The Battle of Neighborhood

By: - Sarthak Mittal

Business Problem

An investor is looking to set up its fast food outlet in United States and therefore looking for the cities that has the high footfall and also has the high per capita income. Best locality has been defined on the basis of: -

➢Population density of a locality

➢Per capita income

➢Population of each location

➢Venues in each locality

Define the best locality in order to set up the fast food outlet

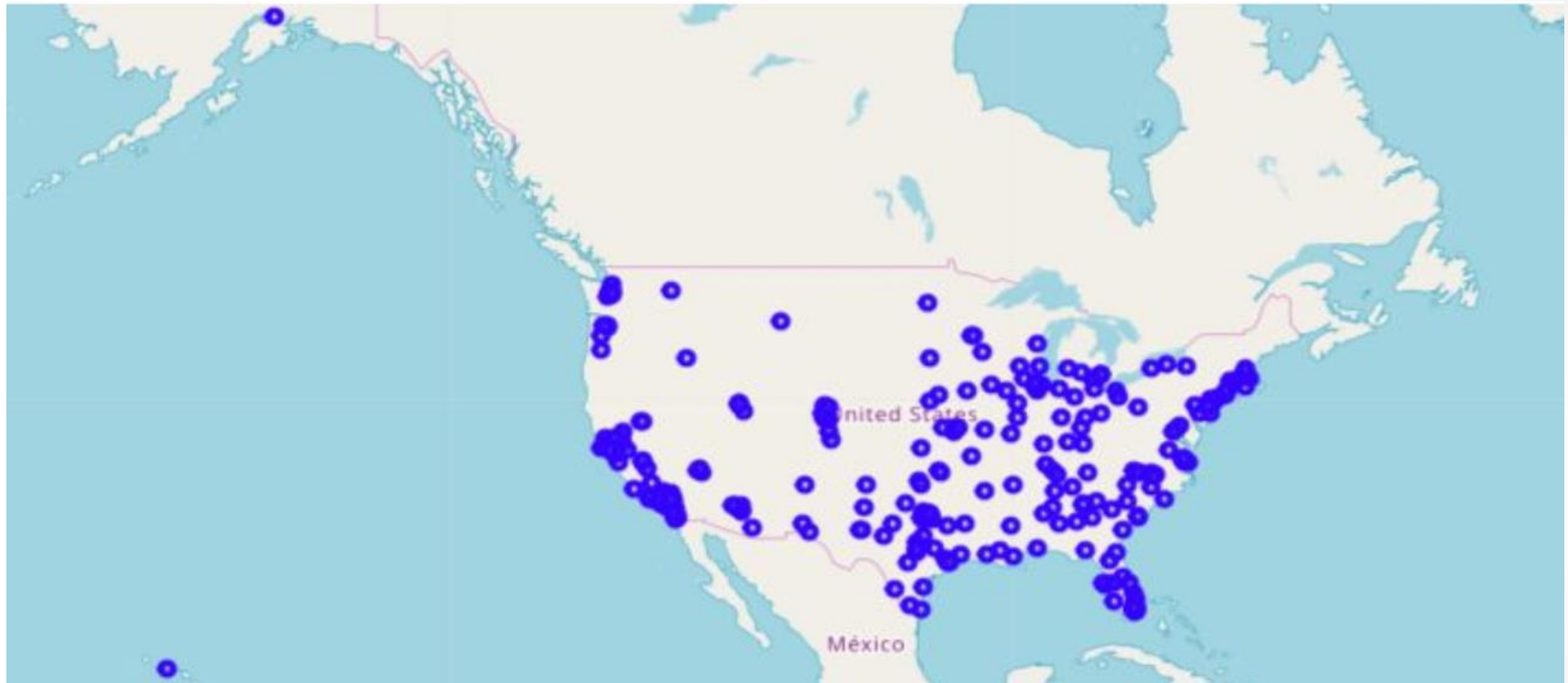# **Data**

➢List of all cities in US with population density and coordinates from the URL(https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population)

➢List of all cities in US with per capita income from the URL (https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income)

➢Using foursquare API to get: -

List of all venues in each city

List of all venues in each locality in selected city

Methodology

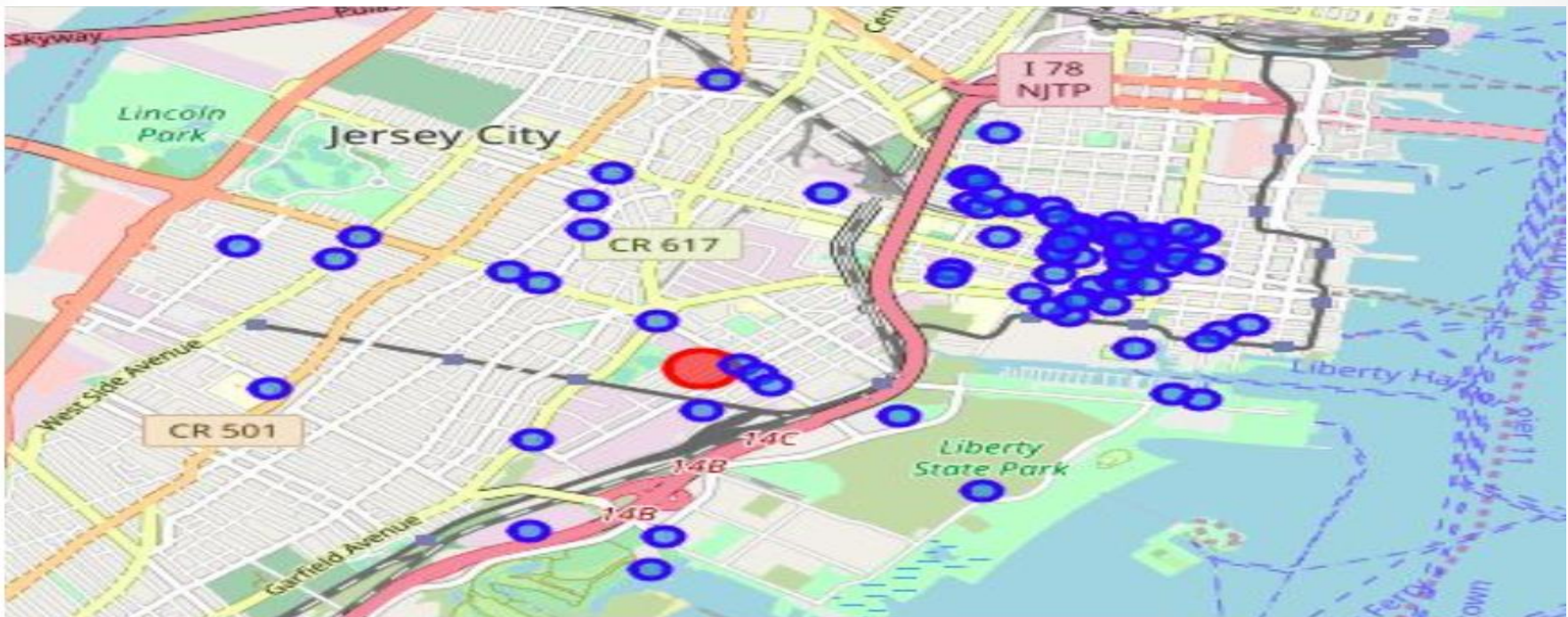Following are the steps we have to follow :

➤ The Wikipedia was scraped using the BeautifulSoup library to build a pandas dataframe listing the cities, states, coordinates, area and population density. The dataframe was cleaned and processed appropriately.

➤ The Wikipedia page was scraped using the BeautifulSoup library to build a pandas dataframe listing the cities, states and per capita income. The dataframe was cleaned and processed appropriately.

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

➤ Based on the categories of each venue, 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 fast food outlet.

Plotting of USA cities

Blue dot indicate the city center in map of US

Venues in Jersey City(Map)

On the basis of constraints Jersey City has been selected as the best city in order to set up the fast food outlet and therefore venues in Jersey City has been plotted on the map

Result

Green Circle indicates the best place to set up the fast food outlet

Recommendations/ 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.