

Venue Data Analysis and Population Density of Yangon

*Sai Sinn Zom Leng
1st December, 2020*

1. Introduction

1.1 Background

Yangon, the commercial capital of Myanmar and a member of '[Asian Network of Major Cities 21](#)', is a densely populated city with an average urban density of 8,600/km² living over an area of 598.75 km². The city is divided into 33 townships or neighborhoods. Being the commercial center, a number of different kinds of business are squeezed in each neighborhood.

1.2 Problem

As most crowded neighborhoods are already dense with shops, stakeholders may have a problem of finding a suitable location to start their business. Depending on their type of business, the owners may want to set up where their type of business is less intense or where related types of business exist. They will also want an appropriate location for specific customers. This may be a complex problem to solve without fine data.

1.3 Target Audience

This report is mainly targeted to stakeholders who want to open a cafe (including coffee shop, bubble tea shop and similar ones) or a bakery (including other dessert shops, donut shops).

2.Data Description

2.1 Data Sources

- Population density of each neighborhood from [Wikipedia](#)
- Population [growth rate](#)
- **Json** file for townships (neighborhood) boundaries
- [Foursquare API](#) for venue data

2.2 Data Wrangling

(A) Population Density

The exact population data of **Yangon** is only available for the **2014 Census**. So, I made a new dataframe for the population of each township in Yangon according to the 2014 Census. Then, I used the **annual population growth rate** to estimate the current population. Lastly, the **population density** of each township was obtained.

(Something quite problematic is that the names of townships are represented differently on different websites as English can't be used to pronounce the exact Myanmar names for specific words.

That's why a custom data frame had to be generated in order to match the names on the geojson file we have got.)

This is how the data frame looks like :

	Township	Latitude	Longitude	Population_2014_Census	Area_sqkm	Population_2020_Estimated	Population_density/sqkm
0	BOTAHTAUNG	16.767500	96.151389	40995	2.40	45062	18775
1	DAGON SEIKKAN	16.856667	96.282778	167448	85.40	184060	2155
2	DAWBON	16.666667	96.183333	75325	3.70	82797	22377
3	EAST DAGON	16.883333	96.283333	165628	91.03	182059	1999
4	MINGALARTAUNGNYUNT	16.783333	96.166667	132494	5.06	145638	28782

(B) Nearby Venues

By using the above data frame and **Foursquare API**, I obtained the nearby venues of each neighborhood. But, some venues were missing and so, I filtered off those which returned less than 30 venues within 2.5 km radius leaving me 20 neighborhoods to analyze.

This is how the filtered data frame looks like :

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BOTAHTAUNG	16.7675	96.151389	Oishii Sushi	16.775120	96.150391	Sushi Restaurant
1	BOTAHTAUNG	16.7675	96.151389	Burma Bistro	16.772176	96.156015	Restaurant
2	BOTAHTAUNG	16.7675	96.151389	DONG JING (DVD,VCD,CD Centre)	16.776719	96.150417	Video Store
3	BOTAHTAUNG	16.7675	96.151389	Rangoon Tea House	16.772273	96.161735	Tea Room
4	BOTAHTAUNG	16.7675	96.151389	Little Yangon Hostel	16.772335	96.163245	Hostel

2.3 Data Selection

- **Population density**

The **demand** of a service or a business greatly depends on the population (the number of customers).

- **Venues**

After generating the nearby venues, it was found that more than **80%** of the venues generated by **Foursquare** for the target location were **hotels, restaurants, cafe(s) & bakeries**, and so are the most common venues of each neighborhood.

Hence, these will be our area of study.

- **Venues in the surrounding**

A restaurant located near a hotel or a landmark or in the city center may have a huge advantage. However, this factor depends greatly on the target customer.

3.Methodology

In accordance with the data obtained, I decided to focus on **cafe(s)** and **bakeries**.

In our country, **Myanmar**, a cafe is usually a place where **residents** often go to for breakfast, for a light snack between lunch and dinner or for supper.

Population density is a great factor as people usually go to those closer to their residence in the morning or evening.

Being close to the city center (where offices and shops) are abundant or locating near Universities or Schools is also a huge advantage for Cafe(s).

For bakeries (including dessert shops and similar ones), the factors are quite similar. As **rice** is the **staple food** of Myanmar people, bakeries are usually like a snack place.

For both of them, locating near a hotel or a shopping center is also beneficial.

So,first of all, the venue data frame is filtered off to contain only the selected venues.

3.1 One Hot Encoding

One hot encoding is used to obtain the abundance of each venue in each neighborhood among the selected venues.

From this, the **most common venues** of each neighborhood are obtained.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	AHLON	Hotel	Café	Bakery	Shopping Plaza	Boarding House	Bubble Tea Shop	Ice Cream Shop
1	BAHAN	Hotel	Café	Shopping Mall	Ice Cream Shop	Tea Room	Bubble Tea Shop	Hostel
2	BOTAHTAUNG	Hotel	Café	Hostel	Bakery	Shopping Mall	Dessert Shop	Bed & Breakfast
3	DAGON	Hotel	Café	Hostel	Bakery	Supermarket	Bed & Breakfast	Boarding House
4	DALA	Hotel	Hostel	Café	Donut Shop	Shopping Mall	Market	Supermarket

3.2 K-mean Clustering

Now, we will use unsupervised learning **K-means algorithm** to create clusters of neighborhoods.

I decided to create **3 clusters** as it is the **optimum value**.

Then, a new merged table containing the **cluster labels** is generated.

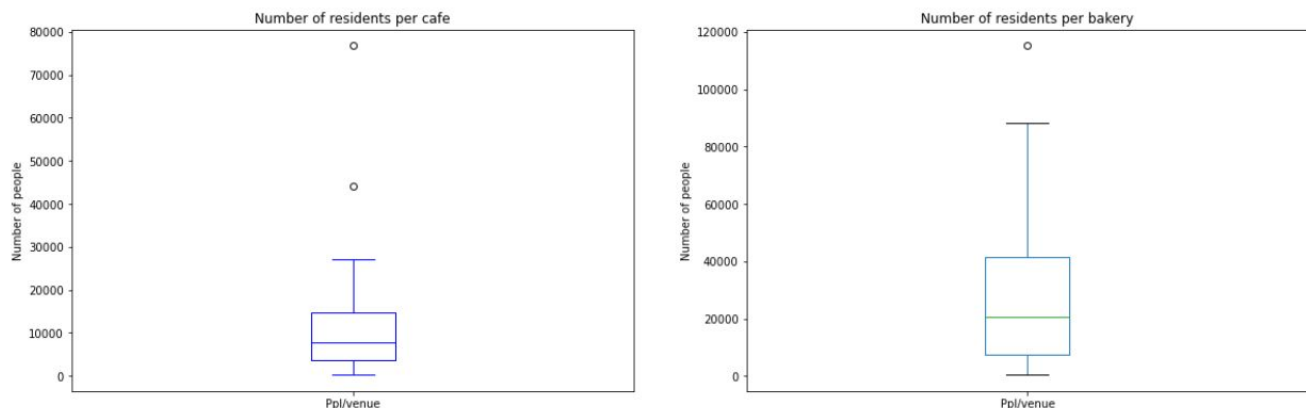
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	BOTAHTAUNG	16.767500	96.151389	Rangoon Tea House	16.772273	96.161735	Tea Room	0	Hotel	Café	Hostel	Bakery	Shopping Mall
1	BOTAHTAUNG	16.767500	96.151389	Backpacker (Bed & Breakfast)	16.772497	96.155875	Bed & Breakfast	0	Hotel	Café	Hostel	Bakery	Shopping Mall
2	BOTAHTAUNG	16.767500	96.151389	Little Yangon Hostel	16.772335	96.163245	Hostel	0	Hotel	Café	Hostel	Bakery	Shopping Mall
3	BOTAHTAUNG	16.767500	96.151389	The Strand	16.769415	96.162671	Hotel	0	Hotel	Café	Hostel	Bakery	Shopping Mall
4	BOTAHTAUNG	16.767500	96.151389	THE STRAND CAFÉ	16.769311	96.162481	Café	0	Hotel	Café	Hostel	Bakery	Shopping Mall

The clusters give a general idea of the neighborhoods . But, they don't point out specific neighborhoods. So, let's combine venue data and population data.

3.3 Residents : Venue ratio

Box plots are generated to obtain which neighborhoods are really in demand.

Let's study more about the neighborhoods which are above the **Interquartile Range** of each boxplots.

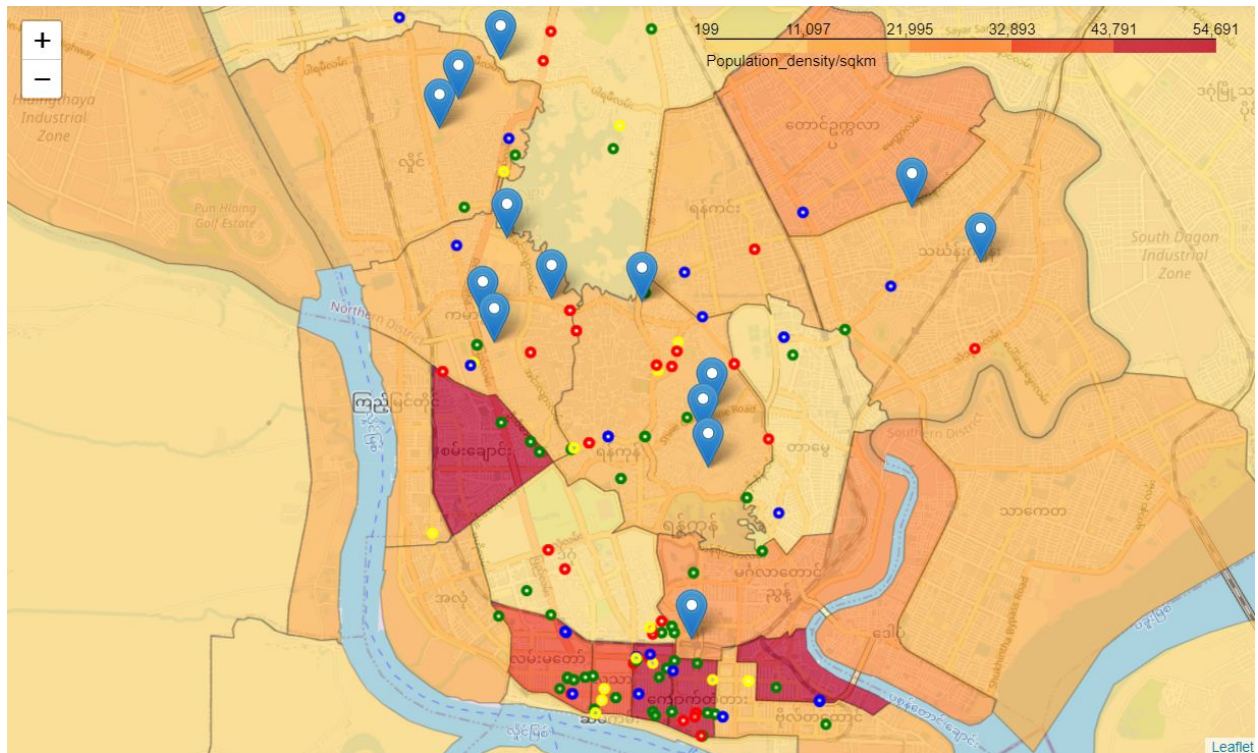


9 neighborhoods are selected for final study.

4. Results

4.1 Visualization

Now, we visualize our selected neighborhoods with venues and some landmarks alongside population density.



We can obviously see some areas with fewer cafe(s) or bakeries(s) but dense with hotels and shopping centers. Stakeholders can select their favourite location by checking the colors.



5. Discussion

Depending on the stakeholders' preference, they can select a location with

- High population density like the city center
 - City centers are often good places for most businesses.
- Average population density but very high resident to venue ratio (high demand)
 - Has a high potential for some specific businesses.
- Other restaurants or similar ones exist
 - Like night markets or street food zones.

6. Conclusion

Myanmar is a developing country and the commercial capital, **Yangon** has a steady growth rate of both population and business. So, it's surely a good place for any startup business.

This report just shows the areas with the highest demand.

However, the location data used has a moderate amount of missing venues and so, the result can't be too accurate. But, it does give a general idea of the neighborhoods. With a better location data, the analysis can be improved a lot.

Moreover, [new city zones](#) are also being developed and these should also be included for analysis.

A lot of Public data is still needed to be published.

Real estate availability and price are also important.

With all these factors, the results will be a lot more precise and this kind of analysis can be used for other types of business which are related to each other.