LOCATION ANALYSIS FOR A NEW RESTAURANT IN BRISBANE

INTRODUCTION

As a capital city of Queensland State, the third most populous city in Australia after Sydney and Melbourne, Brisbane offers a more manageable lifestyle and affordable cost of living. The city is known as one of Australian's major business hubs with strengths in mining, banking, transportation, information technology and food that create jobs and business opportunities. City residents are spoilt by restaurants and dining establishments including outdoor dining, food trucks and roof-top bars.

In today's tight competition, choosing a right location for business can be crucial but it requires research and planning about the area, demographic and market profiles of the location. It becomes a balancing act, for investors who are interested to open a restaurant in Brisbane area, between spending time and money undertaking proper assessment to choose a location and starting a business itself. This location analytics using data science provides a high-level overview of suburb within close proximity to Brisbane City (CBD) presenting information and insight for investors considering locations for their new restaurants.

DATA

Dataset used for this analysis includes:

- Decile Index (suburb profile) originated from Australian Bureau Statistics (ABS)[1] indicates relative socio-economic and demographic profiles of suburbs derived from Australian census data;
- Data on venues in suburbs from Foursquare[2] used to capture information about geographic location, category, and popularity, and provide insight about potential competitions or/and restaurant opportunities.

Geospatial boundary areas for suburbs near Brisbane in a shapefile (.shp) format is sourced and downloaded ABS website, and converted into geojson file. Only suburbs within 10 kilometres radius from Brisbane as a city centre selected and included in this analysis as can be seen in Figure 1. Centroids of polygon boundary geometries were calculated using Python, generating latitude and longitude information used to plot the data.

The data also provides information on distribution of 2016 Australian Socio-Economic Index for Areas (SEIFA) for Statistical Area Level 2 (SA2). The index values used for this analysis derived from aus_decile attribute from SEIFA, suggesting suburb area index ordered from lowest to highest values in terms of relative socio-economic advantage for the whole Australia, which is designed and constructed as a measure for area-based deciles (ten equal group division). Areas with the highest index of 10 has the highest relative socio-economic advantage than the other area groups.[3]

Using those suburb areas, information and data on venues within the suburbs are collated using Foursquare.

Figure 1. Brisbane City and Surrounding Suburbs

Data processing and preparation is performed using Python[4] and Pandas library creating geopandas dataframe for the geospatial data from the ABS and pandas dataframe for the venue data from Foursquare. The shapefile geospatial is converted into geojson file to allow easier interaction with the venue data. Those two dataset, then, is merged allowing seamless data visualisation to present analysis findings which will be discussed in Result and Discussion section later in this report.

METHODOLOGY

Popular unsupervised learning algorithm called K-means clustering[5] is used for the analysis to create cluster groups of suburbs identified based on most popular venue. In order to use the algorithm, categorical values of venue data captured from Foursquare is converted into numerical values using One-Hot-Encoding[6] from SciKit learn library.

The venue data collated from Foursquare is processed to create group rows by suburb and by taking the mean of frequency of occurrence of each venue category, followed by the next step generating a new dataframe recording the top five most common venues identified.

	Cluster	Suburb	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	5	Albion	Café	Liquor Store	Cricket Ground	Sporting Goods Shop	Comedy Club
1	0	Alderley	Pizza Place	Grocery Store	Supermarket	Thai Restaurant	Train Station
2	5	Annerley	Café	Pizza Place	Camera Store	Sandwich Place	Supermarket
3	5	Ascot	Pizza Place	Burger Joint	Thai Restaurant	Flea Market	Café
4	4	Ashgrove	Park	Bus Station	Yoga Studio	Event Space	Flower Shop

Table 1. Venue Data for Analysis

K value identification is undertaken through an iterative process of running the algorithm and reviewing outcomes based on knowledge of Brisbane areas. After such process, k=6 is chosen to train the venue data for venue category clustering.

Figure 3 shows distribution of clusters derived from k-means clustering.

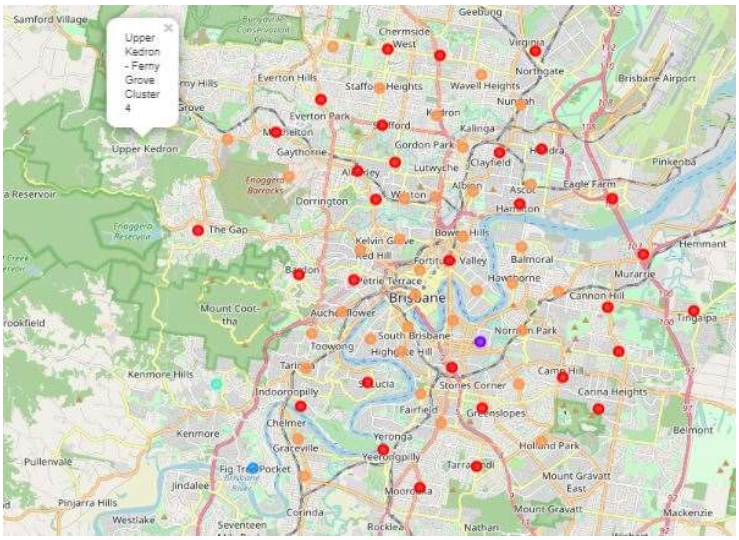

Figure 2. Distribution of Clusters Near Brisbane

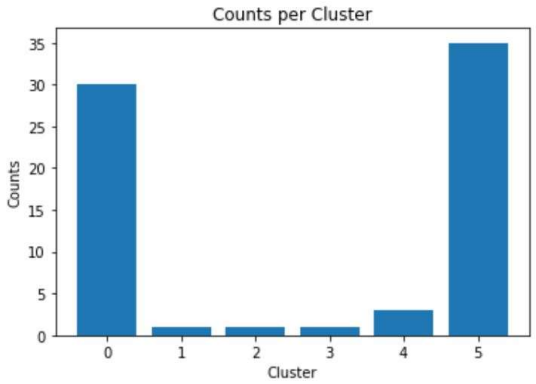

Figure 3. Number of Suburbs in Clusters

Distribution of number of categories (Counts) within each cluster type (Cluster) is demonstrated in Figure 3.

After each cluster is examined, distinguished, and finalised, a name is assigned and added to each cluster category to bring a more meaningful description.

Using Folium[7] library in Python, two data layers of analysis outcome are displayed and presented through thematic mapping, as displayed in Figure 4.

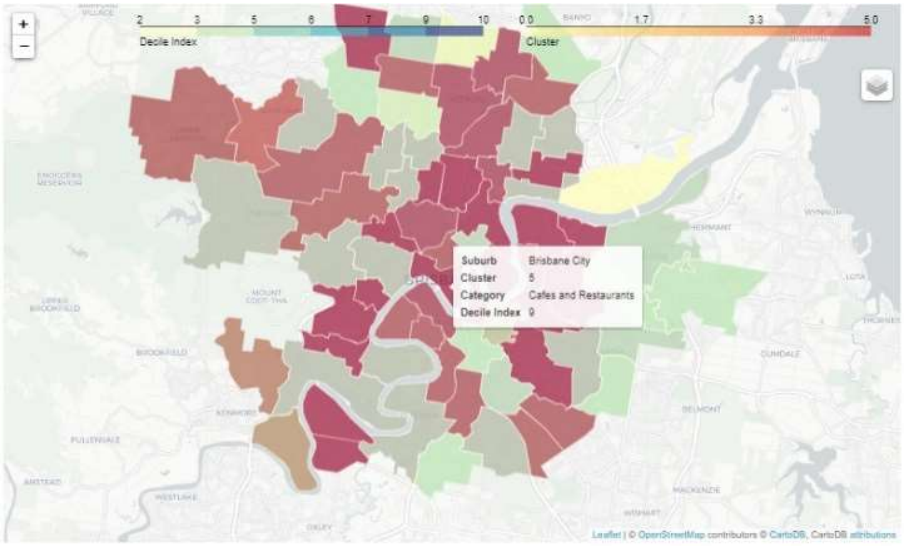

Figure 4. Thematic Map with Decile Index and Cluster Data Layers

The Decile Index layer shows colour-coded suburbs indicating relative socio-economic advantage, and the Cluster layer that displays colour-coded suburbs with different cluster groups.

RESULTS AND DISCUSSION

Suburbs within 10km from Brisbane City are selected resulting of 71 suburbs to be used in the analysis. Using one-hot encoding and Foursquare, 815 venues are found in those suburbs, and categorised generating 169 unique venue types.

Six clusters from k-means clustering analysis are identified, and summarised in Table 2 below.

Cluster	Description	Number of Suburbs
0	Large Stores	30
1	Pubs and Bars	1
2	General Stores	1
3	Special Stores	1
4	Service Stores	3
5	Cafes and Restaurants	35

Table 2. Cluster Analysis Outcome

An investor looking to open a new café will now be able to use the map, find those areas with dark blue colour (the highest decile index value) from the Decile Index map and identify areas in red (Cluster 5 Café and Restaurants) from the Cluster map indicating high concentration of cafés and restaurants. The investor will have the advantage to find other areas that have similar index values with less density of cafés and restaurants (less competition) by understanding distribution of analysis findings plotted on the map.

The application can be described using Figure 5 where the map provides visualisation and overview of potential suburbs for a new café or restaurant based on socio-economic suburb profile and cluster of venues.

Whilst Brisbane City shows a high density of cafes and restaurants (red colour) and has a decile index of 9 (dark blue colour), there are a few suburbs nearby that could be potential locations for a café or a restaurant where those areas still show high decile index such as Paddington-Milton and Fortitude Valley but not in cluster 5 (Cafes and Restaurants).

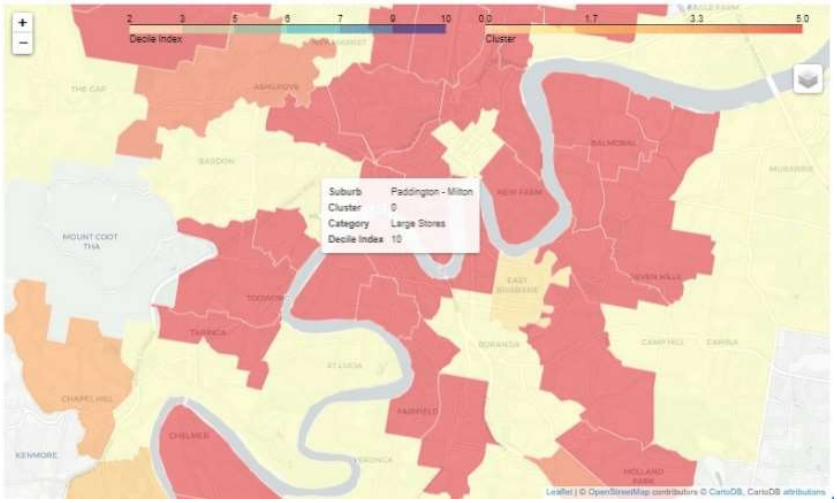

Figure 5. Using Map to View Potential Locations

Improvements can be achieved to get better analysis outcomes including:

- Prepare venue data to have more defined categories (minimise generalisation);
- Merge some cluster's variance that has low values with other cluster;
- Perform more iterative process to find optimal k value for k-means clustering;
- Apply weighted criteria evaluation for Decile Index and Cluster data to provide analysis performance scoring.

Detailed analysis is a necessary step when finalising a decision to choose an area for a new business. However this analysis which findings presented in a thematic map provides information that allow investors to assess areas of interests effectively.

CONCLUSION

This analysis is aimed to provide a high level overview of areas suitable to open a new restaurant for investors who may not be familiar with Brisbane City and surrounding areas. Suburb profile data from ABS combined with information about venues from Foursquare are compiled and analysed to provide insight on potential suburbs to target. With data presented as a thematic map, investors are able to turn on/off the data layers to find areas of interest as a first step when choosing locations for business.

Final decision will be made with a more thorough and detailed analysis, and some consideration on other factors such as competition, access to locations, zoning regulations, and target market.

REFERENCE

1. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~SOCIO-ECONOMIC%20INDEXES%20FOR%20AREAS%20(SEIFA)%202016~1
2. https://developer.Foursquare.com/places
3. Australian Bureau Statistics, 2018, *Technical Paper Socio-Economic Indexes for Areas (SEIFA)*, 2016, 2033.0.55.001.
4. https://en.wikipedia.org/wiki/Python_(programming_language)
5. https://en.wikipedia.org/wiki/K-means_clustering
6. https://en.wikipedia.org/wiki/One-hot
7. https://python-visualization.github.io/folium/