**Predicting the best place in the U.S. to build a new restaurant**

Sienna Spiglanin

May 12, 2020

**1. Introduction**

**1.1 Background** (Source: <https://pos.toasttab.com/blog/restaurant-management-statistics>)

This final project explores the best locations in the United States for a new restaurant. As of May 2019, the Restaurant Performance Index was 101.6. Since this is over 100, it’s considered a time of growth for restaurants. The projected annual sales for the restaurant industry are $863 billion. This comprises 4% of the United States’ gross domestic product.

**1.2 Problem**

A business owner is interested in opening a new restaurant in the United States. They want to know the best neighborhood to open it in. They are concerned with per capita income, location population, population density, and neighborhood venues to define the best neighborhood.

**2. Data**

In order to answer the above question, data on U.S. cities including boundaries, latitude and longitude coordinates, restaurants, venue information, U.S. population and income information is required.

**2.1 Data Sources**

United States per capita income and a list of all cities will be obtained from the data source:<https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income>

United States population density and coordinates and a list of all cities will be obtained from the data source: <https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population>

The FourSquare API will be utilized to obtain the venues in each U.S. city and their respective categories.

**3. Methodology**

In order to perform the analysis and suggest the best location, the following are the steps we will follow:

* The data from <https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population> was collected and scraped using the BeautifulSoup library to build a pandas data frame. This data frame lists the cities, states, latitude and longitude coordinates, area and population density. The data frame was cleaned and processed appropriately.
* Data from <https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income> was also scraped using the BeautifulSoup library to build a pandas data frame listing the cities, states and income per capita. The data frame was then cleaned and processed as well.
* The FourSquare API was then used to call the venues in each U.S. city.
* Based on each venue’s category, we assigned weights to each and got the city with the maximum weight.
* Once that city is finalized, we used the FourSquare API to obtain the venues within that city and assign weights to each category once again.
* Then, we use K means to cluster the venues based on their category and get the coordinates of the cluster with the maximum weight. This is the preferred location to build a new restaurant.

**4. Results**

From the analysis we have completed, our results include the following:

*Plot showing the restaurant location we are suggesting:*

**A close up of a map

Description automatically generated**

**5. Discussions and Further Improvements**

* In the FourSquare API, we’ve queried the venues of a city by specifying a LIMIT and radius. Since using a free FourSquare account allows for less API call, we’ve chosen a lower LIMIT. For more accurate results, the LIMIT can be increased.
* In the venue categories, we’re only choosing a few to give weights and identify the best cluster. Weights must be assigned relatively for each venue category. Then, considering more venue categories would yield an even better output.

**6. Conclusions**

From our results, a new restaurant should be built in Chicago, Illinois. In the above figure, the coordinates in the center of the green circle should attract the most customers and generate a high revenue. This place is located off the Stevenson Expressway. Unlike the majority of locations in Chicago, there are few other restaurants in this location. With the high foot traffic and potential customers with less competition, it makes sense that this was the result.