

Best Location to Open Coffee Shop in Singapore

Applied Data Science Capstone by Ben Sung

July 30, 2019

Table of contents

- Introduction
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

1 Introduction

1.1 Background

According to the www.singstat.gov.sg, the food and beverage services index in Singapore in May 2019 has a total sales value of \$849 million with year-on-year increase of 2%. Indeed, it is a lucrative business to open a restaurant or coffee shop in Singapore.

However, from 2010 to 2015, there were about 3,000 restaurants and coffee shops were opened, but 2,000 closed in each year. Delicious food alone is not enough, there are plenty of other important factors to consider and one of the most important factors is choosing the right location.

1.2 Business Problem

Choosing the best location to open a coffee shop in a big city is not an easy feat. Especially in Singapore, where plenty of coffee shops are available in almost every corner of the streets. Competitions are everywhere. Therefore, it is critical to look for location that are not many coffee shops around.

We will also be looking at the population of each neighborhood, as some of the neighborhood have much less people to generate regular customers. So, in this project, we can narrow down to top 10 neighborhoods for our location.

Lastly, we will be looking at the numbers of train stations and bus stops in the vicinity, as these will help to generate human traffics for the shop. We will find the best locations base on these criteria using data science methodology.

2 Data

2.1 Data Sources

From the problem statements above, these are the data that will be gathered in each neighborhood:

- Top 10 most populous neighborhood
- Number of coffee shops

For the population of each neighborhood, I will be using the table from www.citypopulation.de/Singapore-Regions.html.

Table will be scraped from this website using Python (BeautifulSoup) and sorted based on the most populous and narrow down to top 10 neighborhoods.

SINGAPORE					
Republic of Singapore					
Planning Areas					
The resident population of all Singaporean urban planning areas with more than 2,500 inhabitants according to census results and latest official estimates.					
Name	Adm.	Population Census (C) 2000-06-30	Population Census (C) 2010-06-30	Population Estimate (E) 2015-06-30	Population Estimate (E) 2018-06-30
Ang Mo Kio	NE	180,112	179,297	174,770	165,710
Bedok	E	284,318	294,519	289,750	281,300
Bishan	C	89,746	91,298	90,700	88,490
Bukit Batok	W	126,200	144,198	139,270	144,410
Bukit Merah	C	148,299	157,122	155,840	151,870
Bukit Panjang	W	96,031	128,734	139,030	140,820
Bukit Timah	C	64,361	70,314	74,470	77,280
Changi	E	1,085	2,155	2,530	2,080
Choa Chu Kang	W	136,105	173,291	174,330	187,510
Clementi	W	90,864	91,874	91,630	93,000

From code using the Jupyter Notebook, we got the 10 most populous neighborhood as follows:

```
In [162]: # Reset the index
df = df.reset_index()
df.drop(['index'], axis=1, inplace=True)
df
```

Out[162]:

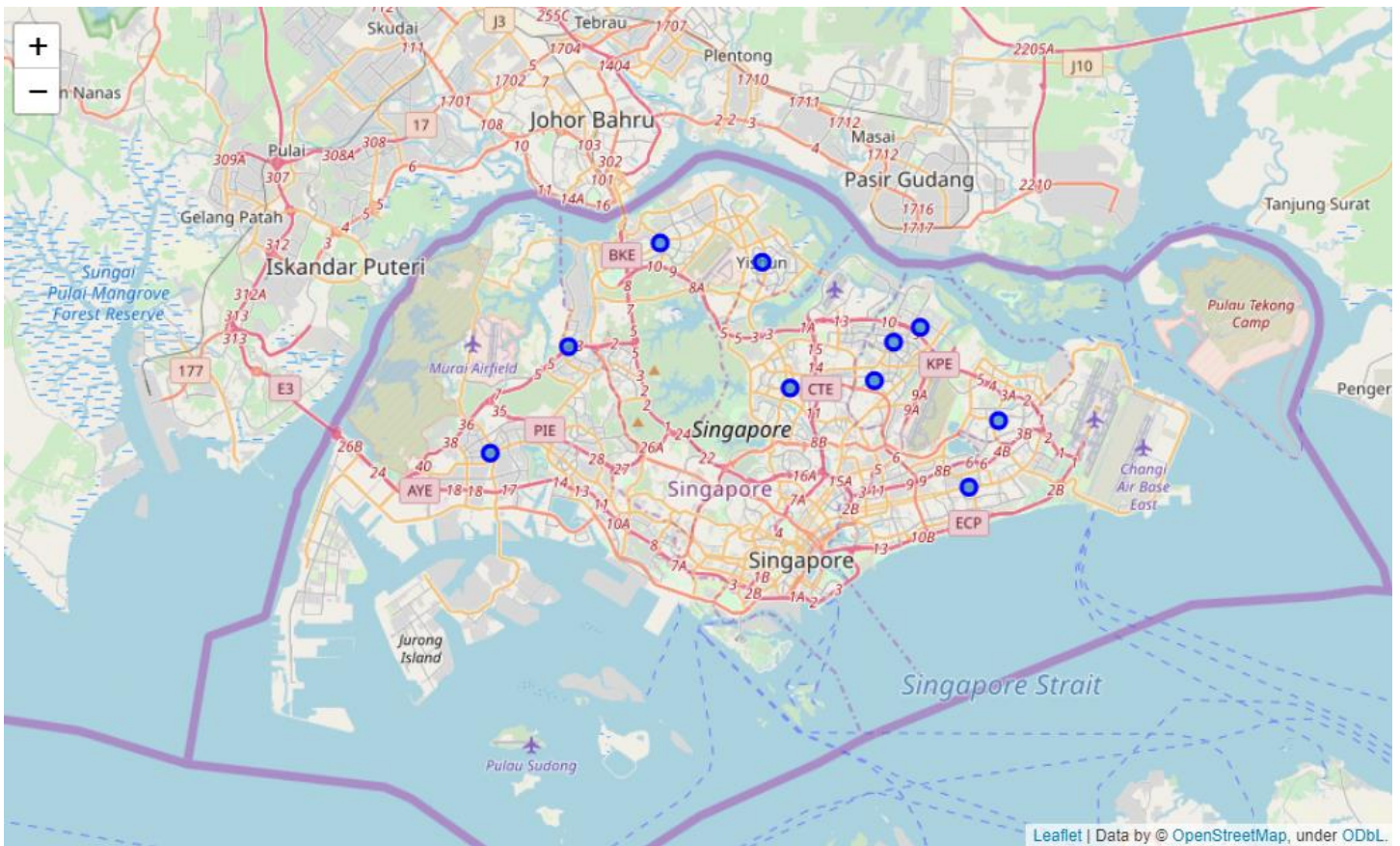
	Neighborhood	Population
0	Bedok	281,300
1	Jurong West	266,720
2	Tampines	257,110
3	Woodlands	252,530
4	Sengkang	240,640
5	Hougang	223,010
6	Yishun	214,940
7	Choa Chu Kang	187,510
8	Ang Mo Kio	165,710
9	Punggol	161,570

Next, we will make use of the Foursquare to find the number of coffee shops, train stations and bus stations in the neighborhood.

Using Google service with using the Jupyter Notebook, we got the geo location 10 most populous neighborhood as follows:

	Neighborhood	Population	Latitude	Longitude
0	Bedok	281,300	1.323976	103.930216
1	Jurong West	266,720	1.339636	103.707339
2	Tampines	257,110	1.354653	103.943571
3	Woodlands	252,530	1.436897	103.786216
4	Sengkang	240,640	1.390949	103.895175
5	Hougang	223,010	1.373360	103.886091
6	Yishun	214,940	1.428136	103.833694
7	Choa Chu Kang	187,510	1.389260	103.743728
8	Ang Mo Kio	165,710	1.369842	103.846609
9	Punggol	161,570	1.398033	103.907331

Using the folium, we can place the circle marker on the 10 neighborhood locations in Singapore, as follows:



From the Foursquare, we sample the first location, Bedok, to find the top 100 within radius of 500 meters:

Out[259]:

	name	categories	lat	lng
0	Bedok Chwee Kueh 勿洛水粿	Chinese Restaurant	1.324903	103.930250
1	Ya Kun Kaya Toast 亞坤	Coffee Shop	1.324095	103.929198
2	Duke Bakery	Bakery	1.324691	103.932514
3	FairPrice Finest	Supermarket	1.324140	103.929260
4	Song Zhou Luo Bo Gao 松洲萝卜糕	Breakfast Spot	1.324836	103.930520

```
In [260]: print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

60 venues were returned by Foursquare.

We can see there are 60 venues provided by Foursquare in Bedok areas. We will explore more on the other neighborhoods in the next sections.

3 Methodology

3.1 Exploring the neighborhood in Singapore

First, we use Foursquare to explore the 10 neighborhoods of Singapore,

```
SingaporeVenues = getNearbyVenues(names=df['Neighborhood'],  
                                  latitudes=df['Latitude'],  
                                  longitudes=df['Longitude']  
                                  )
```

```
Bedok  
Jurong West  
Tampines  
Woodlands  
Sengkang  
Hougang  
Yishun  
Choa Chu Kang  
Ang Mo Kio  
Punggol
```

We can observe the number of venues found by Foursquares with 106 unique categories:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Ang Mo Kio	49	49	49	49	49	49
Bedok	60	60	60	60	60	60
Choa Chu Kang	13	13	13	13	13	13
Hougang	29	29	29	29	29	29
Jurong West	59	59	59	59	59	59
Punggol	5	5	5	5	5	5
Sengkang	28	28	28	28	28	28
Tampines	78	78	78	78	78	78
Woodlands	47	47	47	47	47	47
Yishun	41	41	41	41	41	41

Let's find out how many unique categories can be curated from all the returned venues

```
print('There are {} uniques categories.'.format(len(SingaporeVenues['Venue Category'].unique())))
```

There are 106 uniques categories.

4 Analyze Each Neighborhood

4.1 One Hot Encoding

	Neighborhood	American Restaurant	Arcade	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Basketball Court	Beijing Restaurant
0	Bedok	0	0	0	0	0	0	0	0	0
1	Bedok	0	0	0	0	0	0	0	0	0
2	Bedok	0	0	0	0	0	1	0	0	0
3	Bedok	0	0	0	0	0	0	0	0	0
4	Bedok	0	0	0	0	0	0	0	0	0

And let's examine the new dataframe size.

```
In [268]: Singapore_onehot.shape
```

```
Out[268]: (409, 107)
```

We then group rows by neighborhood and by taking the mean of the frequency of the occurrences of each category:

```
[270]: Singapore_grouped = Singapore_onehot.groupby('Neighborhood').mean().reset_index()  
Singapore_grouped
```

```
Out[270]:
```

	Neighborhood	American Restaurant	Arcade	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Basketball Court	Beijing Restaurant	Bookstore	Breakfast Spot	Bubble Tea Shop
0	Ang Mo Kio	0.000000	0.000000	0.020408	0.000000	0.000000	0.020408	0.020408	0.000000	0.000000	0.000000	0.000000	0.061224
1	Bedok	0.016667	0.016667	0.050000	0.000000	0.000000	0.016667	0.000000	0.000000	0.000000	0.016667	0.016667	0.016667
2	Choa Chu Kang	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Hougang	0.000000	0.000000	0.034483	0.034483	0.034483	0.000000	0.000000	0.000000	0.000000	0.000000	0.034483	0.000000
4	Jurong West	0.016949	0.000000	0.101695	0.000000	0.000000	0.000000	0.000000	0.000000	0.016949	0.016949	0.016949	0.016949
5	Punggol	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Sengkang	0.000000	0.000000	0.035714	0.000000	0.000000	0.071429	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Tampines	0.012821	0.000000	0.012821	0.000000	0.000000	0.051282	0.000000	0.000000	0.000000	0.012821	0.000000	0.038462
8	Woodlands	0.021277	0.000000	0.042553	0.000000	0.000000	0.000000	0.000000	0.021277	0.000000	0.021277	0.000000	0.000000
9	Yishun	0.000000	0.024390	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.024390	0.000000	0.024390

```
[271]: # Let's confirm the new size  
Singapore_grouped.shape
```

```
Out[271]: (10, 107)
```

Finally, we group the top 5 most common venues in each of the 10 neighborhoods:

	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ang Mo Kio	Coffee Shop	Food Court	Bubble Tea Shop	Sushi Restaurant	Japanese Restaurant	Sandwich Place	Supermarket	Fast Food Restaurant	Dessert Shop	Seafood Restaurant
1	Bedok	Coffee Shop	Food Court	Asian Restaurant	Japanese Restaurant	Sandwich Place	Noodle House	Café	Fast Food Restaurant	Chinese Restaurant	Supermarket
2	Choa Chu Kang	Fast Food Restaurant	Coffee Shop	Playground	Sandwich Place	Lingerie Store	Salon / Barbershop	Casino	Food Court	Thai Restaurant	Park
3	Hougang	Coffee Shop	Food Court	Noodle House	Soccer Field	Breakfast Spot	Park	Department Store	Food	Food & Drink Shop	Market
4	Jurong West	Japanese Restaurant	Fast Food Restaurant	Asian Restaurant	Chinese Restaurant	Dessert Shop	Café	Park	Coffee Shop	Food Court	Gym / Fitness Center
5	Punggol	Bus Station	High School	Chinese Restaurant	Bus Stop	Wings Joint	Frozen Yogurt Shop	Electronics Store	Fast Food Restaurant	Food	Food & Drink Shop
6	Sengkang	Fast Food Restaurant	Chinese Restaurant	Food Court	Bakery	Coffee Shop	Café	Shopping Mall	Video Store	Food Stand	Metro Station
7	Tampines	Coffee Shop	Bakery	Café	Japanese Restaurant	Food Court	Fast Food Restaurant	Bubble Tea Shop	Shopping Mall	Sushi Restaurant	Chinese Restaurant
8	Woodlands	Japanese Restaurant	Chinese Restaurant	Café	Coffee Shop	Asian Restaurant	Clothing Store	Indian Restaurant	Frozen Yogurt Shop	Shopping Mall	Fast Food Restaurant
9	Yishun	Food Court	Chinese Restaurant	Coffee Shop	Supermarket	Hainan Restaurant	Noodle House	Fast Food Restaurant	Bus Line	Fried Chicken Joint	Ramen Restaurant

5 Result and Discussion

First, we obtained the top 10 most populous neighborhood in Singapore to narrow down the best place to open our coffee shop, since the more likely we can get regular customers if they live nearby.

	Neighborhood	Population
0	Bedok	281,300
1	Jurong West	266,720
2	Tampines	257,110
3	Woodlands	252,530
4	Sengkang	240,640
5	Hougang	223,010
6	Yishun	214,940
7	Choa Chu Kang	187,510
8	Ang Mo Kio	165,710
9	Punggol	161,570

Next, we look at the top 5 most common venues in each of the 10 neighborhoods.

Not surprisingly, coffee shops are among the most common venues:

- Ang Mo Kio
- Bedok
- Hougang
- Tampines.

There are three other neighborhoods that place the coffee shops or cafe in the top 2 and top 3.

- Choa Chu Kang
- Yishun
- Woodlands

So, we can narrow down to 3 neighborhoods that do not have coffee shops in their top 3:

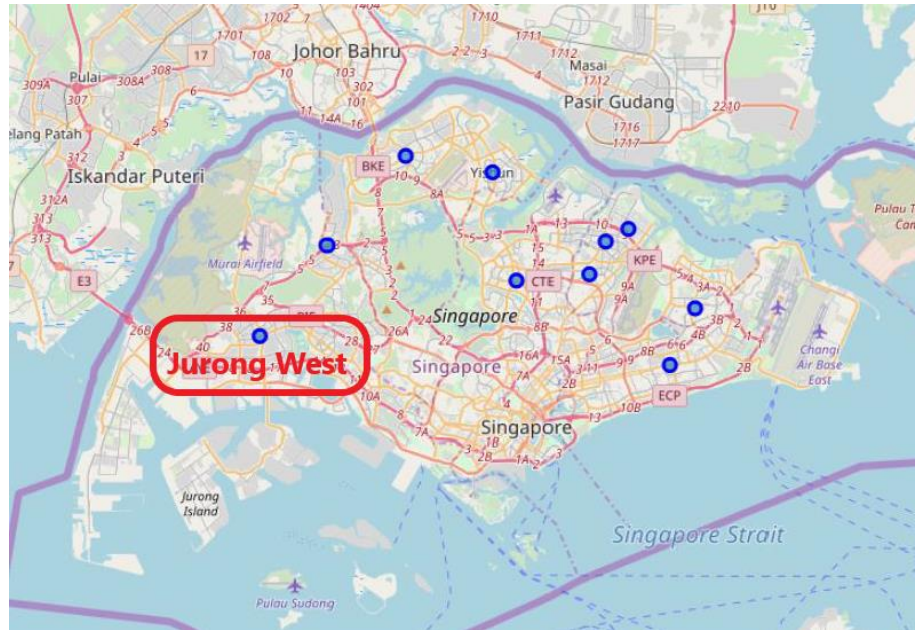
- Jurong West
- Punggol
- Sengkang

We can rank the best location to open our coffee shop in Singapore based on the most populous neighborhood to the least as follows:

1. Jurong West (population of 266,720)
2. Sengkang (population of 240,640)
3. Punggol (population of 161,570)

Jurong West as the best overall place to open a coffee shop.

6 Conclusion



Based on our analysis, **Jurong West** neighborhood is the best overall place to open a coffee shop.

Looking at the top 10 most common venue for Jurong West, coffee shop or cafe is placed as 6th. The first four most common venues are all restaurants (Japanese, Fast Food, Asian and Chinese).

There could be other factors that could be considered to have coffee shop as 6th place. Based on the location that further away from the city center could be a factor, since most working adults might be travelling to the city center and have their coffee there, instead of in their neighborhoods. And, looking at the top 3 (Jurong West, Sengkang and Punggol), indeed, these are some of the furthest away from the city center.

However, until we study further these other factors, having much less competitions in the populous neighborhood, Jurong West, might still be the best location to open the restaurant in Singapore!