**CAPSTONE PROJECT: BATTLE OF NEIGHBOURHOODS**

**Objective:** To succeed in food business its important to choose a location where there is more population and less restaurants. We intend to choose a location best suited for a new Chinese restaurant.

1. **Background: Booming consumerism of Pakistan**

Pakistan is categorized as an emerging country with average age population of 24. With increasing income per capita and youth, consumerism is growing double digits and likely to continue so in coming years. Culturally, South Asians love food and like hanging out over lunch/ dinners and hence decent chunk of disposable income is spent on eating out.

Secondly, with China starting multi billion dollars China Pakistan Economic Corridor (CPEC) with Pakistan, thousands of Chinese are moving to Pakistan with most settling in Karachi as its an economic hub of Pakistan. I intend to open a Chinese restaurant in the economic capital of Pakistan, Karachi. With diversified population of expatriates (mainly Chinese), government officials, businessmen and students, Chinese cuisines are likely to do well.

## Interested Audience

Target audience for the analysis is:

1) Anyone looking to open a new restaurant.

2) Guide for foodies to find the neighbourhoods with different cuisines.

3) Helping Chinese expatriates to choose a location to live or hang out in Karachi.

# Data section

## Data Sources

## Following data will be used for analysis:

* Neighbourhoods/ Towns in Karachi to find number and type of restaurants concentration: <https://en.wikipedia.org/wiki/Towns_in_Karachi#Karachi_Towns>
* Four square API data
* Google API for the coordinates of towns

## How data will be used

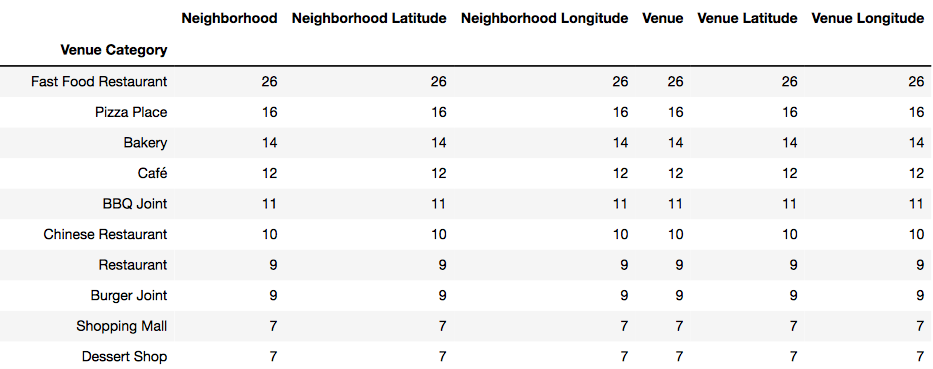
The neighbourhoods data of Karachi contains eighteen towns. The data will be scraped form Wikipedia page to create a dataframe, enabling us to run machine learning algorithms. Once we get the data in required dataframe, Google API will be used to get the coordinates for each town.

Using Foursquare API, each town will be clustered with concentration of restaurants and cuisines. The processing of data will help in identifying 1) Which Council has less concentration of restaurants, 2) Have other amenities and entertainment areas such as Theater, offices etc and 3) Types of Cuisines already available.

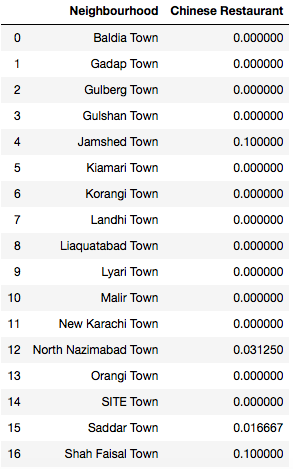
1. **Methodology**

The scraped data contains names of the towns in Karachi. For each town we will first get the coordinates using Google API. Once we have the coordinates for each town, we will clean the data and so that there are no Nan values, missing data etc.

Using foursquare API, we will analyze all the towns and check nearby venues/ amenities.



As the objective is to find a best location for a Chinese restaurant, new data frame will be created containing 1) Towns and 2) Chinese restaurants.

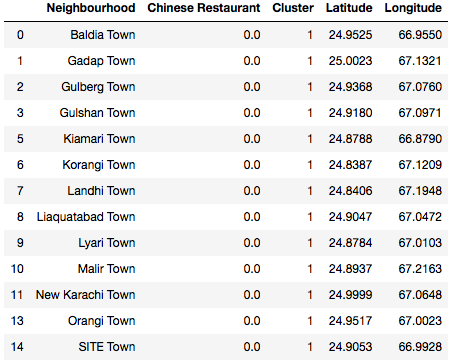


We will use K-Means clustering algorithm of machine learning to cluster the city in 3 clusters to find concentration of Chinese restaurants from lowest, moderate and highest areas.

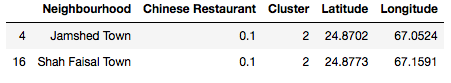
**Cluster 0:**



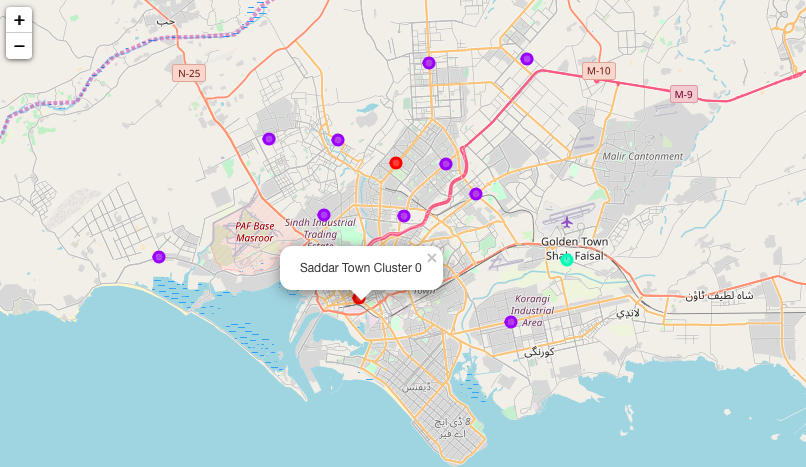
**Cluster 1:**



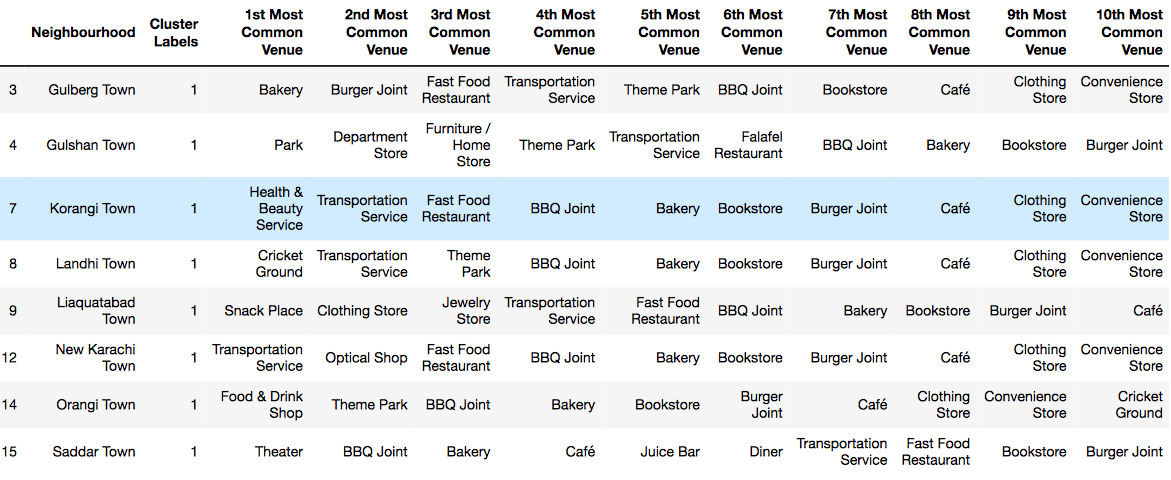
**Cluster 2:**



Mapping the clusters:



Since we also want to choose a location, which is frequently visited by the local inhabitants but also attractive for Chinese expatriates. Hence we need to choose crowded areas where other entertainment avenues are also available such as theatre, other restaurants etc. To achieve that, we will again use K-Means clustering algorithm to cluster the Karachi city in 5 clusters on the basis common venues. Four square API will be used to fetch the most common venues in each town.



## Observation/ Conclusion:

### Most of the Chinese restaurants are concentrated in cluster 2 while cluster 1 is least concentrated. It seems cluster 1 is suitable candidate for opening a Chinese restaurant as it offers less competition. However, it is also important to see if the neighbourhoods in cluster 1 have other amenities and cuisine options to gauge the consumers' traffic. Therefore we will also cluster the city Karachi on "most common venues".

Analyzing the data, it seems Saddar Town in cluster 0 (in data frame Khi\_CH) presents the best neighbourhood for opening a Chinese restaurant because 1) It has a less concentration of Chinese restaurants, 2) its a city center, 3) its nearby/ most common venues include Theatre, Transportation service and several fast food & traditional cuisines. This shows Saddar town has high foot traffic from people coming to the downtown for movies and eating out. If in such a locality, Chinese restaurant option is available to the people, chances of it getting successful are very high.

Despite in cluster 1 (in data frame Khi\_CH), the concentration of Chinese restaurant is lowest, it might not be a best neighbourhood to open a Chinese restaurant as the towns are mainly in outskirts of the city where population belongs to low-mid income and prefer traditional cuisines and fast food. Secondly, Chinese moving to Pakistan are likely to stay in or around the city center due to its proximity to the offices and other amenities.

**HENCE I SUGGEST OPENING OF A CHINESE RESTAURANT IN SADDAR TOWN, KARACHI.**