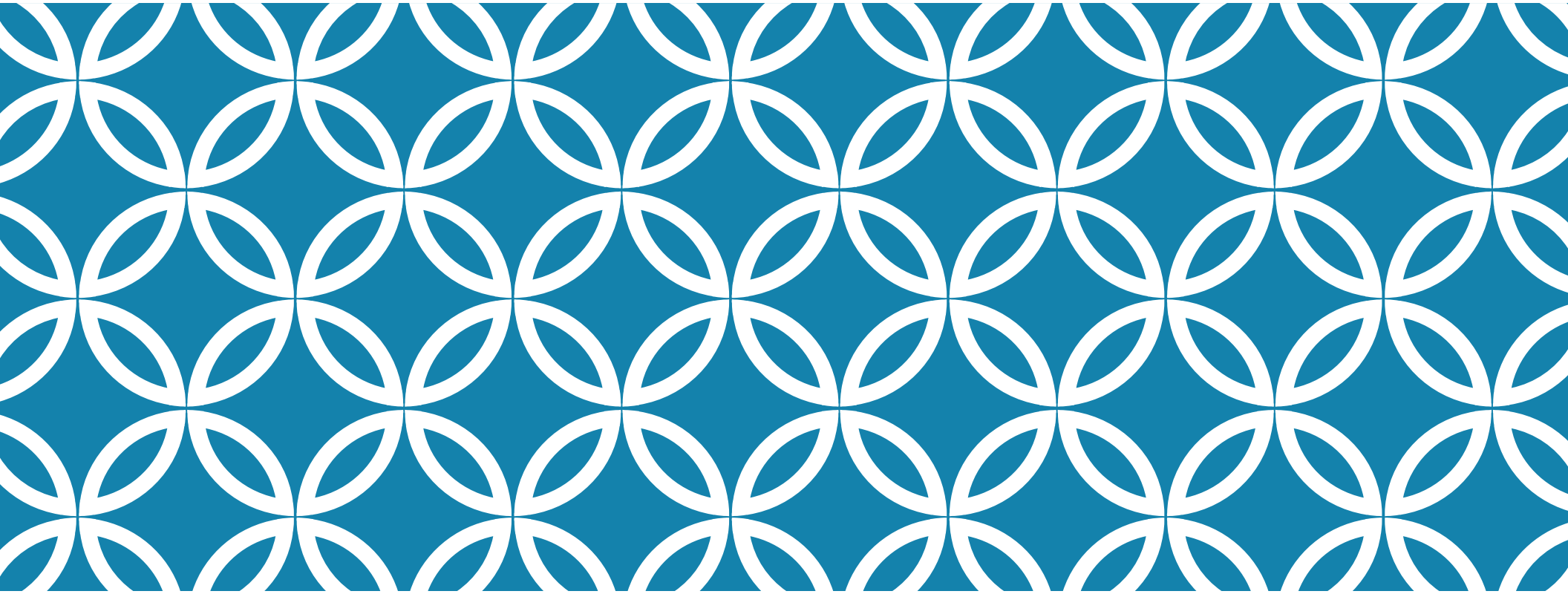


CAPSTONE PROJECT: BATTLE OF THE NEIGHBORHOODS

Recommendation for Gym / Fitness Center installation

By : Alowonou Kowovi Comivi

INTRODUCTION AND BUSINESS PROBLEM

The owner of one Gym / Fitness center in France named Armando, wants to settle in the United States. Due to USA high diversity and very large size, he asked for help in order to find the best place for the installation of his Gym / Fitness center. Building a system for the best location recommendation would help Armando to minimize the competition and have a high income. The best locality is defined bases on the following criteria:

- Location with high average income.
- Location with high population rate.
- Near activity area such as park, playground etc.
- Near residential district, university/school and offices.
- low amount of competition (less or no gym / fitness centers around)

DATA

- List of all the cities in United States with population density and coordinates

  https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population
- List of all the cities in United States with Per Capita Income

  https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income
- Using Foursquare API to get the following
  - List of all venues in each city
  - List of all venues in each locality in the selected city

METHODOLOGY

1. Data collection:

BeautifulSoup, a Python library for pulling data out of HTML and XML files is used to get United States cities by population data from Wikipedia as well as United States counties by per capita income data

The Foursquare API is also used to get the venues in each city of United State, based on the categories of each venue as decided by the Armando, we have assigned weights to each of them

METHODOLOGY

2. Data Processing:

- We drop unnecessary columns
- Adding geographical coordinates of each town location
- Fix data types
- Preprocessing the population density in Km2 column as we have to normalize these values
- For each category of venue obtained with the Foursquare API, a weight (or penalty) has been defined according to what Armando considers the most important

METHODOLOGY

3. Data normalization

Population and Weights were normalized, using MinMaxScaler() function of sklearn. The normalized data will be the one we will used for the segmentation and clustering

4. **Segmentation and clustering**

We cluster every venue using K means algorithm and calculate the weights of each of them

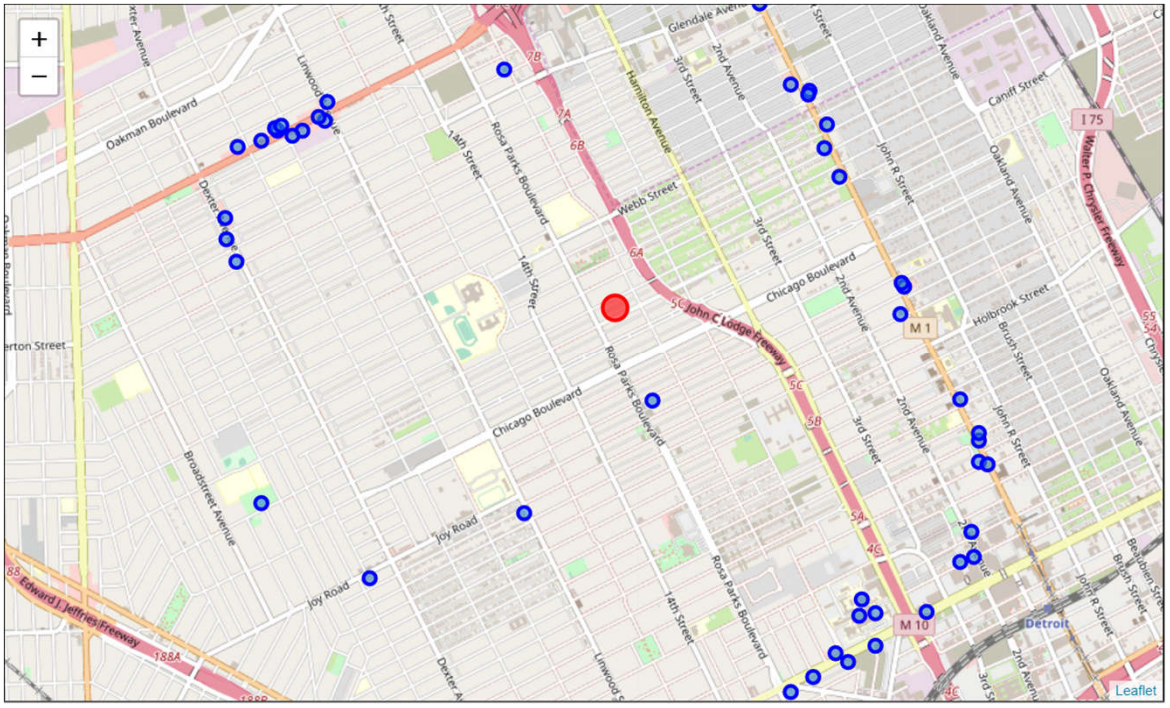
We group venues by clusters and add mean weights to each cluster.

The location with the maximum weight is chosen and then ploted with a circle of 50M

RESULTS

Map of venues in the Michigan:

- Based on the weight we have assigned to the city in USA we find that Michigan is the best place to start the business.



RESULTS

The pink circle indicates the best place to start an Gym center

RECOMMENDATION/ IMPROVEMENTS

In the Foursquare API, we have queried the Venues of a locality by specifying the LIMIT and Radius of our choice. We have chosen less LIMIT as the number of API calls that can be done using a free account in Four Square are less:

- We can increase the limit for more accurate results
- We can increase the Radius for more venue results from each city
- Consider more categories. For example, like "Universities, Schools, Offices, subway station, bus station" which are also a good source.
- In the Locality itself, it can also be computed the distance between all the venues in order to find a place with the greatest number of potential customers