

Battle of the Neighborhoods

Coursera capstone
By:Kislay Srivastava

Introduction

- I want to open up a game arcade in USA.
- I want to determine the best locality for setting up such an arcade.
- I need to have a clear understanding of the customers spending patterns in that locality etc.

DATA Used

- 1. The Wikipedia page
(https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population) .
- 2. The Wikipedia page
(https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income) .

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.