The Battle of Neighbourhoods

-By Sruthi Pedakolimi

Business Problem

- A Customer, Sundar, wants to open a new gaming arcade in USA
- He defines a best locality based on the following constraints,
  - Population Density
  - Popular Venues in the location

We have to suggest the best location to setup the arcade in the United States

Data

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

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

- Using the FourSquare API to get the following
  - List of all venues in each city
  - List of all venues in each locality of the selected city

Methodology

The Following steps were followed to select the best location in United Sates

- The Wikipedia page *https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population* was scrapped using the BeautifulSoup library to build a pandas dataframe listing the city,state,land area,population density and location .

- The Foursquare API was then used to query each of the city to get {venue names,venue category,venue latitude and longitude} within the given radius . The resulting locations for each venue were tabulated and stored in a new dataframe and visualised using Folium.

Methodology

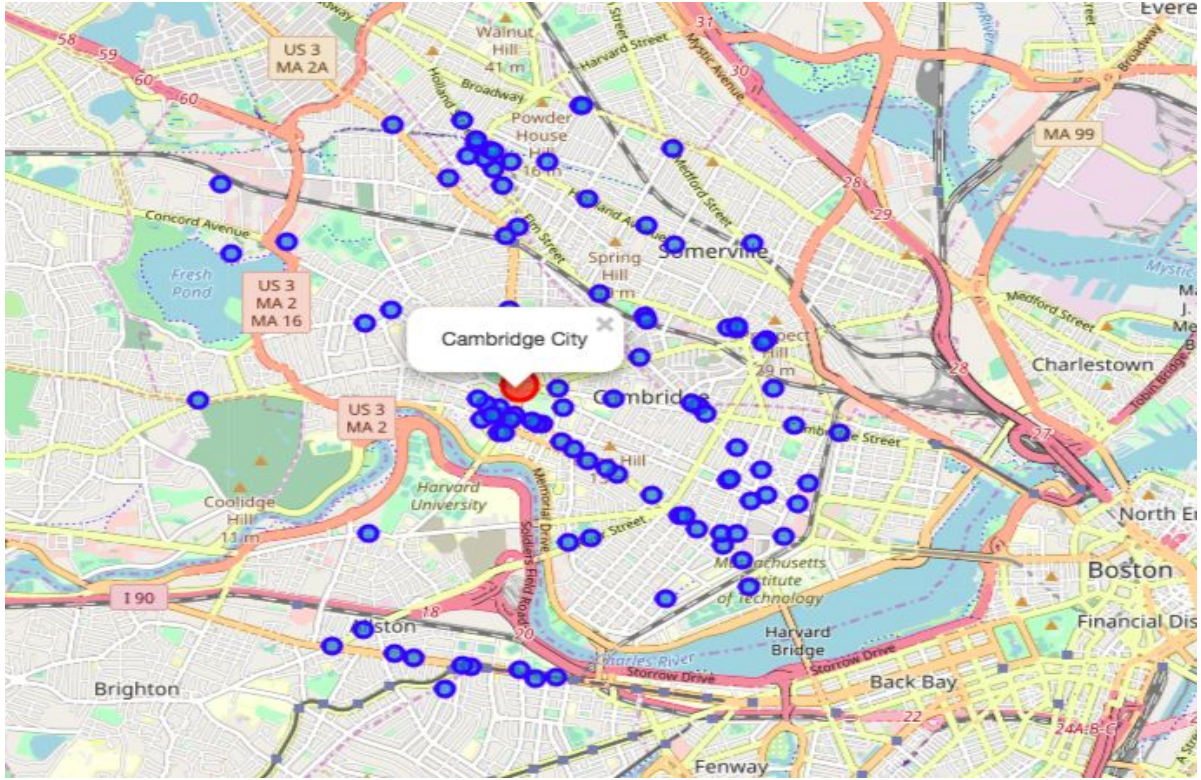
- The K Means Algorithm was used for clustering clustering was applied to the selected categories latitudes and longitudes and calculated the weightage for each cluster

- Then the cluster centroid which has more weightage is selected for the clients gaming arcade

Plot of Cities of United States

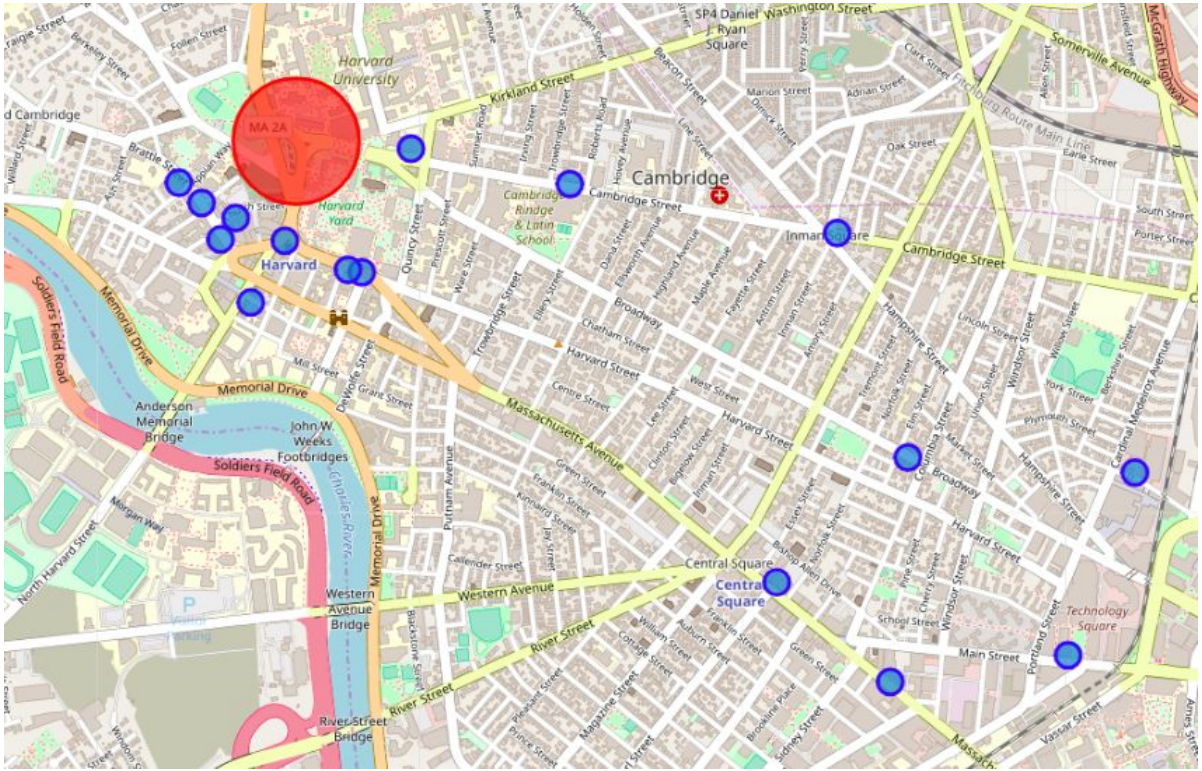
Blue dots represents the cities in the United States

Map of Venues in the Cambridge City



The blue circles represents the venues in the Cambridge city

Result



The red circle represents the best location to start an arcade in cambridge city

Improvements/Recomendations

The Following analysis can be improved with the following extensions

- Here we can consider few more categories. For example Schools,Universities whcih is also a good source for customers
- We can also select multiple locations based on the localities and neighbourhoods to improve the accuracy of the score
- We can extend the area by increasing the radius and the limit value so that we can cover more venues