The Battle of Neighborhoods

1. INTRODUCTION

One of my friend is going to the University of Chicago for his grad school, and now he is stuggling with fiding a new apartment around the school. He has some basic requirements of the new apartment in Chicago. They are:

- The price of the apartment has to be affordable (around $3K/mo.).
- It has to be less than 20 mins away (around 0.6 miles, or 1 km) from the subway station by walking.
- He prefers the neighborhood of his new apartment is similar to the neighborhood of his home in Queens, NY.

Based on his requirements above, I'm planning to help him find a new accommodation in Chicago by using data analysis techniques.

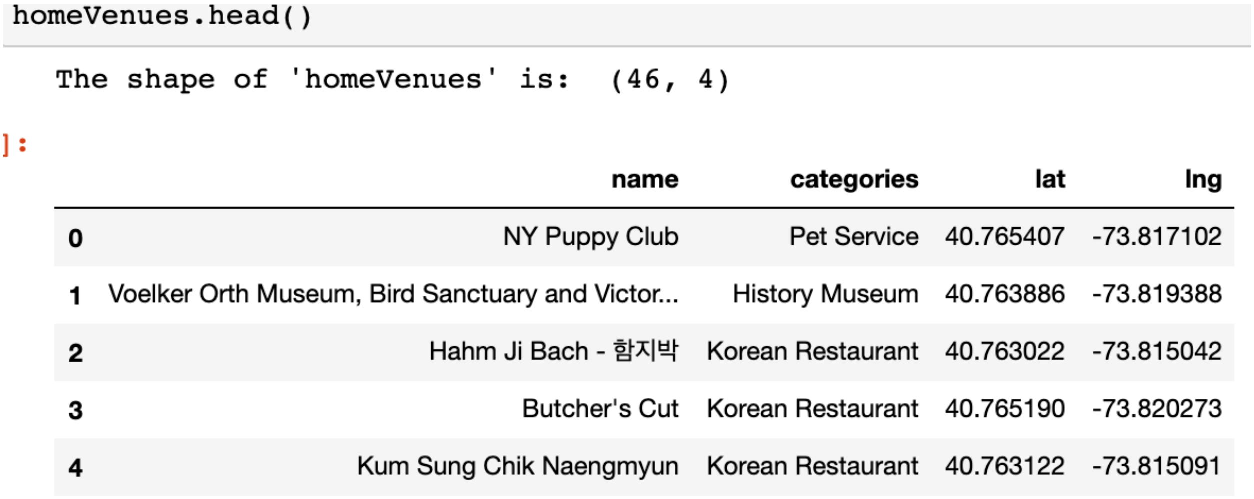
2. DATA

2.1 Accuring Data

The address of my friend's home is in 'Flushing, NY'. This address will be the input to the FourSquare API, and the output will be all the venues around this address, with coordinates. These venues will be stored into a dataframe, called 'homeVenues'. Then, for the neighborhoods in Chicago, it can be found on the Wikipedia website 'https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago. The subway location data can be found in a dataframe, and the coordinates are already provided, 'https://data.cityofchicago.org/Transportation/CTA-System-Information-List-of-L-Stops/8pix-ypme/data'. It can be easily acquired by using API provided on this website. The data of rental apartments available in Chicago will be scraped from a rental website by using the local Python compiler. The data will be stored in a CSV file, which will be uploaded directly to the IBM Watson Studio. For all the coordinates data, it can be found by using the 'geopy.geocoders package'.

2.2 Cleaning Data

There is some missing information in 'ChicagoHoods' and 'ChicagoApt' datasets, especially the missing latitude and longitude. I will delete the whole row containing the missing entry. For dataset called 'sbuway', there are some repeated latitudes and longitudes. I'll delete the repeated information. The cleaned data is shown below.

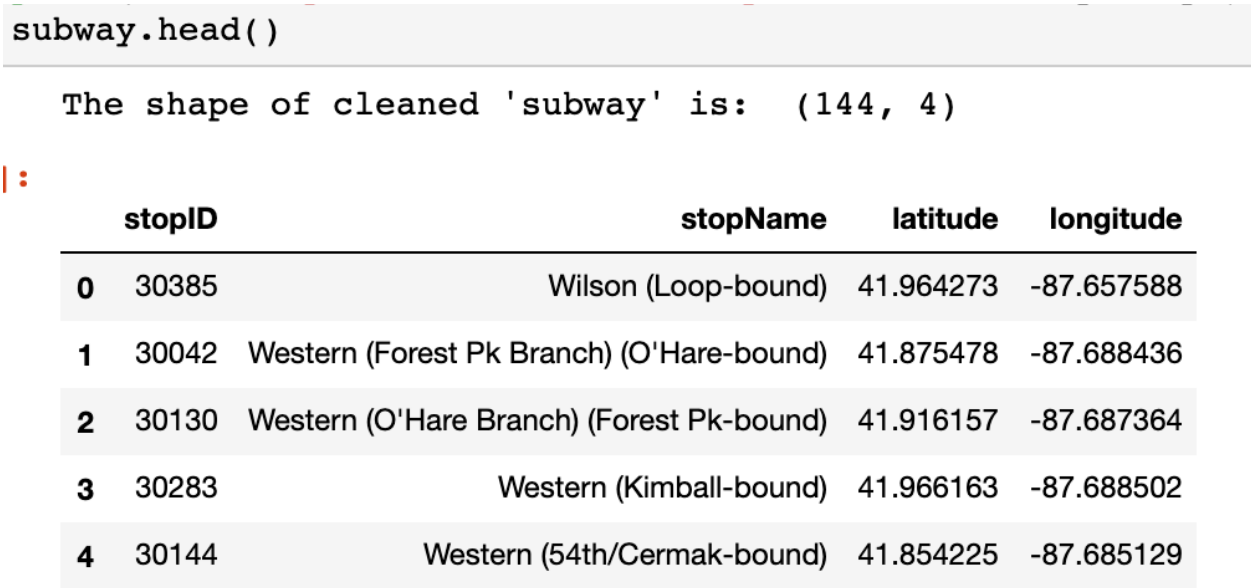


```
homeVenues.head()
```

```
The shape of 'homeVenues' is:  (46, 4)
```

]:

	name	categories	lat	lng
0	NY Puppy Club	Pet Service	40.765407	-73.817102
1	Voelker Orth Museum, Bird Sanctuary and Victor...	History Museum	40.763886	-73.819388
2	Hahm Ji Bach - 함지박	Korean Restaurant	40.763022	-73.815042
3	Butcher's Cut	Korean Restaurant	40.765190	-73.820273
4	Kum Sung Chik Naengmyun	Korean Restaurant	40.763122	-73.815091

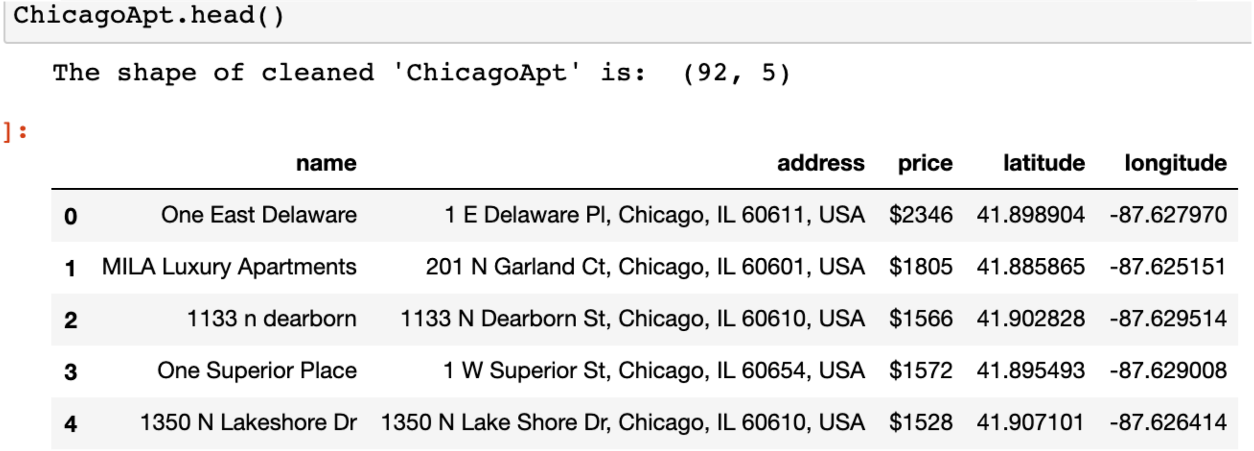


```
subway.head()
```

```
The shape of cleaned 'subway' is:  (144, 4)
```

]:

	stopID	stopName	latitude	longitude
0	30385	Wilson (Loop-bound)	41.964273	-87.657588
1	30042	Western (Forest Pk Branch) (O'Hare-bound)	41.875478	-87.688436
2	30130	Western (O'Hare Branch) (Forest Pk-bound)	41.916157	-87.687364
3	30283	Western (Kimball-bound)	41.966163	-87.688502
4	30144	Western (54th/Cermak-bound)	41.854225	-87.685129



```
ChicagoApt.head()
```

```
The shape of cleaned 'ChicagoApt' is:  (92, 5)
```

]:

	name	address	price	latitude	longitude
0	One East Delaware	1 E Delaware Pl, Chicago, IL 60611, USA	$2346	41.898904	-87.627970
1	MILA Luxury Apartments	201 N Garland Ct, Chicago, IL 60601, USA	$1805	41.885865	-87.625151
2	1133 n dearborn	1133 N Dearborn St, Chicago, IL 60610, USA	$1566	41.902828	-87.629514
3	One Superior Place	1 W Superior St, Chicago, IL 60654, USA	$1572	41.895493	-87.629008
4	1350 N Lakeshore Dr	1350 N Lake Shore Dr, Chicago, IL 60610, USA	$1528	41.907101	-87.626414

3. METHODOLOGY

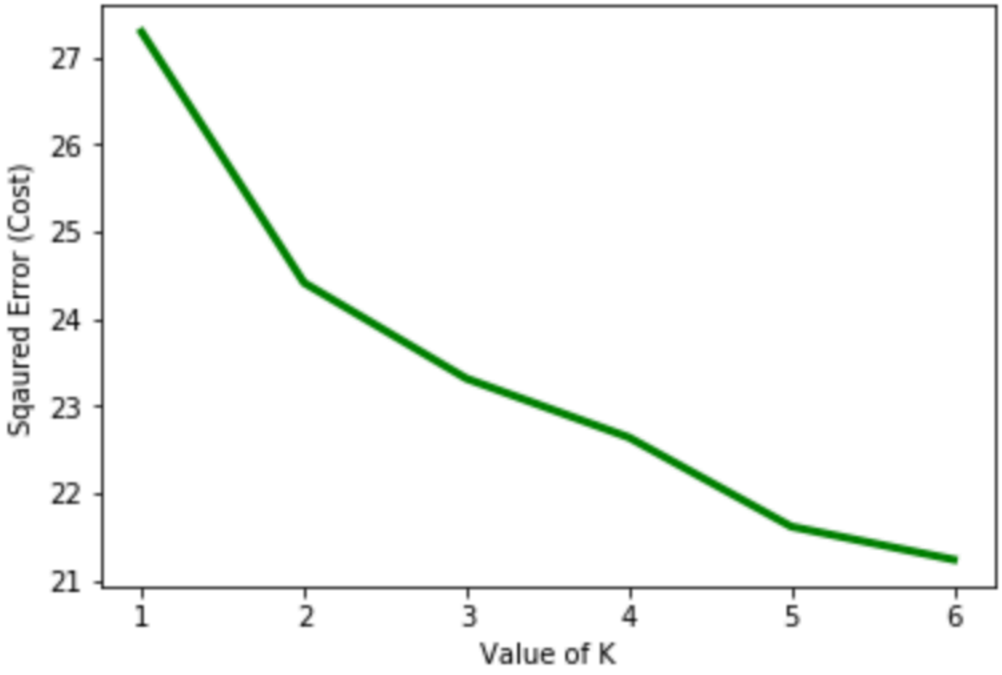
In this section, the focus will be manipulating data and finding useful informations that can be helpful to find the best apartment in Chicago.

3.1 Acquiring Venues in Neighborhoods

With the neighborhood data and the FourSquare API, all the venues in each neighborhood can be found, as well as the categories of these venues, the coordinates of these venues. In total, there are 5818 venues found in all neighborhoods in Chicago, with 339 unique categories. Then, the next step is to set each unique venue category as columns and group the data by the neighborhood. Each entry in the data frame is the percentage of the each category of venues appearing in this neighborhood. Then according to the categories of venues appearing percentage, the 10 most frequently appeared venues in each neighborhood can be defined.

3.2 Clustering Neighborhoods

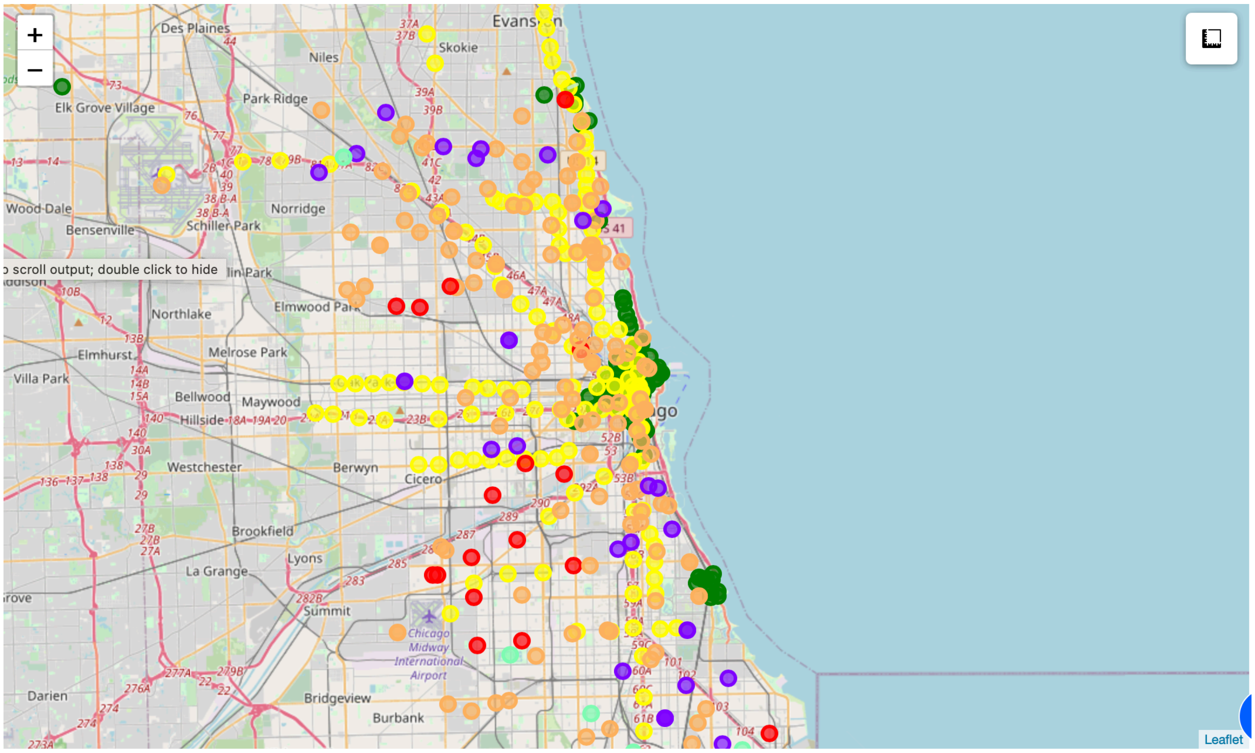
These neighborhoods can be grouped according to the venues in them, by using K-Means Clustering. After testing cost for each K value from 1 to 10, the optimum value of K should be 5 according to the elbow method. Here is the plot of cost for each K value.



Then, all the neighborhoods can be grouped into 5 clusters because the K value is 5, and the cluster labels are appended to the dataframe according to the neighborhood. There are 31 neighborhoods in Cluster0, 3 neighborhoods in Cluster1, 29 in Cluster2, 165 neighborhoods in Cluster3, and 1 neighborhood in Cluster4.

3.3 Plot All Data on One Map

With the Folium package, all the geodata can be ploted on the same map. With colored circle and the pop up labels, the map can be very clear to show all the data very straightforwardly.

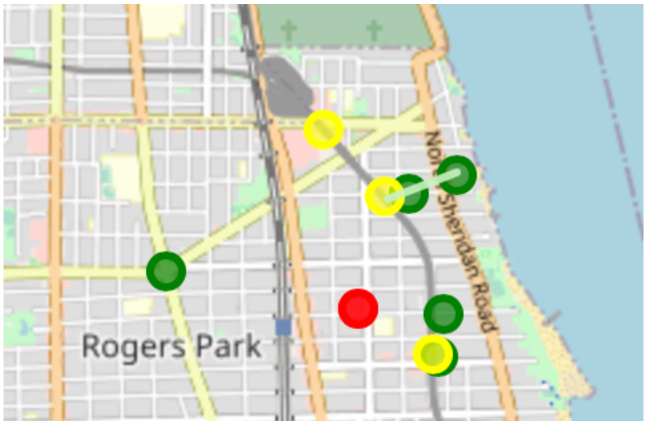


The green circles represent all the apartments for rent in Chicago. Each yellow circles means each subway station in Chicago. And all the rest of the colored circles stand for neighborhoods in Chicago. Different color means different neighborhood clusters.

4. RESULT

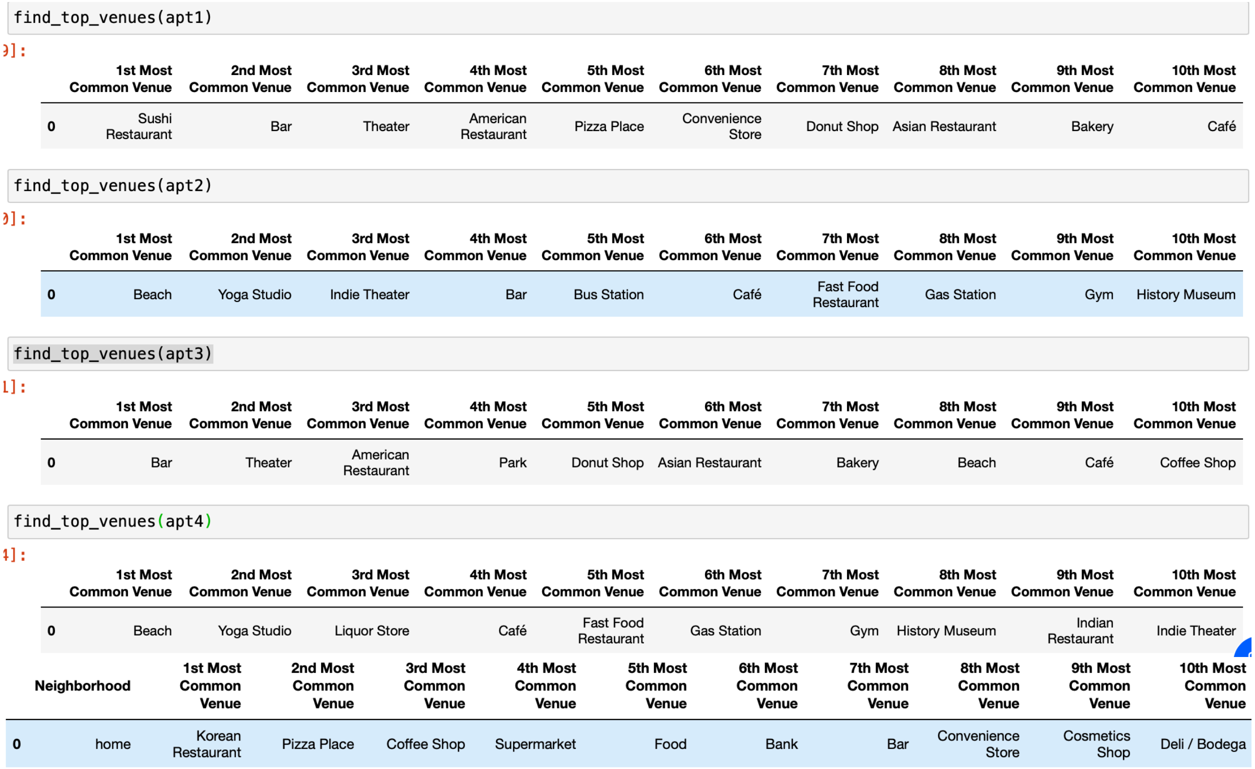
By comparing the neighborhood clusters with top 10 venues around home, it's easily to find that neighborhoods in Cluster 0 is most similar to home. Therefore, the targeted apartments should be in the neighborhood in Cluster 0. Also, Cluster 0 is shown by red circles on the map.

After examining the map, there are some apartments available on the top part of Chicago, and in the neighborhoods in cluster 0.



As shown in this picture, there are 5 apartment available in this neighborhood. There is one green circle below the yellow circle. After measure the distance between each apartment to its nearest subway station, the distance between the first apartment on the left to its nearest subway station exceeds 0.6miles/1km range. However the other 4 are all in the acceptable range.

The next step is to find the top 10 venues around each apartment and compare these top venues list to the home one.



```
find_top_venues(apt1)
```

9]:

	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	Sushi Restaurant	Bar	Theater	American Restaurant	Pizza Place	Convenience Store	Donut Shop	Asian Restaurant	Bakery	Café

```
find_top_venues(apt2)
```

9]:

	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	Beach	Yoga Studio	Indie Theater	Bar	Bus Station	Café	Fast Food Restaurant	Gas Station	Gym	History Museum

```
find_top_venues(apt3)
```

1]:

	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	Bar	Theater	American Restaurant	Park	Donut Shop	Asian Restaurant	Bakery	Beach	Café	Coffee Shop

```
find_top_venues(apt4)
```

4]:

	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	Beach	Yoga Studio	Liquor Store	Café	Fast Food Restaurant	Gas Station	Gym	History Museum	Indian Restaurant	Indie Theater

	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	home	Korean Restaurant	Pizza Place	Coffee Shop	Supermarket	Food	Bank	Bar	Convenience Store	Cosmetics Shop	Deli / Bodega

After the comparison, the most matching apartment is Apartment 1. The address of this apartment is 1340 W Morse Ave, Chicago, IL 60626, USA , and the rent for this apartment is $968/mo.

5. DISCUSSION

There are some data that I can't acquire, such as the area and the bedrooms in the apartments. These missing information, more or less, can affect the result. Therefore, the best apartment I get may not be the best result for my friend. Also, it is very hard to define how many clusters the neighborhoods should be divided into. Hence the cluster of the neighborhoods is not very accurate.

6.COONCLUSIONS

Although, I found the best matched apartment for my friend, there are still some other apartments desired to be considered. Those apartments are much closer to the center of Chicago, so life could be easier with live in the center of the city, since venues gathered more closer in the center of the city generally. Moreover, those apartments are much closer to the campus. That could save a lot of time from commuting to school everyday. However, there are also some disadvantages for living in the center of the city, like the traffic issue and the noise issue.