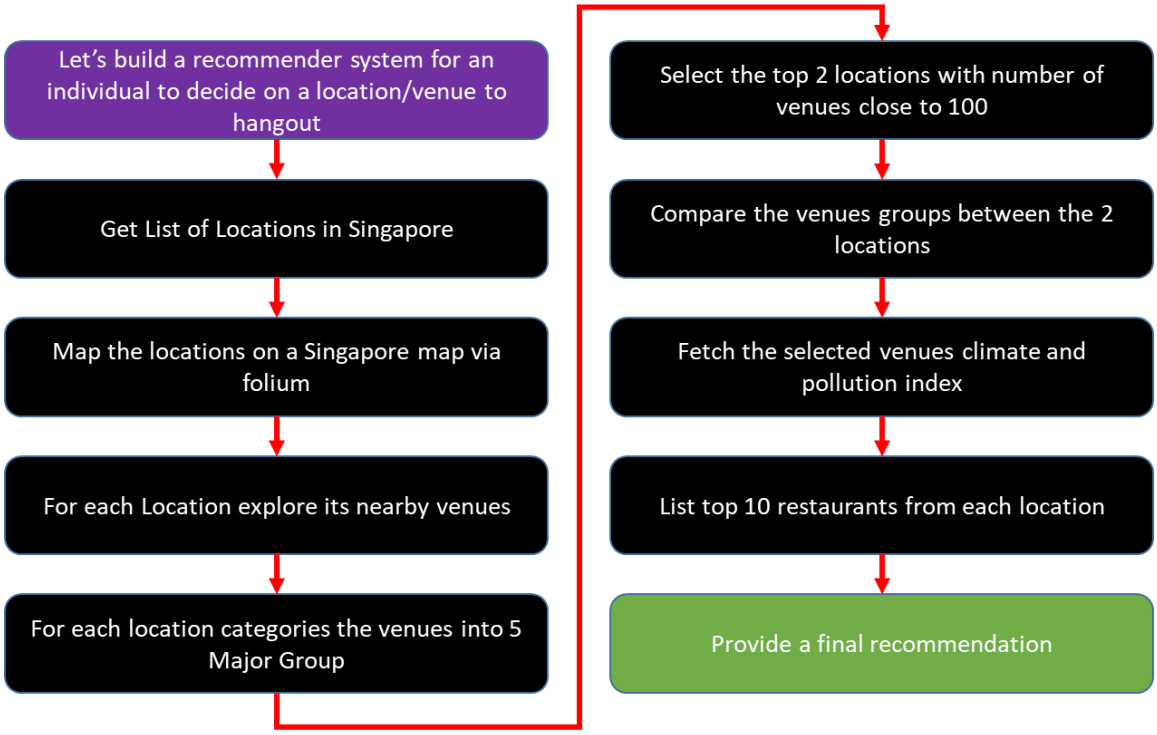
The Battle of the Neighbourhoods

- Data & Approach

High-Level Flow Chart:

The goal is to build a recommendation system for a traveller/resident (who will be referred as customer) decide on which neighbourhood to visit among the available popular locations in Singapore.



Data Sources to Consider:

1. OneMap Singapore API
2. FourSquare Singapore API
3. Data Gov Sg API
4. Euclidian Distance Computations

Data Representation:

1. Tables
2. Folium Maps
3. Data Frame Bar Plots

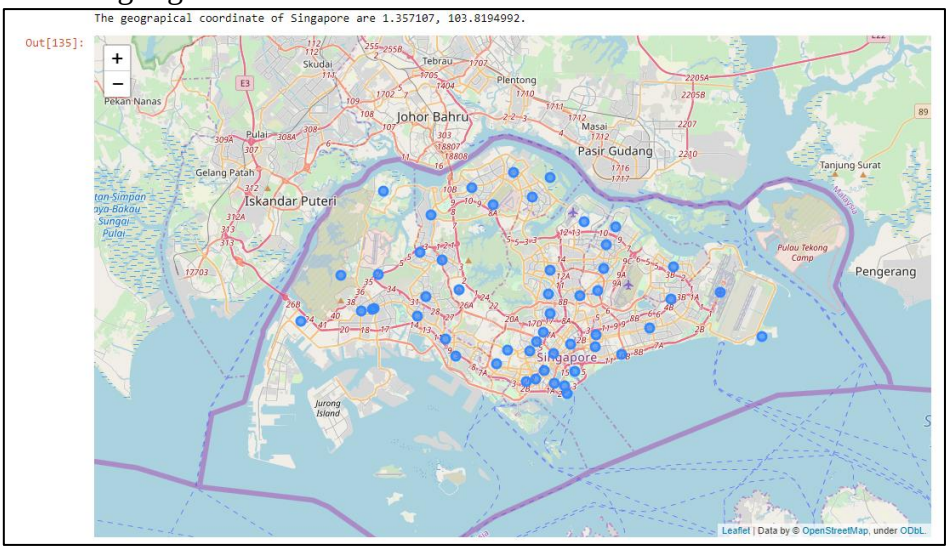
Fetching the Required Data:

1. Getting list of locations in Singapore
    a. To get this data, we use API from OneMapSG that provides list of PlanningAreaNames. The returned data is in Json format which is further processed and cleaned up to extract the Location names.
    b. A total of 53 valid locations with in Singapore are identified
    c. https://docs.onemap.sg/, refer this link for the details on the API supported by One Map SG.

d. In order to identify its LATLON co-ordinates, the python geolocator module is used.
e. Locations with No LATLON information is excluded and the final output is saved in a CSV

2. Map the locations

a. Using the LATLON information and folium python module the map of SG is drawn with the locations highlighted.



3. Explore the venues near the location

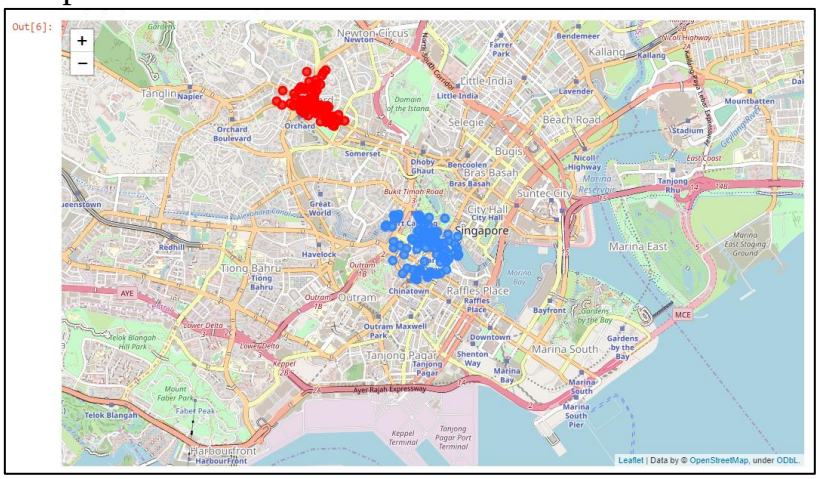
a. The list of venues near each of the 53 location is extracted using the Foursquare API – venues/explore.
b. https://api.foursquare.com/v2/venues/explore
c. The Json output is further processed and dumped into CSV
d. Until now, for the customer we have listed down all the available locations in Singapore and places around these locations to explore

4. Grouping Venues

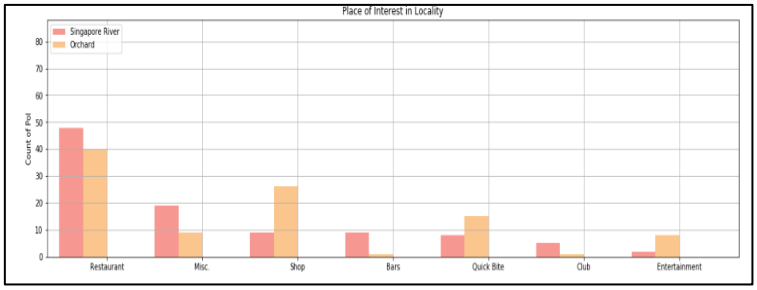
a. The venues returned has a detailed category and hence to enable comparison between any two locations it is critical we group the venues under common denominator.
b. Thus the following groups were created – Restaurant, Bar, Club, Snack, Shop, and Entertainment.
c. Our customer is a traveller/resident who is planning to spend some time at a specific locations. Hence the assumption is he is on a leisure trip and hence the above groups were decided.
d. All other venues that doesn't fall under the above groups are categorized as Misc. e.g. Gym, Education Institution etc.

```
>>> Action 2 : QueryFourSquare
>>> Reading File Into Data Frame : C:\Users\User\Documents\GitHub\BattleOfNeighCapStone\QueryLocations.csv
    $ Exploring Location SOUTHERN ISLANDS, Containing Total of 6 places nearby
    $ Exploring Location SUNGEI KADUT, Containing Total of 4 places nearby
    $ Exploring Location NEWTON, Containing Total of 22 places nearby
    $ Exploring Location ORCHARD, Containing Total of 100 places nearby
    $ Exploring Location KALLANG, Containing Total of 6 places nearby
    $ Exploring Location PASIR RIS, Containing Total of 36 places nearby
    $ Exploring Location STRAITS VIEW, Containing Total of 93 places nearby
    $ Exploring Location MARINA EAST, Containing Total of 4 places nearby
    $ Exploring Location MARINA SOUTH, Containing Total of 5 places nearby
    $ Exploring Location SERANGOON, Containing Total of 41 places nearby
    $ Exploring Location BOON LAY, Containing Total of 71 places nearby
    $ Exploring Location BEDOK, Containing Total of 61 places nearby
    $ Exploring Location BUKIT MERAH, Containing Total of 37 places nearby
    $ Exploring Location BUKIT PANJANG, Containing Total of 36 places nearby
    $ Exploring Location JURONG EAST, Containing Total of 76 places nearby
    $ Exploring Location BUKIT TIMAH, Containing Total of 13 places nearby
    $ Exploring Location CHANGI, Containing Total of 58 places nearby
    $ Exploring Location CHOA CHU KANG, Containing Total of 22 places nearby
    $ Exploring Location QUEENSTOWN, Containing Total of 15 places nearby
    $ Exploring Location SELETAR, Containing Total of 4 places nearby
    $ Exploring Location ANG MO KIO, Containing Total of 40 places nearby
    $ Exploring Location BISHAN, Containing Total of 42 places nearby
    $ Exploring Location BUKIT BATOK, Containing Total of 22 places nearby
```

5. Selecting the top 2 places & comparing them to make recommendation
    a. Based on the number of venues the top two places were selected namely “Singapore River” and “Orchard”.
    b. The list of venues around these two places are clustered and are shown using folium Map



    c. Next a Grouped Bar Chart is used to compare the count of groups in these two locations.
    d. This bar chart will gives an idea on which location of two had more options to spend our leisure time.



6. Recommendation
    a. With the above bar chart we can recommend that Orchard has a variety of venues to explore namely “Restaurants”, “Shops”, and “Entertainments”. This place is suitable to spend an entire day time.
    b. But if our customer is a Night owl, “Singapore River” is a bustling area with “Restaurants”, “Bars” and “Clubs”.
    c. Thus we have a basic interest based recommendation engine built.

7. One Step further
    a. Now our customer has made his choice we believe the following information will be of help.
    b. Weather & Pollution Index: Singapore being a land of surprise weather, its better our customer understand what he can expect when is out at any of this venues.
    c. Weather & PI data is queried on a real-time from Singapore Government Data Bank - https://data.gov.sg/
    d. Since it's tricky to find the weather at the exact location of interest, we use the Euclidian distance formula to find the location with the weather data that is closest to the venue our customer will be at.

```
Full Day Weather :
Location Closer To Clark Quay : City
Location Closer To Orchard    : City
City->Partly Cloudy (Night)
City->Partly Cloudy (Day)
City->Light Rain
City->Showers

Current Weather Info :
Location Closer To Clark Quay : City
Location Closer To Orchard    : City
City->Showers

Pollution Index Info :
Area Clark Quay Belongs To : south
Area Orchard Belongs To    : south
Location : south -> Pollution Index PM2.5 10 and is Good
```

8. Recommending where to go – from selected group
    a. Say if our customer is looking to dine at a restaurant, it's an advantage if he can have the top 10 restaurant suggestion to select from
    b. In order to get this information, I have use web scrapping technique using beautifulsoup python module and extracted information from Trip Advisor website.
    c. https://www.tripadvisor.com.sg/Restaurants-g294265-zfn15622523-Singapore.html

```
Query Link  : https://www.tripadvisor.com.sg/Restaurants-g294265-zfn7291602-Singapore.html
Query Title :  10 Best Orchard Road Restaurants (Singapore) - Tripadvisor

1. Joie by Dozo
2. Lawry's The Prime Rib
3. il Cielo
4. Yan Ting
5. Mitzo
6. Les Amis
7. The Curry Culture
8. Crossroads
9. LaBrezza
10. Opus Bar & Grill

Query Link  : https://www.tripadvisor.com.sg/Restaurants-g294265-zfn15622523-Singapore.html
Query Title :  10 Best Clarke Quay Restaurants (Singapore) - Tripadvisor

1. The RANCH Steakhouse By ASTONS
2. Haidilao Hot Pot
3. Sque Rotisserie & Alehouse
4. RAS The Essence of India
5. Violet Oon Satay Bar & Grill
6. Bayang
7. Brewerkz (Riverside Point)
8. McGettigan's CQ
9. Cafe Iguana
10. Ellenborough Market Cafe
```