

Determining Bakery Locations in Singapore



Table of Content

Introduction	3
Problem Statement	3
Data Gathering	3
Methodology	4
Analysis	4
Results	5
Recommendations	7
Conclusion	7

Introduction

Singapore is a cosmopolitan city near the equator with a varied population mix of about 5.7 million inhabitants¹. Till early 2020, the entertainment and dining scene has been booming. One specific restaurant sector that has seen a lot of growth since the last 5 years are bakeries².

There are still large opportunities for speciality bakeries as families love to eat out. This paper aims to identify a best possible location for a new bakery with the assumption that the location of other popular bakeries is a good indicator of potential success.

Problem Statement

The success of a new bakery is based on many facts, including the location, the decor, the service, the quality of the food and more.

This paper aims to analyse a few of these variables to determine what the best location could be for a new bakery.

OBJECTIVE

***Find the best location for a new bakery in Central Singapore
using a limited number of free data sources***

Data Gathering

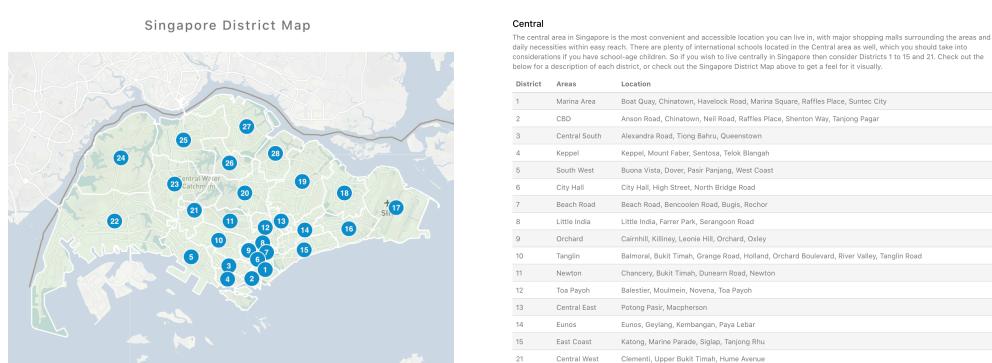
For the analysis, the following data will be needed:

- Location information about Singapore's neighbourhoods/districts
- Location and popularity data of existing bakeries and similar venues in Singapore

One limitation of this paper is that only free data sources have been considered. These sources are:

1. Keylocation.sg (<https://keylocation.sg/singapore/districts-map>)

This website provides an overview of all the Singapore districts with further location information in table form.



¹ <https://www.singstat.gov.sg/modules/infographics/population>

² <https://thepeakmagazine.com.sg/gourmet-travel/singapore-new-bakeries-2015/>

2. FourSquare Developers (<https://developer.foursquare.com>)

This service provides up-to-date geo and other information about venues around the world. A free account is available that allows to query venues categories, locations and popularity (likes). Only end points requiring Userless authentication (no need for a valid OAuth access token) are considered.

3. Singapore Data portal (<https://data.gov.sg>)

The Singapore government has actively been making available numerous online data sources on the Singapore Data portal. However, a quick initial analysis of the available data on this portal indicates that there is no specific data source that would be useful for this paper.

Methodology

Jupyter Notebook with Python 3 and Pandas will be used to perform the analysis. In the first step, the different data sources will be read into Pandas data frames. The data will then be cleaned and normalised so that it can be used for further analysis.

District/neighbourhood data will be scraped from the Keylocation.sg website using BeautifulSoup. Then GeoPy Geocoder is used to find the longitude and latitude of each district.

The following FourSquare API end-points will be used:

- **Categories** (<https://api.foursquare.com/v2/venues/categories>)
 - This end point is used to determine the category ID for bakeries.
- **Search** (<https://api.foursquare.com/v2/venues/search>)
 - This end point is used to get all available restaurants in each district
- **Likes** (https://api.foursquare.com/v2/venues/VENUE_ID/likes)
 - This end point is used to get the number of likes per venue

Using Pandas data frames, the number of bakeries per district and the total number of likes per district is calculated.

Bar charts and map visualisations will be used to determine the top rated bakeries and where they are located in Singapore. Folium is used for the mapping.

Analysis

Link to GitHub Jupyter notebook.

It was identified that using the district name in the location look-up (eg. 'Marina Area Singapore'), does not always provide longitude and latitude information. As such, a decision was made to use the first location listed per district. For example, for Marina Area that would be Boat Quay.

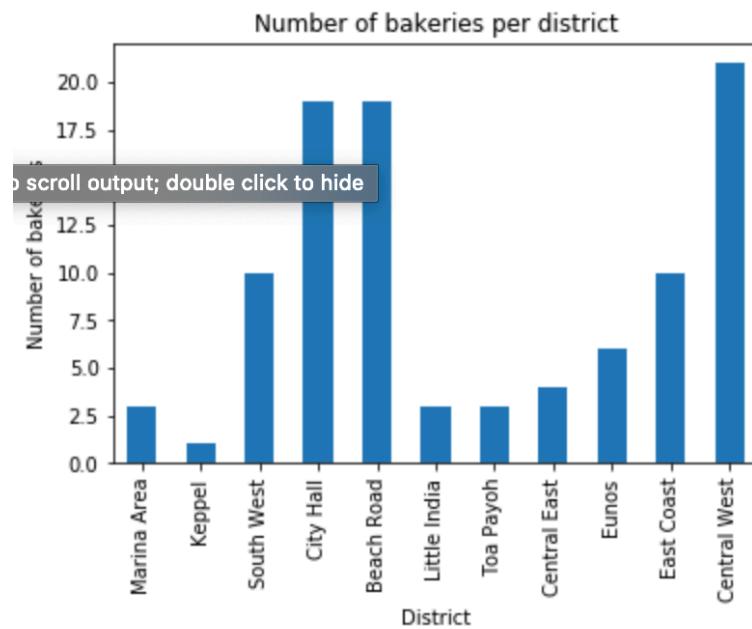
District	Areas	Location
1	Marina Area	Boat Quay, Chinatown, Havelock Road, Marina Square, Raffles Place, Suntec City

This always resulted in lat/lon data using GeoPy Geocoder. The downside is however that it may not result in the actual center location of the district and hence it may impact the bakeries found in that district. It also could result in overlap with some other district locations given the venue search radius of 200 meters.

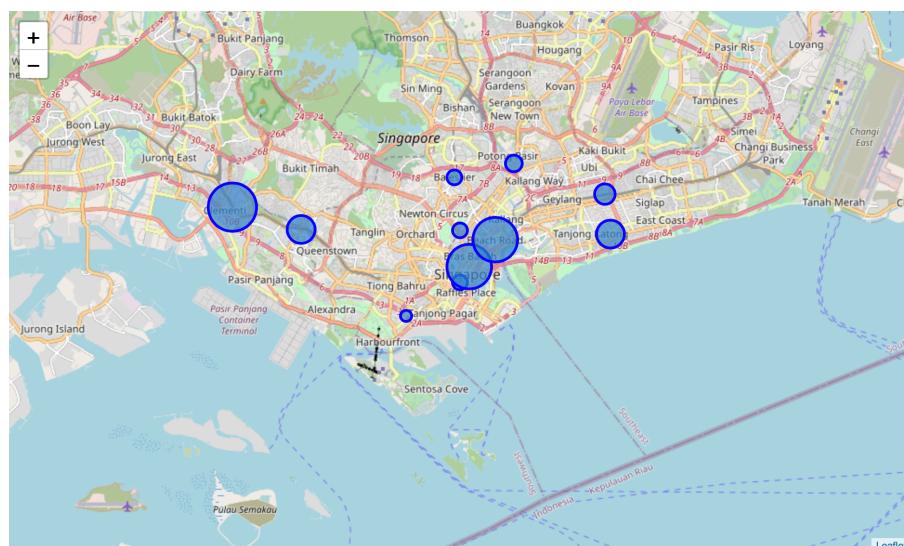
Results

Focusing only on districts in Central Singapore shows that there are 16 districts in total with a total of 99 bakeries in those districts.

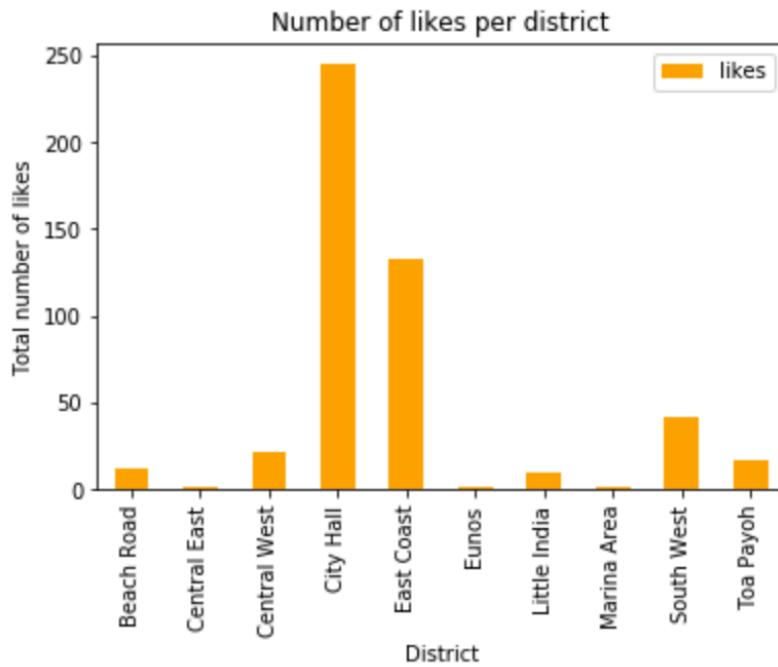
Based on geo-location it is clear that bakeries tend to concentrate in a few districts. The following bar chart shows that City Hall, Beach Road and Central West have the most bakeries.



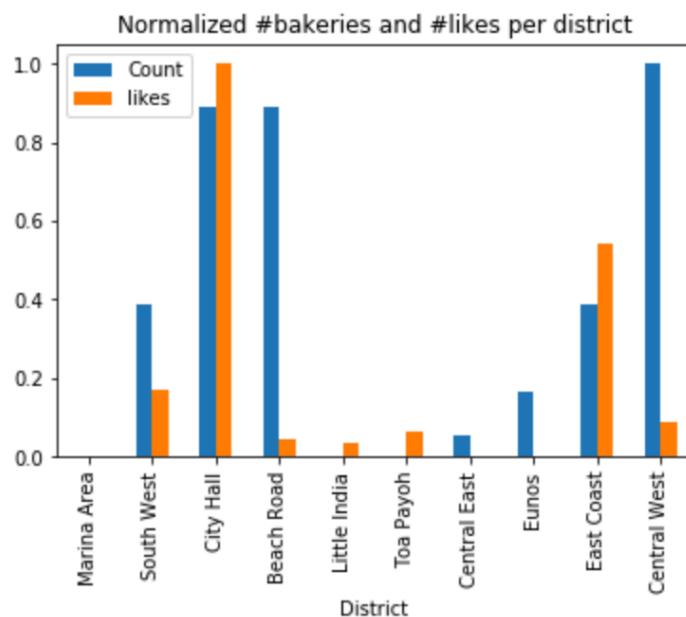
The following map of Singapore shows these districts with the size of the marker representing the number of bakeries in that district.



It is expected that popular districts will also have the most popular bakeries. Looking at the number of 'likes' for all bakeries in each district reveals that bakeries with the most likes can be found in City Hall. However, upon further examination of the individual numbers, it is clear that one specific bakery, Tiong Bahru Bakery, is skewing this data considerably. This bakery has 193 likes, while the mean of likes is just 10.08 per restaurant.



A simple min/max normalisation has been performed on the data to determine which districts have both a large number of bakeries and a lot of likes. The following bar chart depicts the results. From this chart, it is clear that City Hall scores well on both, while Central West has a large count of bakeries but with a rather low number of likes.



Recommendations

A few recommendations to further improve this analysis:

- Optimise the lat/lon calculation of each district. For example, use the average lat/lon of each of the locations in the district instead of just the first one.
- Vary the radius parameters in the venue search to determine if it impacts the overall results.
- Check for any duplicated bakeries due to overlapping search results between districts
- The total number of likes data may not be a good representation as the data is skewed due to some outliers. The two most popular bakeries account for 60% of all likes.
- Include premium FourSquare end-points in the analysis, such as venue stats (unique visitors, busy times, ...)
- Identify other useful data points such as income levels from data sources such as the Singapore Data Portal.

Conclusion

Using free data sources and some basic analysis, it can be concluded that opening a new restaurant in the neighbourhood of City Hall might have a good chance of being successful. With 19 bakeries and the highest score of 245 likes, this area is a very popular area for having breakfast.

At the other hand, the Central West district also has many bakeries, but none of them very popular. It may be interesting to check out if opening a success formula in this part of Singapore may even be more beneficial than opening a bakery near City Hall that needs to compete with a very popular existing bakery venue.

As this study only considered a few data points (number of bakeries per district and number of likes per district), further analysis should be done to include other data sources such as type of bakery, unique visitors per time period, venue tips, average cost of a meal, and more.