

Recommendation of places of open a new restaurant in Canberra

Wenjing Xue

13 Mar 2020

1. Introduction

1.1 Background

Canberra is the capital of Australian, but it is not a tourist city. Canberra has the highest average wage in Australia, but it has very few good restaurants here. Thus, there is a great chance to make profit when opening a restaurant here.

1.2 Problem

There is a large area in Canberra is mountains, government office and foreign consulates, which are not suitable places for restaurant. In this report, I will try to find the population area and with less competition of many other restaurants.

1.3 Interest

People who wish to open restaurants in Australia can be highly interested in this report. Since the restaurants in modern cities in Australia such as Sydney, Melbourne are very competitive and with lower food price. There is a bigger chance to get success in Canberra.

2. Data acquisition and cleaning

2.1 Data source

The postcode and suburb data of Canberra are found here (<http://www.citypostcodes.com.au/Canberra>). The coordinate data of Canberra suburb are found here (<https://www.geonames.org/postal-codes/AU/ACT/australian-capital-territory.html>). We are going to use the suburb data as unique key to identify each area and use the coordinate data to identify the exact place and combine them with the venue source from the Foursquare API. And cluster the area, using the venue data to find which are have more restaurants, café, stores, etc. And this are the place which have more population which are more suitable for opening a new restaurant.

2.2 Data cleaning

In the postcode and suburb dataset, all the name of the suburb has the word *Postcode*, which we need to clean it before we can do any further

task. In the coordinate dataset, which has some duplicate suburb with different postcode, which we only want to keep only one of them, so we can merge them with the postcode dataset. Also, in this coordinate dataset we need to change the column name for the further merge task.

2.3 Feature selection

Since we need the unique *Suburb* name identify different areas, we are going to keep them. And the *Postcode* can help us define which area are the neighborhood, we also want to keep them. The *Latitude* and *Longitude* are the features for us to using the API get the venue data, which means that we need them. Although the *City*, *Country*, *State* are the same, since we want to show the exact place of Canberra for people, that we want to keep these features for a reminder.

3. Exploratory Data Analysis

3.1 Merge and map the suburb data

After cleaning the data and feature selection step, the first step of data analysis is to merge the postcode and coordinate data and map the suburb in a map to show which suburbs in Canberra will we be evaluating later.

3.2 Get and group Venues data

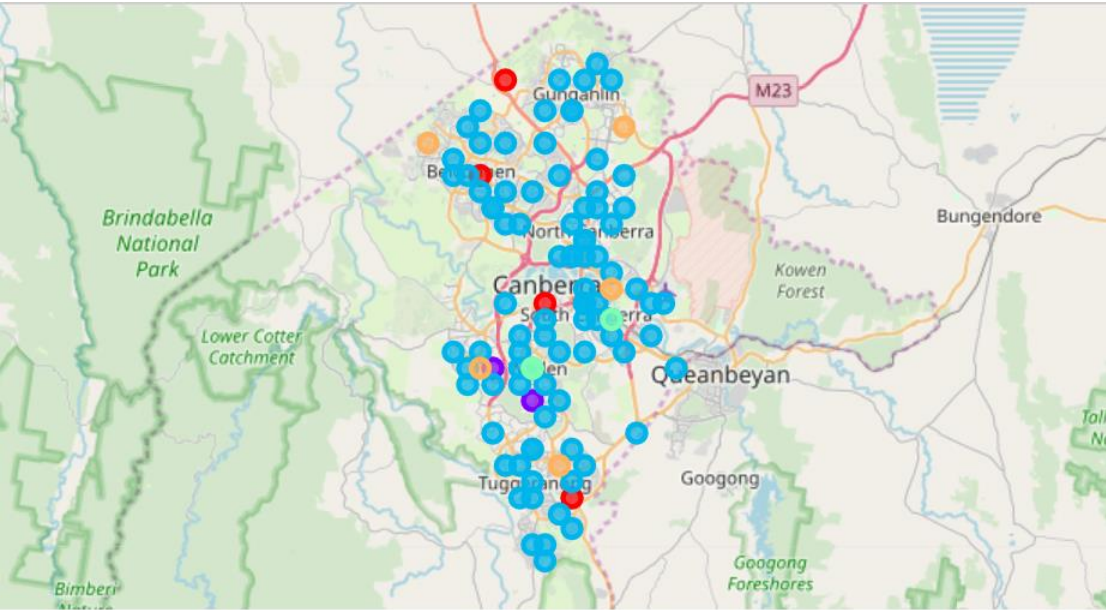
Using the coordinate data to get the Venues data from Foursquare. And then group these venues data by the suburb and choose the top 10 venues of each suburb for further evaluation.

3.3 Cluster

Using K-means to cluster the venues into 5 groups. Cluster can help us to see where the population are. Using the venues in each cluster can helps us to evaluate where is the better location for the new restaurant.

4. Result

We use k-means to category the Canberra suburb into 5 clusters. And using different colors to show the clusters and show them in a map in the following. Different clusters have different colors. Red is cluster 1, Purple represents cluster 2, Blue is cluster 3, Green shows cluster 4, Orange means cluster 5.



I list the five clusters table below.

	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
103	2600	-35.31	149.10	0	Café	Doner Restaurant	Pizza Place	Grocery Store	Winery	Field	Food Court	Food & Drink Shop	Flower Shop	Flea Market
36	2615	-35.23	149.05	0	Bus Stop	Winery	Field	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant
88	2905	-35.43	149.12	0	Grocery Store	Winery	Farm	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant	Field
49	2618	-35.17	149.07	0	Home Service	Market	Trail	Shop & Service	Fast Food Restaurant	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop

Cluster 1

	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
99	2611	-35.35	149.06	1	Grocery Store	Winery	Farm	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant	Field
96	2607	-35.37	149.09	1	Burger Joint	Rugby Pitch	Shop & Service	Bus Station	Field	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market

Cluster 2

	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
70	2607	-35.36	149.10	2	Locksmith	Winery	Field	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant
69	2603	-35.31	149.13	2	Café	Hotel	Coffee Shop	Breakfast Spot	Gym	Japanese Restaurant	French Restaurant	Music Venue	Event Space	Cantonese Restaurant
68	2614	-35.25	149.06	2	Vietnamese Restaurant	Pharmacy	Shopping Plaza	Winery	Farm	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant
66	2904	-35.41	149.13	2	Supermarket	Pizza Place	Gym / Fitness Center	Gym	Café	Sports Club	Shopping Mall	Gas Station	Winery	Fast Food Restaurant
65	2606	-35.34	149.08	2	Café	Skating Rink	Gym / Fitness Center	Italian Restaurant	Asian Restaurant	Grocery Store	Winery	Fast Food Restaurant	Food & Drink Shop	Flower Shop

Cluster 3

	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
102	2606	-35.35	149.09	3	Sandwich Place	Burger Joint	Liquor Store	Fast Food Restaurant	Coffee Shop	Multiplex	Café	Food Court	Sports Bar	Steakhouse
16	2604	-35.32	149.15	3	Go Kart Track	Hotel	Train Station	Liquor Store	Thrift / Vintage Store	Museum	Fast Food Restaurant	Food & Drink Shop	Flower Shop	Flea Market
93	2606	-35.35	149.09	3	Sandwich Place	Burger Joint	Liquor Store	Fast Food Restaurant	Coffee Shop	Multiplex	Café	Food Court	Sports Bar	Steakhouse

Cluster 4

	Postcode	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
92	2611	-35.35	149.05	4	Bar	Australian Restaurant	Optical Shop	Fast Food Restaurant	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop
90	2600	-35.30	149.15	4	Playground	Café	Restaurant	Park	Coffee Shop	Plaza	Farm	Flower Shop	Flea Market	Fish & Chips Shop
67	2615	-35.21	149.01	4	Pool	Bus Stop	Coffee Shop	Field	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop
45	2904	-35.41	149.11	4	Café	Pharmacy	Shopping Mall	Field	Winery	Fast Food Restaurant	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop
51	2914	-35.20	149.16	4	Soccer Field	Park	Winery	Farm	Food & Drink Shop	Flower Shop	Flea Market	Fish & Chips Shop	Filipino Restaurant	Field

Cluster 5

From the 5 clusters top 10 venues that I list above, we can have a brief understanding of the venues of each cluster. Cluster 1 is the center of each area. It has good transport and is good place for opening a new restaurant. Cluster 2 is in a more rural area, which are do not have so many populations. Cluster 3 includes lots of suburbs. There could be some good place for a restaurant, but it is so hard to choose which one is much better. Cluster 4 has lots of fast food stores, which means people here have a fast pace of life. Cluster 5 is suitable for entertainment, which have many playground and swim pool.

5. Discussion

From the result section, we can acknowledge that cluster 1 is the area center, cluster 2 is more like a rural area, cluster 3 have too many locations to consider, people in cluster 4 have a fast pace of life, cluster 5 is for entertainment. Which means that cluster 1 is the best location for people who wish to find a place to open a restaurant. These places have good public transport and is the center which can bring in a lot of people and turnover every day. So, cluster 1 is the best location for new restaurants. And in the map which are naming Yarralumla, Richardson, Florey, Hall.

6. Conclusion and Future work

In this report, we use the k-means cluster to analyze the existing venues in each suburb in Canberra and recommend the location of the new restaurant. Since in this report we only using the existing venues to evaluate the venue categories and predict the population for new restaurant location recommendation. If we analyze more data such as the real population, salary, restaurant cosine, we could have a better recommendation location suggestion.