

# The Dream of Jone Doe or The Battle of Neighbourhoods

BY BORISENKO EVGENY

# Introduction and Business Problem

- ▶ A Jone Doe won a lottery, and decided to make his Dream real. It is a Bar, **Best Bar**. And he started to search to of all the cities in United states. He defines a best locality based on the following constraints,
  - ▶ ● Population density of a locality
  - ▶ ● Per Capita Income
  - ▶ ● Population of each location
  - ▶ ● Venues in each locality
- ▶ We have to suggest the best locality to open Best Bar in the United States

# Data

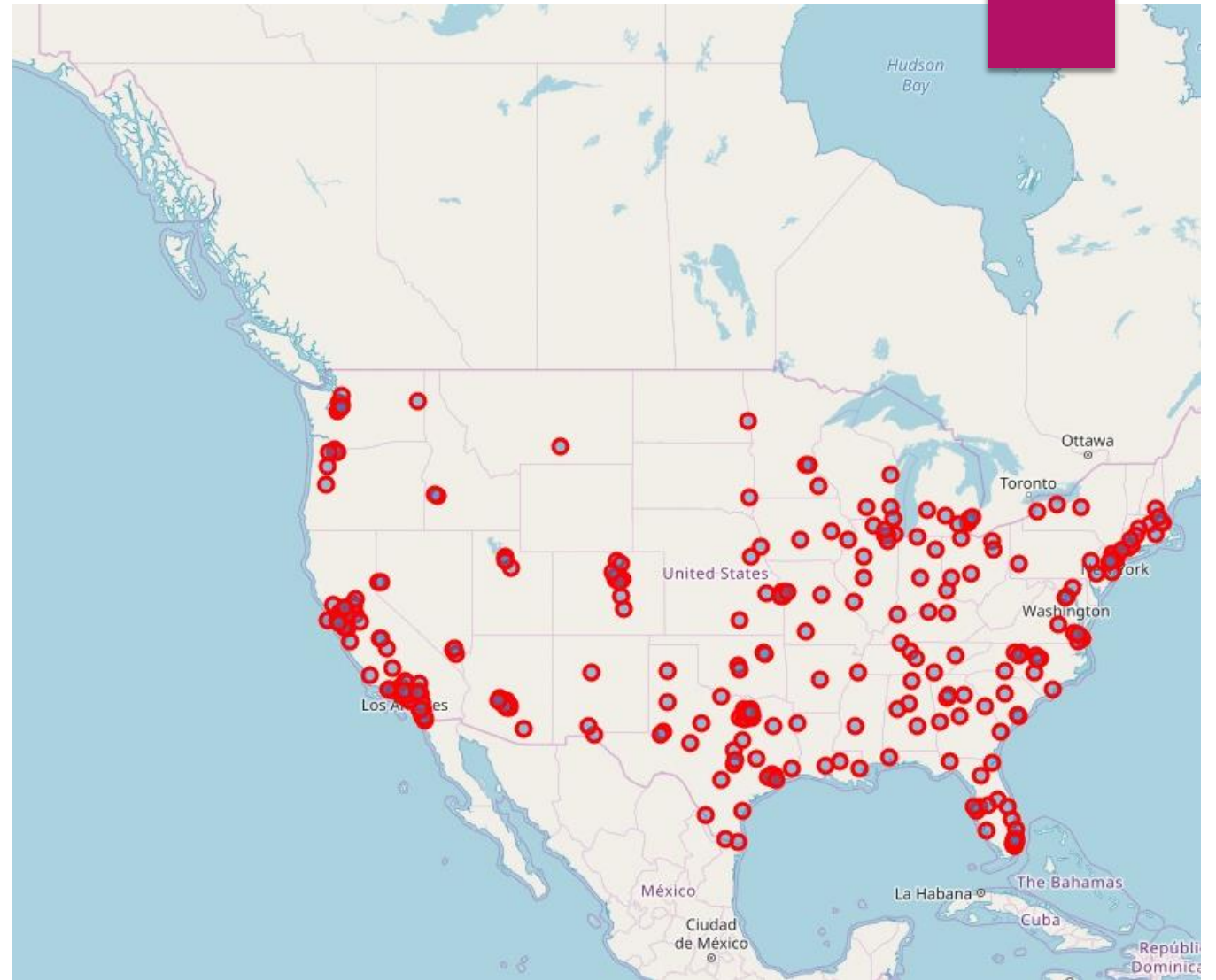
- ▶ ● List of all the cities in United States with population density and coordinates:  
[https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population)
- ▶ ● List of all the cities in United States with Per Capita Income :  
[https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_counties\\_by\\_per\\_capita\\_income](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

- ▶ In order to do the analysis and suggest the best location, following steps were fo
- ▶ In order to do the analysis and suggest the best location, following steps were followed:
- ▶ ● The Wikipedia page ([https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_population](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) and [https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_counties\\_by\\_per\\_capita\\_income](https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income)) was scraped using the BeautifulSoup library to build a pandas dataframe listing the cities, states, coordinates, area, per capita income and population density. The data frame was cleaned and processed appropriately.
- ▶ ● The Foursquare API is then used to get the venues in each city of United State, based on the categories, we have assigned weights to each of them and got the city that has the maximum weight.
- ▶ ● 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 open **Best Bar**:

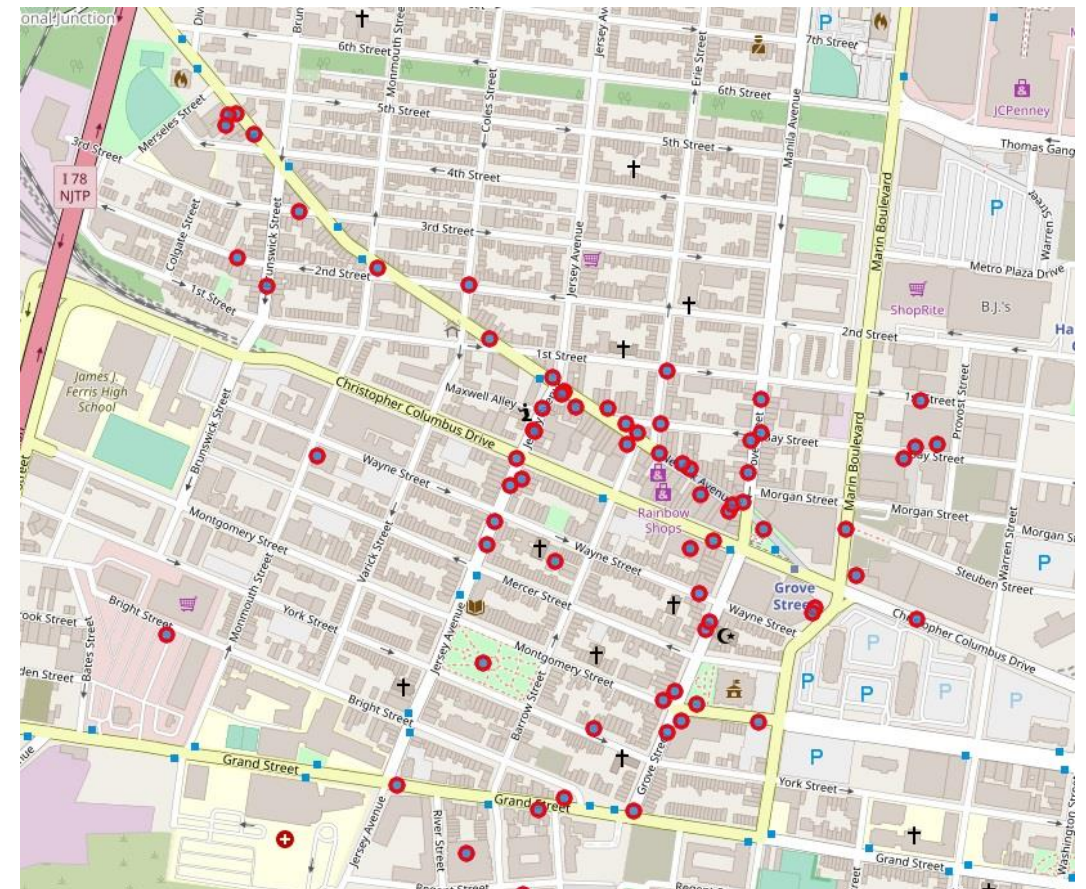
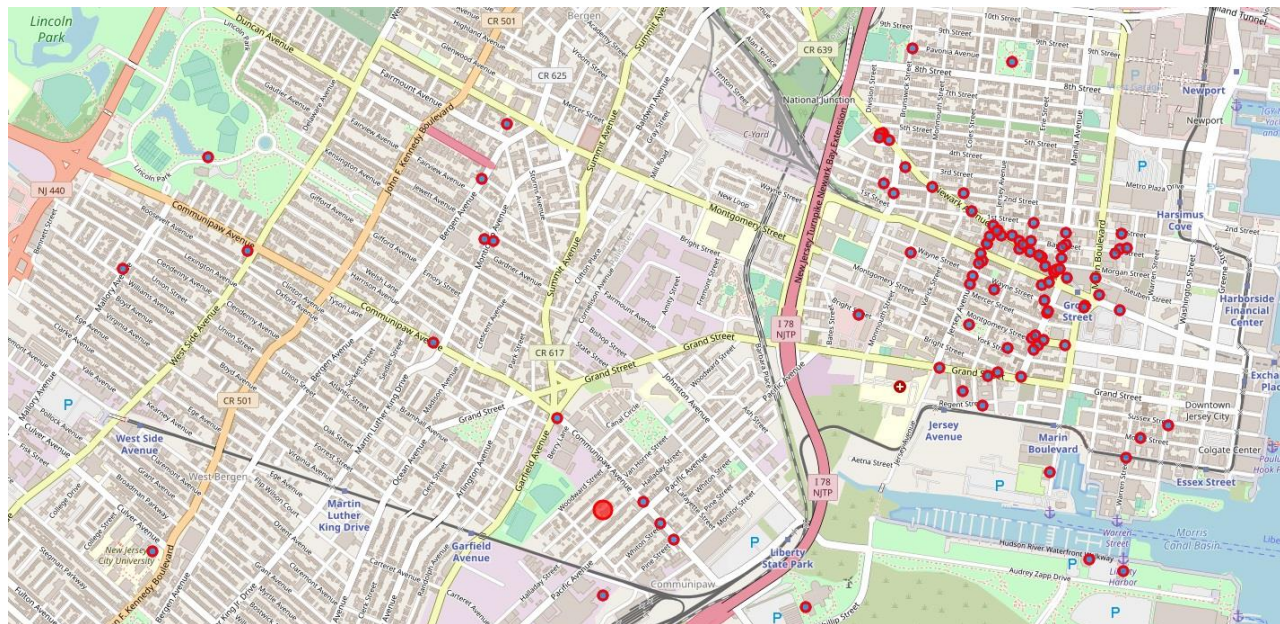
Plot of all cities of USA  
that were extracted:

► Red dot indicate the city center  
in the country map of United  
States.

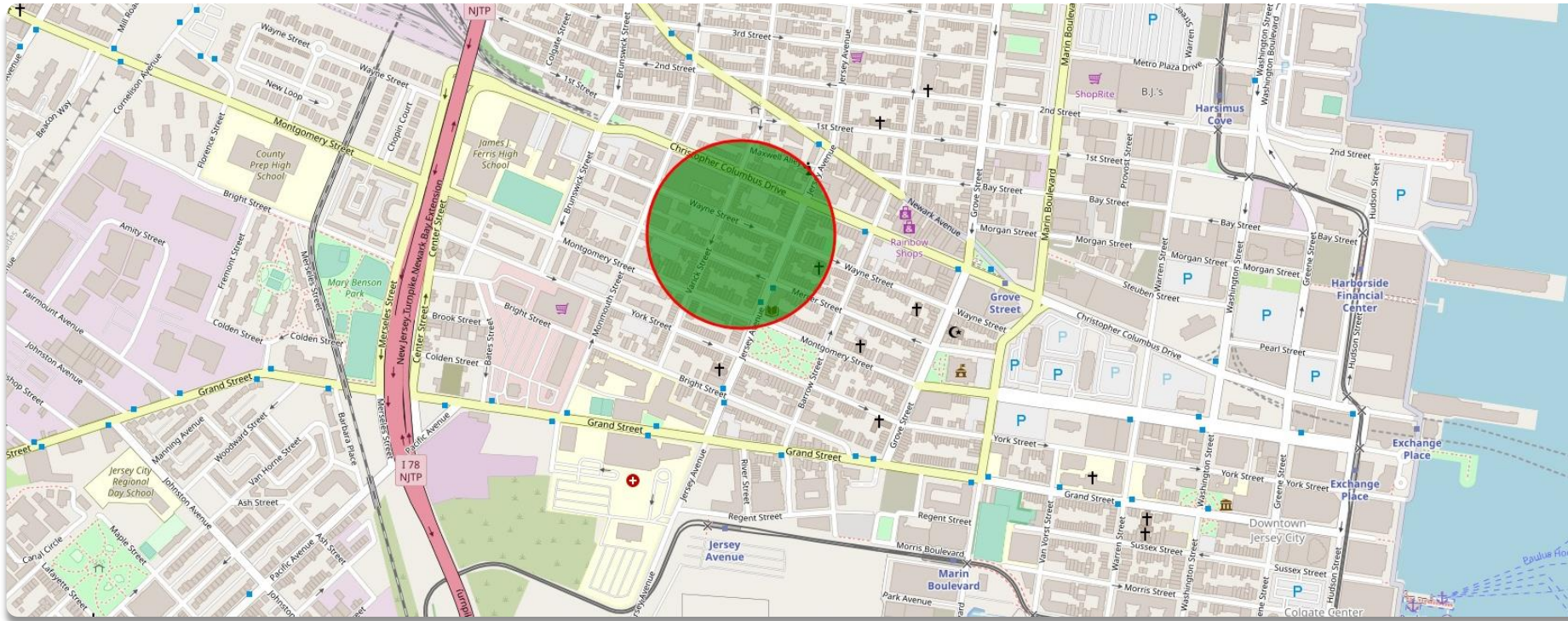




# Map of venues in the Jersey City:







## Result

As a result, we came to the conclusion that the best solution for John Doe would be to open his Best Bar in Jersey city, state of New Jersey, on the corner of Wayne Street and Varick Street.