Final Coursera Capstone Project Report

The Battle of Neighborhoods (Week 2)

Chicago IL, USA

Chicago is the most populous city in the state Illinois, and the 3rd largest city in the U.S. with estimated population of 2.7M (2018). Chicago is famous as 'Windy City'. It's 58M domestic & international visitors in 2018 made it the second most visited city in the U.S. (as compared with NYC's 65M in 2018 – Wikipedia)

1. Description of the problem and a discussion of the background:

A. Background of the City of interest:

Whenever a visitor or proposed investors search for new city, they are mostly interested in the best places – areas of concern city have to offer. They are eager to know about amount and areas of best interesting venues concentrated, and which specific Community Areas (neighborhood) possess more of it.

Chicago as large populated city that has expanded its boundaries over all the sides. As per collected dataset Chicago has 77 Community Areas (neighborhoods). We can explore venues of all Community Areas by API, but it can be very lengthy & confusing. So, we have selected 10 neighborhoods which

are most attractive for tourists and close to Downtown area of city. We will explore these areas for venues that can be more helpful to our audience.

B. Interested Audience:

To solve the problem of our target audience – mass of tourists of Chicago city and proposed investors to start new businesses like Café, any Continental Restaurant Bistro or any kind of new venture of their interest could be interested in our data analysis. This data can be very helpful to reach at worth decision.

2. A description of the data and how it will be used to solve the problem:

The data we have used to create this project is retrieved from various sources:

A. Sources of Data:

- The City of Chicago – Data Portal
- Data of Census – Data Portal
- Boundaries of Neighborhoods
- FourSquare API
- Python Geocoders
- Python Folium

B. Method to extract and use data:

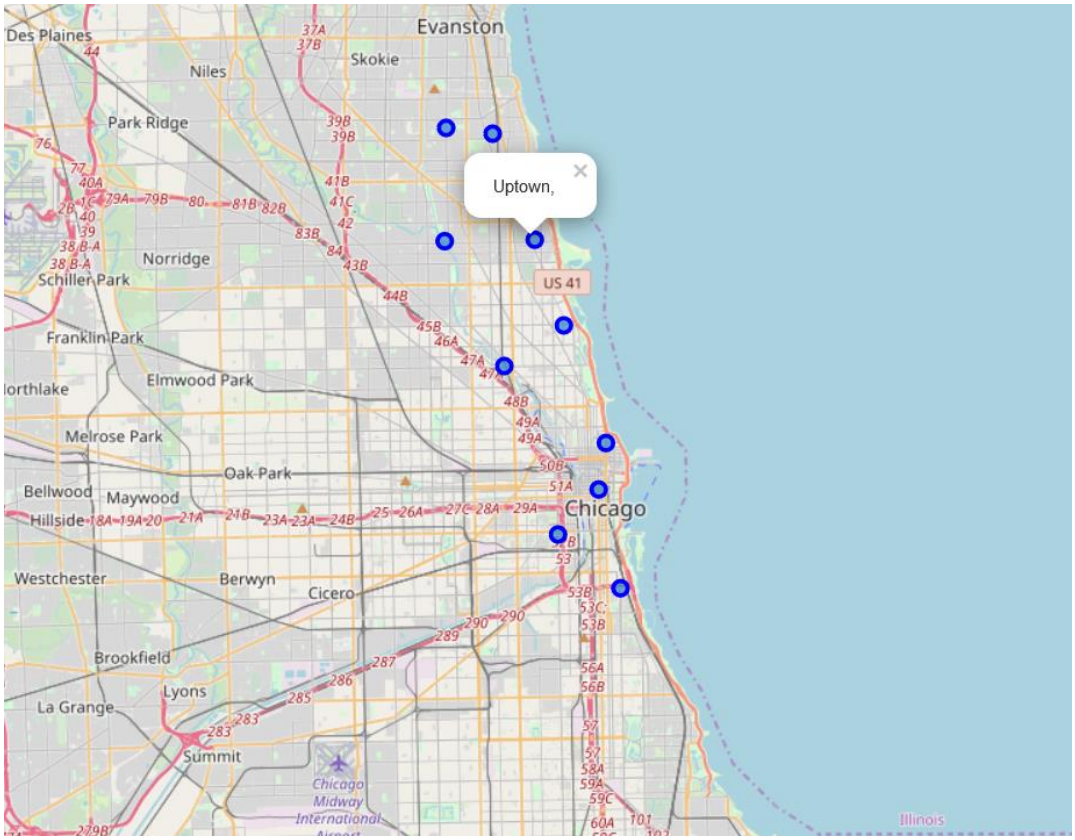
The City of Chicago – Data Portal is free and easily available source to extract names of City's Community Area (neighborhood) with latitudes & longitudes. We can also get the geographical coordinates of the all neighborhoods using Python Geocoders. It can give the Latitude & Longitude of the each neighborhood. Also, we will get help of FourSquare API to have the venue data of neighborhoods. FourSquare have largest database of 150M+ places & widely used by Data Scientists and Developers in world. It provides various categories of the venue data.

In this project we will use different Data Science skills, Python Folium package to create maps of city and neighborhoods and other necessary Machine Learning Techniques.

Data frame containing selected neighborhoods of Chicago city:

	AREA NAME	Latitude	Longitude
0	Rogers Park	42.007012	-87.677996
1	West Ridge	42.008751	-87.699768
2	Uptown	41.969826	-87.657637
3	Lincoln Square	41.969389	-87.700489
4	Lake View	41.939539	-87.644384
5	Lincoln Park	41.925098	-87.672294
6	North Side	41.897983	-87.624096
7	West Side	41.865761	-87.646876
8	Loop	41.881598	-87.627758
9	South Side	41.846860	-87.617324

Map of Chicago with markers of neighborhoods using Folium package of Python:



3. Methodology:

A. Exploratory Data Analysis:

The mentioned functions of Python are used to retrieve the dataset of Chicago Census from City Data Portal. After cleaning it and extraction we have finalize 10 neighborhoods, which have more attractive spots for visitors of the city. They are near the down town and have more attractions of sightseeing. Our goal is to find restaurants – café of our choice in that area. We will try to get data for specific Vegetarian restaurants also, which is the necessity of specific tourists.

Census Data:

[20]:

	COMMUNITY_AREA_NUMBER	COMMUNITY_AREA_NAME	PERCENT OF HOUSING CROWDED	PERCENT HOUSEHOLDS BELOW POVERTY	PERCENT AGED 16+ UNEMPLOYED	PERCENT AGED 25+ WITHOUT HIGH SCHOOL DIPLOMA	PERCENT AGED UNDER 18 OR OVER 64	PER_CAPITA_INCOME	HARDSHIP_INDEX
0	1.0	Rogers Park	7.7	23.6	8.7	18.2	27.5	23939	39.0
1	2.0	West Ridge	7.8	17.2	8.8	20.8	38.5	23040	46.0
2	3.0	Uptown	3.8	24.0	8.9	11.8	22.2	35787	20.0
3	4.0	Lincoln Square	3.4	10.9	8.2	13.4	25.5	37524	17.0
4	5.0	North Center	0.3	7.5	5.2	4.5	26.2	57123	6.0
5	6.0	Lake View	1.1	11.4	4.7	2.6	17.0	60058	5.0
6	7.0	Lincoln Park	0.8	12.3	5.1	3.6	21.5	71551	2.0
7	8.0	Near North Side	1.9	12.9	7.0	2.5	22.6	88669	1.0
8	9.0	Edison Park	1.1	3.3	6.5	7.4	35.3	40959	8.0
9	10.0	Norwood Park	2.0	5.4	9.0	11.5	39.5	32875	21.0
10	11.0	Jefferson Park	2.7	8.6	12.4	13.4	35.5	27751	25.0
11	12.0	Forest Glen	1.1	7.5	6.8	4.9	40.5	44164	11.0
12	13.0	North Park	3.9	13.2	9.9	14.4	39.0	26576	33.0
13	14.0	Albany Park	11.3	19.2	10.0	32.9	32.0	21323	53.0
14	15.0	Portage Park	4.1	11.6	12.6	19.3	34.0	24336	35.0

From this data we have extracted 10 neighborhoods.

B. Analysis:

For the analysis of neighborhoods we used FourSquare to get data of restaurants in it, as well to find count of Veg restaurants.

We used Python Bar Plot to visualize the count of restaurants in each neighborhood, because it's easy to compare watching it. Here we can easily see that all neighborhood haven't more restaurants. WestRidge and WestSide have less than 10 restaurants. Uptown, LakeView, LincolnSquare has plenty of amount for that.

Data frame containing that data:

```
[19]: df_veg_count = get_restaurant_counts(names=chicagodf.Ne
      df_veg_count.head(10)
```

[19]:

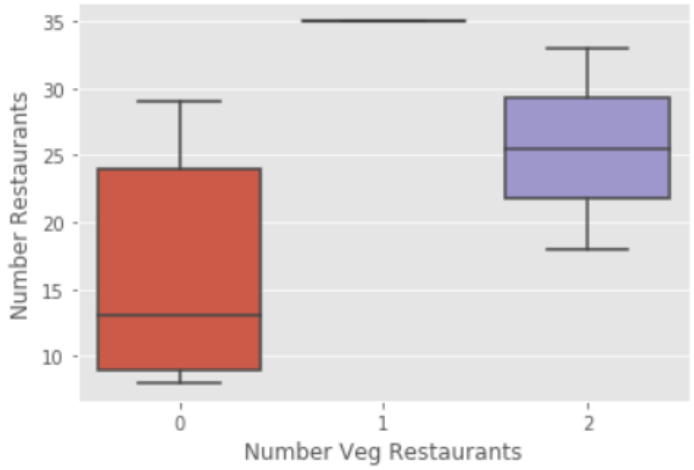
Neighborhood	Number Veg Restaurants	Number Restaurants
LakeView	2	33
LincolnPark	0	13
LincolnSquare	0	29
Loop	2	18
NorthSide	0	24
RogersPark	0	24
South Side	0	10
Uptown	1	35
WestRidge	0	8
WestSide	0	8

Also, there is only 3 neighborhoods – LakeView, Loop and Uptown have least number of Veg restaurants. But some has 0 count. For this we used Python

Box Plot so anybody can easy judge the result when tourists like to visit it. Also, who looks for tentative spot to start new business.

```
[24]: sns.boxplot(x="Number Veg Restaurants", y="Number Restaurants"
```

```
[24]: <matplotlib.axes._subplots.AxesSubplot at 0x7f95a05ddeb8>
```



Also K-means Clustering method to analyze data is used and FourSquare API is used to get all venues of related neighborhoods as it popular for developers and gives plenty of data about.

4. Results:

Here are the results of received data of all venues in neighborhoods and after Clustering we can see the most popular kind of venues in all area.

```
[45]: chicago_grouped = chicago_onehot.groupby('Neighborhood').mean().reset_index()
chicago_grouped
```

[45]:

	Neighborhood	American Restaurant	Arepa Restaurant	Art Museum	Arts & Crafts Store	Asian Restaurant	Bakery	Bookstore	Boutique	Bubble Tea Shop	Building	Burger Joint	Café	Camera Store	Chinese Restaurant	Coffee Shop
0	LakeView	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
1	LincolnPark	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
2	LincolnSquare	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
3	Loop	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
4	NorthSide	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
5	RogersPark	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
6	South Side	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
7	Uptown	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
8	WestRidge	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08
9	WestSide	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.08

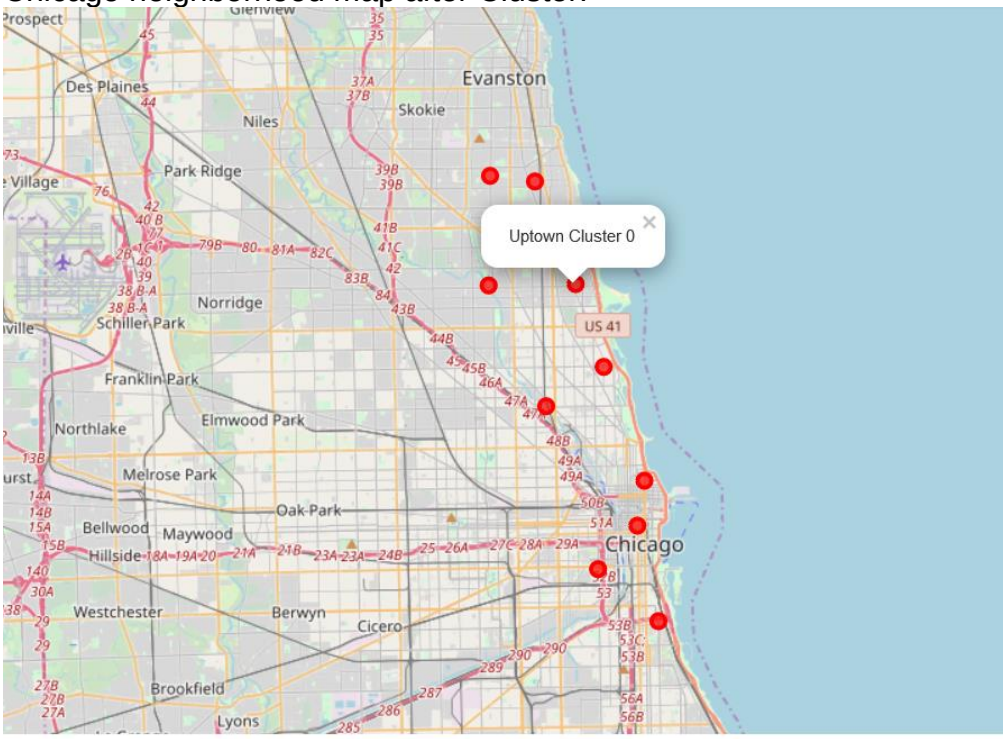
Most Common 10 venues of each neighborhood of Cicago:

```
neighborhoods_venues_sorted.head()
```

[49]:

	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	LakeView	Coffee Shop	Sandwich Place	Hotel	Pub	Bookstore	Pizza Place	Museum	Donut Shop	Hot Dog Joint	Garden
1	LincolnPark	Coffee Shop	Sandwich Place	Hotel	Pub	Bookstore	Pizza Place	Museum	Donut Shop	Hot Dog Joint	Garden
2	LincolnSquare	Coffee Shop	Sandwich Place	Hotel	Pub	Bookstore	Pizza Place	Museum	Donut Shop	Hot Dog Joint	Garden
3	Loop	Coffee Shop	Sandwich Place	Hotel	Pub	Bookstore	Pizza Place	Museum	Donut Shop	Hot Dog Joint	Garden
4	NorthSide	Coffee Shop	Sandwich Place	Hotel	Pub	Bookstore	Pizza Place	Museum	Donut Shop	Hot Dog Joint	Garden

Chicago neighborhood map after Cluster:



5. Discussion:

With the help of these data, it's visualization and analysis tentative visitors of Chicago can manage and plan for own food during the memorable visit. If need specific Veg food, they have to plan visit of LakeView, Chicago Loop or Uptown to have food during the visit of other area's tourist spots. Though, they can get it in other venues also if they like it.

For the discussion of other side as anybody like to start business of the restaurant in the city, they can look at WestSide, SouthSide, WestRidge or LincolnPark if other circumstances match. If someone like to jump in competitive business, they can start new venture at LakeView or Uptown because there are most chances of having more customers there due to more attractions.

6. Conclusion:

This project can be helpful to new tourist-visitors of Chicago city. We have explored hot favorite neighborhoods of visit. This analysis and technological skill of filtering can help to plan visit or new business to make decision. As we mentioned above, this is a City of mass population and 2nd hot favorite to visit again and again!

7. References & Libraries Used to Develop this Project:

- Pandas: For creating and manipulating data frames
- Data Portal – City of Chicago.
- Folium: Python visualization library.
- Scikit Learn: For k-means clustering.
- Geocoder: To retrieve Location Data.
- Matplotlib: Python Plotting Module.
- FourSquare API to get venues of related neighborhoods.

Thanks for Reading and review it!!!