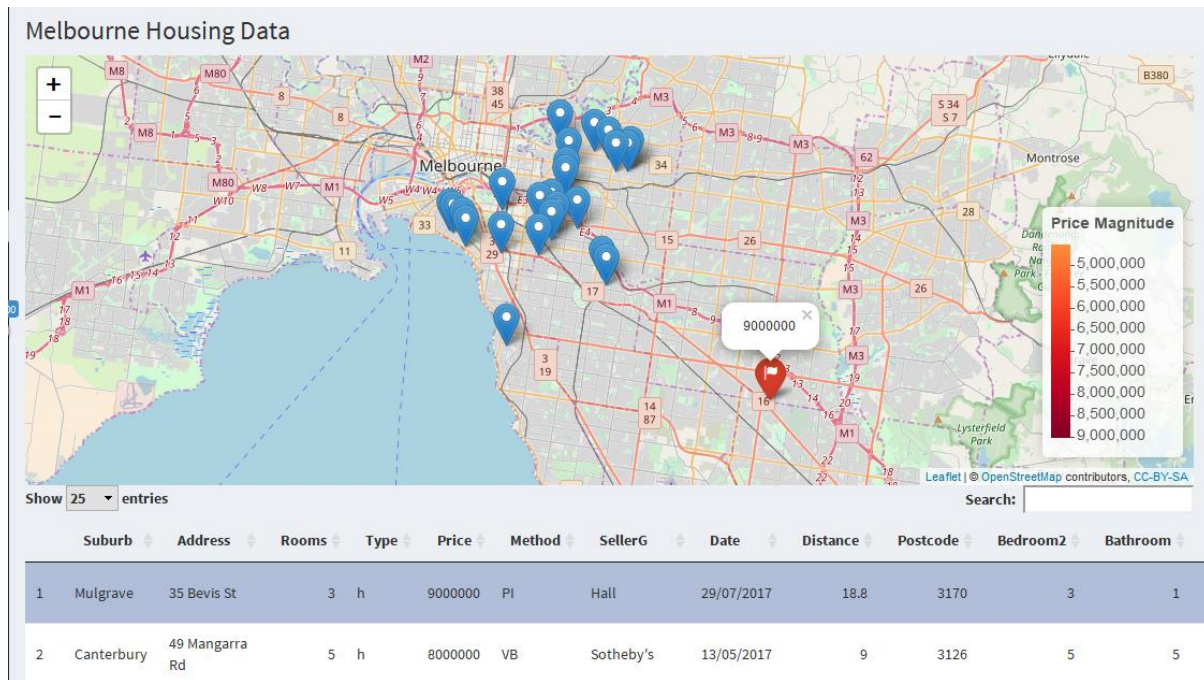
Figure 5. Melbourne Housing Data

5. **CONCLUSION**

- This visualization provides more detail into various housing options by showing the different suburbs with respect to the prices of the houses in those suburbs. These visualizations help you to study the various suburbs with respect to their prices in detail. Houses which are close to CBD are pretty expensive. Additionally, the number of bedrooms, rooms, bathrooms and Car parking are less in suburbs and houses close to CBD. However, as one moves away from the CBD, private dwelling are less expensive and has larger building area.
- Overall, it can be concluded that there are several other factors other than Price which play a significant role in buying a house. Also, from the Visualisation we can say that price varies depending on the various factors.

6. REFERENCES

melbourne-housing-market. (2018). Retrieved from www.kaggle.com:

https://www.kaggle.com/anthonypino/melbourne-housing-market

https://stackoverflow.com/questions/36709441/how-to-display-widgets-inline-in-shiny

https://shiny.rstudio.com/gallery/widgets.html

https://www.htmlwidgets.org/