Establishing a restaurant in Canberra, Australia

Liam Minagall

12 March 2019

# Introduction

Starting a restaurant is tricky. You need to hire good staff and keep them happy, source great produce, pay your bills, manage your clientele. Importantly, you also need to set up your restaurant in an area that gets you an adequate supply of diners.

The aim of this report is to detail prospective locations for establishing a new Indian themed restaurant. This analysis will explore whether it is better to be away from other restaurants because they might take your customers, or be close because the number of people looking to eat will be higher?

# Data acquisition and cleaning

The data I used comes from the Foursquare API. From Foursquare I was able to obtain a number of features of restaurants around Canberra:

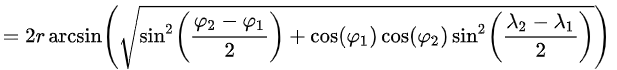
* Name
* Location (latitude and longitude)
* Restaurant type (e.g. North Indian, Chinese, Thai)
* The number of people who have liked the restaurant

The Foursquare “api.foursquare.com/v2/venues/search?...” query only returns up to 50 restaurants for a given location. To overcome this, six separate locations where chosen around Canberra. These restaurants were compiled into a list, and duplicates were removed.

Restaurants were plotted on a map of Canberra using the Folium library. The ‘Shortname’ feature of the restaurant was examined to determine if the restaurant was Indian themed. Often the shortname was ‘North Indian’ or ‘South Indian’, so if the word ‘Indian’ was within the shortname, the restaurant was classed as Indian. In the restaurant was Indian, it was plotted as a red dot, otherwise it was plotted as blue. This can be seen in Figure 1.

To obtain the number of likes, an “api.foursquare.com/v2/venues/{}/likes?...” query was performed for each restaurant. This feature was added to the respective restaurants, and the map was re-plotted with bigger dots representing restaurants with higher likes (see Figure 2).

As the distance between restaurants was of interest, a function to calculate distance from Longitude and Latitude was written. It was based on the Haversine formula, where:



In this case, r is the radius of the Earth. The function was used to create a square distance matrix detailing the distance between each restaurant and every other restaurant.

|  |  |  |  |
| --- | --- | --- | --- |
| **Restaurant** | **A** | **B** | **C** |
| **A** | 0 | 123 | 2047 |
| **B** | 123 | 0 | 256 |
| **C** | 2047 | 256 | 0 |

Table 1: A representation of a distance matrix. The matrix is square, with a 0 diagonal and is a mirror image across the diagonal.

From this distance matrix, I was able to determine how many restaurants were within a distance from each other restaurant. For each restaurant I plotted the number of restaurants within 2km of it against the number of likes. This can be seen in Figure 3.

The machine learning operation DBSCAN used the distance matrix to cluster the restaurants. A cluster was defined with epsilon = 500m and needed to contain at least three restaurants.

Following the DBSCAN operation, clusters were plotted using Folium and were each assigned a different colour. Each cluster was then examined, and ‘likes’ for each restaurant in the c luster were tabulated. This data was compiled in a box plot with the same colours as the Folium map, as this provided a clear representation of the cluster mean, standard deviation and outliers. This can be seen in Figure 4.

# Results

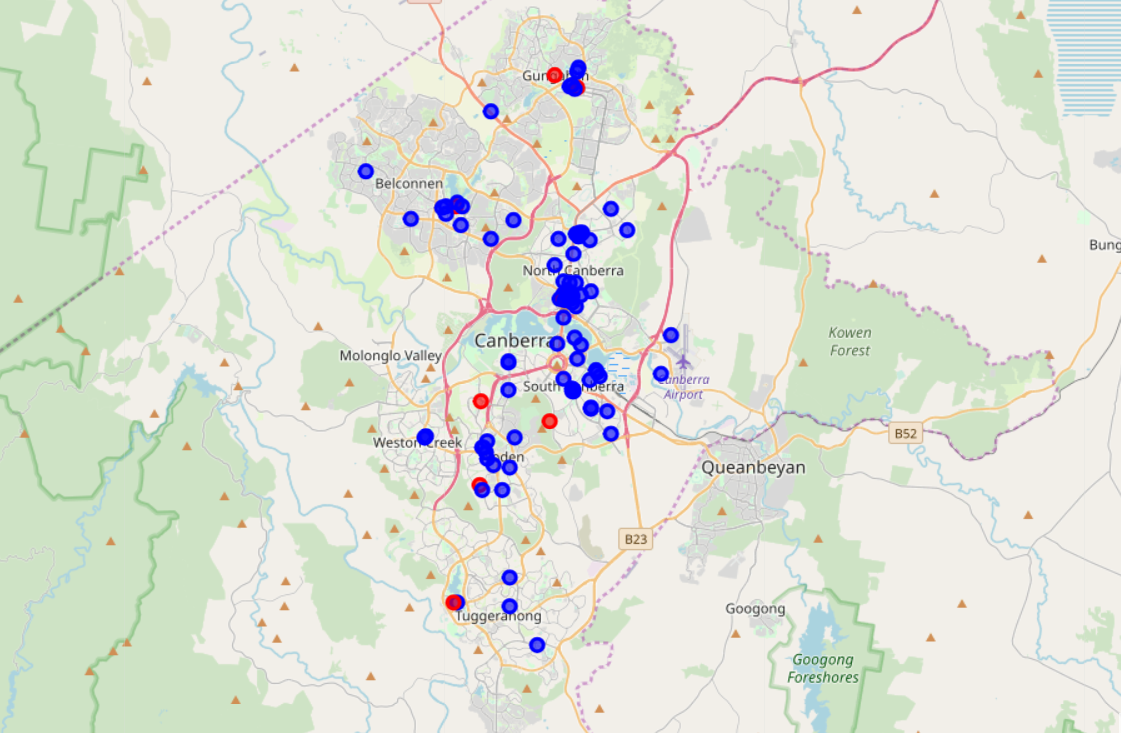


Figure 1: Restaurants in Canberra: Red icons represent Indian restaurants and blue icons represent all other restaurants.

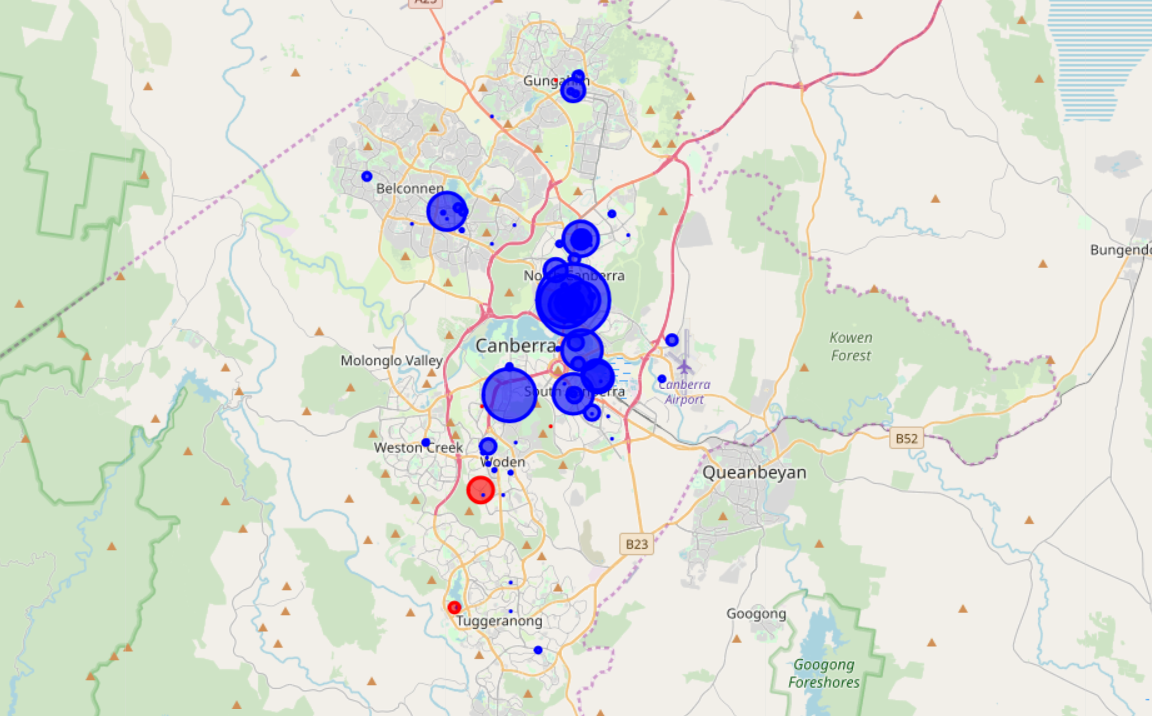


Figure 2: Restaurants in Canberra (including likes): The markers in this figure vary on size depending on how many likes they receive (larger icons have received more likes)

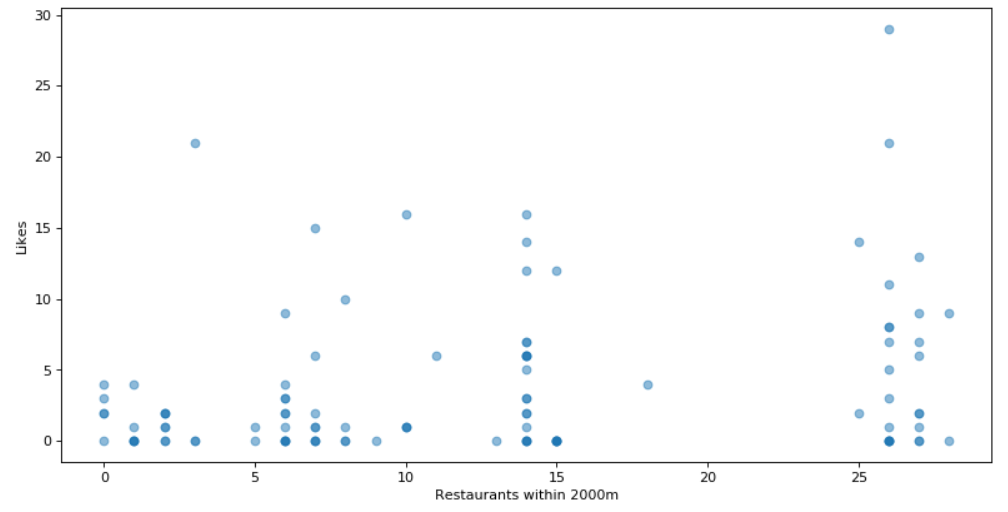


Figure 3: Number of nearby restaurants vs likes: This analysis shows a scatter plot of likes a restaurant receives against how many restaurants are near it. This is different from DBSCAN as it only looks at each restaurant’s neighbours, and doesn’t necessarily represent clusters.

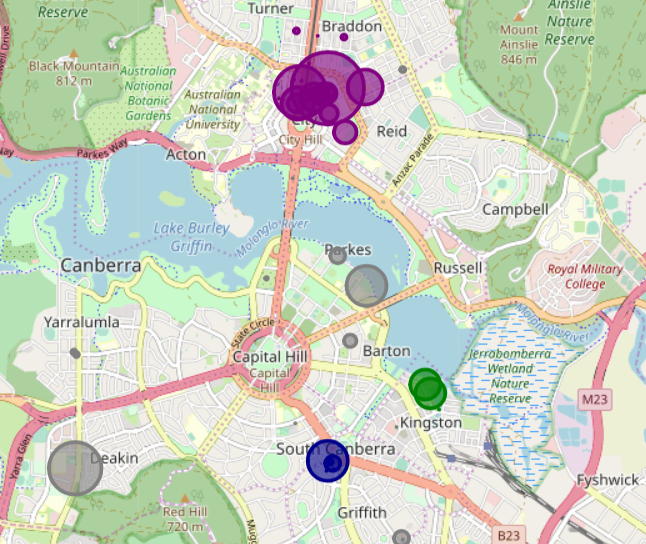
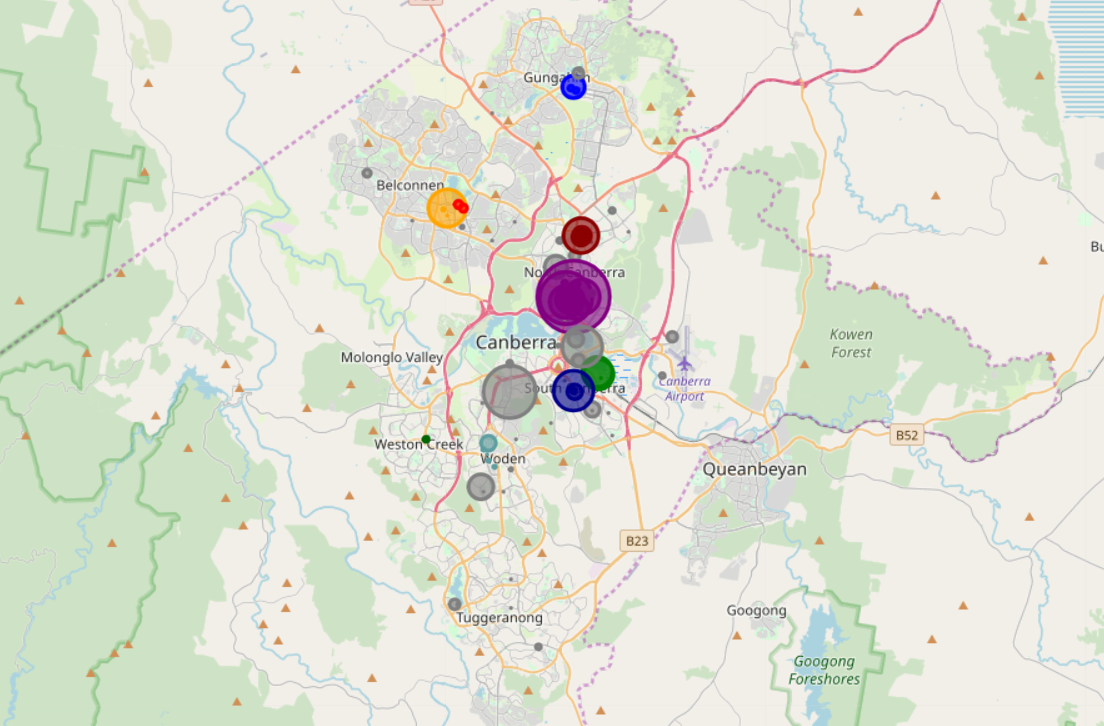


Figure 4: DBSCAN Clustering: This shows a DBSCAN of Canberra Restaurants. Epsilon = 500m and a cluster requires a minimum of 3 restaurants.

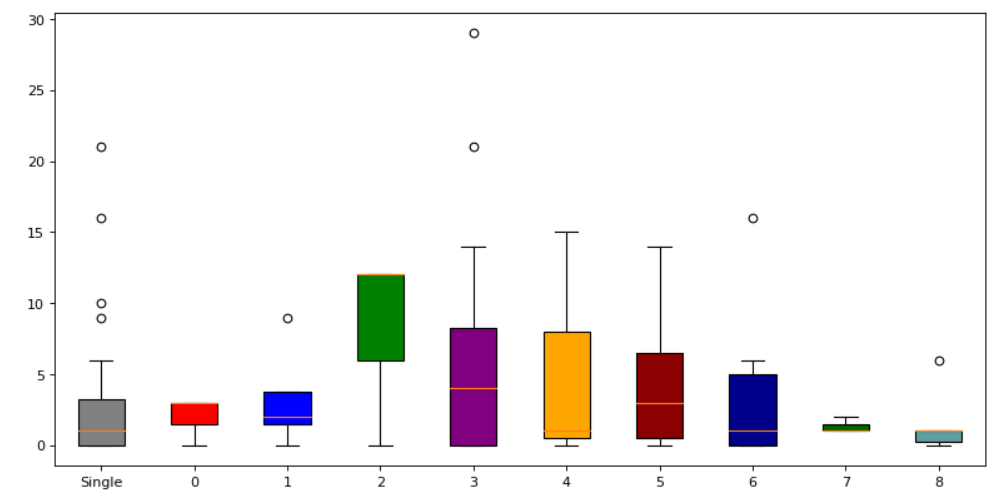


Figure 5: Box plot of Likes for each cluster: The colour of the box plot corresponds to the colour of the cluster in Figure 4. The ‘Single’ column represents restaurants not in a cluster.

# Discussion

Canberra’s restaurants have a number of clusters throughout the territory, and also a significant number of restaurants away from city centres and other restaurants. This provides a good environment to determine if location affects the number of likes received by a restaurant.

Restaurants not in a cluster tend to have a lower number of likes than those within a cluster. Similarly, restaurants with fewer restaurants near but not necessarily within a cluster tend to have less like. This can be seen in the scatter plot in Figure 3. The ‘Green’ cluster near ‘Kingston’ has the highest mean in the box plot, but has a small sample size. The ‘Purple’ Cluster has the second highest value and a much larger sample size, although it still has a large amount of variability. Restaurants not in a cluster, shown as ‘Grey’ in the Folium map and ‘Single’ in the Box plot, have a low mean score but a significant number of outliers. It must also be noted that not all restaurants in Canberra are shown in Foursquare. It should also be noted that new restaurants may have less likes than established restaurants.

# Conclusion

Where should you set up a restaurant in Canberra? This analysis only examines location and ‘likes’ based on Foursquare data, and establishing a successful restaurant requires a lot more analysis than location and ‘likes’. If all other factors were equal, I would recommend setting up a restaurant in the ‘Green’ Kingston cluster. The ‘Green’ cluster has restaurants with a high median number of ‘likes’, and no other Indian Restaurants nearby.