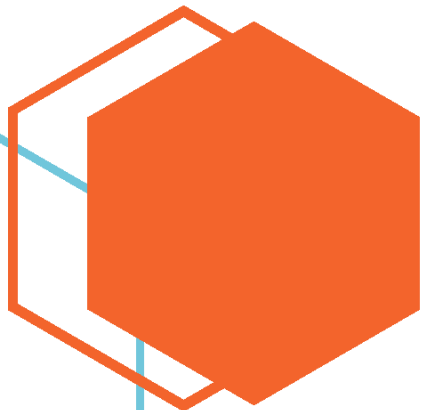# BATTLE OF NEIGHBORHOODS

FINAL REPORT

by Siddhesh Kankekar

[This report seeks to answer the question – Which City in United States of America is the best and most profitable for opening a Shopping Mall/ Casino with the help of Data science and Machine Learning.]

[Introduction]

Business Problem

Problem brought up by Client -
Which City and Locality in the City is best in United States of America for opening a business of Shopping Mall / Casino?

Analytic Approach
To solve this problem first we need to setup some benchmarks for the City to be eligible for Business. After getting the desired results of the City, we need to find the Locality in that City.

1. For discovering the City following benchmarks must meet
   - Nightlife
   - Services
   - Transport
   - Residence
   - Universities
   - Food
   - Entertainment
   - Travel
   - Outdoor Activities

2. Benchmarks for the Locality in City
   - Per Capita Income
   - Population
   - Population Density
   - Venues

[About the Client and the Problem]
. . .

[Client is a Successful Entrepreneur in Asia. He wishes to expand his business overseas. Since United States of America is known as the land of opportunity client wants to set up his business in USA by opening a Shopping Mall or a Casino. Since he's never been to USA before his problem is to choose the location which will be most beneficial for conducting the business.]

[DATA]

Data Requirements

Data that will be required for solving the problem -

+ List of all the cities in United States with population density and coordinates
+ List of all the cities in United States with Per Capita Income
+ List of all venues in each city
+ List of all venues in each locality in the selected city

Data Source

Wikipedia

https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income

https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

Four Square API

- For Venues in City and in Each locality in selected city.

Data Description

As defined by the United States Census Bureau, an "incorporated place" includes a variety of designations, including city, town, village, borough, and municipality. A few exceptional census-designated places (CDPs) are also included in the Census Bureau's listing of incorporated places. Consolidated city-counties represent a distinct type of government that includes the entire population of a county, or county equivalent. Some consolidated city-counties, however, include multiple incorporated places.

About the State income levels: State income levels and income data for the United States as a whole are included for comparison. Note that county-equivalents in Louisiana are called "parishes" and in Alaska are called in "boroughs," and also that in Alaska census areas in the Unorganized Borough are county-equivalents. For states where independent cities are county-

equivalents, the word "city" is included to identify the independent cities and to differentiate them from counties with identical names; the counties with the identical names have the word "county" following them. The word "county" is included in the names of counties that have names identical to the names of U.S. states or cities to differentiate them.

2

...

[Methodology]

1. Using Beautiful Soup Library Wikipedia pages containing information about Cities of USA by Population and Per Capita income are scraped into Pandas Dataframe.

2. Dataframe contained data about Cities, Coordinates, Area, Per capita Income and Population Density. Than Dataframe was Cleaned and Processed according to requirement of the problem to be solved. Proper benchmarks were set to obtain the best results.

3. The List of Venues in a City were obtained using Foursquare API and the city with maximum weight according to the Model is selected.

4. With the help of Unsupervised Machine Learning Algorithm (K Means Algorithm) Locality in the city is obtained where opening a Shopping Mall/Casino will be most benefitted.
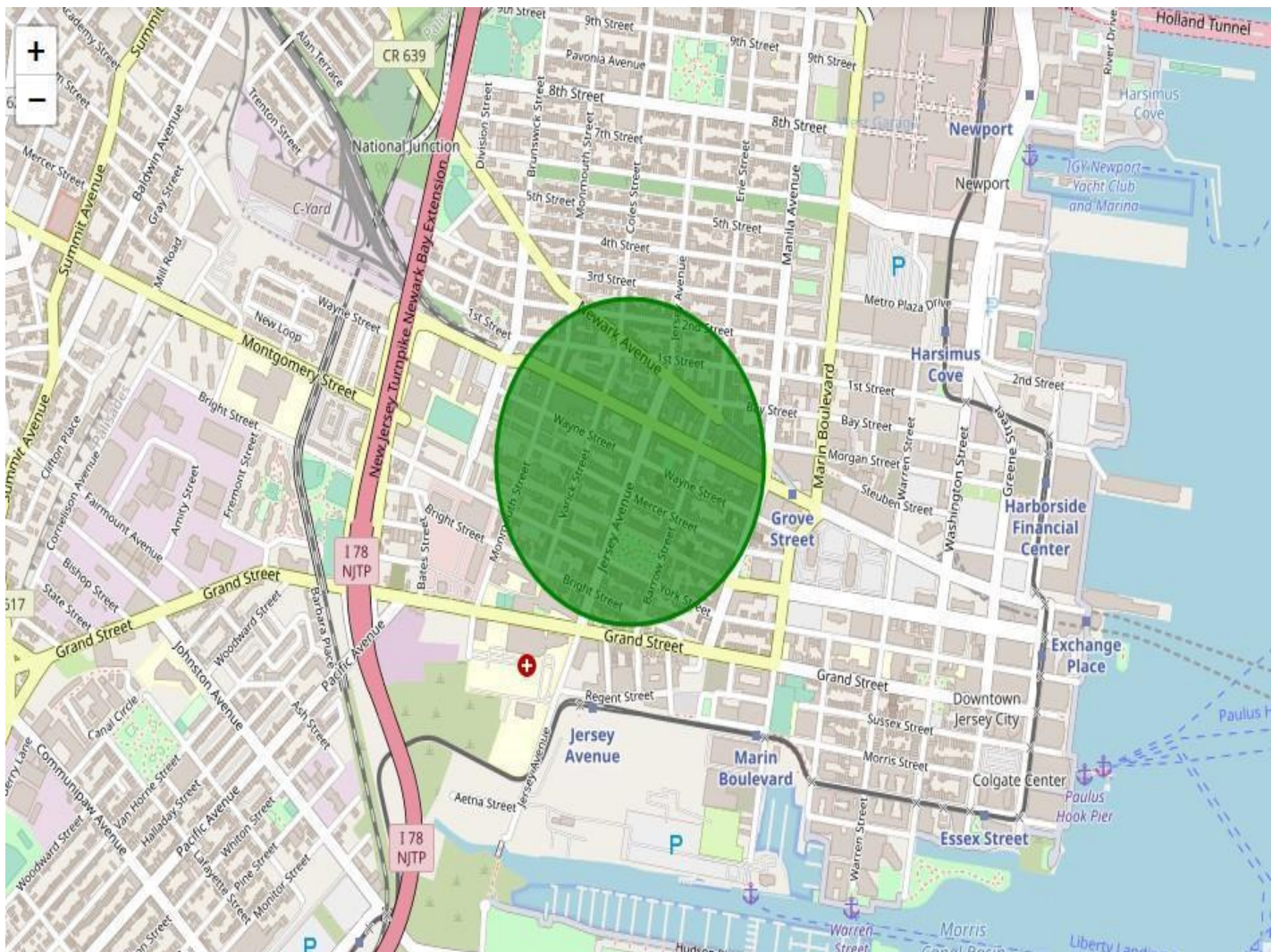
[Results]

Based on analysis we have done the results we obtain are – The plot shows the final location predicted by the Machine Learning Model –

For opening the Shopping mall/Casino

The Best city in USA – New Jersey

The best locality in New Jersey – Jersey Avenue

[Discussions and 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 a greater number of venue categories would actually yield better output.

[Conclusion]

Based on the given constraints, Shopping mall/Casino can be open in a place closer to the center of the circle to attract a greater number of diverse customers and get huge revenue.

[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_capit a_income

- **Four Square API: https://foursquare.com**