**The Battle of Neighbourhoods**



Author

Heera Mechery Kambil

# Comparing Suburbs in Brisbane to find where to buy a house.

**Introduction**

Buying a house is everyones dream.But with the numerous options avialable in terms of suburbs,type of residence,location,proximity to venues etc it is hard to decide on which one to buy The decision made on intuition may not always be the right one.

I recently moved to Indooropilly which is a suburb in Brisbane, Australia.Living in Indoorooppilly has been a pleasant experience and it also has a great shopping mall and plenty of breakfast and dining options.I have plans of buying a house but don’t know where to buy as I am new to Brisbane.I like Indooroopilly but I would also like to have more choice when it comes to buying a house.

In this project I would like to make use of the insights gained from data to narrow down my choices to a small number of suburbs which would then make it easier to take a decision.

***Audience***

People who are looking to buy a home in Brisbane.They can get an idea of which suburbs are similar.It also gives a general idea of the suburbs from the type and number of venues present.

**Data**

Brisbane Suburbs names will be scraped from wikipedia .As the post code information is not directly available on wikipedia I have copied it from a file on github that had all the postcode and geocodes of Australia.It will be copied to an excel file converted to csv and then using pandas appended to my original Brisbane suburbs file.Now I have my dataframe with Suburb name,Postcode,Suburb latitude and Suburb longitude.

Next step is to find the different venues in each suburb.Foursquare location data will be used to acheive this.Sending a request to the Foursquare with suburb latitude and longitude will return all the venues in that suburb.For each suburb,I will be extracting venues names,venue category and venue location.

**Methodology**

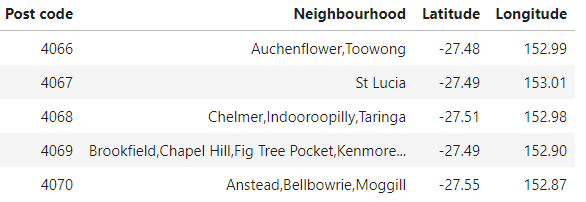
How do I find suburbs similar to Indooroopilly? The main assumption I’m making for this project is:

The presence of certain venues in a suburb depend on the lifestyle of people living in the suburb.Hence the venues in a suburb can give you an idea of the community living in the suburb.For example:a suburb that has restaurants serving wide variety of cuisines indicate a multicultural neighbourhood.

The project has to be done using Foursqare data.Foursquare API is used here to get details about venues around a given location.

In this project I intend to find suburbs of Brisbane that are similar to Indooroopilly based on venues present in the suburbs.

Suburb names are scraped from wikipedia using python package BeautifulSoup.Suburbs with the s

ame post code are combined to form one row.The resulting dataframe looks like this: 

The suburb with postcode 4068 is the suburb I am interested in. There are 75 different suburbs based on postcodes .

I have pulled out venues which are inside a radius of 1500m around the geocoordinate of a suburb.With a radius of 1000m the API was not returning venues in the main shopping area of Indooroopilly.Hence I had to choose a larger radius.

Shown below are 10 rows of the venue data extracted from Foursquare for Indooroopilly :



62 venues were returned for Indooroopilly

An average of 31 venues were returned per suburb.Altogether there were 223 unique categories.

The categories are too specific. Most of the categories it returned will have one or none of these venues in each suburb. For example it had categories like Japanese restaurnt,Ramen restaurant etc.Hence I decided to encode them into broader categories.I decided to go with 7 generalalised categories

1.Restaurant

2.Caffe\_Dessert

3.Pub

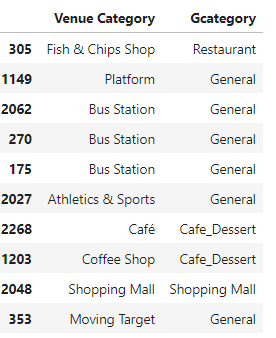
4. Grocery

5.Activities

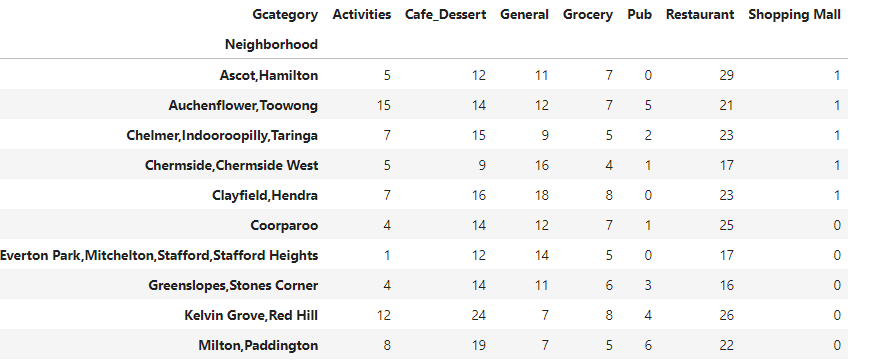
6.Shopping Mall

7.General

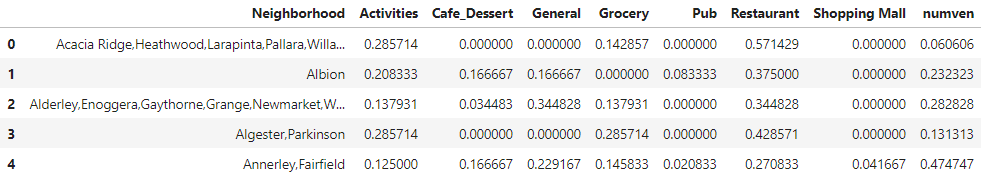
All restaurants were combined into one category.All café ,coffe shops ,icecream and dessert shops were put in second category and so on.Anything that does not fall in any of the first 6 categories go into the general category.I haven’t included a category for transport and that too has benn put in the general category.All suburbs may not have shopping Malls.As I would like a shopping Mall in the suburb I live,I have included that too as a category.



This general category is then onehot encoded and normalised. The number of venues in each category looks like this:

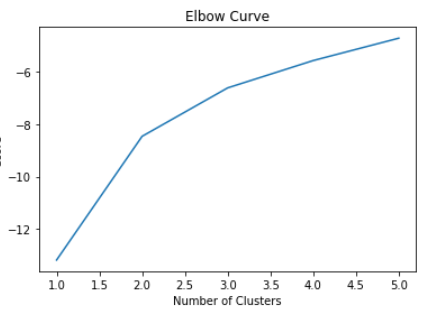


I have also added the total number of venues in a suburb as another category so that it can be given weightage while clustering.I wouldn’t want a suburb with the similar ratio of venues as Indooroopilly ,but with a considerable lower number of total venues be shown as a similar one.These total venue values are then normalised.The resulting dataframewith that would be used for clustereing looks like below:The neighbourhood column will be dropped and It would have 8 columns .



**Clustering Suburbs**

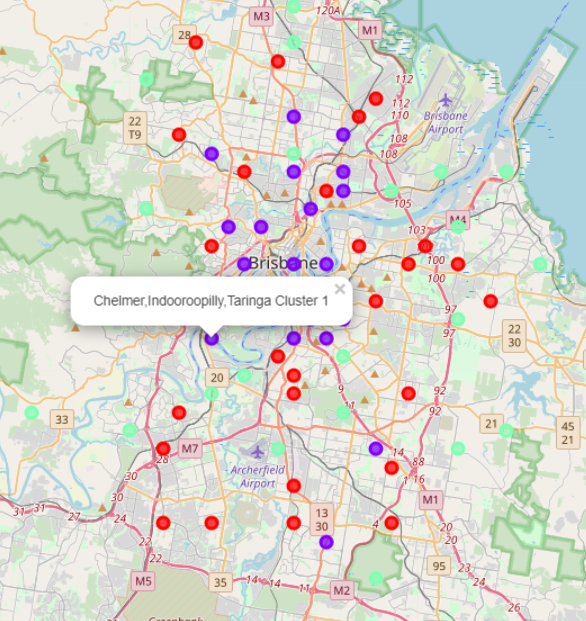
Clustering is done by K Means clustering .Elbow Plot Method is used to identify ideal number of clusters.The elbow plot shown below has an elbow at 3 and hence I have chosen 3 clusters .



After clustering I would like to shortlist 10 suburbs that are most similar to my current suburb.I will be using distance matrix function of scipy libraby to achieve this.The distance of all suburbs to the remaining suburbs will be calculated.From this matrix the row corresponding to Indooroopilly would be selected and the suburbs that are closer in distance are displayed. The most similar suburb would have the least distance in the distance matrix.

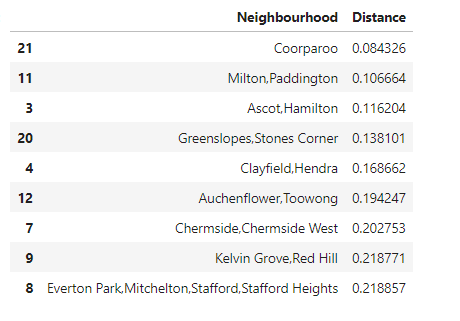
**Results**

The visualisation of the three clusters using Folium is shown below.

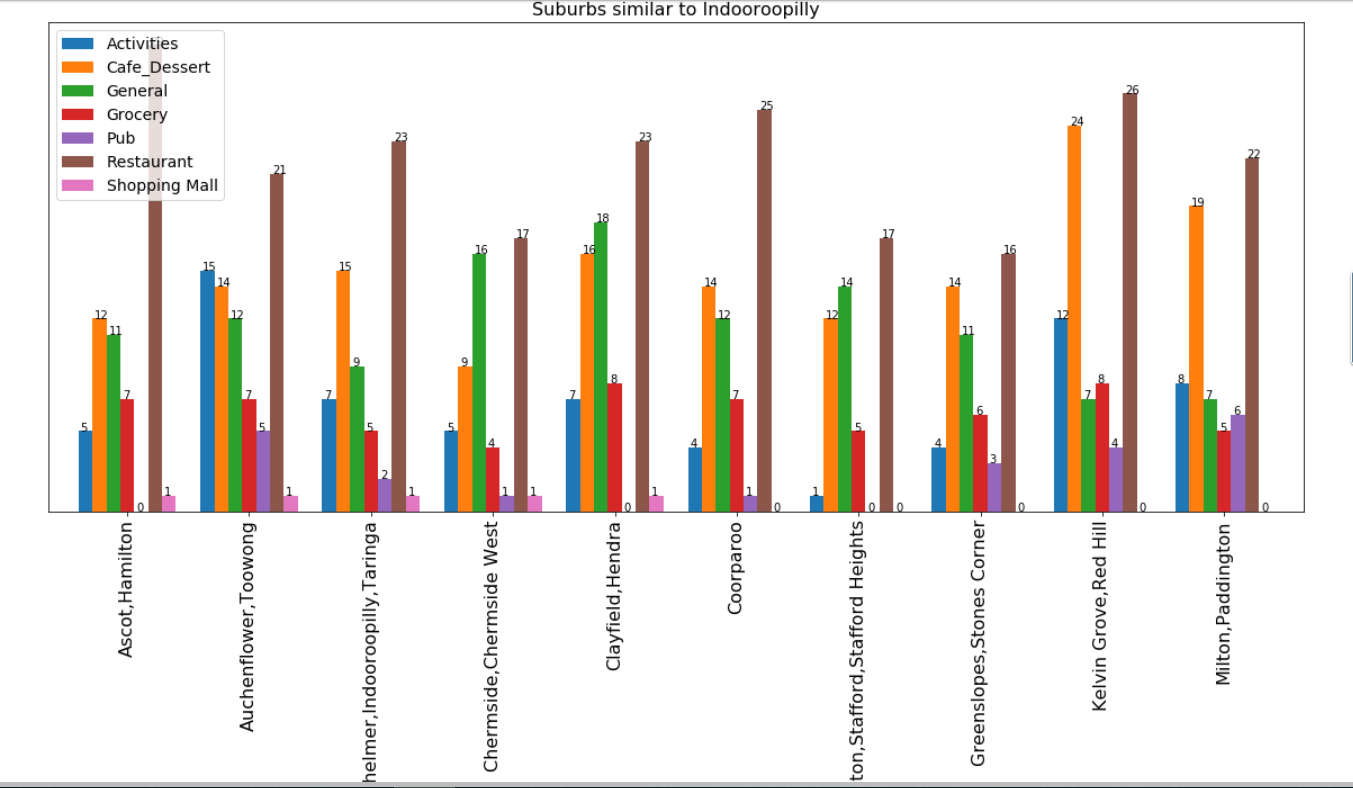


The cluster of interest is the blue one.This cluster with Indoorooppilly has around 23 suburbs.The second and thirds clusters have 24 and 27 suburbs respectively.To order the suburbs similar to Indooroopilly I have calculated the distance matrix for the suburbs. matrix.Based on this distance I have shortlisted 10 suburbs in the order of similarity.So Coorparoo is the closest and Everton park is

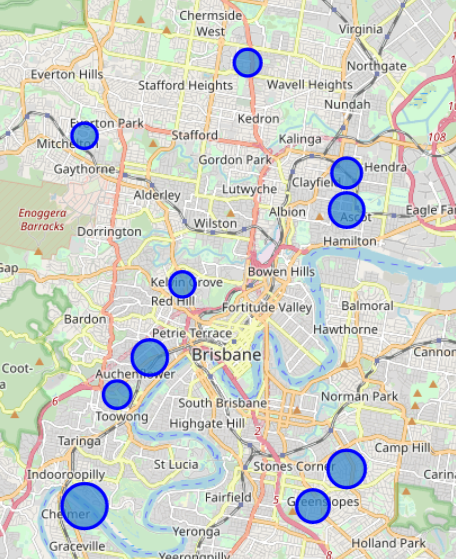
The farthest.



A bar chart of the 10 similar suburbs are shown below with the number of venues in each category shown separately.A bubble chart.The suburbs are also represented using Folium.



The bigger the size of the bubble the more the similarity.



**Discussion.**

I am not sure whether the data returned by Foursquare is a true representation of the suburb.From my knowledge of my suburb the shopping centre itself has more than 100 stores.Hence the reliability of the the results completely depends on whether Foursquare has equal representaion of all suburbs in the venues returned by it.

Transport venues were returned by Foursquare for some suburbs but not all bus stations or train stations were returned for all the suburbs.Addition of trasport details would have given a better idea of each suburb.

**Conclusion.**

Brisbane suburbs were analysed and suburbs similar to Indooroopilly were identified. Foursquare location data was used to solve the problem.