Project for Predicting location preference for a fine dining Restaurant by Pooja Prasad Srivastava

July 11th, 2019

1. Introduction

This project is created to give recommendation on most profitable area to setup a fine dining restaurant.

* 1. Problem Definition:

Restaurants open and shut down, people lose thousands or millions dollars of their investments due to a wrong location selections of the restaurants. A person could have a great concept in mind, best cook and floor staff money could buy but his/her restaurants could still fail if the location is not selected with care.

In today’s world customer is the king. The disposable income of any household is at an all time high. Today’s consumer has a high appetite to spend money for the right kind of goods and services. It is the restaurant owners responsibility to make the restaurant accessible to the right kind of customer to ensure the success of a restaurant.

A fantastic restaurant in a wrong location will fail to attract the right kind of customers and/or create the right image perception and thus not sustain.

Given the problem statement, now an entrepreneur with decent amount of capital to invest wants to setup a Spanish fine dining restaurant in Chicago. He wants to identify a location which would be most suitable for his business.

* 1. Background:

To identify an area which would be suitable for a fine dining Spanish restaurant we need to evaluate the selected area on the following aspects:

* High spendable income: Identify areas with high income families.
* Population age: People with age between 18-64 years are more likely to go to fine dining restaurants.
* Literacy rate: People with high level of education are more likely to be aware about availability of various cuisines.
* Population Lifestyle: Area with the people who have the ability and desire to finer things in life. This can be judged if it have places like other specialty restaurants, boutique stores, health and fitness centers, yoga studios, etc.
  1. Interest

Any entrepreneur wanting to open a fine dining restaurant in Chicago. Currently the project is limited to Chicago. A similar study can be done for any location using the same steps

1. Project Data acquisition and cleansing:
   1. Data Source:

Data about Chicago socio economic indicators is available on the Chicago Data Portal. (https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2/data )

The dataset downloaded from Chicago Data Portal contained a selection of six socioeconomic indicators of public health significance and a “hardship index,” by Chicago community area, for the years 2008 – 2017. The indicators are the percent of occupied housing units with more than one person per room (i.e., crowded housing); the percent of households living below the federal poverty level; the percent of persons in the labor force over the age of 16 years that are unemployed; the percent of persons over the age of 25 years without a high school diploma; the percent of the population under 18 or over 64 years of age (i.e., dependency); and per capita income.

* 1. Data Cleaning:

The data available where in % for all indicators showing the poverty and illiteracy in the Chicago area. These % had to be converted to show the prosperity.

All the records with no data in the vital columns where removed. Un-necessary columns where removed.

* 1. Feature Selection:

After data cleaning following socio economic indicators were consider to identify the area/community most suitable for a fine dining restaurants:

* Percent Households Above Poverty
* Percent Aged 25+ With High School Diploma
* Percent Aged Between Age 18 And 64
* Per Capita Income
  + 1. Data Description:

|  |  |  |
| --- | --- | --- |
| **Feature Name** | **Sample Values** | **Data Type** |
| Community Area Number | 1,2,3.. | Int |
| Community Area Name | A.B.C… | Alpha Numeric |
| Percent Households Above Poverty | 76.4, 52.3, …. | Float |
| Percent Aged 25+ With High School Diploma | 76.4, 52.3, …. | Float |
| Percent Aged Between Age 18 And 64 | 76.4, 52.3, …. | Float |
| Per Capita Income | 23939, 764833 | Int |

* + 1. Sample Data:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Community Area Number | COMMUNITY AREA NAME | PERCENT HOUSEHOLDS ABOVE POVERTY | PERCENT AGED 25+ WITH HIGH SCHOOL DIPLOMA | PERCENT AGED BETWEEN AGE 18 AND 64 | PER CAPITA INCOME |
| 1 | Rogers Park | 76.4 | 81.8 | 72.5 | 23939 |
| 2 | West Ridge | 82.8 | 79.2 | 61.5 | 23040 |
| 3 | Uptown | 76 | 88.2 | 77.8 | 35787 |
| 4 | Lincoln Square | 89.1 | 86.6 | 74.5 | 37524 |
| 5 | North Center | 92.5 | 95.5 | 73.8 | 57123 |

The socio economic indicators will be analyzed to identify the top community.

Once the top community is identified, nature of all the business establishments in it will be explored to narrow down on the location most suitable for a fine dining Spanish restaurant.

Project would use Foursquare API to get information on all the venues/business establishments in the community:

* Venue ID
* Venue Name.
* Venue Category of business
* Venue Latitude
* Venue Longitude
  + 1. Data Description:

|  |  |  |
| --- | --- | --- |
| **Feature Name** | **Sample Values** | **Data Type** |
| Venue ID | 4aa01fd8f964a5206a3e20e3.. | Alpha Numeric |
| Venue Name | A.B.C… | Alpha Numeric |
| Venue Category | Restaurant, Gym, Bar, …. | Alpha Numeric |
| Venue Latitude | 76.4, 52.3, …. | Float |
| Venue Longitude | 76.4, 52.3, …. | Float |

* + 1. Sample Data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **name** | **categories** | **lat** | **lng** | **id** |
| Kiki's Bistro | French Restaurant | 41.89895 | -87.6359 | 4aa01fd8f964a5206a3e20e3 |
| BIG & little's Restaurant | Fast Food Restaurant | 41.89836 | -87.63709 | 4aef2171f964a520ded521e3 |
| Hi-Fi Personal Fitness | Gym | 41.89737 | -87.63734 | 4b9e1e3cf964a520accb36e3 |
| Headquarters Beercade | Bar | 41.89722 | -87.63488 | 52f139a1498e481bc2d24e0d |
| Fitness Formula Club (FFC) | Gym / Fitness Center | 41.90141 | -87.63174 | 4b257751f964a5205d7224e3 |

1. Methodology:
   1. Libraries:

Following libraries were used in the project:

* pandas: Data frames, json\_normalize
* request
* matplotlib: cm, colors
* sklearn: cluster, datasets.samples\_generator
* json
* geopy.geocoders: Nominatim
* folium
  1. Identification of Top community in Chicago land:
     1. Background:

The Socio economic status indicates a communities access to collectively desired resources, be they material goods, money, power, friendship networks, healthcare, leisure time, or educational opportunities. And it is access to such resources that enable individuals and/or groups to prosper in the social world.

Social hierarchy, or stratification, appears to be intuitively impact almost everyone, everything and every aspect of life.

Over the past 60 years, restaurants gained that significant food dollar share by adapting to consumer [tastes and behaviors](https://rewardsnetwork.wpengine.com/blog/understanding-dining-decisions-helps-increase-restaurant-sales/).

In 1955, only 25 percent of American dollars spent on food was spent at restaurants. Today, the restaurant industry’s share of the American food dollar has risen to 47 percent and continues to grow. (*Source:* [*https://smallbusiness.chron.com/economic-trends-restaurants-71804.html*](https://smallbusiness.chron.com/economic-trends-restaurants-71804.html)*)*

The sector of restaurants is positioned to continue growing. Many restaurants are slotted between fast food and casual dining restaurants . Also the demand for high end fine dining restaurants are on rise with more ethnic foods becoming mainstream and high-end chefs focusing on the sector. The high end fine dining sector has even more upward pressure

The National Restaurant Association reports that one of the key drivers for growth of the restaurant industry is the Socio economic status of Americans.

* + 1. Criteria to evaluate socio economic of a community:

According to various studies residents in the immediate area are typically the most important demographic for restaurants. Therefore the socio economic indicators of the resident in the immediate community will determine the suitability of that community for the fine dining restaurant in question.

Project would evaluate the following criteria:

* Per Capita Income: Per capita income(PCI) or average income measures the average income earned per person in a given area (city, region, country, etc.) in a specified year. It is calculated by dividing the area's total income by its total population. Project selects the community with high PCI as they are most affluent and high disposable income
* Population Structure: Population structure is the composition of a given population, which is broken down into categories such as age. Project is targeting on the community with high population between the age of 18 and 64 years as they are more likely to appreciate a fine dining cuisine and pay extra for the right ambiance and services.
* Economic Health: Project used ‘Percent Households Above Poverty’ indicator to find communities economic position. The Higher the % , more affluent the community is
* Literacy Rate: Higher educated people are more aware and interested in experimenting multiple cuisine. Project used ‘Percent Aged 25+ With High School Diploma' to determine the same.
  + 1. Prioritizing socio economic indicators:

The indicators are sorted is the following priority order:

1. Per Capita Income
2. Percent Aged Between Age 18 And 64
3. Percent Households Above Poverty
4. Percent Aged 25+ With High School Diploma

Based on the above prioritization one top community was selected.

Note: The project can be expanded to include top 2 to 3 communities.

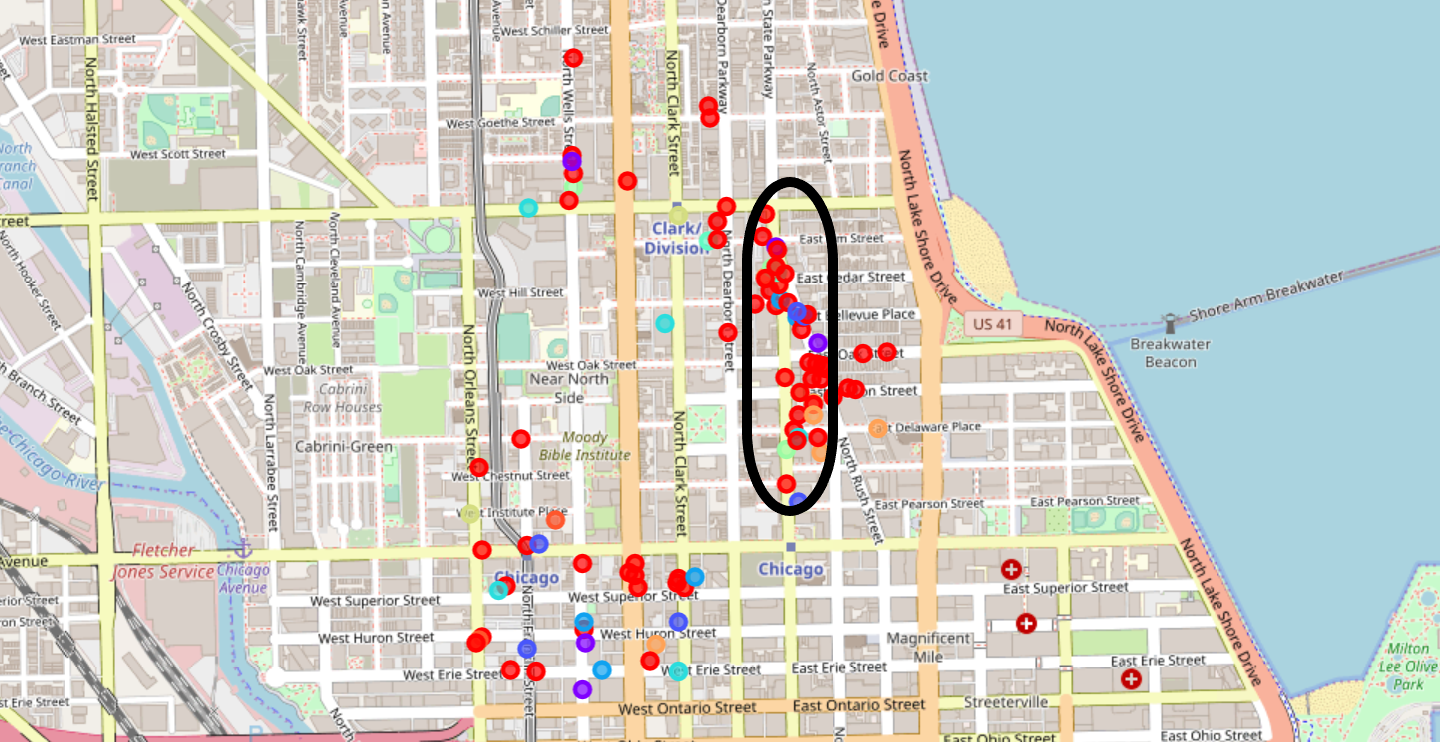
* 1. Clustering and analyzing Venues in selected top community:
* Used geolocator to seek the latitude and longitude of the top community
* Foursquare api was used to get the details of all the business establishment in the community with their geo-locations (lat and long). Four square API rutrned a JSON file.
* Analyzed the JSON and retrieved the relevant part of Json file which gave the listing of all the venues.
* Converted the records into data frame
* Cleansed the data and columns to to seek only the required data columns and relevant column names.
* Used data slicing and dicing to divide the venues in clusters. K-mean was used to cluster the venues based on their categories.
* The clusters were studied to identify locations having the most number of venues offering lifestyle related services. Life style services are like fitness centers, yoga centers, boutiques , etc. Those clusters are preferred as people around those business establishments already have more appetite for high end services like that of fine dining restaurants. Also the look and feel of these places would be more in accordance to the affluent image which needs to be created for a fine dining restaurant.
* The venues were plotted on the maps to see the concentration of the selected clusters on the map of Chicago. Location with highest population of the venues falling in the cluster with lifestyle service providers is identified and recommended for the opening of the fine dining restaurant.

1. Results:

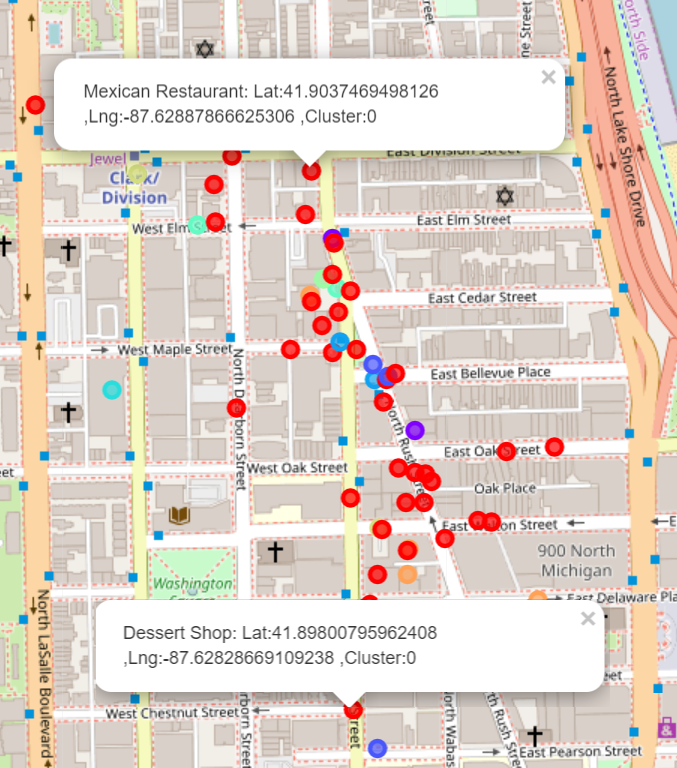
* The top community in Chicago land based on the socio economic indicators is : Near North Side, IL
* Latitude and longitude for North Side is 41.9000327 and -87.6344975.
* The top 100 venues is specific area around this community where analyzed.
* Following unique categories of the venues were identified:

French Restaurant', 'Fast Food Restaurant', 'Gym', 'Bar','Fitness Center', 'Paper / Office Supplies Store', 'American Restaurant', 'Pizza Place', 'Beer Bar', 'Cycle Studio', 'Dance Studio', 'Yoga Studio', 'Hotel', 'Restaurant', 'Spa', 'Steakhouse', 'BBQ Joint', 'New American Restaurant', 'Massage Studio', 'Lounge', 'Dessert Shop', 'Cocktail Bar', 'Grocery Store', 'Boutique', 'Optical Shop', 'Vietnamese Restaurant', 'Hotel Bar', 'Breakfast Spot', 'Coffee Shop', 'Italian Restaurant', "Men's Store", 'Seafood Restaurant', 'Pub', "Women's Store", 'Diner', 'Department Store', 'Farmers Market', 'Burger Joint', 'Korean Restaurant', 'Clothing Store', 'Latin American Restaurant', 'Tapas Restaurant', 'Café', 'Health Food Store', 'Juice Bar',‘Cupcake Shop', 'Bakery', 'Mexican Restaurant','Furniture / Home Store', 'Salad Place', 'Sushi Restaurant', 'Speakeasy', 'Brazilian Restaurant', 'Bookstore', 'Salon / Barbershop', 'Shoe Store'

* The venues where grouped in 10 clustered based on the categories. The cluster with most numbers of life style service providers and high end restaurants where considered: Cluster 0, Cluster 3, Cluster 4 and Cluster 7.
* Clusters (Each cluster with different color marker) where populated on the map of Chicago to find the most densely populated location for the above mentioned clusters. (Circled in the below fig)



* The lat and lng for the densely populated area was identified and that would be the recommended location



1. Conclusion:

Based on the analysis of the project any location between the following Latitude and Longitude would be great for opening a fine dining Spanish Restaurants:

Lat: 41.9037469498126 and Lng:-87.62887866625306

Lat: 41.8800795962408 and Lng:-87.62828669109238