

# Final Capstone Project – Opening a new Pizza Place near Chicago

## Introduction

The primary objective of this project is to locate the ideal area to start a Pizza Place in one of the many suburbs of Chicago. Chicago is the most populous city in Illinois, and the third most populous city in the United States. Chicago is an international hub for finance, culture, commerce, industry, education, technology, telecommunications, and transportation. Since it's very expensive to live in the city, many people working in the city choose to live in the suburbs of Chicago. Some of the nearby suburbs such as Aurora, Naperville, Joliet etc have transits to Chicago so that people can get to the city for work. These places have phenomenal malls, supermarkets, restaurants etc.

Chicago already has lots of pizza places and the cost of setting up a new restaurant in Chicago will be more expensive compared to its nearby cities. Rather than opening a Pizza place in Chicago, one can look into opening one in the nearby expanding cities. First, we have to find the nearby cities and number of pizza places already in these areas. Then we have to look into how populated these cities are. Opening a restaurant in a place that is less populated may not bring much revenue, whereas opening one in a populated area will bring more competition to business.

## The Target Audience

The target audience for this project is any business owner who wishes to open a Pizza Place near Chicago. Starting a new business in near by places of Chicago will be less expensive than starting one in the main city. For example the rent for a building in Chicago will be more compared to its nearby cities.

## Data

In order to find the ideal location to open a new Pizza Place we will need the following data,

- List of the neighboring cities of Chicago with the population.  
The above data is obtained from Wikipedia page [https://en.wikipedia.org/wiki/Chicago\\_metropolitan\\_area](https://en.wikipedia.org/wiki/Chicago_metropolitan_area). This page gives the nearby cities of Chicago with their population. Using data scrapping we get the city names and the population of each city. This will give us an idea about how thickly or thinly are these places populated.
- Venue Data of the above cities  
We will use Foursquare API for getting this information. By using the API we can get all the venues for the above cities and conduct an analysis on the Pizza Places in each of the cities.

## Methodology

The first step done for the analysis is getting the information regarding Chicago suburbs. This was available on Wikipedia website. The required data was extracted from the website using Python request and BeautifulSoup package. The extracted data had city names and population in bracket. So created a new data frame with columns and population as two columns. This will give us an idea about how populated is a city. As mentioned earlier starting a new pizza place in a less populated area will not bring much revenue. Below is the sample data frame that was created after extracting and cleaning the data,

suburb\_population\_format

	suburb	population
0	Aurora, Illinois	198870
1	Naperville, Illinois	149196
2	Joliet, Illinois	148227
3	Elgin, Illinois	111401
4	Kenosha, Wisconsin	101124
5	Waukegan, Illinois	85720
6	Cicero, Illinois	79943
7	Bolingbrook, Illinois	76468
8	Arlington Heights, Illinois	74593
9	Hammond, Indiana	74423
10	Gary, Indiana	74879
11	Evanston, Illinois	72836
12	Schaumburg, Illinois	72387
13	Palatine, Illinois	67143
14	Skokie, Illinois	62088
15	Des Plaines, Illinois	58243

Now we will have to get the geographical coordinates of each of the city. To do so, we will use the Geocoder package that will help us get the latitude and longitude of each of the city. For this we loop through the above data frame and get the coordinates. This is saved into a new data frame as below

	Suburb	Latitude	Longitude
0	Aurora, Illinois	41.757170	-88.314754
1	Naperville, Illinois	41.772870	-88.147928
2	Joliet, Illinois	41.526360	-88.084021
3	Elgin, Illinois	42.037260	-88.281099
4	Kenosha, Wisconsin	42.584677	-87.821226
5	Waukegan, Illinois	42.363633	-87.844794
6	Cicero, Illinois	41.845540	-87.754020
7	Bolingbrook, Illinois	41.700330	-88.071771
8	Arlington Heights, Illinois	42.081156	-87.980216
9	Hammond, Indiana	41.583366	-87.500043

Next we will use the Foursquare API to explore each of the places in the above dataframe and save the data . Below is the resulting data

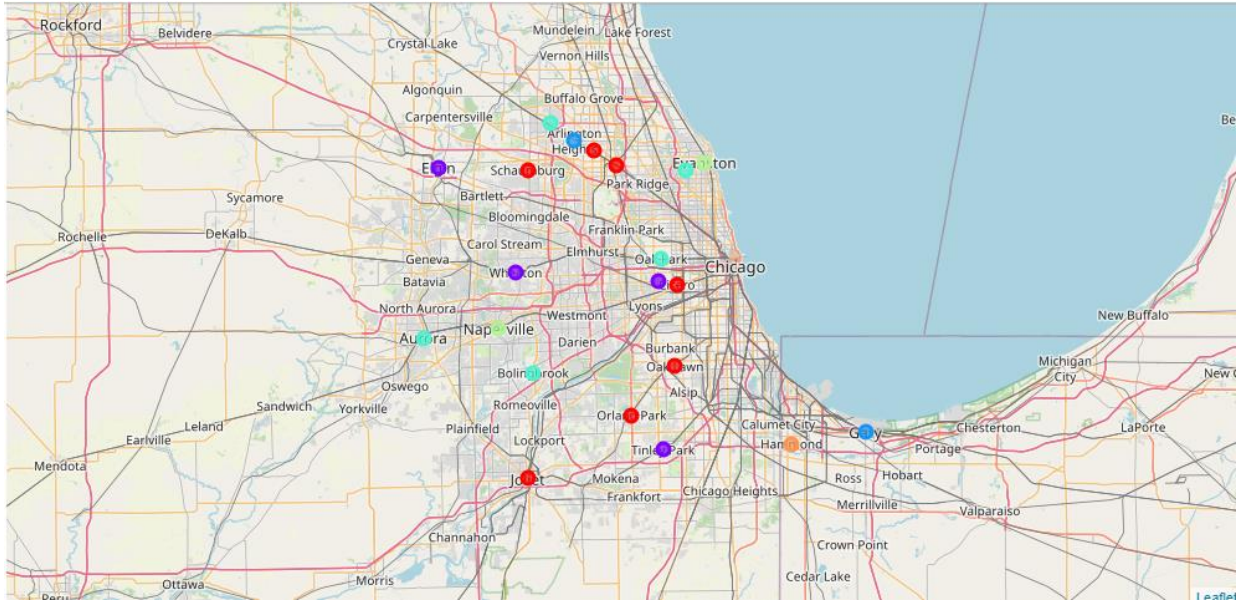
	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aurora, Illinois	41.75717	-88.314754	Paramount Theatre	41.757414	-88.314938	Theater
1	Aurora, Illinois	41.75717	-88.314754	Gillerson's	41.759606	-88.315031	Pub
2	Aurora, Illinois	41.75717	-88.314754	Endiro Coffee	41.759559	-88.314812	Café
3	Aurora, Illinois	41.75717	-88.314754	Tecalitlan Restaurant	41.756192	-88.313986	Mexican Restaurant
4	Aurora, Illinois	41.75717	-88.314754	Ballydoyle Irish Pub	41.759348	-88.315126	Pub
5	Aurora, Illinois	41.75717	-88.314754	Taqueria El Tio & Restaurant	41.757515	-88.319157	Mexican Restaurant
6	Aurora, Illinois	41.75717	-88.314754	Jake's Bagels & Deli	41.760742	-88.310206	Bagel Shop
7	Aurora, Illinois	41.75717	-88.314754	Two Brothers Roundhouse	41.760639	-88.308788	Brewery
8	Aurora, Illinois	41.75717	-88.314754	La Quinta De Los Reyes	41.758603	-88.312294	Mexican Restaurant
9	Aurora, Illinois	41.75717	-88.314754	Holiday Inn Express & Suites	41.759361	-88.309838	Hotel
10	Aurora, Illinois	41.75717	-88.314754	Riverfront Playhouse	41.757061	-88.313964	Theater
11	Aurora, Illinois	41.75717	-88.314754	U Samba?	41.759762	-88.315110	Brazilian Restaurant
12	Aurora, Illinois	41.75717	-88.314754	SciTech Hands On Museum	41.756344	-88.318203	Science Museum
13	Aurora, Illinois	41.75717	-88.314754	Dairy Queen	41.761321	-88.316895	Ice Cream Shop

The API call resulted in 1377 venues of which there are 228 unique venues.

From the above data we have to analyze the number of pizza places in each of the city. For this we use one hot encoding to all the venues so that we can look at the data with pizza places. Created a new data frame with city names and the number of pizza places in it.

Next we will cluster the areas with same number of pizza places. Based on the number of pizza place we can locate the ideal area to open a new pizza place. K means algorithm was used for clustering. First kmeans algorithm was ran for values of k from 1 to 15 to find the optimum k value. As a result got 6 as the optimum value for k.

Using the value of k as 6 , we cluster each area into one of the cluster using kmeans algorithm. The resulting data was combined with the data frame that had the coordinates for each city. The result was visualized on a map



## Results

Cluster 4 has most number of pizza places which is five where as Cluster 5 has Zero pizza place. The only place in Cluster 5 is Hammond, Indiana. Since our objective is to start a Pizza place in Illinois near Chicago , lets look at the cluster that has at least one pizza place. Below are the list of places that has at least one pizza place.

Arlington Heights, Illinois

Gary, Indiana

Kenosha, Wisconsin

Arlington Heights, Illinois has an estimated population of 74,593 and has many other attractions like theaters, train stations, stores etc. So an ideal location to start a new Pizza place will be Arlington Heights, Illinois.

## Discussions

Before deciding on a location to start a new pizza place we may also have to consider some other factors like what other attractions, restaurants etc are there in these cities. We may also have to compare the setup cost, rent etc in each of the city where there are no pizza place.

## Conclusion

Since our objective is to start a Pizza place in Illinois near Chicago , lets look at the population of Joliet and Waukegan. Joliet has an estimated population of 148,227 and Waukegan has an estimated population of 85,720. So an ideal location to start a new Pizza place can be either of the two cities.

The above approach to find an ideal location to start a pizza place can be applied to other datasets like gym, coffee place etc. In this project we consider only the number of a particular venue in each city.