**The Battle of Neighborhoods – Final Report**

**Introduction**

Spas have become a pertinent cultural force, influencing not only how consumers manage their health, appearance, and stress, but also how consumers socialize, spiritualize, travel, and work. Rising levels of income, education, and sophistication among travelers and consumers worldwide have dramatically elevated the consciousness and desirability of spa treatments. The market potential of spa development is being captured by global and premium-brand spas that have expanded their service menus. As interest in physical well-being increases, spa therapy has become popular among consumers and has been recommended by many medical specialists

In 2017, there were over 149,000 spas, earning $93.6 billion in revenues and employing nearly 2.5 million workers. The spa sector has been growing by 9.9 percent annually from 2015–2017, and it is projected to reach $128 billion in 2022.

**Business Problem**

One of the leading multinational Hotel and resort groups in Europe is interested in starting Luxury spa in Unites States of America and they would like to analyze the data pertaining to main cities and its localities of the country. The suitable locality can be identified based on the inputs such as Per Capita Income of the Main cities, Density of Population in the city, Population of each location and Various venues in the location

As a data scientist we have to take lot of venues into consideration which can directly or indirectly influence the customers at the locality to visit the Luxury spa.

In this project we will consider various factors such as Travel & Transport, Residence, community, Work and offices, Events, Food and recreation, Arts & Entertainment, Shops & Service, College & University, Nightlife Spot and Outdoors & Recreation. Ultimate aim is to find the suitable locality which will attract more customers and it is highly dependent on the location and earning capacity of individuals visiting the location and residing at the vicinity.

**Data Acquisition and Preprocessing**

In this project, I will be using datasets from Wikipedia and Foursquare which will help us to identify the optimal locality to start the Luxury spas in the city of USA.

* population density and coordinates : <https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population>
* Per Capita Income :  <https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income>

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

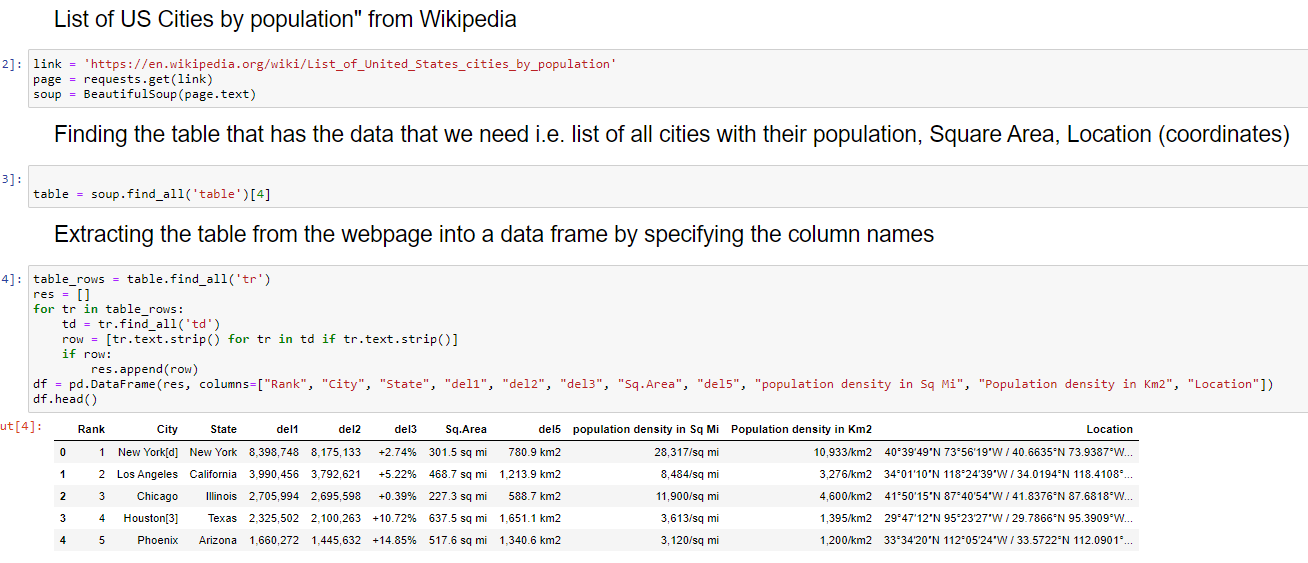
Using the above data, we will first select best city to proceed with based on the values like Population density, per capita income of the state, number of venues (as we are giving weights to each venue based on its category).

The first stage is to select the city in the USA based on available datasets, then to find the locality in the city which is influenced by the events and other above said factors.

**Methodology**

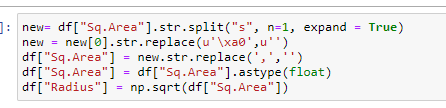
To identify optimal vicinity to start new Luxury SPA venture below methods are followed:

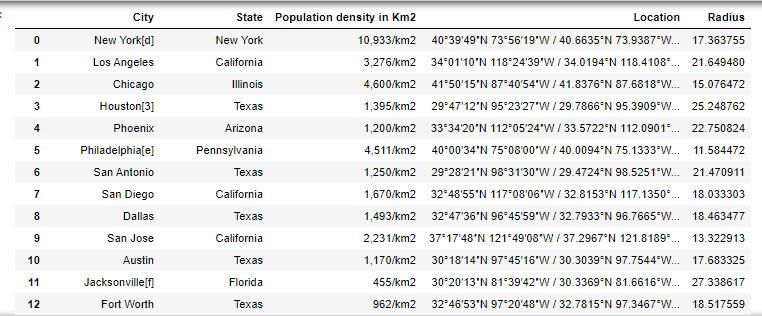
* The data pertaining to various factors are acquired from Wikipedia pages. The acquired data has been scraped using beautiful soup into panda data frames which are categorized under cities, per capital income, states, co-ordinates, area and density of population
* Appropriate clean-up and processing of Data frame has been done.
* Venue details of various cities are extracted using Foursquare API. Each category has been weighted based on our preferences. City with maximum weightage is selected for our venture.
* Venues are clustered based on category using K-Means algorithm. Co-ordinates of the location having maximum weightage is found out.

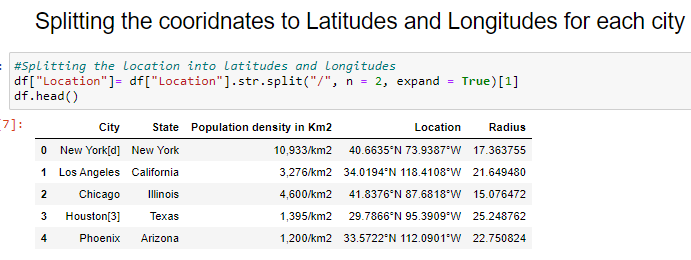


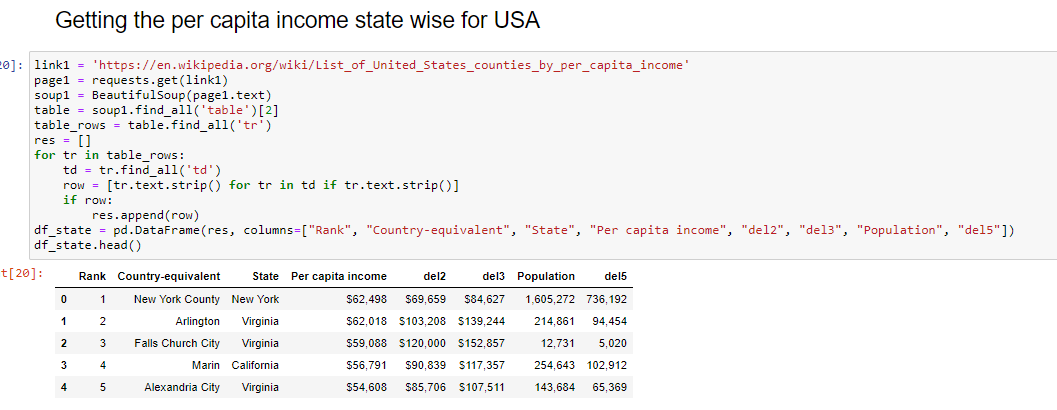


Finding the radius of each city with the help of Sq.Area, this step involves in preprocessing of the the column Sq.Area (changing its data type to float) then finding its square root

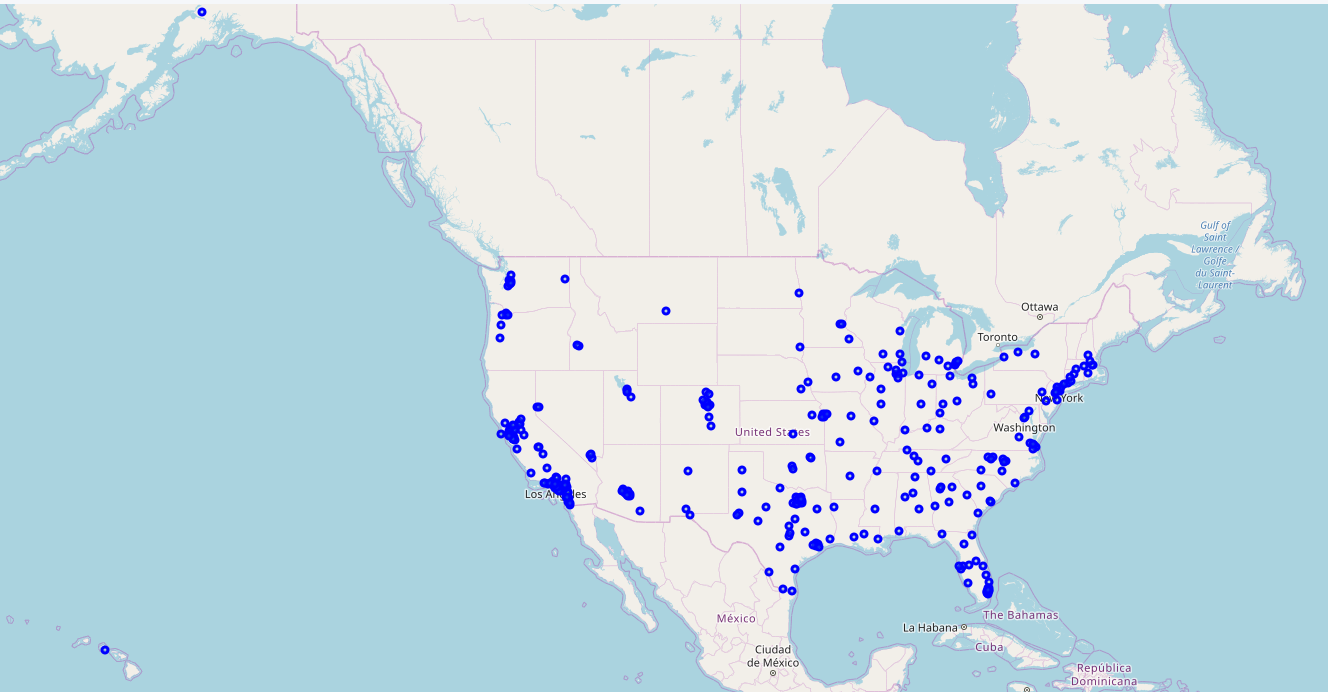




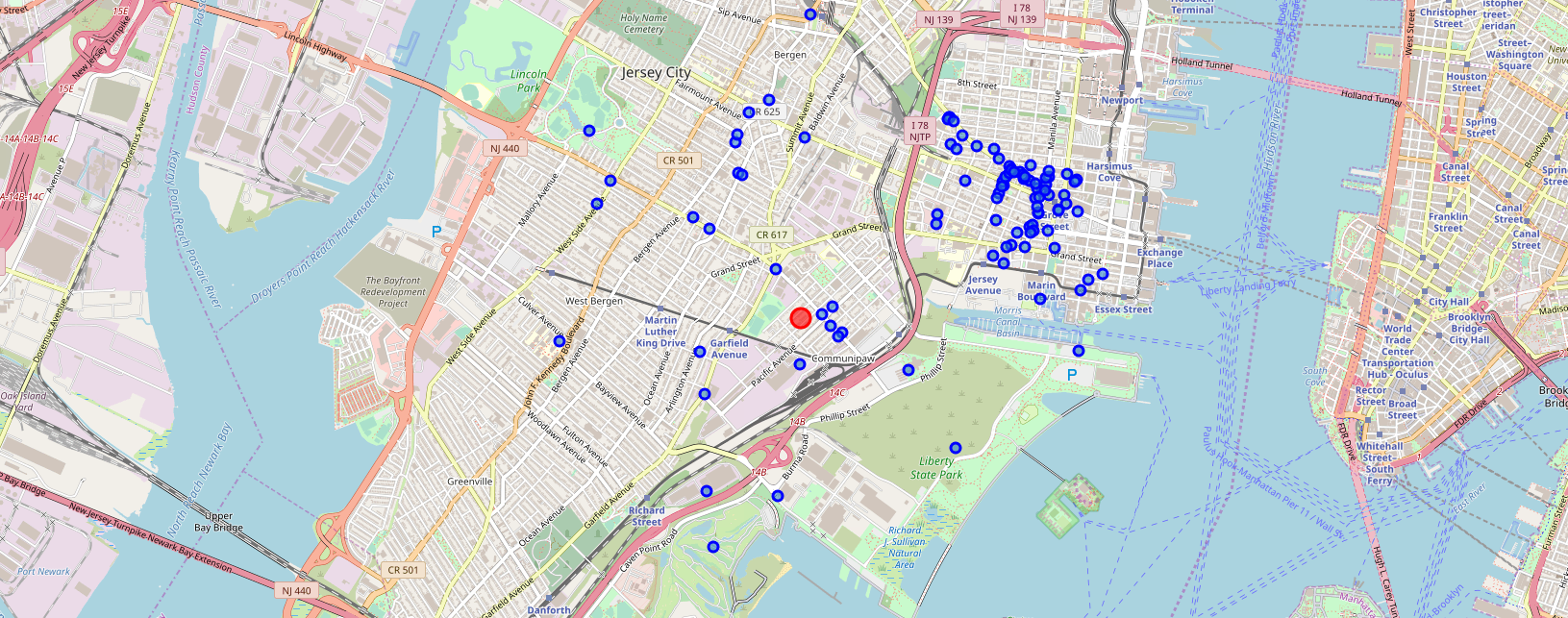




Plotting all the cities of USA that we have extracted from Wiki page, using their coordinates



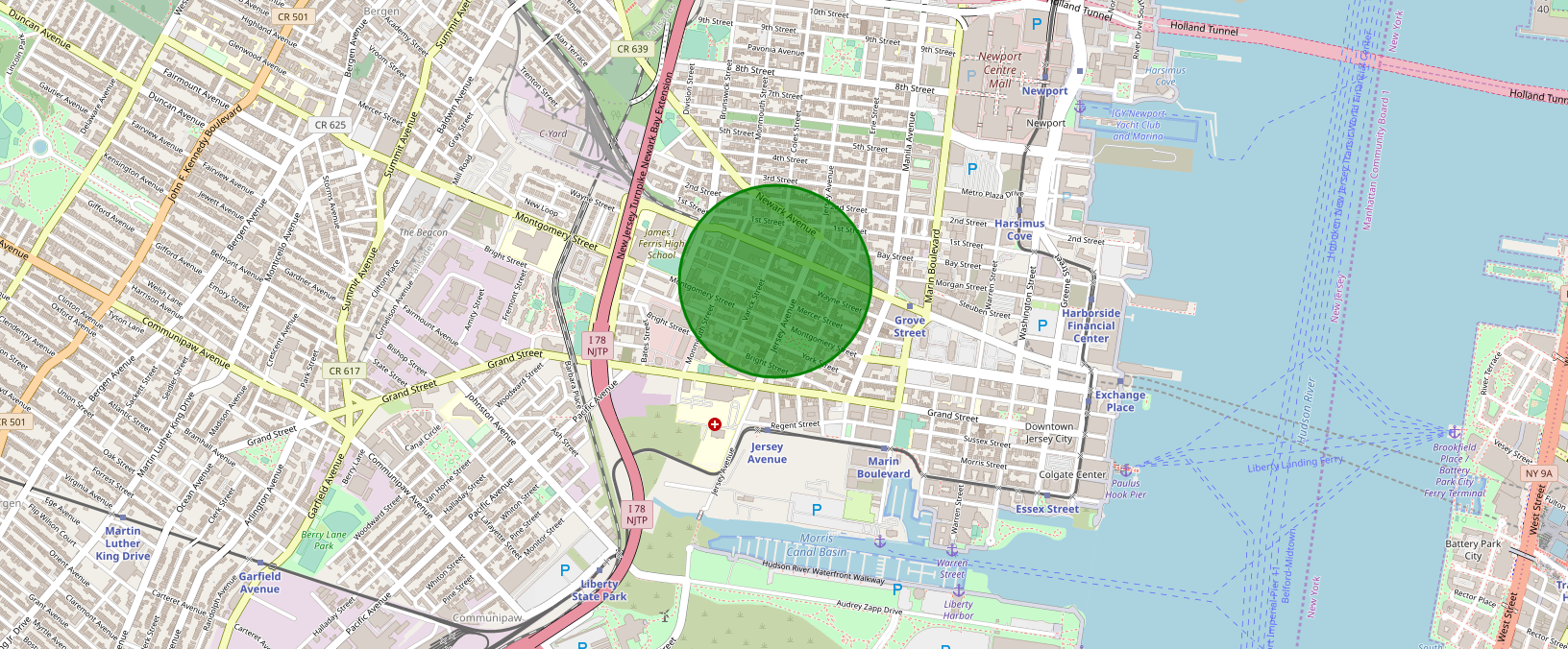
Plotting all the venues that we have got from the Four Square API



**Result/Conclusion:**

Based on the analysis we have done, the following is the result we have got:

Plot to show the location of final arcade that we are suggesting:



We have arrived at the suggestion of better spot in the best city of USA Based on the dependent factors and events we considered in this project, we suggest to start the Luxury spa in the location where there will be more visitors be it tourists, residents or professionals in that vicinity.

This analysis can be further expanded deeply by considering other factors like crime rate, floating populations, customer data similar to other businesses such as Gyms, health and wellness centers, Saloons etc.

### Resources :

* 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>
* Four Square API : [https://foursquare.com](https://foursquare.com/)