

# BATTLE OF NEIGHBORHOOD SINGAPORE

# Problem Statement:

Prospects of a Lunch  
Restaurant, Singapore.

# Target Audience

- ▣ Business personnel who want to invest or open a restaurant.
- ▣ Freelancer who loves to have their own restaurant as a side business.
- ▣ New graduates, to find reasonable lunch/breakfast place close to office.
- ▣ Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyze it, and, finally be able to tell a story out of it.

# Data Acquisition And Cleaning

- ▣ Data is acquired from Wiki Page Planning\_Areas\_of\_Singapore to scrap the table to create a data-frame using requests and BeautifulSoup4 library
- ▣ Getting Coordinates of Areas : Geopy Client
- ▣ Cleaning The Data :for wrong coordinates of Areas in Singapore. Cleaned data have 55 areas of Singapore
- ▣ Using Foursquare Location Data

# Screenshot of Final Dataframe

**Final Data-Frame with Coordinates of the Major District**

Singapore\_data\_Final

	Name	Region	Area_SqKm	Latitude	Longitude
1	Ang Mo Kio	North-East	13.94	1.369842	103.846609
2	Bedok	East	21.69	1.323976	103.930216
3	Bishan	Central	7.62	1.351452	103.848250
4	Boon Lay	West	8.23	1.345640	103.711802
5	Bukit Batok	West	11.13	1.349057	103.749591
6	Bukit Merah	Central	14.34	1.281905	103.821711
7	Bukit Panjang	West	8.99	1.377921	103.771866
8	Bukit Timah	Central	17.53	1.354690	103.776372
9	Central Water Catchment	North	37.15	1.355205	103.795011
10	Changi	East	40.61	1.345005	103.981011
11	Changi Bay	East	1.7	1.316850	104.020649
12	Choa Chu Kang	West	6.11	1.389260	103.743728
13	Clementi	West	9.49	1.314026	103.762410
14	Downtown Core	Central	4.34	1.286705	103.851311
15	Geylang	Central	9.64	1.318186	103.887056

# Visualization and Data Exploration

## ▣ Folium Library and Leaflet Map:

Folium is a python library that can create interactive leaflet map using coordinate data

- ▣ Only Restaurants are considered from all the venue category. A snippet of code is shown below

```
# Create a Data-Frame out of it to Concentrate Only on Restaurants
```

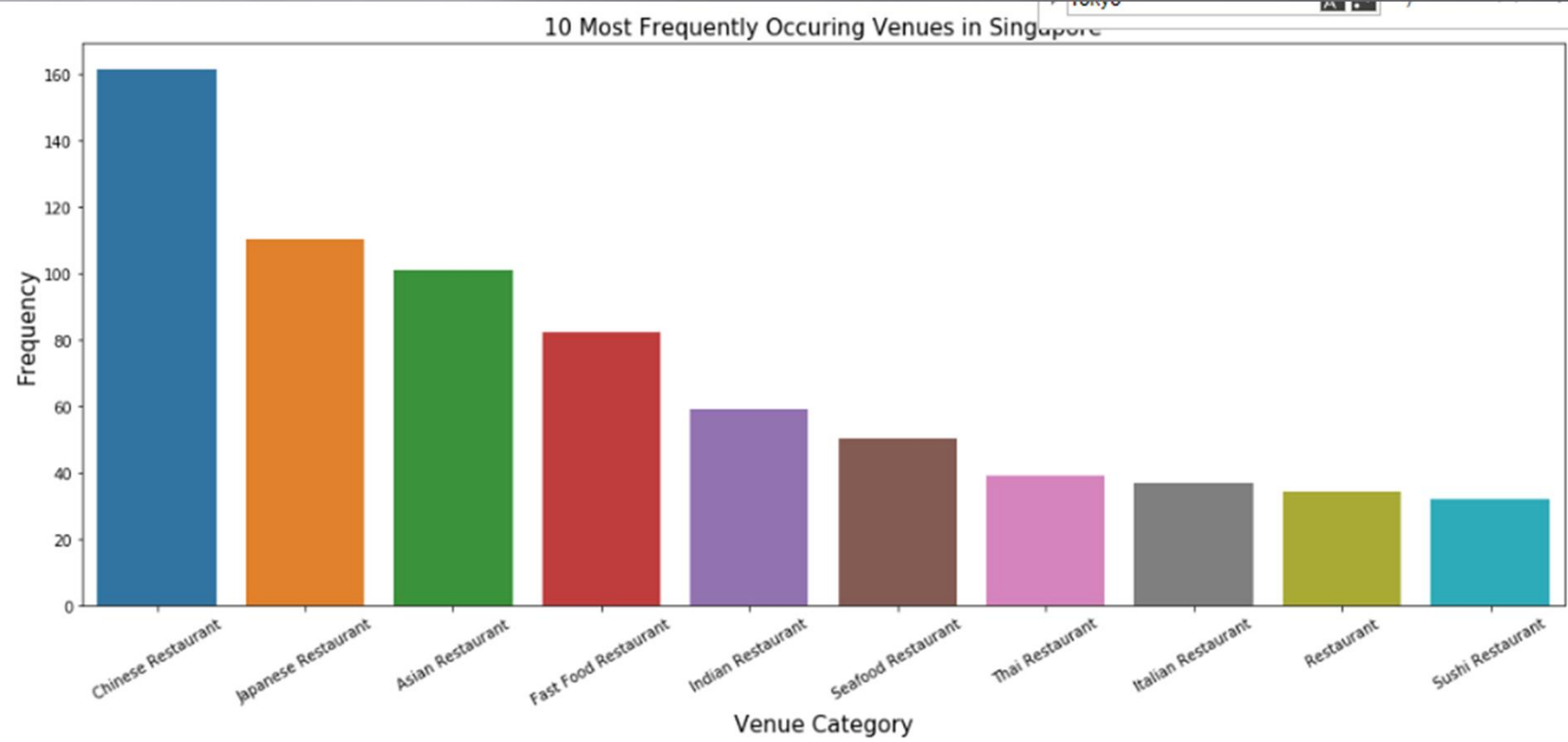
```
Singapore_Venues_only_restaurant = singapore_Venues[singapore_Venues['Venue_Category']\
                                                         .str.contains('Restaurant')].reset_index(drop=True)
Singapore_Venues_only_restaurant.index = np.arange(1, len(Singapore_Venues_only_restaurant)+1)
print ("Shape of the Data-Frame with Venue Category only Restaurant: ", Singapore_Venues_only_restaurant.shape)
Singapore_Venues_only_restaurant.head(3)
```

Shape of the Data-Frame with Venue Category only Restaurant: (964, 7)

	Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
1	Ang Mo Kio	1.369842	103.846609	Kam Jia Zhuang Restaurant	1.368167	103.844118	Asian Restaurant
2	Ang Mo Kio	1.369842	103.846609	Collin's Grille . Bento	1.371713	103.847526	Modern European Restaurant
3	Ang Mo Kio	1.369842	103.846609	Xi Xiang Feng Yong Tau Foo 喜相逢 三寶	1.371975	103.846408	Chinese Restaurant

# Exploratory Data Analysis

- ▣ There are 59 unique venue categories and Chinese Restaurants top the charts as we can see in the plot.





# Clustering the Districts

- ▣ Cluster the Areas based on the venue categories and use K-Means clustering. So our expectation would be based on the similarities of venue categories, these Areas will be clustered.

## Results and Discussion:

- ▣ Chinese restaurants top the charts of most common venues in Singapore.
- ▣ Geylang (Central region) and Novena (Central region) are dominated by restaurants as the most common venue.