Final Report- Battle of Neighbourhoods

Problem-

opening a Indian sweet shop in United States of America. The main problem is *to find the best location in USA* to establish the sweet shop.

1). Business Understanding:-

1.1 Scenario and Background-

Ramesh is businessman and currently lives in Singapore. He runs a sweet shop in Singapore near metro station. He earns good amount of money per month. His brother Vikas lives in New York, USA. He wants to open a Indian sweet shop in USA. Opening a shop depends on many things and circumstances such as population of that city, income of people and nature of the people who lives there. Landmarks and Tourist spots are the main things, so city should be chosen wisely and should have great tourist attraction.

Venues can be categorized into the following-

- Schools
- Hospitals
- College and universities
- Travel and transport
- Shops and services
- Residence

Aspect that should be considered-

- Firstly, we should know why to start off with a place in particular?
- If a place is chosen, we must have a very clear understanding of who would your potential customers be?
- We need to have a clear understanding of the customers spending patterns in that locality etc.
- Wou need to take care of the legal aspects involved in that locality.

2). Data Section:-

Description of the data and its sources that will be used to solve the problem-

In order to make a good choice of city in USA, we will use the following data, that is required.

- 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
  (Per capita income means average salary of a person in a region, city or country.)

https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income

We will use Foursquare API to create the map using Folium.

Using Four Square API to get the following-

- List of all venues in each city.
- List of all venues in each locality in the selected city.

3). Methodology:-

Tools we will use-

- Foursquare API
- Jupyter Notebooks
- IBM Watson Account
- List of United States countries by per capita income
- List of United States cities by population

In order to do the analysis and suggest the best location, following are the steps we have to follow :

- The Wikipedia page (https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) was scraped using the Beautiful Soup library to build a pandas data frame listing the cities, states, coordinates, area and population density.
-
- The Wikipedia page (https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income) was scraped using the Beautiful Soup library to build a pandas data frame listing the cities, states and per capita income.

- Four Square API : https://foursquare.com

- The Foursquare API is then used to get the venues in each city of United States.

- Based on the categories of each venue as decided by the Ramesh, we have assigned weights to each of them and got the city that has the maximum weight.

- Once the city is finalized, we again use Four Square API to get the venues within that city and assign weights to each category.

- We will now use K means to cluster the venues based on the category and get the coordinates of the cluster that has maximum weight which is also our preferred location to setup a Shop.

4). Discussions and improvements:-

- In the Four Square 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.

- In the venue categories we are choosing only few out of 2000 that are available to give weights and identify the best cluster. Hence, assigning weights must be done relatively for each category and then considering more number of venue categories would actually yield better output.

5). Conclusions:-

Based on the given constraints, a gaming arcade can be installed in a place closer to the center of the circle to attract more number of diverse customers and get huge revenue.