BATTLE OF NEIGHBORHOOD

MANISH GUPTA

INTRODUCTION

- Singapore is an economic hub though it is small in the size.
- Most modern city in Southeast Asia over a century.
- Became a flourishing country for trade & tourism due to efficient and determined Govt.
- Vast job opportunities in the country.
- Immigration to Singapore is the main source of population growth in the country.
- Immigration & immigration workers have been closely associated with Singapore's economic development.

BUSINESS PROBLEM

- Finding the right accommodation in Singapore is a really tough.
- Depending upon the location & places to eat, people have different preferences.
- Some people prefer to live near work-place even with limited availability of resources.
- Some of the problems addressed by this project:
  - Finding the right accommodation for person who is moving to new location.
  - Finding the right location where there have the required facilities available.
  - Finding the neighborhood near to particular attractions such as parks, galleries & restaurants.

DATA USED IN PROJECT

- Singapore Neighborhood data list
  - Important parameters – Building Names, latitudes & longitudes, postal code.
  - Data Source: Open Repository
- Venue Data: Exploring various neighborhood for each locality within radius of 500m.
  - Data Source: Foursquare API

DATA CLEANING & PRE-PROCESSING

- Libraries used – pandas, numpy, geopy, sklearn, folium, requests, json
- Building Data from Json file.
- Total 141726 samples with 11 columns.
- Only 4 are selected – postal code, building name, latitude & longitude.
- Building names with NIL is ignored.
- Duplicate building names are removed and only 1 entry is kept.
- Entries related to park are removed.
- Final dataset to be analyzed is of shape (16150,4)

ANALYSIS

- Random 600 samples selected from the whole dataset.
- Selected samples are printed on Singapore Map using folium library.

ANALYSIS

- Initially first building is selected and its near venues are fetched using foursquare API
- Building analyzed is TAMPINES GREENWOOD and total 7 unique venues are fetched for this building.

Out[41]:

	name	categories	lat	lng
0	Tampines Central Park	Park	1.354111	103.936393
1	NTUC Fairprice	Supermarket	1.355541	103.934758
2	Playground @ Tampines Blk 869	Playground	1.354625	103.933893
3	Madison's	Sandwich Place	1.354242	103.933217
4	Bus Stop 75139 (Blk 863)	Bus Station	1.355861	103.936358

[42]:
```
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
7 venues were returned by Foursquare.
```

ANALYSIS

- Now, venues for all 600 buildings are fetched and merged together in the dataset.
- There are total 373 unique venues for entire neighborhood.
- Further data is processed to created sorted venues dataset with top 10 venues for each neighborhood.

Out[56]:

	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	57 @ KOVAN	Vegetarian / Vegan Restaurant	Food Court	Asian Restaurant	Noodle House	Indian Restaurant	Supermarket	Bar	Sandwich Place	Park	Furniture / Home Store
1	ACACIA WELFARE HOME	Bus Station	Harbor / Marina	Thai Restaurant	Baby Store	Café	Automotive Shop	Kids Store	Bus Stop	Asian Restaurant	Zoo Exhibit
2	AL - ISTIQAMAH MOSQUE KINDERGARTEN	Asian Restaurant	Pet Store	Coffee Shop	Supermarket	Playground	Noodle House	Chinese Restaurant	Breakfast Spot	Food Court	Electronics Store
3	ALJUNIED COMMUNITY CENTRE	Coffee Shop	Noodle House	Bakery	Food Court	Café	Cafeteria	Breakfast Spot	Gym	Bus Stop	Bus Station
4	ALKAFF MANSION	Bakery	Bus Station	Scenic Lookout	Trail	Bus Line	Karaoke Bar	Gym	Miscellaneous Shop	Clothing Store	Club House

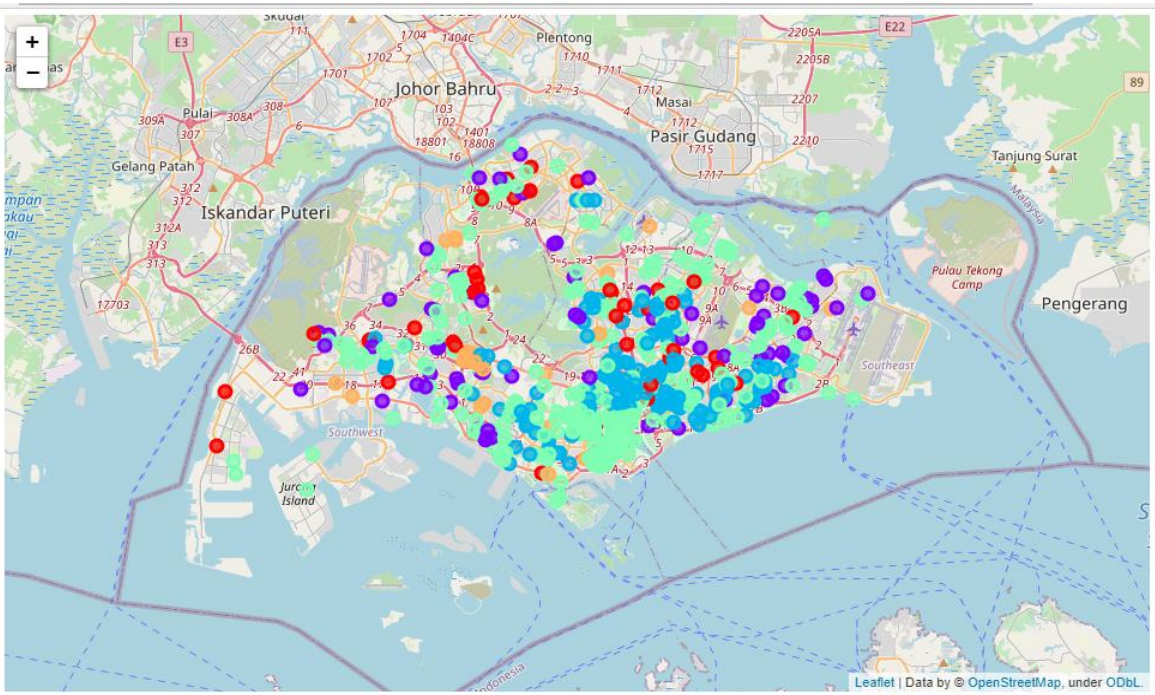
In [57]:
```
neighborhoods_venues_sorted.shape
```

Out[57]: (596, 11)

MODELING

- K-Means Algorithm is used to create clusters.
- No of clusters selected = 5
- Clustered data is merged with neighborhood data and plotted on the map using folium library

Out[77]:



USER CASES

- Case-1: Person coming to Singapore for first time – which place would be best to move in?
  - Cluster 3 is the one which has most number of neighborhood. Best place to settle down initially.
- Case-2: Person wants to move from place as it is expensive for me but likes the neighborhood so which would be best place for him similar to previous place?
  - Previous place “PAN PACIFIC SERVICED SUITES” belongs to cluster-3.
  - So he can move to FERNVALE FLORA which belongs to same cluster sharing similar neighborhood.