Battle of the Neighbourhoods

Sydney, New South Wales,

Australia.



Gopalrathnam Venkataraman, 07/08/2019.

IBM Coursera Capstone Project.

Contents

[List of tables 2](#_Toc16070511)

[List of Figures 2](#_Toc16070512)

[Summary 2](#_Toc16070513)

[Introduction 2](#_Toc16070514)

[Data Collection 2](#_Toc16070515)

[Suburb data: 2](#_Toc16070516)

[Data consolidation and preparation 3](#_Toc16070517)

[Neighbourhood Location Data 4](#_Toc16070518)

[Planned Methodology and Workflow 4](#_Toc16070519)

[Data Visualization 4](#_Toc16070520)

[Location data and analysis 5](#_Toc16070521)

[Apply Clustering algorithm 5](#_Toc16070522)

[Clustering analysis: 7](#_Toc16070523)

[Summary and Recommendations 7](#_Toc16070524)

[Python packages and dependencies 7](#_Toc16070525)

[Appendix: 8](#_Toc16070526)

# List of tables

[Table 1 -Methodology and workflow 4](#_Toc16070535)

[Table 2 - Packages and dependencies 8](#_Toc16070536)

[Table 3 - Links to data sources 8](#_Toc16070537)

# List of Figures

[Figure 1 - Suburb list and distance from the city 3](#_Toc16070527)

[Figure 2 - Geo locations for all suburbs 3](#_Toc16070528)

[Figure 3 - Cleaned data set containing distance and geo co-ordinates 4](#_Toc16070529)

[Figure 4 - Sydney suburbs within 50 kilometres from city centre 5](#_Toc16070530)

[Figure 5 - Most popular venues for a suburb 5](#_Toc16070531)

[Figure 6 - Elbow curve 6](#_Toc16070532)

[Figure 7 - Sydney suburb clustering 6](#_Toc16070533)

[Figure 8 - Comparison of suburbs 7](#_Toc16070534)

# Summary

The project is aimed at solving the problem of a Sydney local resident who has recently moved to a new suburb and wants to confirm the following

* Is the new suburb similar to the old suburb?
* Is the new suburb conducive for an active life in physical fitness activities?

The old suburb is Parramatta, Sydney, Australia and the new suburb is Hornsby, Sydney, Australia.

The result is based on the most common venues found in both the suburbs and whether the suburbs are classified under the same cluster.

# Introduction

Sydney has many suburbs which are unique in their culture. The project is aimed at suggesting whether the suburb the person has moved into is similar to the suburb he was living earlier. This suburb similarity comparison was done by taking 765 suburbs that are within 50 kilometres form the centre of the city

# Data Collection

## Suburb data:

The suburbs that are within 50 kilometres from the city centre is taken from 'Freemaptools' website. Please see appendix for the link.

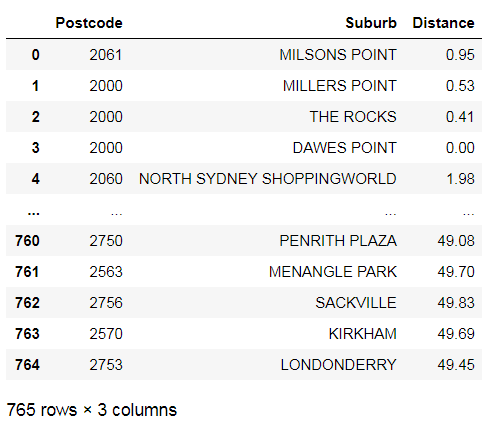


Figure - Suburb list and distance from the city

In addition, the geo location data, such as latitude and longitude for all suburb is taken from public website 'matthewproctor'. Please see the appendix for the link

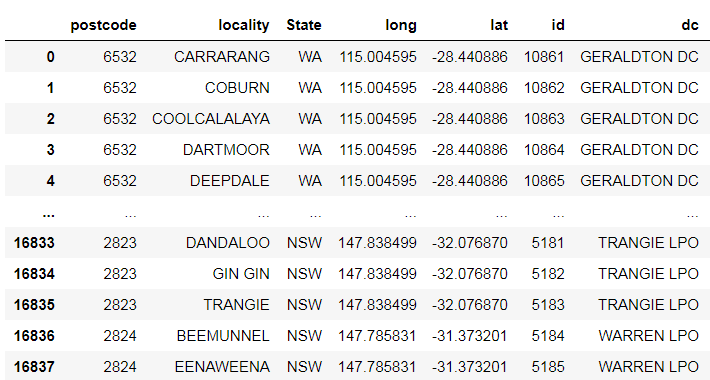


Figure - Geo locations for all suburbs

# Data consolidation and preparation

Both datasets are combined and cleaned to obtain the dataset of all suburbs within 50 kilometres from Sydney city

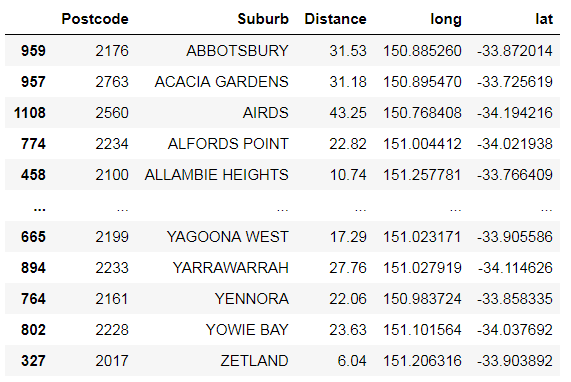


Figure - Cleaned data set containing distance and geo co-ordinates

## Neighbourhood Location Data

Neighbourhood data is sourced to identify the similarities of all the Sydney suburbs that are within 50 kilometres from the city, by using the Foursquare APIs. Neighbourhood data for each suburb was sourced via the Foursquare API, including the most favourite places, such as parks, gyms, cafes, restaurants, stores etc to compare the resident’s current suburb to new suburb and consider similarities.

# Planned Methodology and Workflow

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Project phase | Status | Date |  |
| 0 | Scope Verification | Completed | 04/08/2019 |  |
| 1 | Data collection | Completed | 04/08/2019 | From public sites |
| 2 | Data consolidation | Completed | 05/08/2019 | Python |
| 3 | Data preparation | Completed | 05/08/2019 |  |
| 4 | Data Visualization | Completed | 06/08/2019 | Folium |
| 5 | Clustering | Completed | 06/08/2019 |  |
| 6 | Recommendations | Completed | 07/08/2019 |  |

Table -Methodology and workflow

# Data Visualization

For suburb data visualization, used folium choropleth maps to show all the suburbs in Sydney that are within 50 kilometres from the city.

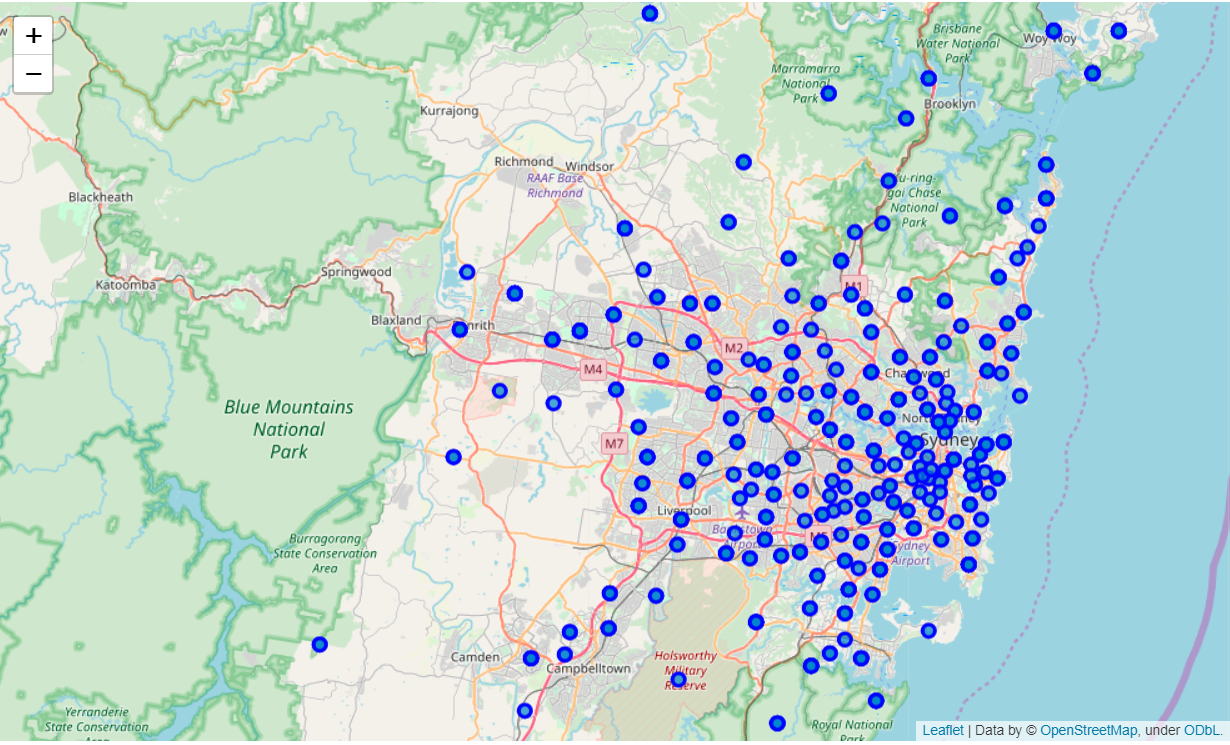


Figure - Sydney suburbs within 50 kilometres from city centre

## Location data and analysis

Location based data is was sourced from Four Square, with the popular venues for all suburbs listed in a pandas dataframe, for all suburbs

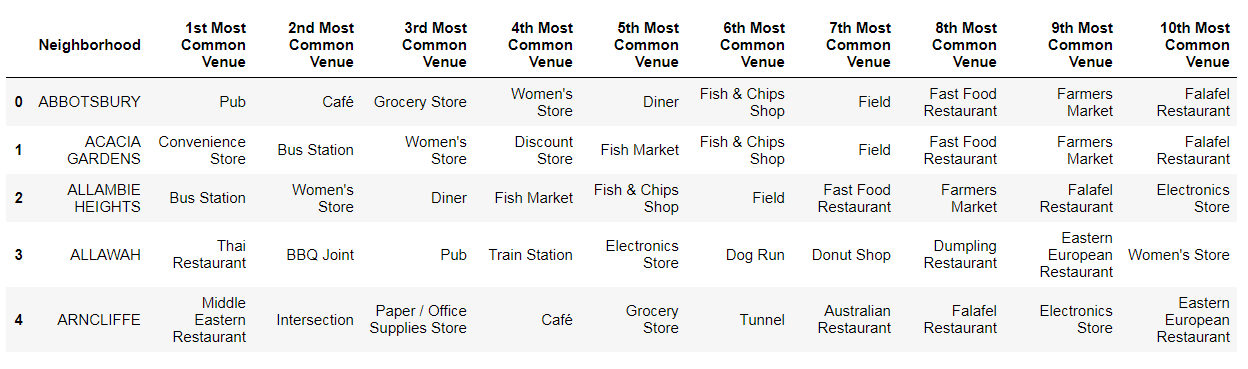


Figure - Most popular venues for a suburb

# Apply Clustering algorithm

To find similar suburbs across all suburbs in Sydney, a Kmeans algorithm was applied taking into consideration all the common venues found in each of the suburb.

In order to find the optimum K, an elbow method is used and arrived at an optimum K values to be 5, to cluster around 600 suburbs. The output is not so much of an elbow, but an optimum value is chosen

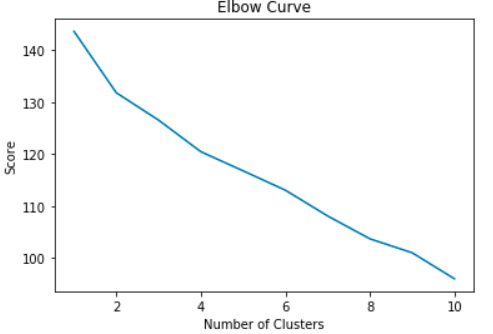


Figure - Elbow curve

With the K value, the output of the suburb cluster is shown below

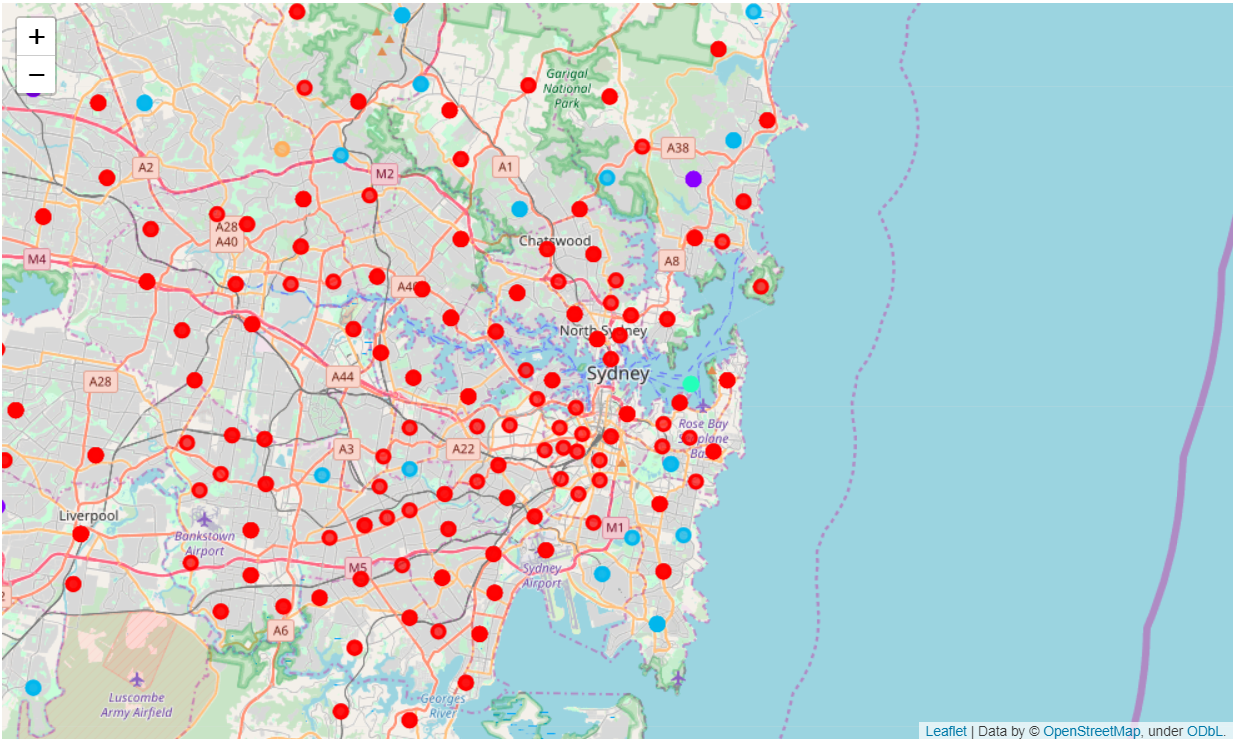
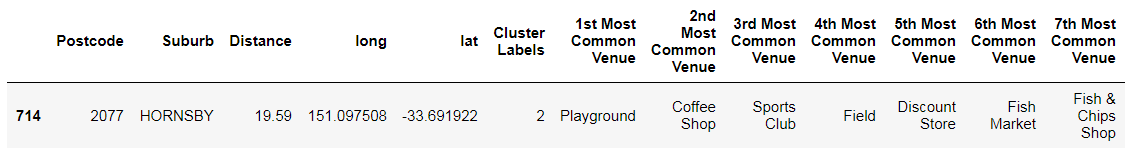


Figure - Sydney suburb clustering

# Clustering analysis:

Looking at the cluster, we find the following

* Most of the Sydney clusters are similar in nature, as this is evident from plethora of red dots
* Parramatta and Hornsby are a different suburb, as they belong to a different clusters.
* Looking closely into the common venues for Parramatta and Hornsby, it is evident that the culture and lifestyle of the suburb are different. Parramatta is largely into Food and eating out whereas Hornsby is about fitness and outdoor activities.



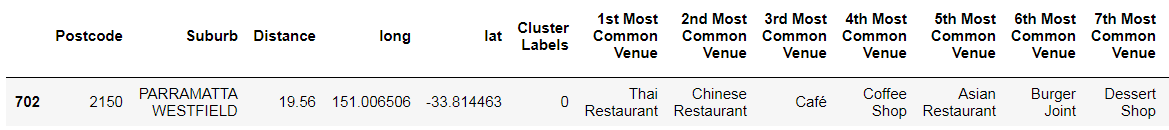


Figure - Comparison of suburbs

# Summary and Recommendations

We started the exercise with two questions to be answered. Let us revisit the questions and check whether the clustering algorithm has answered these.

**1. Is the new suburb similar to the old suburb?**

From the clustering algorithm, we find that the Parramatta and Hornsby are not similar. Parramatta's common venues are related to cafes and restaurants, whereas Hornsby is all about playgrounds, sports clubs and fields.

**2. Is the new suburb conducive for an active life in physical fitness activities?**

From the clustering algorithm, we can see that Hornsby will satisfy the person's appetite for outdoor and fitness activities as three of the top four common venues are related to fitness activities.

# Python packages and dependencies

|  |  |
| --- | --- |
| **Python package** | **Dependency** |
| Pandas | Library for data dictionary and analysis |
| Numpy | Library to handle arrays |
| Geopy | Library to retrieve location data |
| Requests | Library to handle http requests |
| Matplotlib | Library for data visualization |
| Folium | Library for rendering maps |
| Sklearn | Library for Kmeans algorithm |

Table - Packages and dependencies

# Appendix:

|  |  |
| --- | --- |
| **project Data** | **Data source** |
| Australian postcodes with geo co-ordinates | https://www.matthewproctor.com/australian\_postcodes |
| Australian postcodes with distance from city centre | https://www.freemaptools.com/find-australian-postcodesinside-  radius.htm |

Table - Links to data sources