IBM Data Science Capstone Project- The Battle of the Neighborhoods

2.0 Data Required

In order to address our previously stated business problem, we would require various kinds of data from numerous sources which have been listed and briefly described below:

- Crime Data from the Chicago City Data Portal which maintains an extensive record of all crimes that have been reported in Chicago over the last two decades including- date, type of crime, exact location, district, police ward etc.
    - Data Source: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2
    - Description: As the original dataset is extremely large and cannot be handled properly for analysis, only the data from the year 2016 has been used as representative. This data has been used to build a Choropleth map of the city of Chicago showing the intensity of crime in various neighborhoods.
    - Example of the Data:

[2]:

	Case Number	Date	Primary Type	District	Ward_Number	Community Area	Year	Latitude	Longitude
0	HZ250496	05/03/16 23:40	BATTERY	10.0	24.0	29.0	2016	41.864073	-87.706819
1	HZ250409	05/03/16 21:40	BATTERY	3.0	20.0	42.0	2016	41.782922	-87.604363
2	HZ250503	05/03/16 23:31	PUBLIC PEACE VIOLATION	15.0	37.0	25.0	2016	41.894908	-87.758372
3	HZ250424	05/03/16 22:10	BATTERY	15.0	28.0	25.0	2016	41.885687	-87.749516
4	HZ250455	05/03/16 22:00	THEFT	15.0	28.0	25.0	2016	41.886297	-87.761751

- Chicago city data which includes the names and boundaries of all neighborhoods was obtained from the Chicago Data Portal in ‘csv’ format as well as Geospatial data was obtained in ‘json’ format.
    - Data Source: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9
    - Description: Using the above data in conjunction with the *GeoPy* geocoding libraries for Python, latitudes and longitudes for each neighborhood was obtained and organized into a *Pandas* dataframe.
    - Example of the Data:
      Dataframe from the imported CSV File:

[11]:
```
# Reading the data into a Pandas Dataframe
df = pd.read_csv("chicago_neighborhoods.csv")
df.head()
```

[11]:

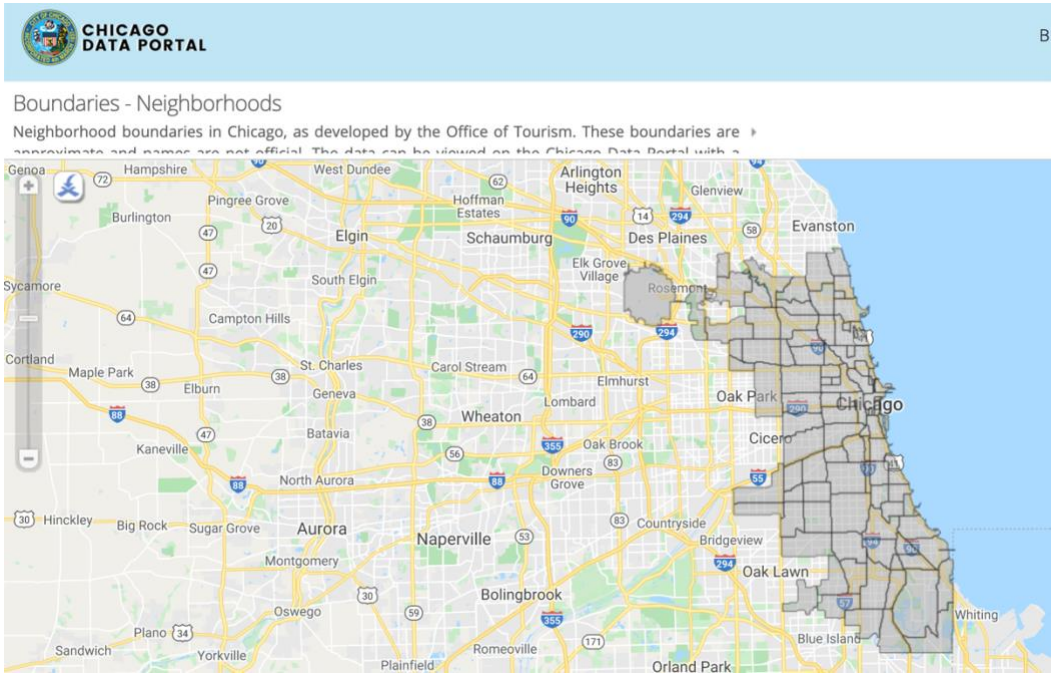
	PRI_NEIGH	the_geom	SEC_NEIGH	SHAPE_AREA	SHAPE_LEN
0	Grand Boulevard	MULTIPOLYGON (((-87.60670812560372 41.81681377...	BRONZEVILLE	4.849250e+07	28196.837157
1	Printers Row	MULTIPOLYGON (((-87.62760697485348 41.87437097...	PRINTERS ROW	2.162138e+06	6864.247156
2	United Center	MULTIPOLYGON (((-87.66706868914602 41.88885187...	UNITED CENTER	3.252051e+07	23101.363745
3	Sheffield & DePaul	MULTIPOLYGON (((-87.65833494805533 41.92166144...	SHEFFIELD & DEPAUL	1.048259e+07	13227.049745
4	Humboldt Park	MULTIPOLYGON (((-87.74059567509266 41.88782316...	HUMBOLDT PARK	1.250104e+08	46126.751351

After Cleaning the Data and Adding Latitude and Longitude:

[18]:

	Borough	Neighborhood	Latitude	Longitude
0	Grand Boulevard	BRONZEVILLE	41.813923	-87.617272
1	Printers Row	PRINTERS ROW	41.873787	-87.628900
2	United Center	UNITED CENTER	41.880683	-87.674185
3	Sheffield	SHEFFIELD & DEPAUL	41.912203	-87.654327
4	Humboldt Park	HUMBOLDT PARK	41.905767	-87.704174

Geospatial Data:



- Data about the venues and amenities within a 1km radius of each neighborhood and the locations of established Japanese restaurants in Chicago was obtained using the *Foursquare API*.
    - Data Source: Foursquare credentials can be obtained by creating a developer account on https://developer.foursquare.com/

Defining Foursquare API Credentials

[22]:
```
CLIENT_ID = '4W0W0WHMGIC31GMXQZWOWVXQQGTP5KCLWDBC33U5K0SDYKRY'
CLIENT_SECRET = 'XXHAOM32AVNZDYWS1EHRWIGLOKVCCOI3XZZSLL1KBASRDIPX'
VERSION = '20180605' #Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: 4W0W0WHMGIC31GMXQZWOWVXQQGTP5KCLWDBC33U5K0SDYKRY
CLIENT_SECRET:XXHAOM32AVNZDYWS1EHRWIGLOKVCCOI3XZZSLL1KBASRDIPX
```

    - Description: Foursquare's API was used to obtain the names of venues as well as related information such as location, customer ratings, tips etc.
    - Example:
      Defining Foursquare credentials:

Example of Venue Data obtained through the Foursquare API:

[59]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BRONZEVILLE	41.813923	-87.617272	Ain't She Sweet Cafe	41.816817	-87.613004	Coffee Shop
1	BRONZEVILLE	41.813923	-87.617272	Chicago Blues District	41.810071	-87.614105	Jazz Club
2	BRONZEVILLE	41.813923	-87.617272	Sip & Savor	41.816817	-87.612876	Coffee Shop
3	BRONZEVILLE	41.813923	-87.617272	Parkway Ballroom	41.813142	-87.616064	Food
4	BRONZEVILLE	41.813923	-87.617272	Blues Brothers Mural / Shelly's Loan & Jewelry...	41.809391	-87.619517	Plaza