**The Battle of Neighborhoods (Week 1)**

**-- Su Yiping**

**Introduction:**

Singapore is recognised as the world’s safest country and it’s a playground for rich, only truly sovereign city-state does have a certain sheen of wealth in the world. Singapore is an island country with a high population density, over 5.6 million residents living in the land area of 725.7 km2, (i.e. approximately 7870 people/ km2,).  In fact, Singapore is one of the easiest and most comfortable countries to navigate in Southeast Asia. It is a global hub for economy, finance, education, tourism, entertainment and healthcare, among many other areas.

If a group of stakeholders have a plan that is acquisition and sell of real estate for white collar. So, we need to analysis of various venues for finding useful insights into the kind of business thriving in difference locality.

**Object:**

This project aims at profiling the neighborhoods to come up with the best location for acquiring some real estates. The project of analysis will be based on the number and category of venues of various types present in Singapore.

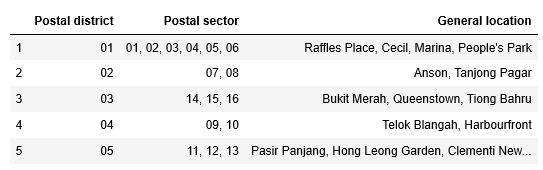
Now let me explain more detail the context of this project through a scenario. Assume my company want to invest a real estate for residential, Every household want to have the great amenities that a lot of useful feature or facility of place exist in the neighborhood such as supermarket, shopping mall, gourmet fast food joints, cinema, park, pharmacies, graduate schools and so on. If everything around the real estates, maybe we need higher price acquiring, but if only some function in the neighborhood, maybe we need invest surrounding, otherwise we can’t get higher profit.

**Tools:**

1. Using Panda package to create data frame and analyze the data
2. Using the BeautifulSoup package to scrape the following Wikipedia page,   [<https://en.wikipedia.org/wiki/Postal_codes_in_Singapore>,](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) In order to obtain the data that is in the table of postal district (Postal sector: 1st 2 digits of 6-digit postal codes) and to transform the data
3. Using Geocoder package to get the latitude and the longitude coordinates of each neighborhood and create the new data frame.
4. Using folium package to show the map
5. Using k-means package from hierarchical clustering algorithm sklearn.cluster to cluster the neighborhood
6. Using Numpy package for scientific computing.

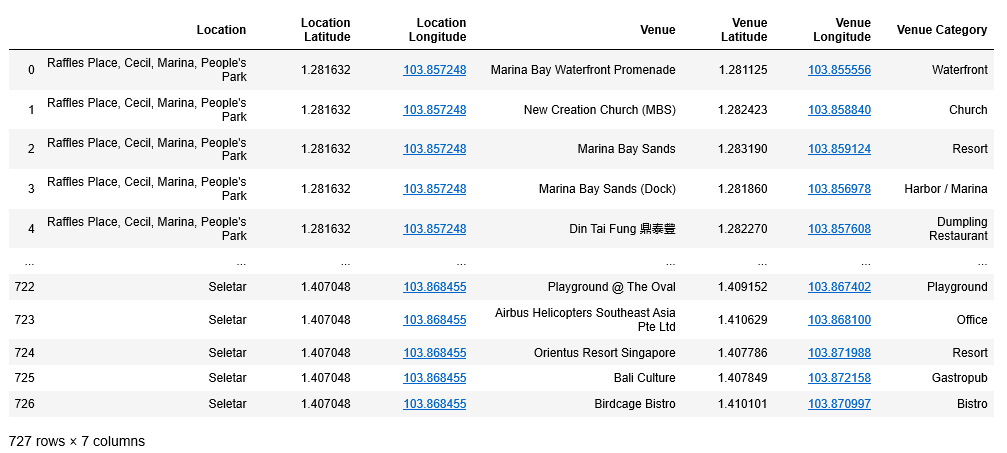
**Data Frame:**

1. Scrape from website:



2. Add latitude and longitude of postal district to data frame

3. Top venue around the general location



4. top 10 places of each general location



**Methodology:**

Collect the data

Explore the data

Visualize the map

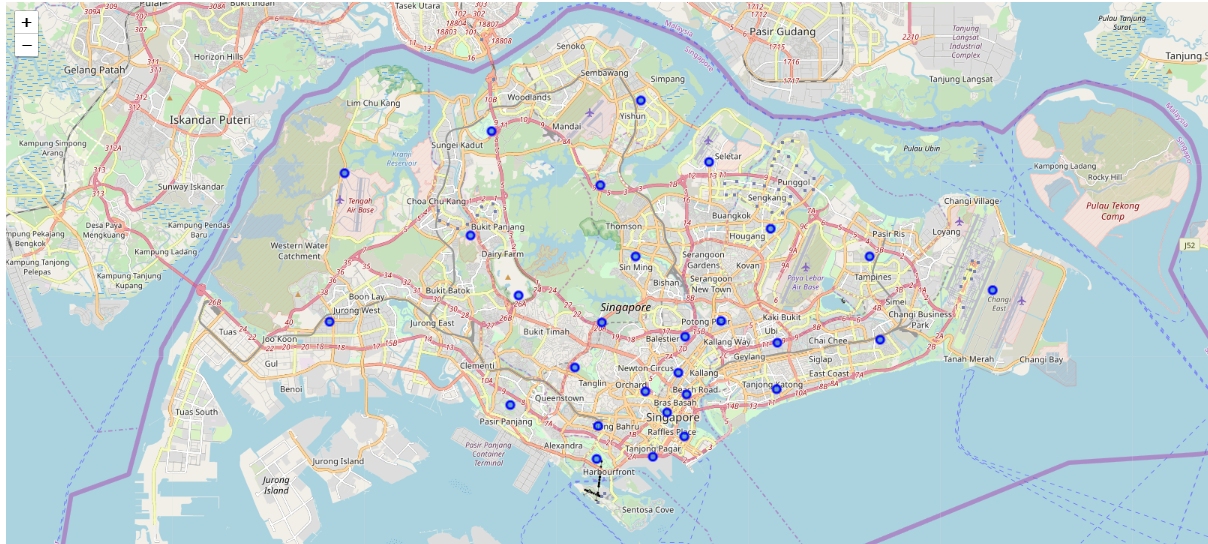
Clusters in Singapore

Data pre-processing

Analysis the data

**Visualize the map:**

1. The geographical visualization of the different location for Singapore



1. The geographical visualization of the different types of clusters created using K-Means for Singapore

