Rent Pricing & Venue Data Analysis of Chicago
IBM Data Science Professional Certificate - Capstone Project

Alex Johnson

June 2020

1. **Introduction**
   a. **Background**

Chicago, IL is the 3rd largest city by population in the United States of America, home to 2.7 million residents. It is comprised of 51 neighborhoods and 77 community areas covering 234 square miles of land. It is home to countless attractions such a major sporting events and world-class restaurants. Average rental housing can cost anywhere from $600 to more than $2,500 per month.

   b. **Problem**

The size and diversity of the city can be daunting for potential immigrants, whether domestic or international. It is difficult to narrow down a neighborhood within your budget and interests. This project aims to ease this process by displaying average rent costs and popular venues within each neighborhood.

   c. **Interest**

I'm interested in this topic because sometime in the near future I hope to intern or work in Chicago and will have to go through the housing process. Others going through this process may be interested as well.

2. **Data**
   a. **Data Sources**

Foursquare API was used to obtain the most common venues for each neighborhood of Chicago

Rent Jungles' rent trend database was used to analyze and display average rent prices

Spatial data from the Chicago Data Portal was used to display neighborhood boundaries on the choropleth map

Google Maps was used in order to get center coordinates of each Neighborhood

b. Data Cleaning

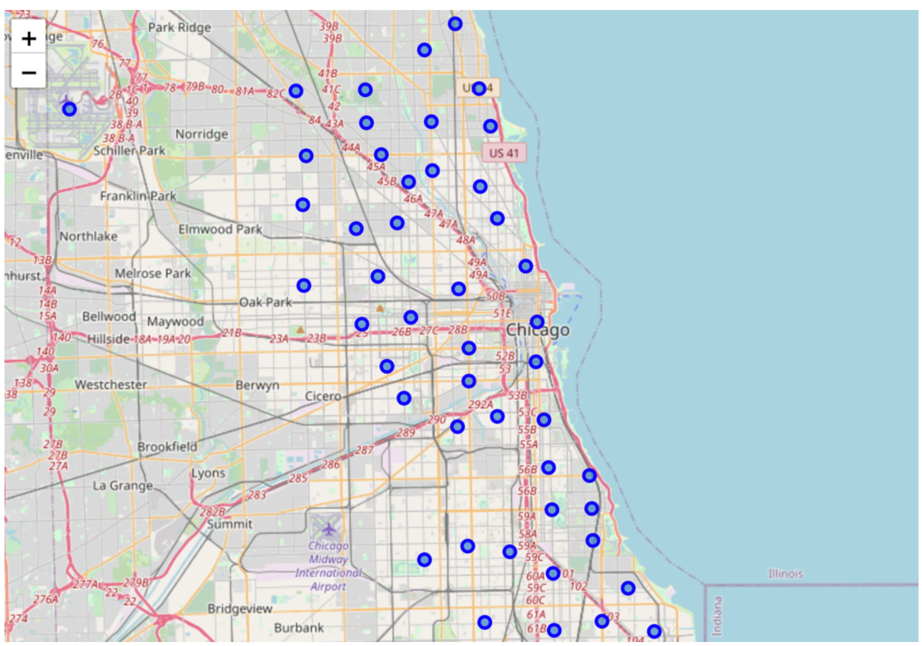
The majority of data cleaning was done prior to preprocessing in Microsoft Excel. Average Rent statistics were appended to neighborhood central coordinates and processed into the notebook. There were a few initial problems with the data frame. The column now labeled "Average" was previously named " Average Rent". The added space caused future errors.

3. Methodology

My GitHub repository was used as a database for this project. My master data frame contained *Neighborhood, Average Rent, Latitude,* and *Longitude* information of the city of Chicago.

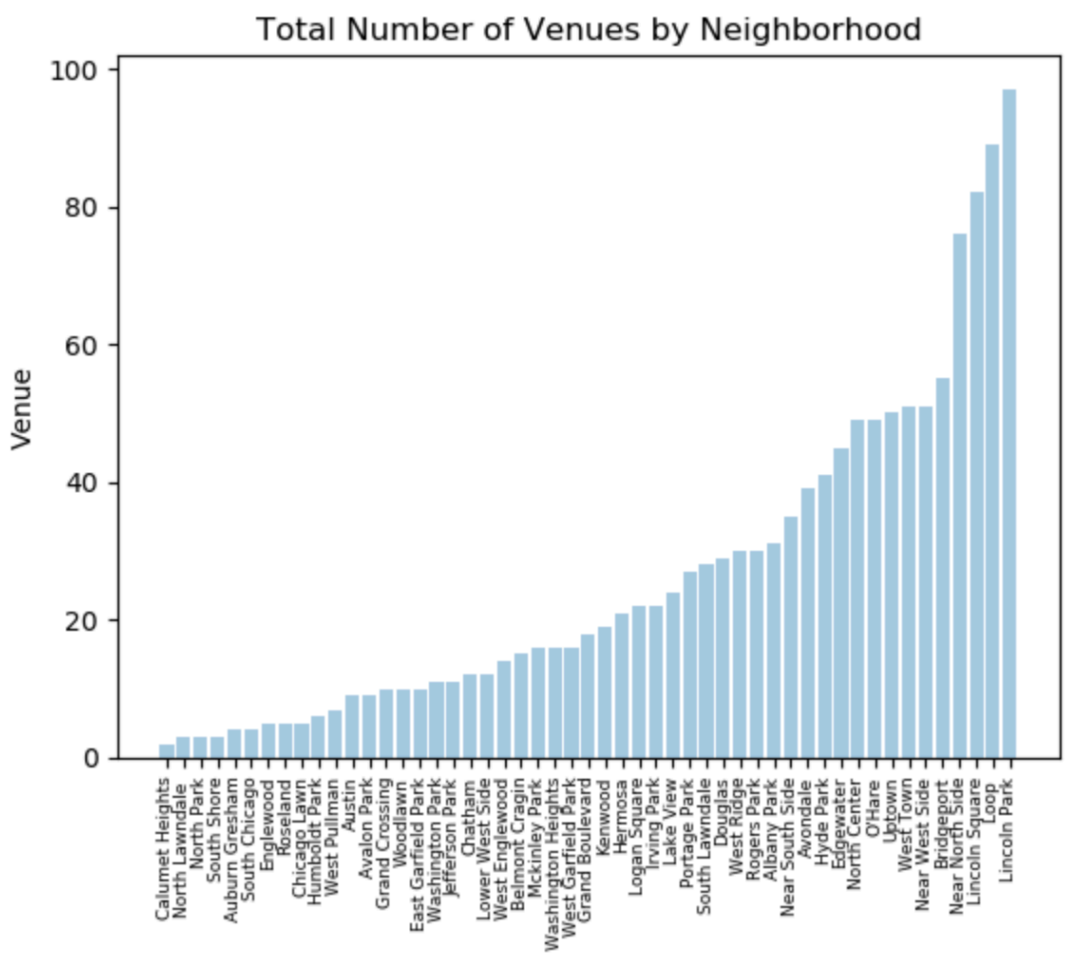
	Average Rent	Latitude	Longitude
Neighborhood			
Near South Side	$2,803	41.8608	-87.6257
Near North Side	$2,623	41.9039	-87.6315
Loop	$2,470	41.8786	-87.6251
Near West Side	$2,386	41.8668	-87.6664
West Town	$2,122	41.8936	-87.6722
Lincoln Park	$2,016	41.9255	-87.6488
North Center	$1,925	41.9467	-87.6883
Hyde Park	$1,752	41.7948	-87.5917
Avondale	$1,730	41.9415	-87.7025
Lake View	$1,703	41.9398	-87.6589

In order to visualize my analysis, I utilized the python folium library. The graph below displays a map of Chicago with a marker on each neighborhood, using the latitude and longitude values.



The Foursquare API database was used in order to explore popular venues in each distinct neighborhood. I designed the getNearbyVenue function to limit 100 venues within a 500 meter radius from the central latitude and longitude coordinates of each neighborhood.

The graph below displays the distribution of venues by neighborhood in Chicago. We can see central neighborhoods such as Loop, Lincoln Park, and Lincoln Square each contain more than 80 venues whereas the bottom 50 percent has less than 20 venues per neighborhood. A total of 1324 venues were found within these neighborhoods.

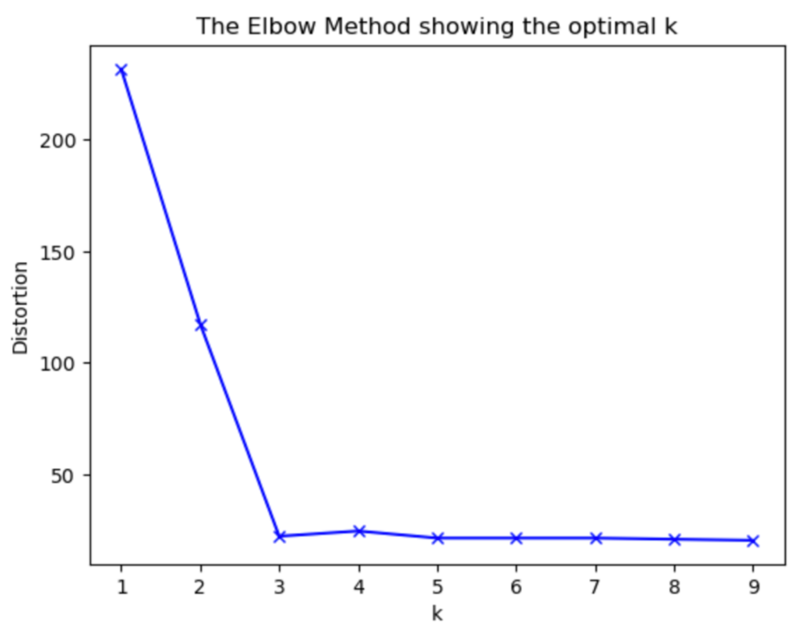


To summarize the venue data, I compiled a data frame of the top 10 most common venues by neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Albany Park	Mexican Restaurant	Pizza Place	Korean Restaurant	Donut Shop	Bank	Discount Store	Dive Bar	Mobile Phone Shop	Fried Chicken Joint	Taco Place
1	Auburn Gresham	Discount Store	Pool	Basketball Court	Park	Dog Run	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant	Fast Food Restaurant
2	Austin	Discount Store	Pizza Place	Breakfast Spot	Salon / Barbershop	Café	BBQ Joint	Food	Cosmetics Shop	Fish & Chips Shop	Filipino Restaurant
3	Avalon Park	Burger Joint	Pizza Place	Grocery Store	Boutique	Fast Food Restaurant	Sandwich Place	Diner	Cajun / Creole Restaurant	Deli / Bodega	Dance Studio
4	Avondale	Bar	Coffee Shop	Mexican Restaurant	Food Truck	Italian Restaurant	Pub	Beer Store	Big Box Store	Donut Shop	South American Restaurant

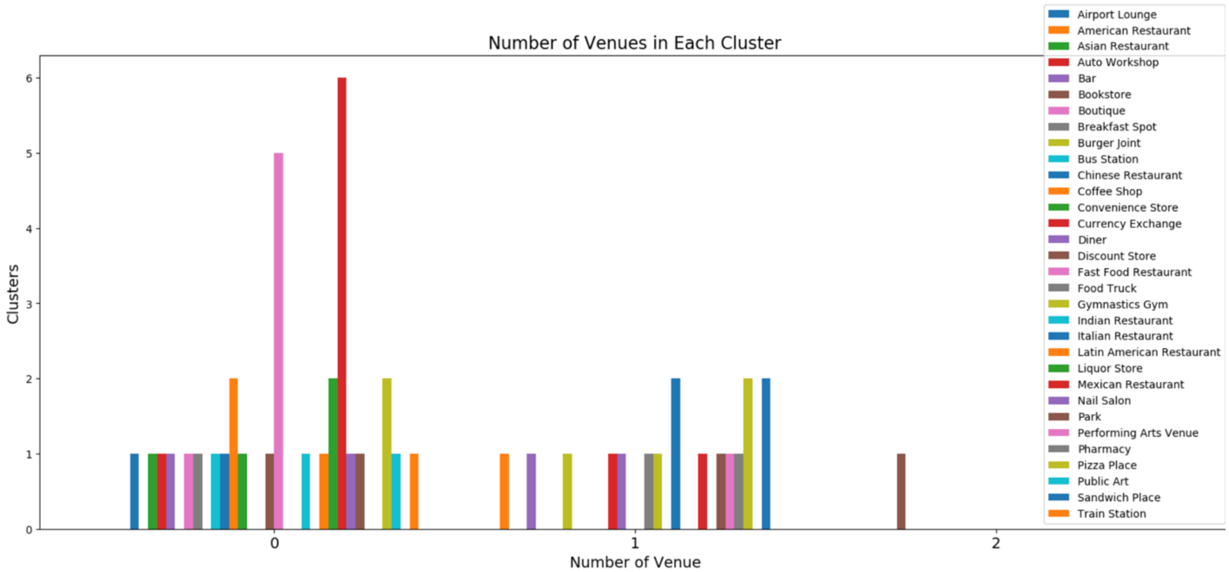
With a brief visual analysis, we observe some clear similarities between neighborhoods. For this reason, I used k-means clustering to divide the data. The neighborhoods will be divided into non-overlapping subsets with no cluster-internal structure.

In order to determine the correct value for k, I ran the algorithm with a k value of 2 and used the elbow method to display the optimal value.



Using the elbow method, I could rule that 3 is the correct k-value and I re-ran the clustering algorithm.

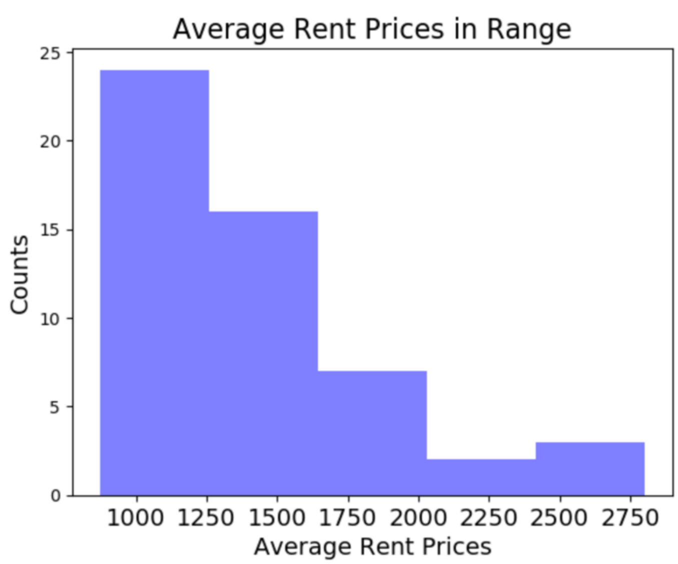
Once we have identified our clusters, using pyplot, we can analyze the distribution of most common venue types by cluster.



Through the cluster venue type distribution we are able to label each cluster:

Cluster 0 – Mexican/Fast Food Restaurants
Cluster 1 – Multiple Social Venues
Cluster 2 – Parks

We can also use a histogram to display the average rent prices throughout the city. As we can see the vast majority of properties are between $1000 - $1750 per month.



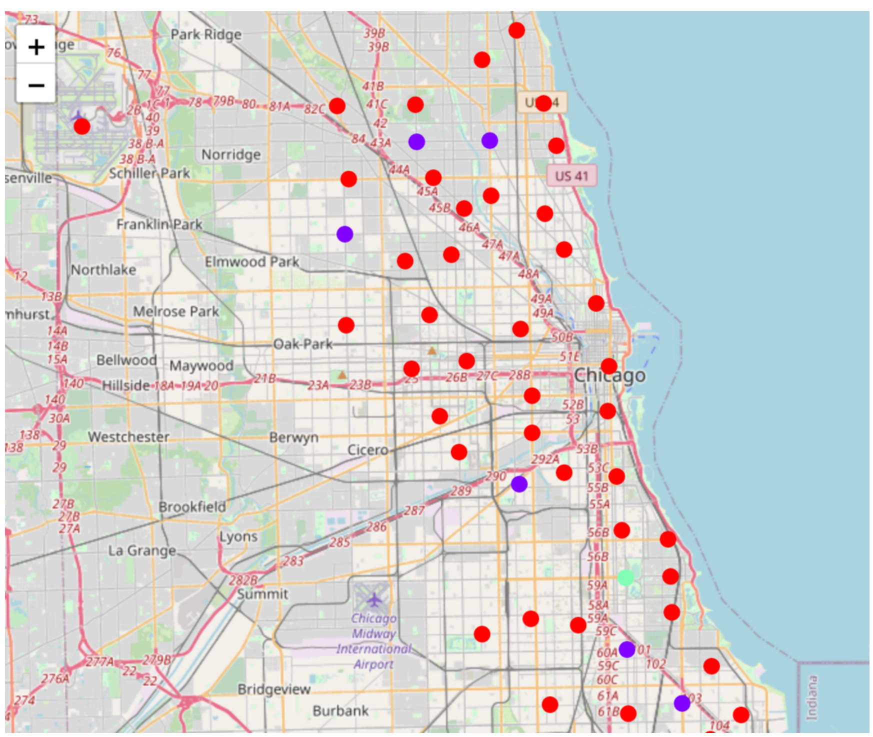
To provide a thorough summary of each neighborhood on the map I used the join function to display the count of the top three venue types for each neighborhood.

	Neighborhood	Join
0	Albany Park	4 Korean Restaurant, 2 Mexican Restaurant, 2 P...
1	Auburn Gresham	1 Basketball Court, 1 Discount Store, 1 Grocer...
2	Austin	2 Discount Store, 1 Athletics & Sports, 1 BBQ ...
3	Avalon Park	2 Burger Joint, 2 Pizza Place, 1 ATM
4	Avondale	2 Coffee Shop, 2 Food Truck, 2 Mexican Restaurant

I then merged those results with all related cluster and venue information to create one cumulative data frame.

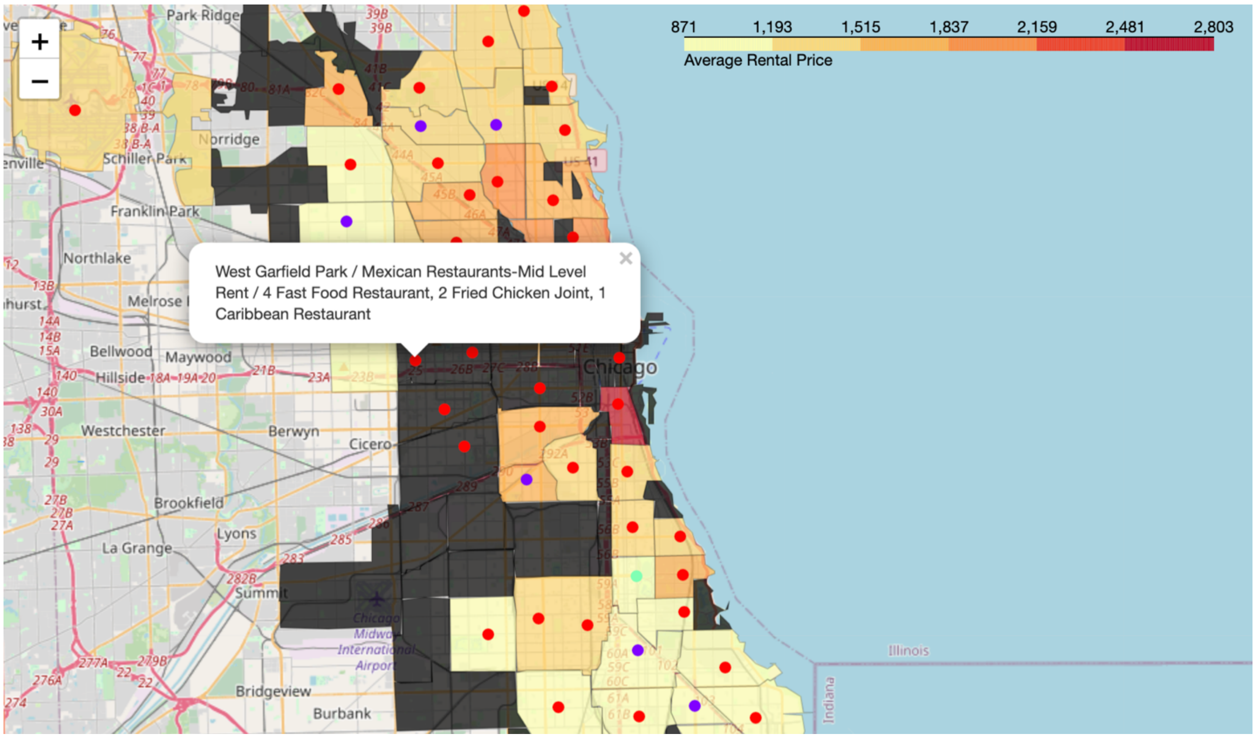
1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Join	Labels	Level_labels
Park	Bar	Grocery Store	Dog Run	Rental Car Location	Sushi Restaurant	Italian Restaurant	Steakhouse	Donut Shop	Candy Store	3 Park, 2 Bar, 2 Dog Run	Mexican Restaurants	High Level Rent
Italian Restaurant	Gym / Fitness Center	Coffee Shop	Gym	Café	Pizza Place	Steakhouse	Pub	Restaurant	Bar	5 Italian Restaurant, 4 Café, 4 Coffee Shop	Mexican Restaurants	High Level Rent
Sandwich Place	Plaza	Coffee Shop	American Restaurant	Theater	Bar	Art Museum	Public Art	Sculpture Garden	Boutique	5 Sandwich Place, 4 Coffee Shop, 4 Plaza	Mexican Restaurants	High Level Rent

Using the cumulative data frame, I recreated the folium neighborhood graph using results from the k-means clustering.



4. Results

Finally, I used the Geojson file from the Chicago Data Portal to display the boundaries of each neighborhood. I created a choropleth map in order to overlay the Average Rent Prices by neighborhood. When a neighborhood is selected the Neighborhood Name, Cluster Name, Rent Level, and Top 3 Venues are displayed.



5. Discussion

Chicago, as previously noted, is an extremely large, incredibly diverse city. Yet, the range in average rent prices still shocked me, with a difference of almost $2000 per month in some neighborhoods. Also, the sheer number of venues, 1324 was quite astonishing.

I was a bit disheartened at the effectiveness of my k-means analysis. The optimal value of k at 3 left a single neighborhood in the final cluster. But, the distribution of venue types came out nicely.

In order get obtain central neighborhood coordinated I searched nearby locations for each neighborhood. This was a long, annoying process that I'm sure could be automated by a more experienced programmer.

6. Conclusion

As with anything, there is much deeper analysis to be done on this topic. For myself, the primary goal of this project wasn't to see what type of restaurant were near Grant Park but to strengthen and develop the data science tools I've learned throughout this course. But if it helps someone find their new apartment, fantastic.

Alex Johnson