# Opening a New Bakery in Brisbane, Australia

**IBM Applied Data Science Capstone**

**Coursera Capstone**

19.04.20

# Introduction

For many shoppers, visiting café and bakeries is a great way to relax and enjoy themselves during weekends and holidays. Property developers are also taking advantage of this trend to build more Bakeries to cater to the demand. As a result, there are many Bakeries in the city of Brisbane and many more are being built. Opening Bakeries allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new Bakery requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the Bakery is one of the most important decisions that will determine whether the mall will be a success or a failure.

The objective of this capstone project is to analyse and select the best locations in the city of Brisbane, Australia to open a new Bakery. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Brisbane, Australia, if a property developer is looking to open a new Bakery, where would you recommend that they open it?

This project is particularly useful to property developers and investors looking to open or invest in new Bakeries in Brisbane. This project is timely as the city is currently suffering from oversupply of Bakeries.

To solve the problem, this report will need the following data:

• List of neighbourhoods in Brisbane. This defines the scope of this project which is confined to the city of Brisbane, the capital city of the country of Australia.

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to Bakeries. This report will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Brisbane) contains a list of neighbourhoods in Brisbane, with a total of 104 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Bakery category in order to help us to solve the business problem put forward.

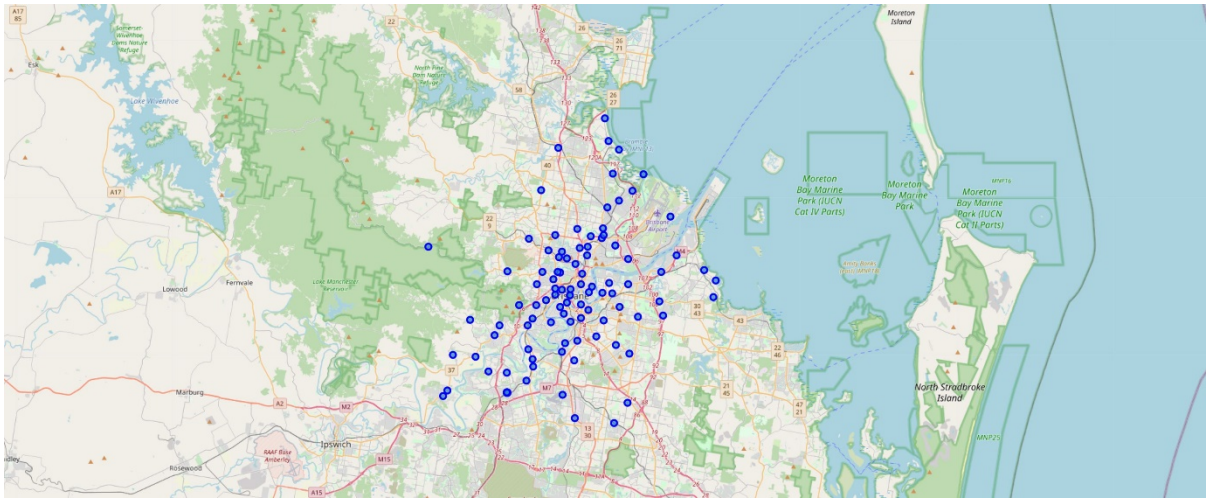
This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

# Methodology

Firstly, we need to get the list of neighbourhoods in the city of Brisbane. Fortunately, the list is available in the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Brisbane). This report will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. However, this is just a list of names. This report need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, This Report will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, This Report will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Brisbane.

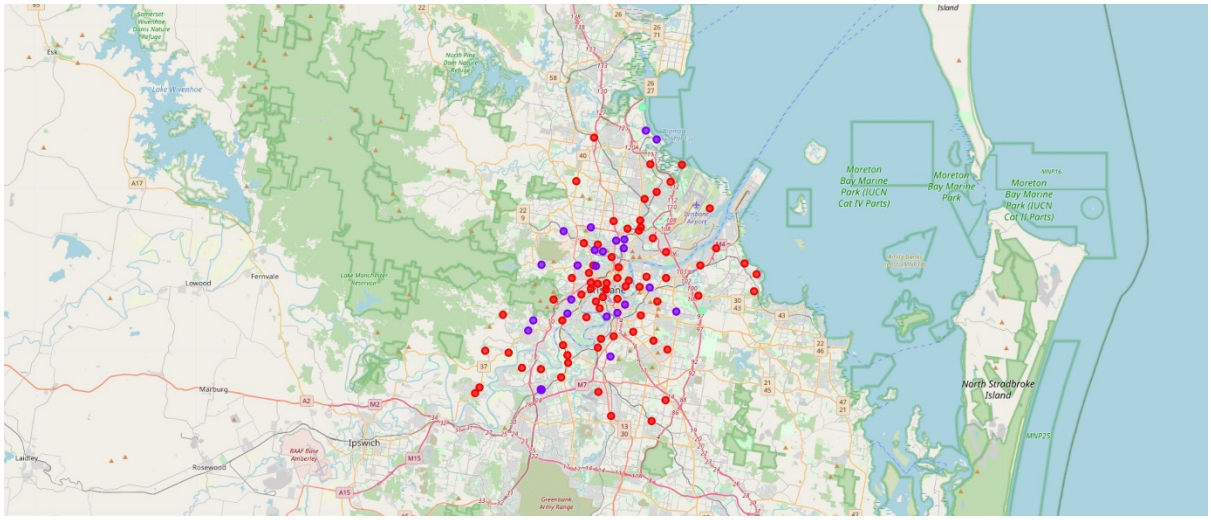
Next, This Report will use Foursquare API to get the top 50 venues that are within a radius of 1000 meters. This Report need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. This Report then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and this Report will extract the venue name, venue category, venue latitude and longitude. With the data, this report can check how many venues returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, this report will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, this report is also preparing the data for use in clustering. Since This report are analysing the “Bakery” data, this report will filter the “Bakery” as venue category for the neighbourhoods.

Lastly, this report will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. This report will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “Bakery”. The results will allow us to identify which neighbourhoods have higher concentration of Bakeries while which neighbourhoods have feThis reportr number of Bakeries. Based on the occurrence of Bakeries in different neighbourhoods, it will help us to ansThis reportr the question as to which neighbourhoods are most suitable to open new Bakeries.




## Results

The results from the k-means clustering show that this report can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Bakery”:




• Cluster 0: Neighbourhoods with low number or no existence of Bakeries

• Cluster 1: Neighbourhoods with moderate number of Bakeries

• Cluster 2: Neighbourhoods with high concentration of Bakeries The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.

• Cluster 0

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **0** | ► Acacia Ridge, Queensland (3 P) | 0.000000 | 0 | -27.590410 | 153.028480 |
| **68** | ► Nudgee Beach, Queensland (2 P) | 0.000000 | 0 | -27.348830 | 153.105180 |
| **67** | ► Norman Park, Queensland (6 P) | 0.000000 | 0 | -27.480033 | 153.078572 |
| **65** | ► New Farm, Queensland (20 P) | 0.020000 | 0 | -27.465890 | 153.044510 |
| **64** | ► Mount Ommaney, Queensland (2 P) | 0.000000 | 0 | -27.544310 | 152.931650 |
| **63** | ► Mount Gravatt East, Queensland (3 P) | 0.000000 | 0 | -27.526740 | 153.089920 |
| **62** | ► Mount Coot-tha, Queensland (5 P) | 0.000000 | 0 | -27.478210 | 152.965920 |
| **59** | ► Moggill, Queensland (3 P) | 0.000000 | 0 | -27.568640 | 152.880670 |
| **57** | ► Milton, Queensland (1 C, 14 P) | 0.000000 | 0 | -27.468070 | 153.006790 |
| **56** | ► Manly, Queensland (6 P) | 0.000000 | 0 | -27.454210 | 153.186210 |
| **55** | ► Lytton, Queensland (5 P) | 0.000000 | 0 | -27.428980 | 153.142310 |
| **53** | ► Lota, Queensland (4 P) | 0.000000 | 0 | -27.470230 | 153.183550 |
| **101** | ► Yeerongpilly, Queensland (3 P) | 0.000000 | 0 | -27.524640 | 153.014470 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **50** | ► Kedron, Queensland (8 P) | 0.000000 | 0 | -27.402920 | 153.031540 |
| **49** | ► Kangaroo Point, Queensland (28 P) | 0.000000 | 0 | -27.477720 | 153.035710 |
| **48** | ► Kalinga, Queensland (3 P) | 0.000000 | 0 | -27.410050 | 153.046270 |
| **46** | ► Indooroopilly, Queensland (23 P) | 0.000000 | 0 | -27.498260 | 152.975740 |
| **45** | ► Holland Park, Queensland (5 P) | 0.000000 | 0 | -27.517730 | 153.074290 |
| **69** | ► Nudgee, Queensland (3 P) | 0.000000 | 0 | -27.364760 | 153.092860 |
| **70** | ► Nundah, Queensland (1 C, 18 P) | 0.000000 | 0 | -27.401960 | 153.059980 |
| **71** | ► Oxley, Queensland (5 P) | 0.000000 | 0 | -27.553280 | 152.974090 |
| **72** | ► Paddington, Queensland (19 P) | 0.020000 | 0 | -27.461920 | 153.006690 |
| **100** | ► Wynnum, Queensland (1 C, 16 P) | 0.000000 | 0 | -27.443690 | 153.173650 |
| **97** | ► Windsor, Queensland (1 C, 15 P) | 0.031250 | 0 | -27.437860 | 153.029590 |
| **95** | ► West End, Queensland (15 P) | 0.020000 | 0 | -27.480120 | 153.012220 |
| **93** | ► Virginia, Queensland (6 P) | 0.000000 | 0 | -27.381690 | 153.064630 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **91** | ► Toombul, Queensland (6 P) | 0.000000 | 0 | -27.408740 | 153.060500 |
| **90** | ► Tingalpa, Queensland (3 P) | 0.000000 | 0 | -27.474570 | 153.122990 |
| **88** | ► Teneriffe, Queensland (15 P) | 0.020000 | 0 | -27.460280 | 153.047480 |
| **85** | ► St Lucia, Queensland (10 P) | 0.027778 | 0 | -27.495211 | 153.001729 |
| **44** | ► Highgate Hill, Queensland (5 P) | 0.020000 | 0 | -27.486990 | 153.016360 |
| **84** | ► Spring Hill, Queensland (46 P) | 0.020000 | 0 | -27.462530 | 153.023780 |
| **82** | ► Sinnamon Park, Queensland (3 P) | 0.000000 | 0 | -27.545290 | 152.952070 |
| **80** | ► Sherwood, Queensland (6 P) | 0.000000 | 0 | -27.532160 | 152.981070 |
| **78** | ► Runcorn, Queensland (3 P) | 0.000000 | 0 | -27.595570 | 153.072740 |
| **77** | ► Red Hill, Queensland (12 P) | 0.020000 | 0 | -27.452610 | 153.004340 |
| **76** | ► Pullenvale, Queensland (3 P) | 0.000000 | 0 | -27.527950 | 152.891690 |
| **75** | ► Pinkenba, Queensland (5 P) | 0.000000 | 0 | -27.390620 | 153.135350 |
| **74** | ► Pinjarra Hills, Queensland (2 P) | 0.000000 | 0 | -27.529440 | 152.917630 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **73** | ► Petrie Terrace, Queensland (5 P) | 0.000000 | 0 | -27.463243 | 153.014480 |
| **83** | ► South Brisbane, Queensland (1 C, 55 P) | 0.000000 | 0 | -27.475890 | 153.019600 |
| **42** | ► Hendra, Queensland (2 P) | 0.000000 | 0 | -27.419400 | 153.073680 |
| **51** | ► Kelvin Grove, Queensland (11 P) | 0.025000 | 0 | -27.445260 | 153.009540 |
| **19** | ► Brisbane localities (2 C, 15 P) | 0.000000 | 0 | -27.468440 | 153.023340 |
| **22** | ► Cannon Hill, Queensland (4 P) | 0.000000 | 0 | -27.457650 | 153.088220 |
| **20** | ► Brookfield, Queensland (3 P) | 0.000000 | 0 | -27.493000 | 152.910890 |
| **41** | ► Hemmant, Queensland (6 P) | 0.000000 | 0 | -27.445400 | 153.125530 |
| **18** | ► Brisbane central business district (16 C, ... | 0.000000 | 0 | -27.468440 | 153.023340 |
| **16** | ► Bridgeman Downs, Queensland (2 P) | 0.000000 | 0 | -27.364620 | 152.990960 |
| **15** | ► Bowen Hills, Queensland (11 P) | 0.000000 | 0 | -27.447600 | 153.036740 |
| **14** | ► Boondall, Queensland (1 C, 7 P) | 0.000000 | 0 | -27.347720 | 153.071430 |
| **26** | ► Clayfield, Queensland (10 P) | 0.000000 | 0 | -27.412230 | 153.058940 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **12** | ► Bellbowrie, Queensland (3 P) | 0.000000 | 0 | -27.563020 | 152.885900 |
| **10** | ► Banyo, Queensland (5 P) | 0.000000 | 0 | -27.374730 | 153.078100 |
| **9** | ► Balmoral, Queensland (3 P) | 0.000000 | 0 | -27.456200 | 153.067220 |
| **8** | ► Bald Hills, Queensland (4 P) | 0.000000 | 0 | -27.322100 | 153.009950 |
| **7** | ► Auchenflower, Queensland (1 C, 8 P) | 0.020000 | 0 | -27.473420 | 152.996140 |
| **4** | ► Archerfield, Queensland (5 P) | 0.000000 | 0 | -27.567170 | 153.015120 |
| **3** | ► Annerley, Queensland (9 P) | 0.028571 | 0 | -27.513910 | 153.031140 |
| **2** | ► Alderley, Queensland (4 P) | 0.000000 | 0 | -27.424490 | 152.998770 |
| **11** | ► Bardon, Queensland (1 C, 9 P) | 0.000000 | 0 | -27.458010 | 152.986050 |
| **27** | ► Coorparoo, Queensland (1 C, 16 P) | 0.000000 | 0 | -27.493600 | 153.061010 |
| **102** | ► Yeronga, Queensland (9 P) | 0.000000 | 0 | -27.516420 | 153.017530 |
| **31** | ► Eagle Farm, Queensland (7 P) | 0.000000 | 0 | -27.432570 | 153.088040 |
| **28** | ► Corinda, Queensland (8 P) | 0.000000 | 0 | -27.539380 | 152.981690 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **37** | ► Grange, Queensland (2 P) | 0.000000 | 0 | -27.425470 | 153.014340 |
| **33** | ► Eight Mile Plains, Queensland (3 P) | 0.000000 | 0 | -27.575450 | 153.087570 |
| **35** | ► Fortitude Valley, Queensland (41 P) | 0.020000 | 0 | -27.457880 | 153.035580 |
| **38** | ► Greenslopes, Queensland (4 P) | 0.000000 | 0 | -27.509730 | 153.052650 |
| **36** | ► Graceville, Queensland (6 P) | 0.000000 | 0 | -27.522480 | 152.976480 |
| **40** | ► Hawthorne, Queensland (3 P) | 0.000000 | 0 | -27.466030 | 153.059660 |

• Cluster 1

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **94** | ► Wacol, Queensland (10 P) | 0.076923 | 1 | -27.565002 | 152.952512 |
| **96** | ► Wilston, Queensland (3 P) | 0.068966 | 1 | -27.432200 | 153.019630 |
| **92** | ► Toowong (1 C, 25 P) | 0.040000 | 1 | -27.478510 | 152.985620 |
| **5** | ► Ascot, Queensland (10 P) | 0.076923 | 1 | -27.565002 | 152.952512 |
| **6** | ► Ashgrove, Queensland (8 P) | 0.041667 | 1 | -27.445650 | 152.991960 |
| **89** | ► The Gap, Queensland (5 P) | 0.055556 | 1 | -27.444900 | 152.952910 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **98** | ► Woolloongabba (24 P) | 0.040000 | 1 | -27.490950 | 153.035340 |
| **47** | ► Ithaca, Queensland (7 P) | 0.076923 | 1 | -27.565245 | 152.953137 |
| **87** | ► Taringa, Queensland (6 P) | 0.038462 | 1 | -27.491700 | 152.981500 |
| **86** | ► Stafford, Queensland (2 P) | 0.066667 | 1 | -27.409070 | 153.006780 |
| **99** | ► Wooloowin, Queensland (5 P) | 0.058824 | 1 | -27.420420 | 153.043060 |
| **1** | ► Albion, Queensland (6 P) | 0.035714 | 1 | -27.429260 | 153.042280 |
| **39** | ► Hamilton, Queensland (16 P) | 0.076923 | 1 | -27.564669 | 152.952295 |
| **61** | ► Morningside, Queensland (1 C, 4 P) | 0.045455 | 1 | -27.467060 | 153.070520 |
| **34** | ► Enoggera, Queensland (11 P) | 0.076923 | 1 | -27.564930 | 152.953060 |
| **52** | ► Kenmore, Queensland (6 P) | 0.038462 | 1 | -27.508030 | 152.938660 |
| **60** | ► Moorooka, Queensland (3 P, 2 F) | 0.090909 | 1 | -27.533210 | 153.027750 |
| **25** | ► Chelmer, Queensland (10 P) | 0.076923 | 1 | -27.565002 | 152.952512 |
| **66** | ► Newmarket, Queensland (8 P) | 0.050000 | 1 | -27.430920 | 153.010480 |
| **24** | ► Chapel Hill, Queensland (2 P) | 0.090909 | 1 | -27.498730 | 152.944040 |
| **23** | ► Carina, Queensland (3 P) | 0.083333 | 1 | -27.489740 | 153.099340 |
| **29** | ► Darra, Queensland (3 P) | 0.076923 | 1 | -27.565610 | 152.952420 |
| **81** | ► Shorncliffe, Queensland (4 P) | 0.041667 | 1 | -27.324400 | 153.078280 |

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **21** | ► Bulimba, Queensland (10 P) | 0.076923 | 1 | -27.565002 | 152.952512 |
| **30** | ► Dutton Park, Queensland (10 P) | 0.033333 | 1 | -27.494580 | 153.023830 |
| **32** | ► East Brisbane, Queensland (13 P) | 0.040000 | 1 | -27.483450 | 153.043350 |
| **54** | ► Lutwyche, Queensland (7 P) | 0.047619 | 1 | -27.421610 | 153.033990 |
| **79** | ► Sandgate, Queensland (9 P) | 0.047619 | 1 | -27.315450 | 153.066040 |
| **58** | ► Mitchelton, Queensland (6 P) | 0.058824 | 1 | -27.412630 | 152.977100 |
| **43** | ► Herston, Queensland (1 C, 10 P) | 0.047619 | 1 | -27.446038 | 153.012298 |

• Cluster 2

| Neighborhood | Bakery | Cluster Labels | Latitude | Longitude | |
|---|---|---|---|---|---|
| **17** | ► Brighton, Queensland (2 P) | 0.250000 | 2 | -27.29293 | 153.06216 |
| **13** | ► Belmont, Queensland (2 P) | 0.166667 | 2 | -27.48904 | 153.12734 |

# Discussion

In terms of facility concentration, most of the bakery are placed in the central area of Brisbane city, with the highest number in cluster 2. On the other hand, cluster 0 has very low number of Bakery in the neighborhoods. This represents a great opportunity and high potential areas to open new bakery as there is very little to no competition from existing bakeries. Meanwhile, Bakeries in cluster 2 are likely suffering from intense competition due to oversupply and high concentration. From another perspective, this also shows that the oversupply of Bakeries mostly happened in the Asian town. Therefore, this project recommends property developers to capitalize on these findings to open new bakeries in neighborhoods in cluster 0 with little to no competition. However, this is only an analysis based on the concentration of facilities, and additional variables such as population density or change of population density must be considered. Therefore, this report recommends cluster 0 in opening a new bakery but highlights the need for further research.

As observations noted from the map in the Results section, most of the Bakeries are concentrated in the central area of Brisbane city, with the highest number in cluster 2 and moderate number in cluster 1. On the other hand, cluster 0 has very low number to no Bakeries in the neighbourhoods. This represents a great opportunity and high potential areas to open new Bakeries as there is very little to no competition from existing facilities. Meanwhile, Bakeries in cluster 2 are likely suffering from intense competition due to oversupply and high concentration. From another perspective, the results also show that the oversupply of bakeries mostly happened in the central area of the city, with the suburb area still have very few Bakeries. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new Bakeries in neighbourhoods in cluster 1 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 2 which already have high concentration of Bakeries and suffering from intense competition.