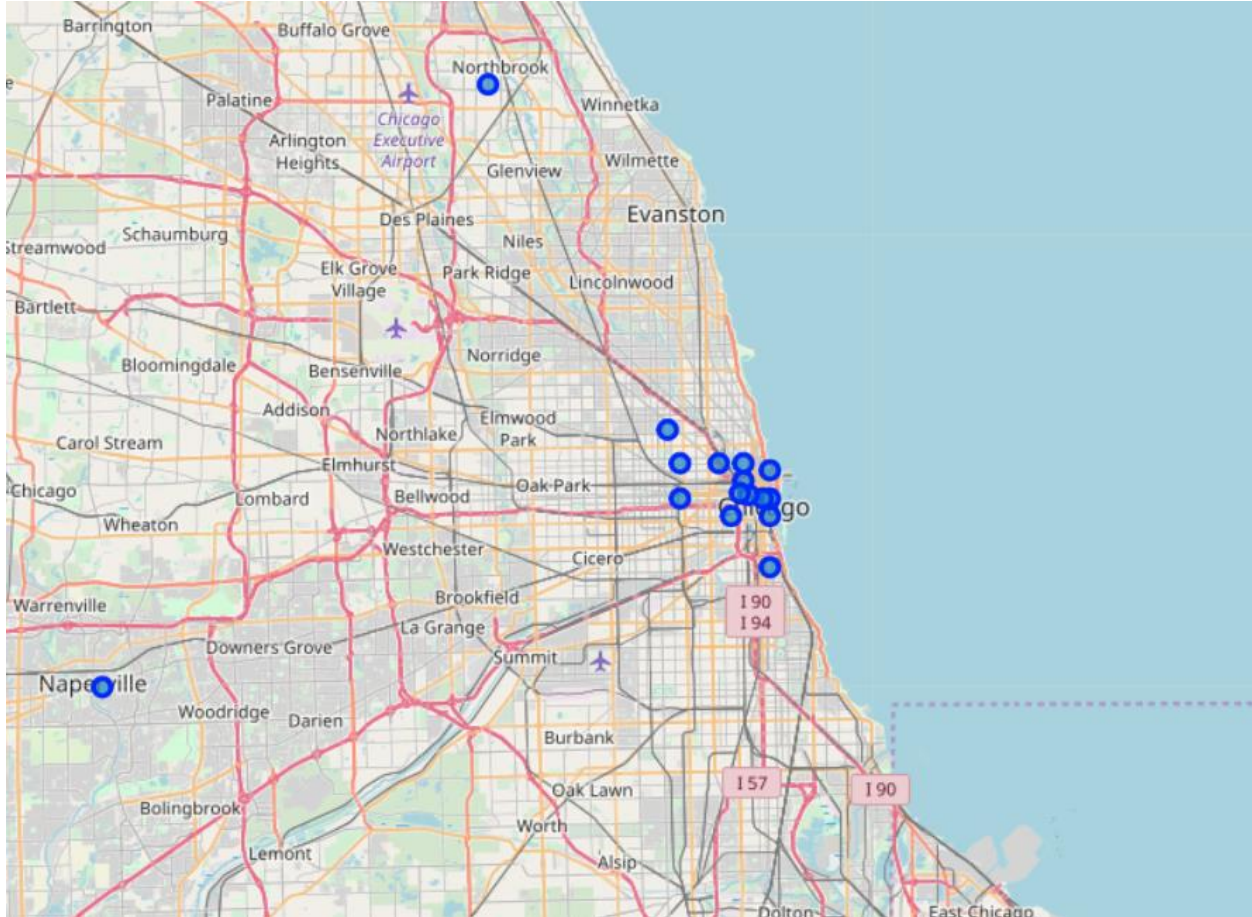


The Battle of Neighborhoods

Chicago, Northbrook, or Naperville?



1. Project Description:

For midclass who move to Chicago, they will always ask themselves questions about where should they live? The downtown area? The north suburb Northbrook? Or the west suburb Naperville? People usually work in downtown, so live in town will give them lots of conveniences such as easy commute and the quick access to tons of museums, etc., but they must pay a very high price for the crowded apartment. Of course, they can buy the big house in the suburb area, such as Northbrook and Naperville, two of the most welcomed townships by the middle class. How similar are those neighborhoods? What other attributes should we consider further to make the conclusion?

In this project, we explored those problems and hope that we can find solution for people based on the housing availability, house price, median income level, nearby facilities, and in sum rating scores. Specifically, we selected 14 downtown neighborhoods and one neighborhood of Naperville and neighborhood of Northbrook. Then, by applying the

'uszipcode' package, we got the coordinates, total housing units, occupied housing units, median home value, and median household income information for each neighborhood. After that, we used the Foursquare API to get the most common venues for each region. After converting the most common venues data into dummies, we applied the unsupervised machine learning K-means algorithm to group the neighborhoods into 5 clusters to explore the similarities among neighborhoods. Finally, we scraped ranking data from areavibes.com for the city of Chicago, Northbrook, and Naperville to conclude the best living area among the three.

2. Data Description

Generally, my data comes from four different sources:

1. Downtown Chicago neighborhoods basic information: <https://www.seechicagorealestate.com/chicago-zip-codes-by-neighborhood.php>
2. Neighborhood demographic information data: uszipcode package
3. Most popular venues data: Foursquare API
4. City livability ranking data: <https://www.areavibes.com/>

3. Methodology Section

Web scraping method would allow us to acquire data with similar html pattern.

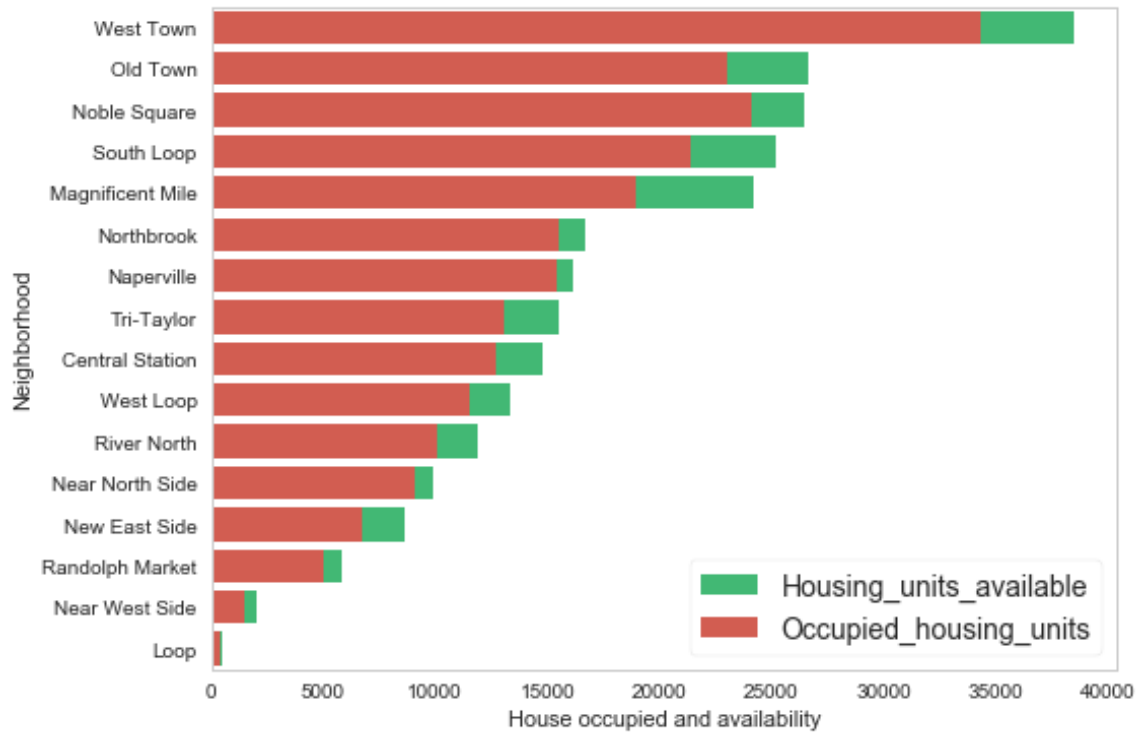
Seaborn and Matplotlib are used to visualize the demographic data.

Foursquare API would enable us to collect the nearby venues information for the neighborhoods. But free users have many limitations, and some data like the trends, stats requires authorization.

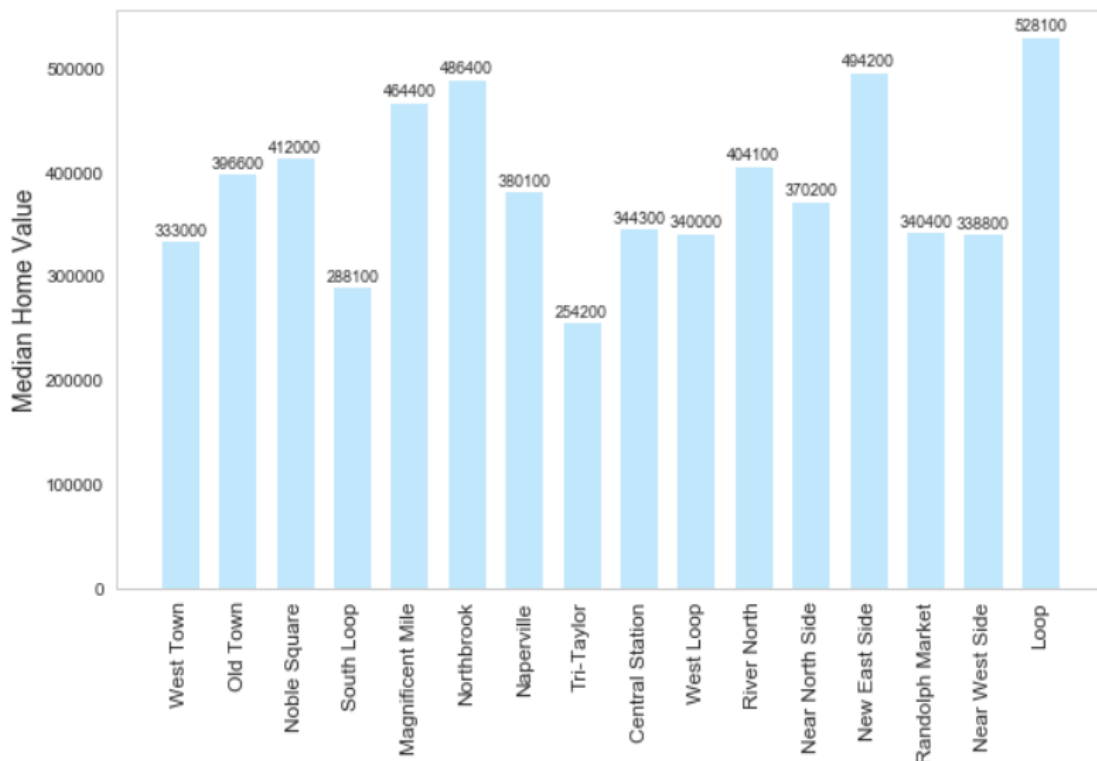
Map visualization: by applying the Folium package, we can easily plot the neighbors into the interactive map

K-means clustering algorithm: This is an unsupervised machine learning model used to handle non-labeled data and to separate all entities into K groups by measuring the similarities (individual distances and group distances).

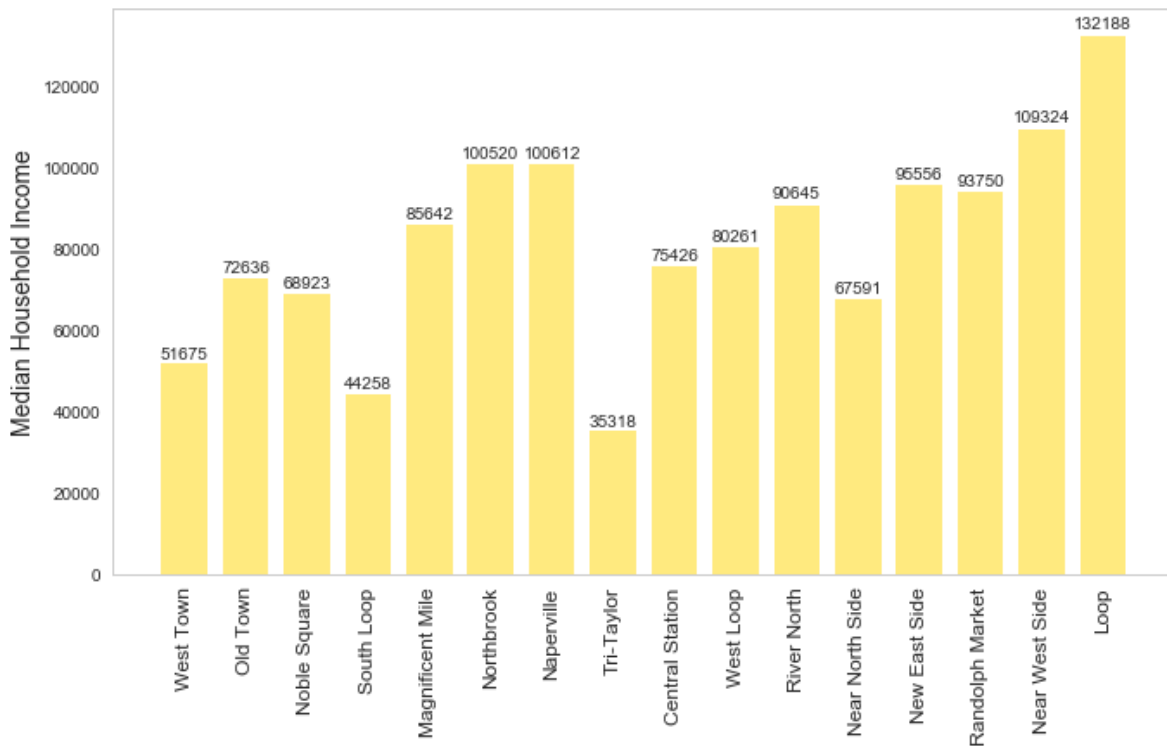
4. Results section



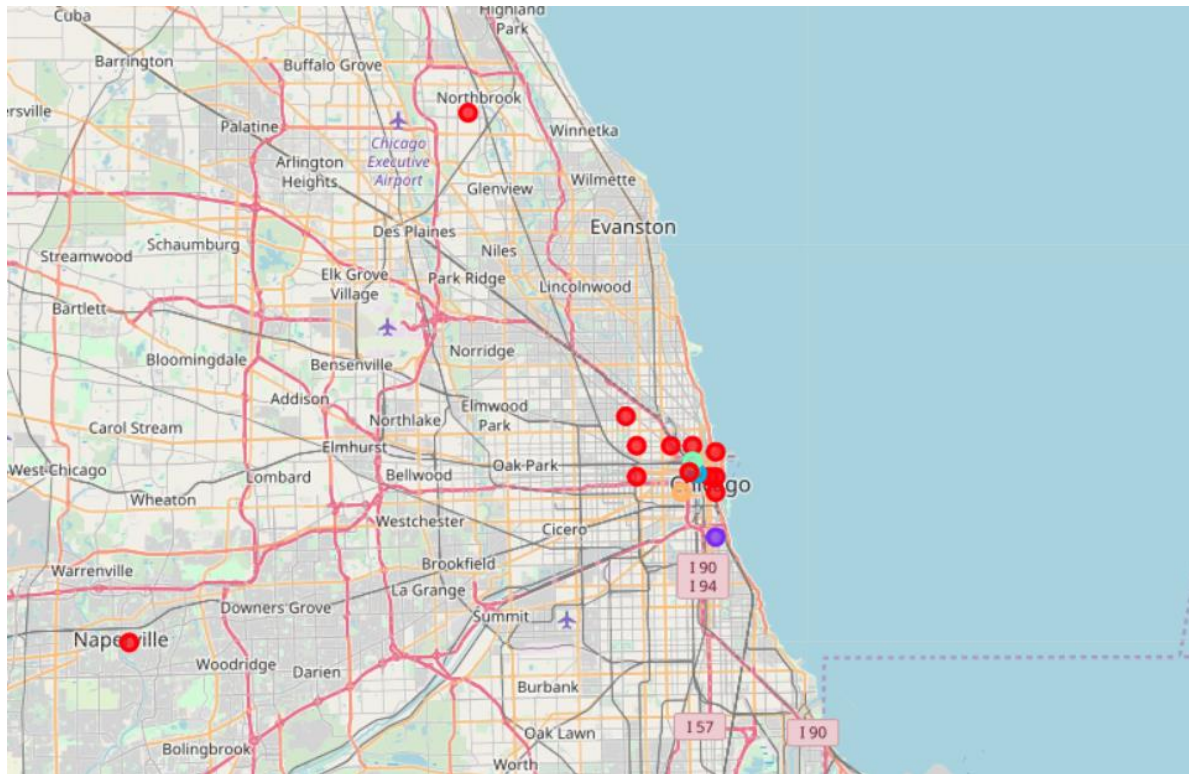
As we noticed from the figure above that the housing availability is very tight in Northbrook and Naperville, but highly different among downtown neighborhoods. This somehow proved that Northbrook and Naperville are very popular place to live.



For the median house price, the loop area is the most expensive place, and the price is highly different among downtown neighborhood, for example, the median price in Tri-Taylor is only \$254,200. The median house price in Northbrook is \$486,400 which is higher than the \$380,100 median price at Naperville. The price shows another support for Naperville.



The household median income in Loop area is the highest and the income level for Naperville and Northbrook are in par.



From the cluster results, we noticed that Northbrook and Naperville both have been grouped into a same cluster, which basically says that based on the most common nearby facilities, those two suburb towns are no different to downtown area.

5. Discussion section

	Neighborhood	Latitude	Longitude	Livability	Amenities	Cost_of_Living	Crime	Education	Housing	Weather
0	Naperville	41.74917	-88.16202	75	A+	F	A+	A-	F	D-
1	Northbrook	42.12965	-87.83514	74	A+	F	A+	A+	F	D-
2	Chicago	41.87811	-87.6298	57	A+	F	F	F	C	D

According to the data scraped from the areavibes.com, Naperville scored the highest among the three cities and got 75/100 Livability scores. The score is based on the quantify of different measurement variables such as Amenities, Cost of Living, Crime, Housing, and Education, etc.

6. Conclusion section

Naperville is the best choice for people to live, not only because the relative low median house price, but also the downtown-like nearby popular facilities and stores. And our conclusion is align with areavibe's livability rankings.