The Battle of Neighbourhoods

- By Ankita kumari

Introduction and Business Problem

A CEO of a company is interested in starting a gaming arcade in the best locality of all the cities in United states.

"An arcade game or coin-op game is a coin-operated entertainment machine typically installed in public businesses such as restaurants, bars and amusement arcades. Most arcade games are video games, pinball machines, electro-mechanical games, redemption games or merchandisers.

Funspot Family Entertainment Center (or simply Funspot) is a video arcade which features one of the largest collections of late-1970s to mid-1980s games in the world. It is located in the village of Weirs Beach in Laconia, New Hampshire, 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

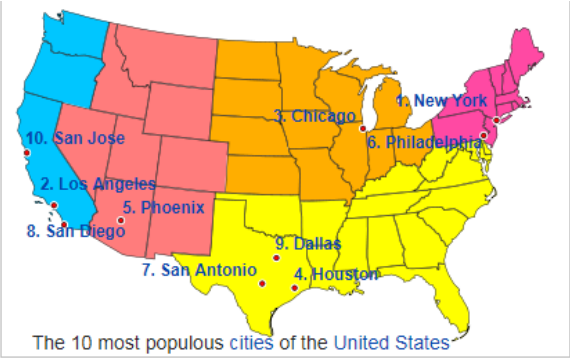
We have to suggest the best locality to setup a gaming arcade in the United States.

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

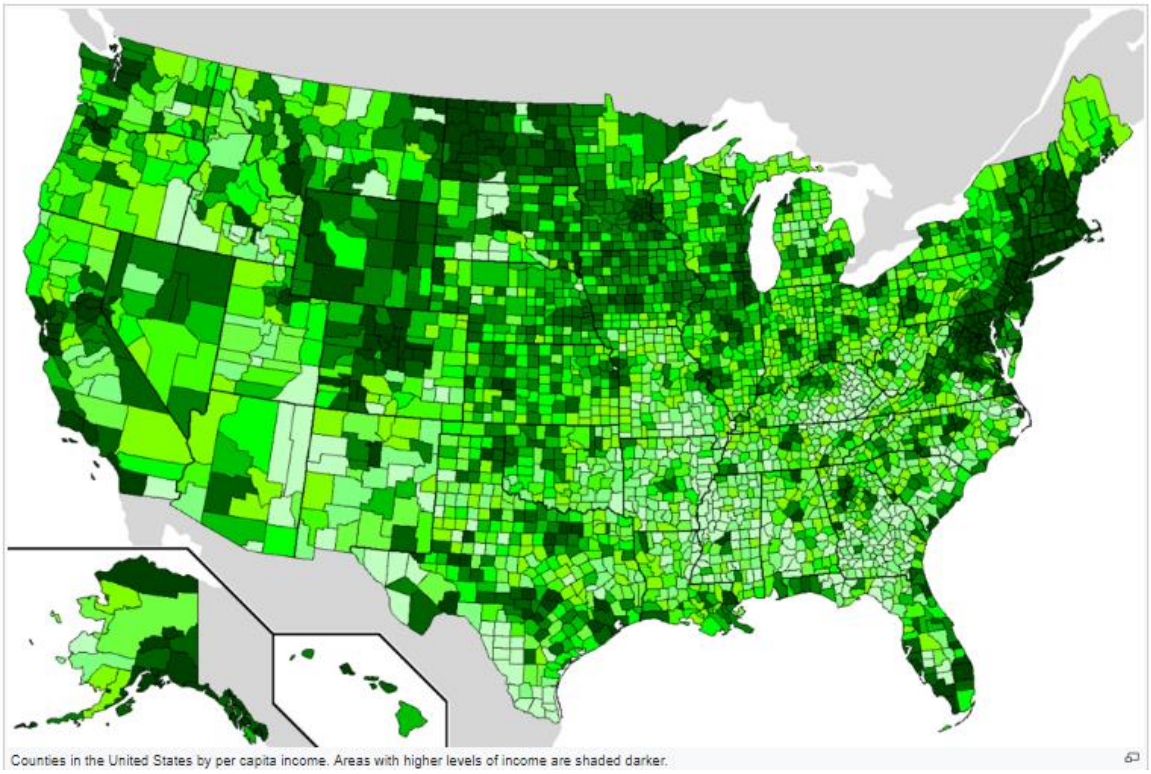
The following table lists the 314 incorporated places in the United States with a population of at least 100,000 on July 1, 2018, as estimated by the United States Census Bureau. A city is displayed in **bold** if it is a state or federal capital, and in *italics* if it is the most populous city in the state. Five states—Delaware, Maine, Vermont, West Virginia and Wyoming—have no cities with populations of 100,000 or more. The table below contains the following information:

1. The city rank by population as of July 1, 2018, as estimated by the United States Census Bureau[1]
2. The city name[1]
3. The name of the state in which the city lies[1]
4. The city population as of July 1, 2018, as estimated by the United States Census Bureau[1]
5. The city population as of April 1, 2010, as enumerated by the 2010 United States Census[1]
6. The city percent population change from April 1, 2010, to July 1, 2018
7. The city land area as of January 1, 2016[2]
8. The city population density as of July 1, 2016 (residents per unit of land area)[2]
9. The city latitude and longitude coordinates[2]



The 10 most populous cities of the United States

- 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



- 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 followed:

- The Wikipedia page (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) 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 of each venue as decided by the CEO, 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 setup a gaming arcade.

Plot of all cities of USA that were extracted:

Map of venues in the Jersey City:

Result

The circle indicates the *best place* to start an arcade in the **Jersey City**

Recommendation/ Improvements:

- In the Foursquare 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.
    - We can increase the Radius for more venue results from each city.
- 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 a better output.