Expanding a Bakery to Brisbane

Introduction

Bourke St Bakery is a popular bakery chain that started in Surry Hills, Sydney and has since expanded to twelve venues across Greater Sydney. The owners wish to expand their business to Brisbane, the capital of Queensland, Australia, but are not familiar with the city and are unsure of the best location to expand their business.

The purpose of this study is to identify candidate locations for the business owners to open their first bakery in Brisbane. The aim is to identify suburbs that are similar to those where already existing bakeries have already found success.

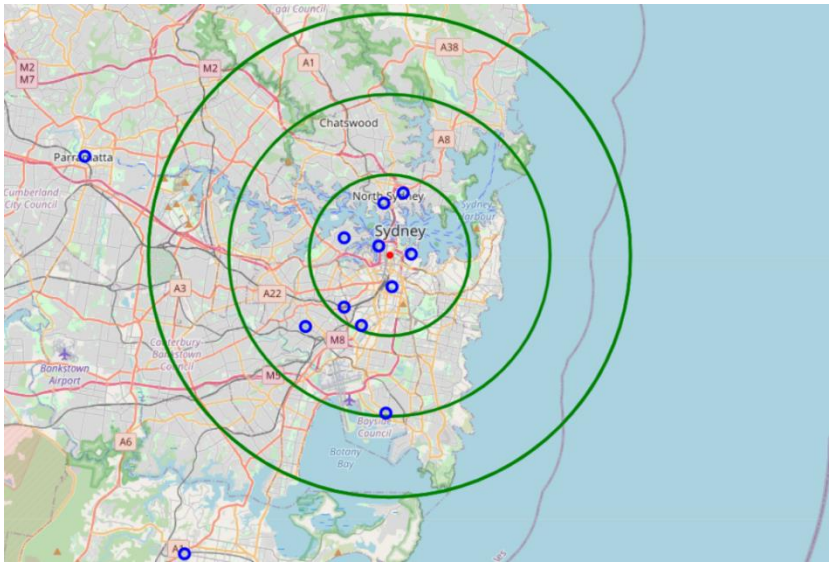
Data Sources

The following data sources were used to create the datasets used for this study:

- Wikipedia for a list of suburbs in Sydney and Brisbane
- Bourke St Bakery's website for a list of bakery locations
- Google's Geocoding API for converting suburb names into geospatial coordinates
- Geopy package in Python for calculating distances between two locations
- Foursquare API for finding venue data in each location

Analysis

The existing can be visualised overlayed on a map of Sydney using the Folium package in Python. The green circles identify 5km, 10km and 15km radii from the city centre.



We can see that most bakeries are within 5km of the city centre and only two are more than 15km away. Therefore, we will limit out study to suburbs within 15km of the city centre.

Suburbs were first parsed from Wikipedia for both cities using Beautiful Soup, a Python library for pulling data out of webpages. They can visualised in the next two figures, red markers indicate suburbs within a 15km radius of the centre and will be considered in the study.

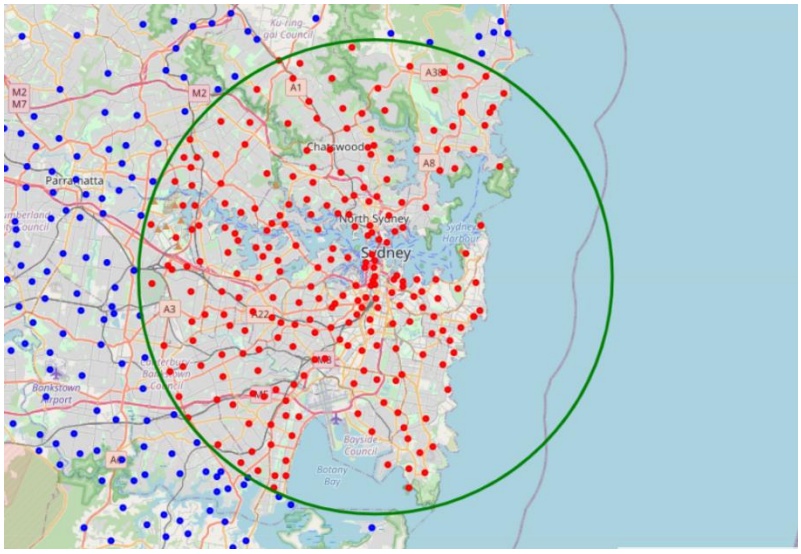


Figure 1: Suburbs in Sydney

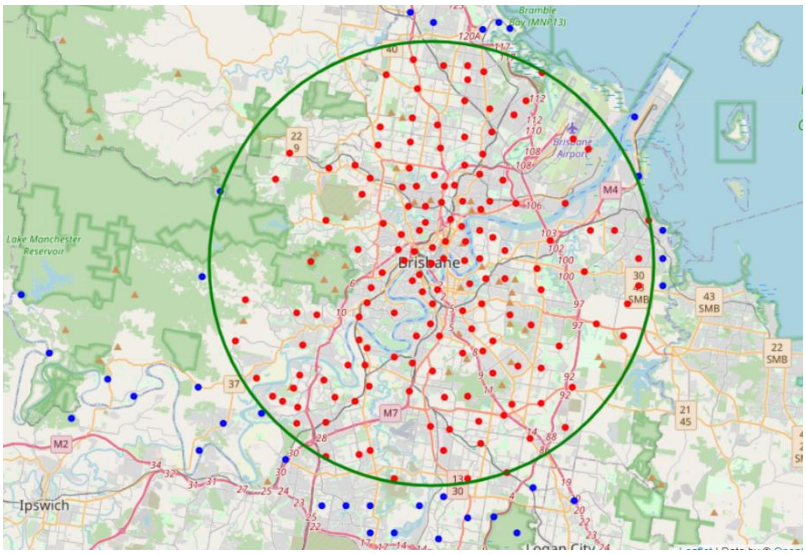


Figure 2: Suburbs in Brisbane

The Foursquare API was used to explore venues in each suburb. The coordinates of each suburb was used as a location, a radius of 1000m and limit of 100 (max for the access I have) were used for the searches. The ten most common venues in each city are shown in the table below.

Sydney Venues	Quantity	Brisbane Venues	Quantity
Café	2301	Café	620
Pub	660	Pub	128
Park	476	Supermarket	114
Thai Restaurant	365	Pizza Place	109
Japanese Restaurant	300	Sandwich Place	93
Pizza Place	300	Bakery	92
Bakery	294	Fast Food Restaurant	89
Italian Restaurant	284	Park	85
Chinese Restaurant	182	Thai Restaurant	80
Supermarket	176	Liquor Store	78

The venues were grouped into suburbs and the data was converted to represent the frequency occurrence of each venue type in each suburb. In this form the data can be passed into a clustering algorithm. The distance from each suburb to the city centre was also considered and clustering was performed with and without distance.

Using the Sydney suburb data to fit the clusters we get the following clusters. For this case a K-Means clustering algorithm with five clusters was used. Five clusters was found to be the optimum number using the elbow method.

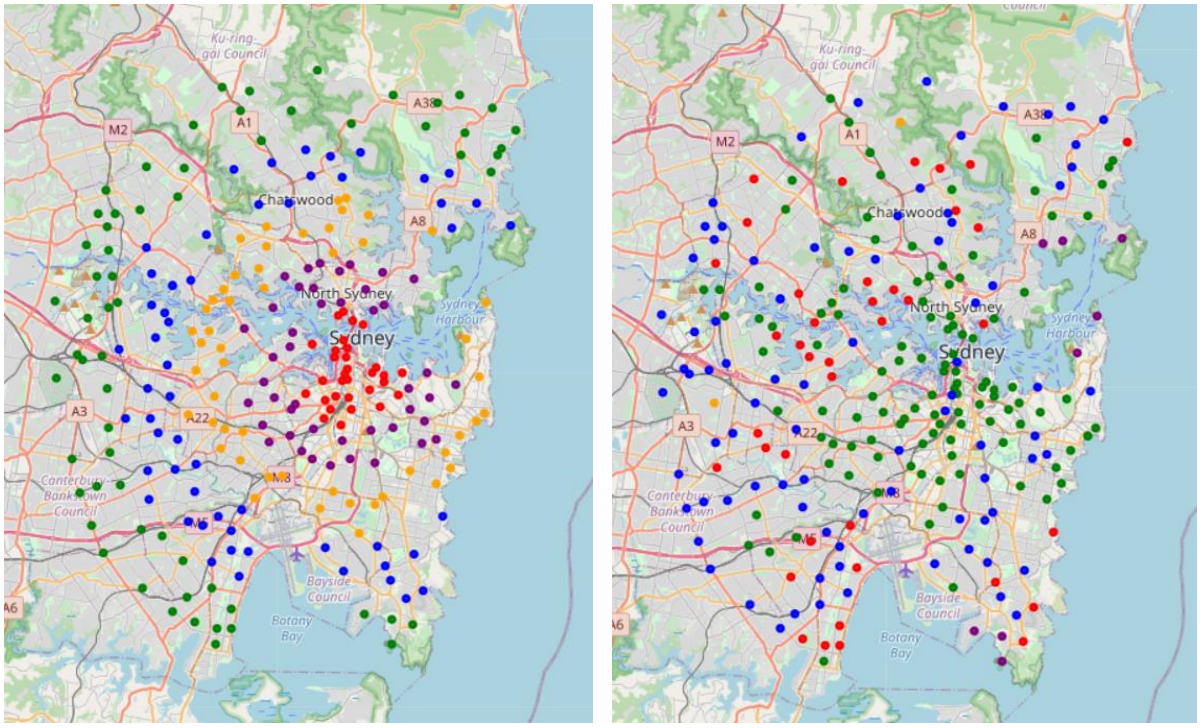


Figure 3: Sydney clusters generated when considering distance to city centre (left) and when not considering distance to city centre (right).

Interstingly, there are two dominant clusters in Sydney when distance to the city centre is not considered, the green and the blue cluster. When examining the venues in each it appears that both have a relatively high number of cafés when compared to other venues.

The blue cluster's venues are made up of approximately 10% of cafés whereas the cafés make up approximately 20% or more of the venues in the green cluster. A similar conclusion that was made with the distance parameter could be made regarding the number of cafés: that the number of cafés is now the driving factor for the clustering algorithm.

Once these clusters were established, the clusters the Brisbane suburb data belonged to could be predicted using the clustering model trained by the Sydney data.

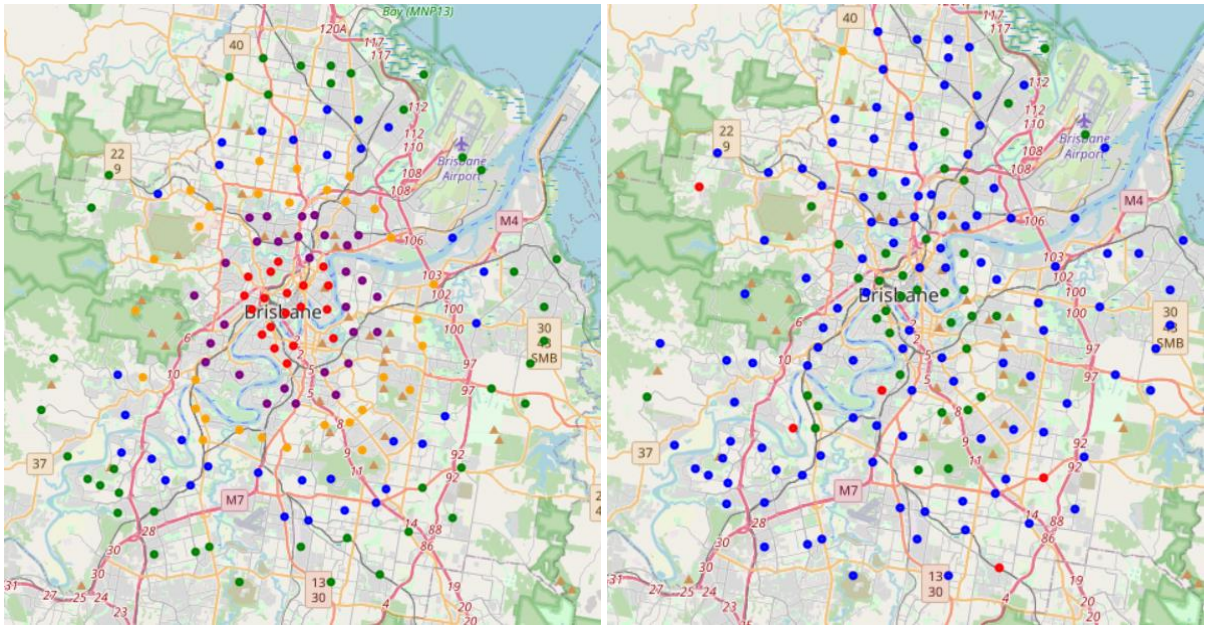


Figure 4:Predicted clusters for Brisbane suburbs when considering distance to city centre (left) and when not considering distance to city centre (right).

We can see that when distance is considered the clusters appear to be grouped based on distance from the city centre, whereas when distance in not considered the clusters appear to be a lot more random. This is expected as venue data is generally not correlated to geographic location.

From these two sets of clustering, we can reduce the eligible suburbs in Brisbane by considering the location of the existing bakeries in Sydney. This is shown in the table below. It should be noted that the colours are only to identify each cluster.

Cluster Colour	Qty: Distnace (left plots)	Qty: Without Distance (right plots)
Red	3	0
Purple	5	0
Yellow	1	0
Blue	1	2
Green	0	8

We see that most existing bakeries are in the red and purple clusters when distance is considered. We also see that most bakeries are in the green cluster when distance is not considered. Therefore, we can infer that the most eligible suburbs in Brisbane will belong to both of these groups in each analysis.

This results in the suburbs shown in the figure below. We see that all suburbs are mostly inner city and surrounding suburbs.

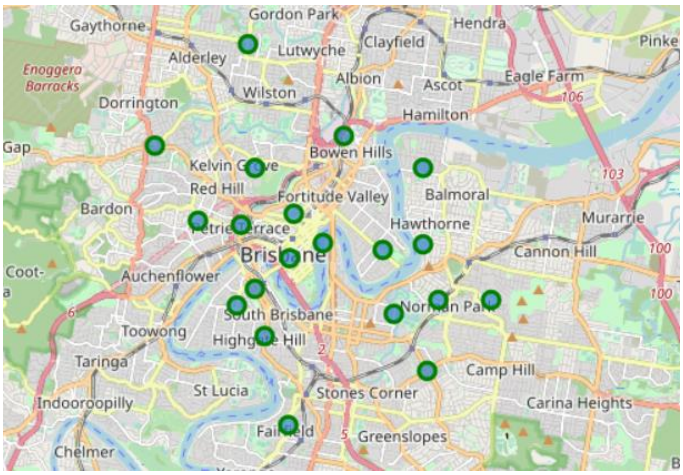


Figure 5: Final suburb choices for the new bakery.

Conclusion

From the above analysis we have arrived at candidate suburbs for the expansion of Bourke St Bakery into Brisbane. These suburbs are all inner city, so would expect higher foot traffic and are all in suburbs where there is already a high number of cafés, bars and restaurants, when compared to other venues.

Ashgrove, Bowen Hills, Brisbane CBD, Bulimba, Coorparoo, East Brisbane, Fairfield, Grange, Hawthorne, Highgate Hill, Kangaroo Point, Kelvin Grove, New Farm, Norman Park, Paddington, Petrie Terrace, Seven Hills, South Brisbane, Spring Hill, West End

The recommendation for the business owners of Bourke St Bakery is that they should open their new bakery in one of these suburbs after conducting further research into aspects that are not included here, such as: public transport options, businesses that may not show up as venues and population density.