Capstone project: Similar neighborhoods in Ho Chi Minh city

1.Introduction: Business Problem

Business problem:

Currently I live in Distric 9 (Wards/Neighborhoods: Phuoc Long B, district 9), Ho Chi Minh city, Viet Nam. I will move to another location in distric 9 or Ho Chi Minh city in the future depend on my new job, so I 'd like to know:

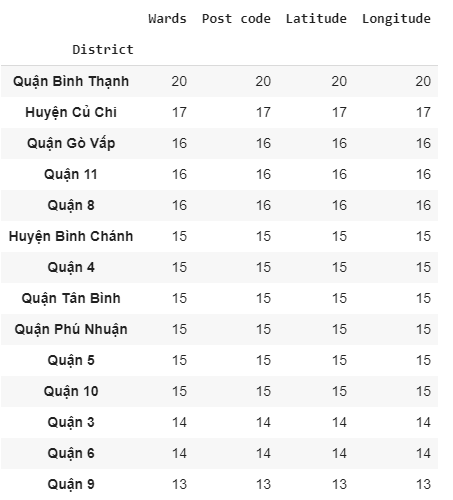
* Which neighborhoods in district 9 are similarities to where I live (about utilities and services)
* Which neighborhoods in Ho Chi Minh city are similarities to where I live (about utilities and services)

This report will be targeted to stakeholders interested in neighborhoods similarities for living/transfer location

2.A description of the data and how it will be used to solve the problem.

My excel data file: Postal code Ho Chi Minh city

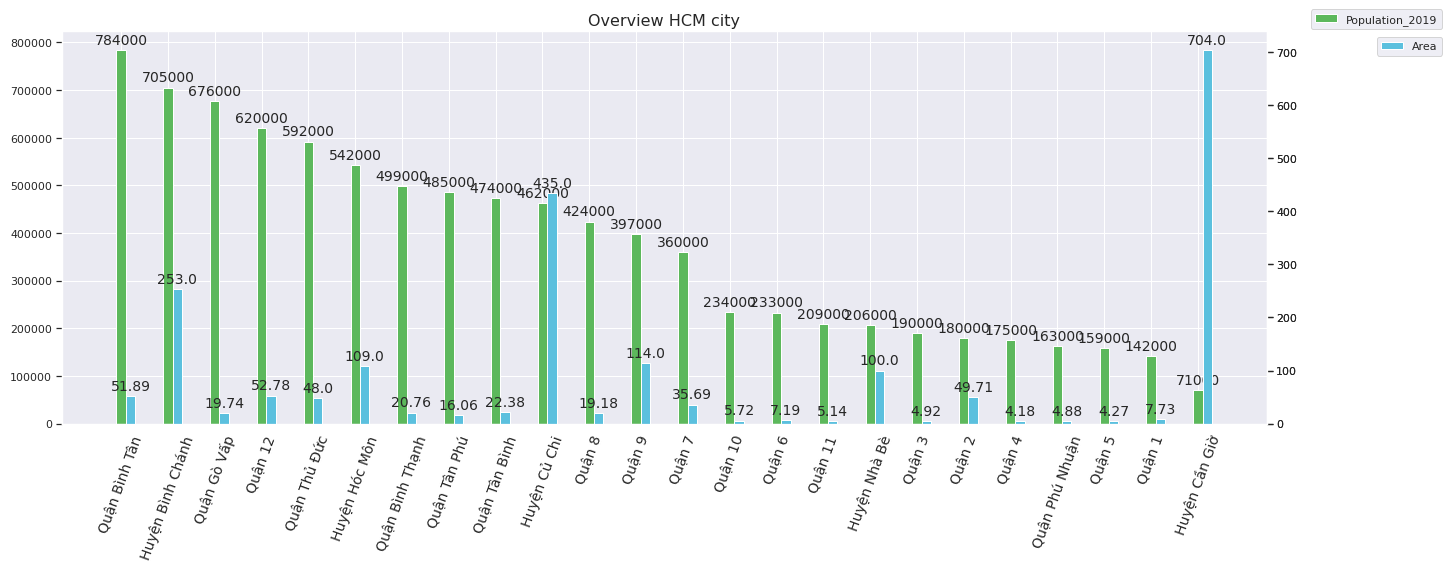
* This file contains all district and wards name (with post code) in Ho Chi Minh city (update 2019).
* Based on this data information, I will find location information (lat/lng) of each district/ward by googlemap API reverse geocoding service.
* Then, I will use Foursquare API to explore all venue (utilities and services) in each ward (neighborhoods) base on location information
* When I know all information each wards (neighborhoods), I will use clustering method to group similar wards together. It will help me know which neighborhoods are similarities to where I live (based on number of utilities and services)



Information about post code, lat/lng of all district and neighborhoods (wards) in HCM city (lat/lng get from google API

Based on this information, I can get all venue in each neighborhood from Foursquare API. This information will be feed to clustering method

Overview population and area of all district in HCM city



My current location: District 9 has average population (397,000) and area (114 sqm)

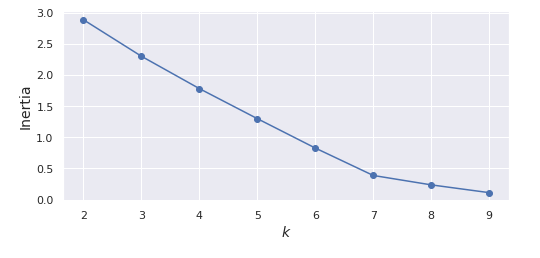
Compare to another district

3.Methodology

In this project, I will find all venue in each neighborhood based on foursquare API, this information will be used for clustering method. After that I will use K means to cluster similarity neighborhoods. I use two method to find optimize optimal value of cluster: elbow and silhouette\_score. Optimal cluster number is very important to find similarity neighborhoods, so we will look closely to these method

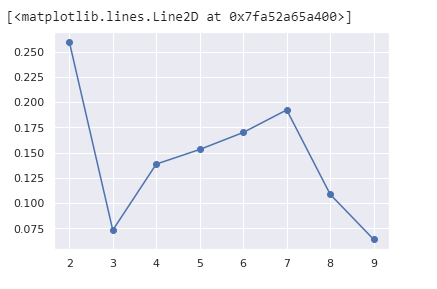
1. Elbow method: find elbow value in chart (hard, depend on human visual)

Result: number of clusters is 7



1. Silhouette\_score method: high score is better (easy to find optimal value but the speed is low)

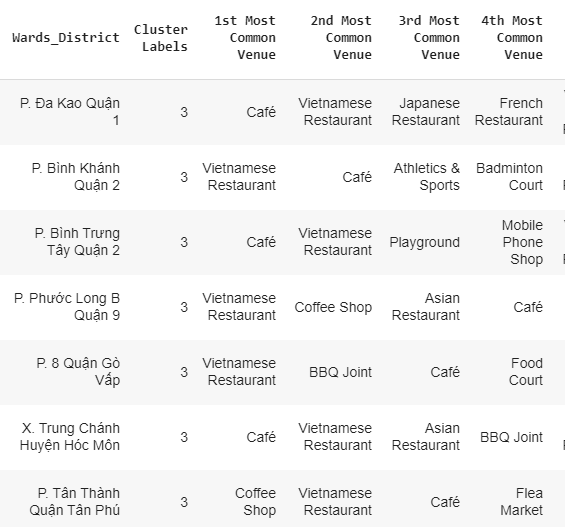
Result: number of clusters is 7



## 4.Results and Discussion

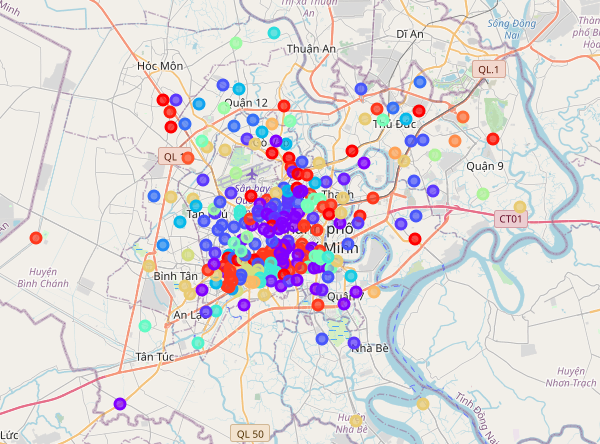
My current location: P. Phuoc Long B (Distric 9) is similar as many neighborhoods in district 2, 3, 7,8, 10, Phu Nhuan.... So I can decide go to these location for my job and it will not effect to my life so much

Moreover, based on this information, I know the reason why these neighborhoods are similar: most common venues are coffee shop and restaurant. If I want to open the business in these locations, I will consider coffee shop and restaurant first

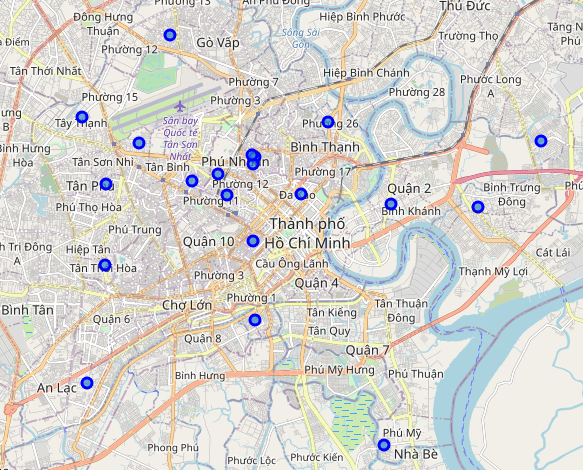


Cluster information: almost neighborhoods that similar to my current location (cluster 3) have common venue is Coffee shop and restaurant

All 28 cluster neighborhoods in HCM city:



All 21 neighborhoods that similar to my current neighborhoods: Phuoc Long B , district 9 (cluster 3) :



# **5.Conclusion**

Purpose of this project to identify neighborhoods similar to my current neighborhoods (P. Phuoc Long B, Distric 9). These findings in this result can help me a lot to find similar neighborhoods for moving in the future

Point for improvement:

The foursquare API currently not supply enough information about some important venue each neighborhood such as hospital, school, market...These things are very important to compare similarity of each neighborhoods. I think the reason is that users of foursquare are not much in VN