Premium Fitness Centre Location Recommender (Using K-Means)

Data Acquisition and Preparation

Based on definition of our problem, factors that will influence our decision are:

- Number of existing gyms in the neighbourhood or nearby neighbourhood
- Population of the neighbourhood
- Average Income of the neighbourhood

In our Project we will :

- Acquire the names and boroughs of the neighbourhoods by scrapping a wikipedia page.
- After we have got the names of all the neighbourhoods, we will geocode them using the library geopy.geocoder (Nominatim) to get the latitude, longitude of each neighbourhood.
- Use Kaggle and internet to get Population data of each neighbourhood of Bangalore.
- Similarly to get Average Income of each Neighbourhood.

After we combine and merge all data into one DataFrame, we have to clean it, process it to make it useful for Data Understanding and further processes ahead. So here is a sample of how our clean DataSet looks like on which we going to work.

```
df.head()
```

Out[3]:

	Borough	Neighborhoods	Latitude	Longitude	Population	AverageIncome
0	Central	Cantonment area	12.972442	77.580643	866377	18944.099792
1	Central	Domlur	12.960992	77.638726	743186	56837.022198
2	Central	Indiranagar	12.971891	77.641151	474289	41991.817435
3	Central	Jeevanbheemanagar	12.962900	77.659500	527874	6667.447632
4	Central	Malleswaram	13.003100	77.564300	893629	53270.063892