

The Battle of the Neighbourhoods

(*Singapore*)

1. Introduction & Business Problem

High-Level Project Scope:

The goal is to build a recommendation system to help traveller/resident decide on which neighbourhood to visit among the available popular locations in Singapore.

Introduction & Business Problem:

Singapore is a diverse, multi-cultural and multi-racial society. It's an urban jungle and always a challenge for travellers and residents to decide on which place to visit and where to spend time, especially if you are thinking about a place to hang-out. Hence, setting up an application which can recommend the user possible locations and venues one could spend time and also provide him with necessary information pertaining to his choice of location will help him in making informed decisions.

With the Data Analytics & visualization I would like to build an application concept which will use the available public data about Singapore from different sources to find list of places available in Singapore, explore the places and categorize the venues, select top TWO places and compare the available venues among the places and also setup a framework to provide more information pertaining to that location and help figure out why one location is better than other.

Targeted Audience:

The targeted audience is individuals who have very less information about Singapore but want to make the most out of their visit. The individual could be a traveller visiting Singapore for the very 1st time or a resident who wants to explore and compare different places in and around Singapore.

Final Solution:

Select 2 locations the user can explore and assist him in knowing more about the locations.

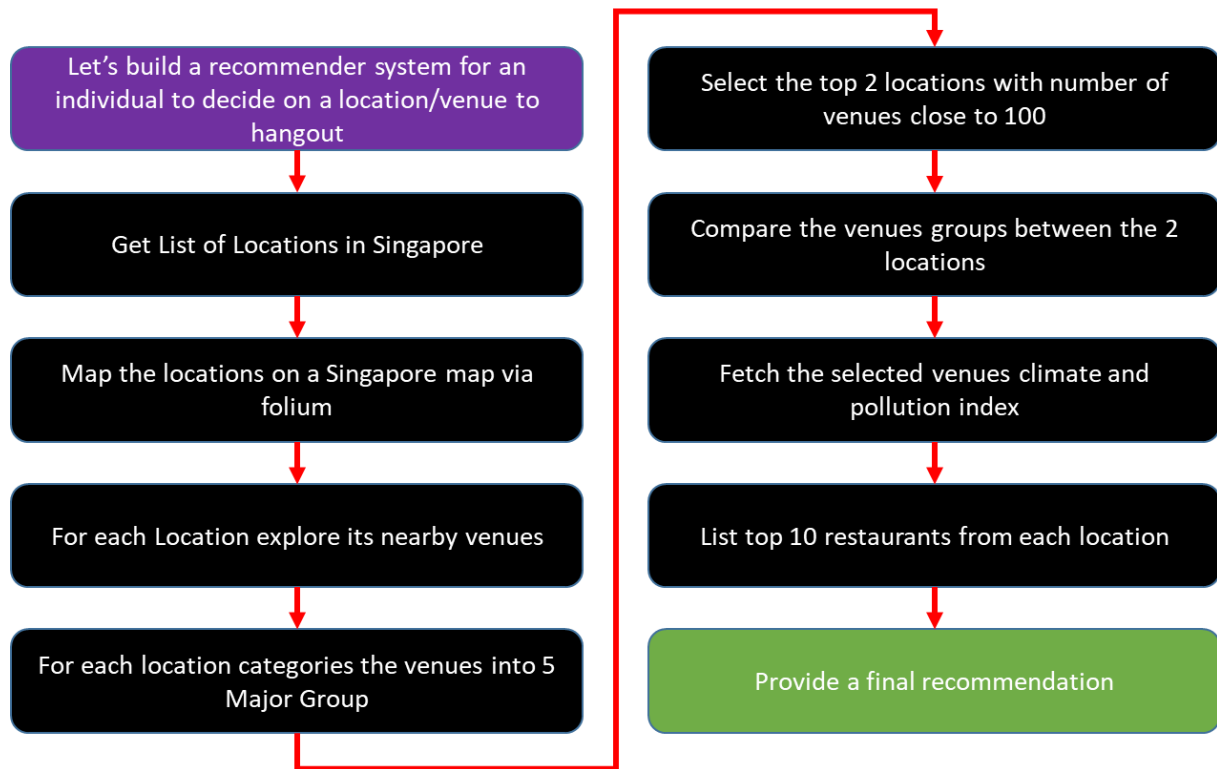
Customer Requirements:

1. My Customer would like to visit a location in Day Time under Clear Weather Conditions
2. My Customer would like to have options of dining, Shopping and Movies
3. My Customer would like to have an option to select from multiple cuisine
4. My Customer would like to be recommended on the Top Restaurants and its Operational Status Real-Time.

2. Data & its Source:

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:

1. OneMap Singapore API - <https://docs.onemap.sg/>
2. FourSquare Singapore API - <https://api.foursquare.com/v2/venues/explore>
3. Data Gov Sg API - <https://data.gov.sg/>
4. Web Scrapping of Trip Advisor - <https://www.tripadvisor.com.sg/>

Data Representation:

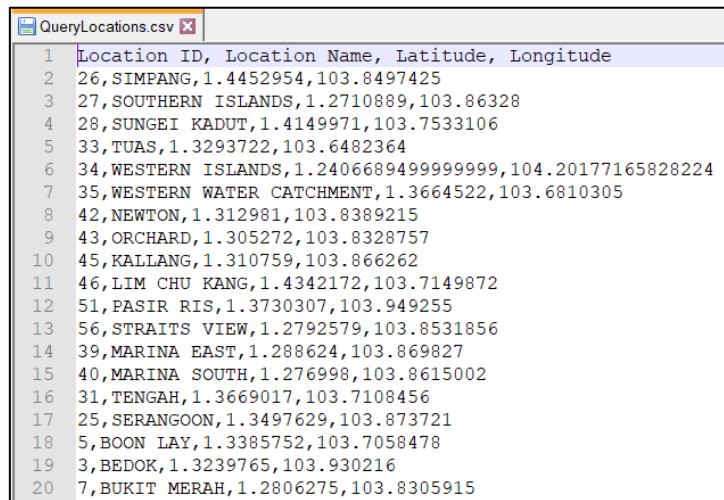
1. Tables
2. Folium Maps
3. Data Frame Bar Plots
4. Word Cloud

3. Methodology

Fetching the Required Data:

1. Getting list of locations in Singapore

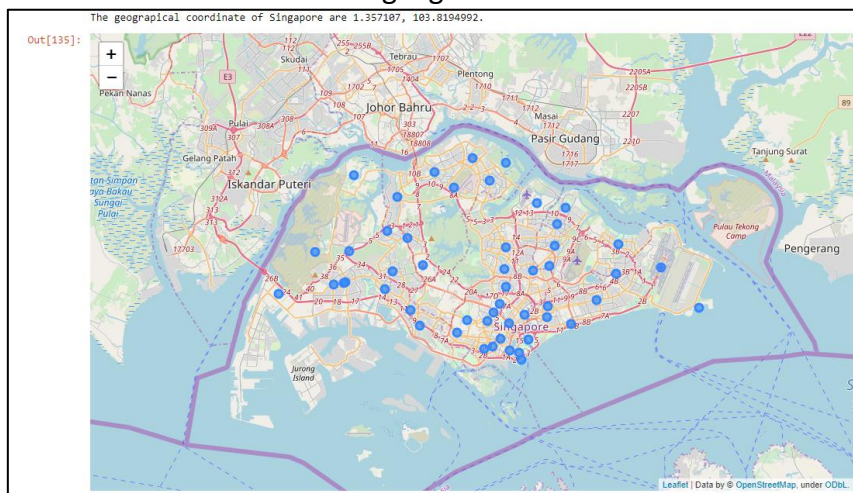
- To get this data, we use API from OneMapSG that provides list of PlanningAreaNames. The returned data is in Json format which is further processed and cleaned up to extract the Location names.
- Cleaning is required as some locations are not a valid location of exploration.
- <https://docs.onemap.sg/>, refer this link for the details on the API supported by One Map SG.
- In order to identify its LATLON co-ordinates, the python geolocator module is used.
- Locations with No LATLON information is excluded and the final output is saved in a CSV
- A total of 53 valid locations with its LATLON information was extracted



	Location ID	Location Name	Latitude	Longitude
1	26	SIMPANG	1.4452954	103.8497425
2	27	SOUTHERN ISLANDS	1.2710889	103.86328
3	28	SUNGEI KADUT	1.4149971	103.7533106
4	33	TUAS	1.3293722	103.6482364
5	34	WESTERN ISLANDS	1.2406689499999999	104.20177165828224
6	35	WESTERN WATER CATCHMENT	1.3664522	103.6810305
7	42	NEWTON	1.312981	103.8389215
8	43	ORCHARD	1.305272	103.8328757
9	45	KALLANG	1.310759	103.866262
10	46	LIM CHU KANG	1.4342172	103.7149872
11	51	PASIR RIS	1.3730307	103.949255
12	56	STRAITS VIEW	1.2792579	103.8531856
13	39	MARINA EAST	1.288624	103.869827
14	40	MARINA SOUTH	1.276998	103.8615002
15	31	TENGAH	1.3669017	103.7108456
16	25	SERANGOON	1.3497629	103.873721
17	5	BOON LAY	1.3385752	103.7058478
18	3	BEDOK	1.3239765	103.930216
19	7	BUKIT MERAH	1.2806275	103.8305915
20				

2. Mapping the locations to enable visualization

- 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

- The list of venues near each of the 53 location is extracted using the Foursquare API – venues/explore.
- The URL is configured with LATLON values for each location and venues surrounding a radius of 500 and LIMIT of 100 are considered
- <https://api.foursquare.com/v2/venues/explore>
- The Json output is further processed and dumped into CSV
- Until now, for the customer we have listed down all the available locations in Singapore and places around these locations to explore
- Out of the 53 locations only 2 locations had 100 venues in and around it.
- These locations are categorized as popular location and shall be used for analysis and provide a recommendation

```
>>> 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
```

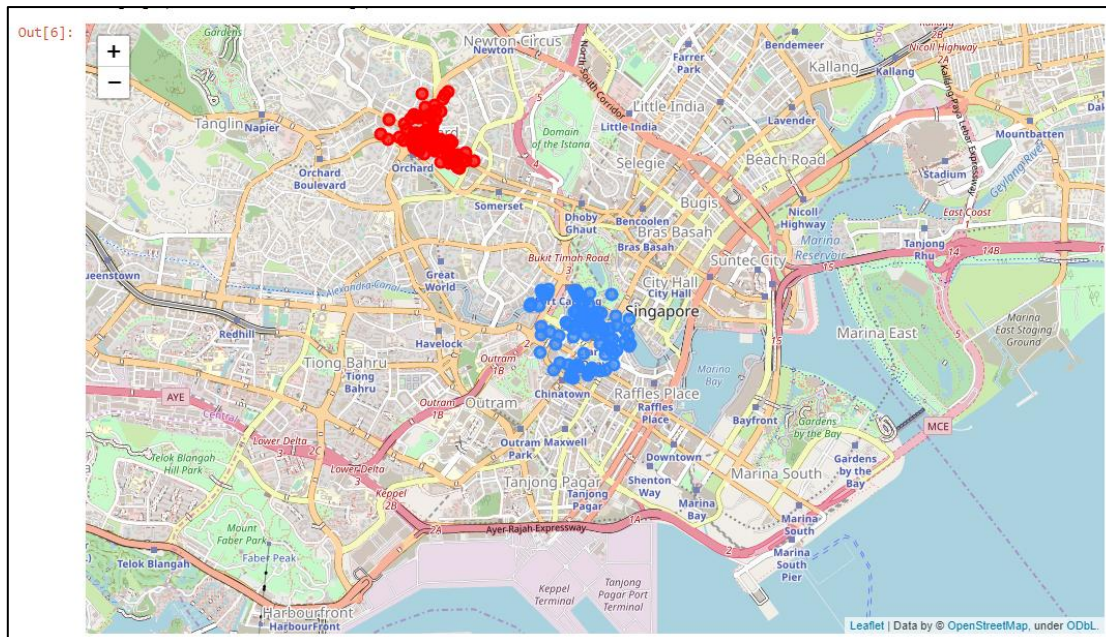
The Prime Locations Are !!!

```
In [97]: for Loc in PrimeLocations:
         print(Loc)

ORCHARD
SINGAPORE RIVER
```

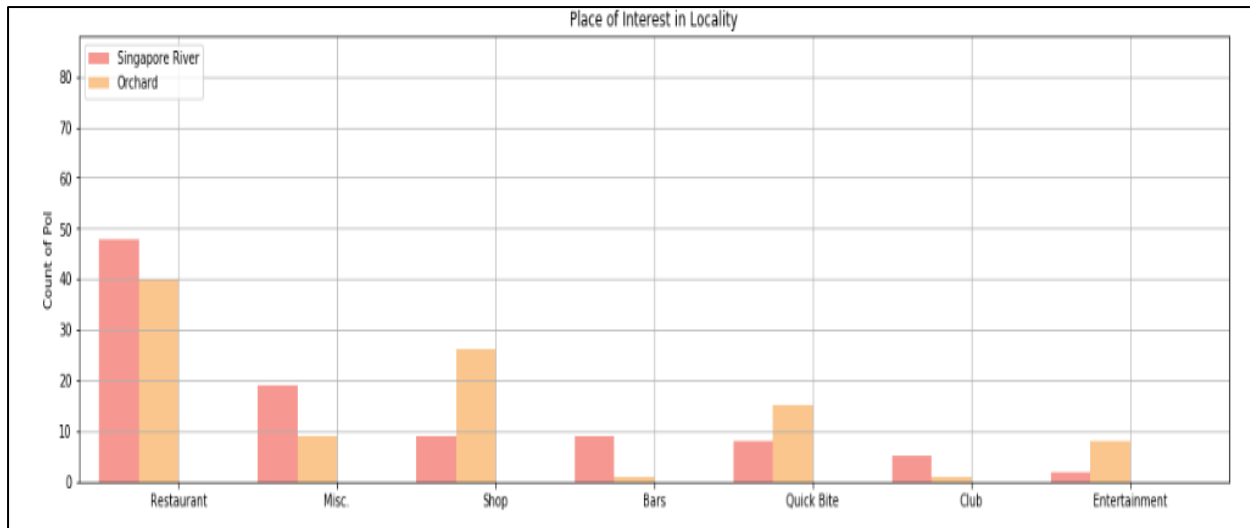
4. Visualizing Selected Locations To Explore Further

- Based on the number of venues the top two places were selected namely “Singapore River” and “Orchard”.
- The list of venues around these two places are clustered and are shown using folium Map



5. Grouping Venues

- 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.
- Thus the following groups were created – Restaurant, Bar, Club, Quick Bite, Shop, and Entertainment.
- 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.
 - If customer is willing to spend his day time - He can opt for venues under Restaurant, Shop, Entertainment
 - If customer is willing to spend his Night time - He can opt for venues under Restaurant, Bar, Club
- All other venues that doesn't fall under the above groups are categorized as Misc. e.g. Gym, Education Institution etc.
- Next a Grouped Bar Chart is used to compare the count of categorized venues under these two locations.
- This bar chart will gives an idea on which location of two had more options to spend and further selection can be made based on his time of visit.



6. Finalizing a Location based on Supporting Data

- 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.
- But if our customer is a Night owl, “Singapore River” is a bustling area with “Restaurants”, “Bars” and “Clubs”.
- Thus we have a basic interest based recommendation engine built.
- Let's assume he is willing to spend his day time and hence he selects “Orchard”*

- Explore : Venues @ Orchard
- Time : Day Time

```

: Orchard = 'Orchard, SG'
geolocator = Nominatim(user_agent="TheProject")
location = geolocator.geocode(Orchard)
Orchard_lat = round(location.latitude, 3)
Orchard_lon = round(location.longitude, 3)
print("Orchard : " + str(Orchard_lat) + ", " + str(Orchard_lon))

Orchard : 1.305,103.833

```

7. Providing More Data to validate his choice of location

- Now our customer has made his choice we believe the following information will be of help.
- 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.
- 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 Orchard : City
City->Partly Cloudy (Night)
City->Fair (Night)
City->Fair (Day)
City->Fair & Warm


Current Weather Info : Time - 2020-03-17 22:50:34
Location Closer To Orchard : City
City->Fair (Night)

Pollution Index Info :
Area Orchard Belongs To : south
Location : south -> Pollution Index PM2.5 11 and is Good
```

8. Recommending him a Restaurant

- My customer would looking to dine, it's an advantage if he can have the top 10 restaurant suggestion to select from and also information on the popular cuisine in his locality
- In order to get this information, I use web scrapping technique using beautifulsoup python module and extracted information from Trip Advisor website.

Query Link : <https://www.tripadvisor.com.sg/Restaurants-g294265-zfn15622523-Singapore.html>
 Query Title : 10 Best Clarke Quay Restaurants (Singapore) - Tripadvisor
 Output File : C:\Users\User\Documents\GitHub\BattleOfNeighCapStone\Orchard.csv



Top 10 Restaurants :

	Restaurant Name	Operating
0	The RANCH Steakhouse By ASTONS	Closed Now
1	Haidilao Hot Pot	Open Now
2	Sque Rotisserie & Alehouse	Open Now
3	RAS The Essence of India	Open Now
4	Violet Oon Satay Bar & Grill	Closes in 26 min
5	Bayang	Closed Now
6	Brewerkz (Riverside Point)	Open Now
7	McGettigan's CQ	Open Now
8	Cafe Iguana	Open Now
9	Ellenborough Market Cafe	Open Now

4. Results

- *We found the list of locations in Singapore using the OneMapSG API results*
- *Using the Foursquare API we explored the available venues in each locations and selected the Top 2 venues with number of venues equal to 100*
- *The venues were further categorized and compared across both the locations to select the best location for our customer to visit*
- *Once our customer has decided the locations we provided him the real-time weather conditions, as Singapore's climate at times become unpredictable. Singapore Governments database was used to retrieve this information real-time.*
- *Assume our customer is planning on visiting a restaurant we use the trip advisor page to retrieve the top 10 restaurants and popular cuisine information at his selected location*
- *Also, we provided him the real time operational status of the restaurants*
- *Thus a basic recommendation system was built using the available web data to recommend a suitable location to explore and then dine at a restaurant of the customer choice, satisfying his requirements.*

5. Discussion

Customer, a traveller or resident can use our recommendation system to select his choice of location for exploration and further drill down to a venue level (in this case it will be a restaurant) with confidence using the data and visualization provided by the project.

6. Conclusion

This analysis is performed on limited data and few assumptions were made to arrive at the final recommendation. This can be further expanded in to a real – time interactive application with more granular representation of the data and recommendations.