**Table of Contents**

1. **Introduction**
2. **Target Audience**
3. **Data**
4. **Methodology**
5. **Results**
6. **Discussion**
7. **Conclusion**

**Map

Description automatically generated**

1. **Introduction**

Consider the situation where I am having to relocate because of my job and I'm moving to Chicago. Chicago will always be famous for “The Great Fire” of 1871, when a fire in the barn of Catherine and Patrick O'Leary spread across the city, killing hundreds. (Contrary to the legend, it was probably not started by a cow kicking over a lantern.)

CHICAGO TODAY

The thriving commercial and financial “City of Broad Shoulders” is spiked with gorgeous architecture and set with cultural and recreational gems including the Art Institute, Millennium Park, 250 theater companies, and 30 miles of shoreline. Approximately 2.8 million residents live within the city limits and tens of thousands commute from the ever-sprawling suburbs to work downtown.

Chicago has 246 different community areas. I would like to figure out where I would like to move and into what area neighborhood. I am going to base that choice on my main interests. What I would like is a nice selection of Restaurants, Music venues and Pool Halls.

**2. Target Audience**

This project is aimed towards anyone who is interested in moving to Chicago. The analysis could also provide vital information that can be used by the someone interested in opening a business in Chicago, since we will be exploring the different Neighborhoods and Community Areas of Chicago.

**3. Data**

* 1. Data Overview

I’ve web-scraped the table for Chicago Neighborhoods and Community Areas and appended the latitude and longitude for each Community Area using the geopy library in Python. Venue data pertaining Restaurants, Music venues and Pool Halls was obtained via Foursquare. The Venue data will help find which neighborhood is best suitable for my interests and help me decide on a suitable location for my move.

3.2 — Data acquisition:

Source 1: Table of Chicago Neighborhoods and Community Areas via Wikipedia

Graphical user interface, application

Description automatically generated

Figure 1 Table of Chicago Neighborhoods and Community areas

<https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>

Data read into a Dataframe:

Graphical user interface, application

Description automatically generated

Figure 2 Dataframe of Data

3.3 — Adding Latitude and Longitude:

Using geolocator and geocode I looped through each row of the dataframe and found the corresponding latitude and longitude. This took considerable time, so I saved the result as a csv file and then read that file when needed.

Graphical user interface, text, application

Description automatically generated

Figure 3 Dataframe with Latitude and longitude added

3.4 — Data Cleansing:

Four of the Community Areas had incorrect names and geolocator locator could not find a latitude and longitude for these. The names had to be corrected in the original data frame so that the latitude and longitude could be found.

3.5 — Adding Venue information using Foursquare:

Graphical user interface, application

Description automatically generated

Figure 4Chicago Data with Venue information added

1. **Methodology**
   1. — Exploring the data:

The following is a map Of Chicago’s Neighborhood’s made with a folium map.

Map

Description automatically generated

Figure 5Map of Chicago's Neighborhoods

Using Foursquare we found venues that met our criteria (Restaurants, Music venues and Pool Halls).

4.2 — One Hot Coding

Categorical Data is transformed into Numerical Data for Machine Learning algorithms.

Table

Description automatically generated

Figure 6 Data after One Hot Coding

Then we group rows by neighborhood and by taking the mean of the frequency of occurrence of each category.

Graphical user interface, application, table

Description automatically generated

Figure 7 Group by Neighborhood and then Standardize the Data

And finally, we find the Top 10 Venues for each Neighborhood:

A picture containing table

Description automatically generated

Figure 8 Top 10 Venues for each Neighborhood

4.3— Machine Learning with k-means Clustering

We used k = 4 clusters and added the Cluster Label to our Dataframe:



Figure 9 Added Cluster Labels

The following is the resulting clusters:

Map

Description automatically generated

Figure 10 Map of Clusters

**5. Results**

Cluster Analysis

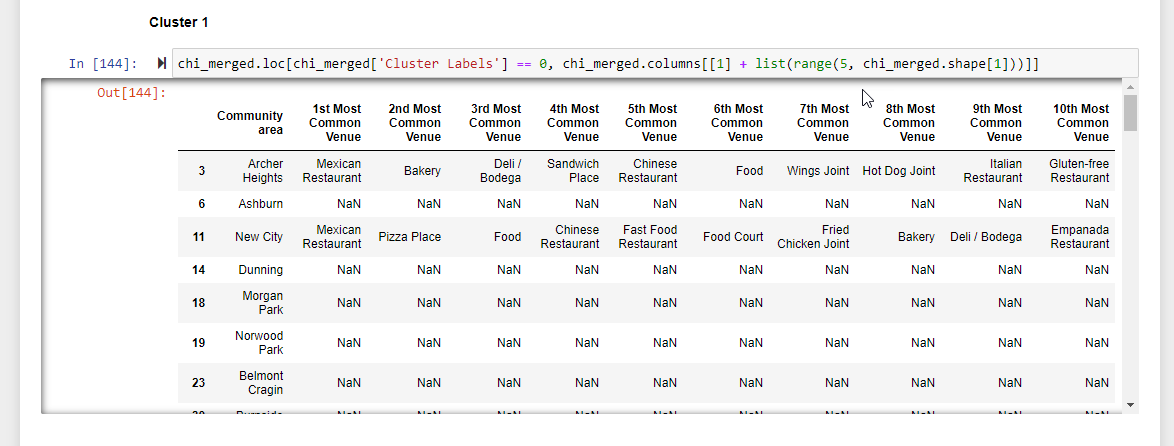


Figure 11 Cluster #1

Table

Description automatically generated

Figure 12 Cluster #2

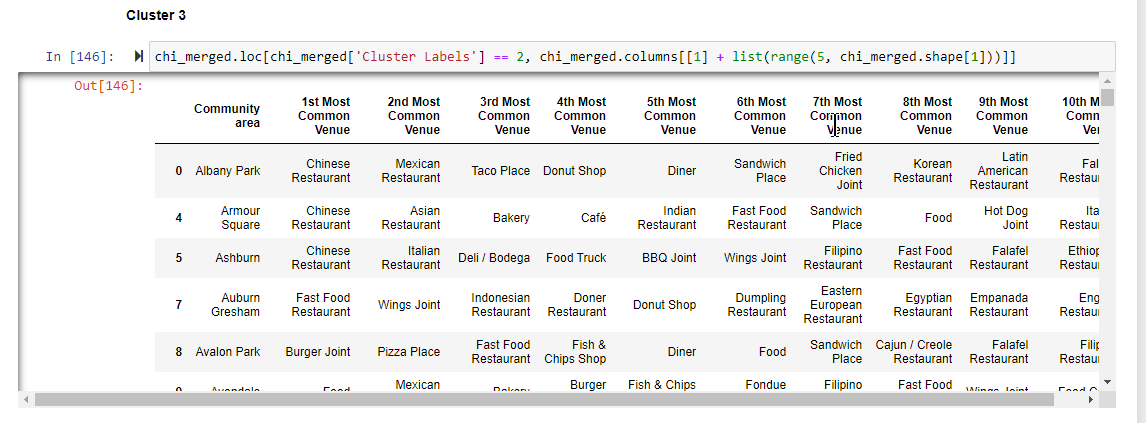


Figure 13 Cluster #3

Table

Description automatically generated

Figure 14 Cluster #4



Figure 15 Cluster #5

**5. Discussion**

Cluster #1 was the biggest, but most neighborhoods had no venues that I was interested in (NAN’s).

Cluster # 3 was the second biggest and had a very nice variety of restaurant venues and food venues.

Clusters # 2, # 4 and # 5 were all rather small and did not have much variety in feud venues.

None of the clusters or at least very few of them had music venues or pool halls in the top 10.

**6. Conclusions**

Cluster # 3 is my choice for an area to start looking into the Real-estate market to find a place to live!!

Though I am disappointed that there are not more Music venues and Pool Halls.

6.1— Further Investigations

So, I did do a sub-search where I eliminated the Food venues and just checked Music venues and Pool Halls. Now there was some overlap because many of the Music venues and Pool halls also serve food and they showed up in the food category. But as evidenced in our clusters the Music venues and Pool Halls are so few and far between that I do not feel they would be a good criterion for a unique search. They would just limit the neighborhoods too much. See Figure 16.

A picture containing text

Description automatically generated

Figure 16 A snippet of results with Foursquare search for Just Music Venues and Pool Halls