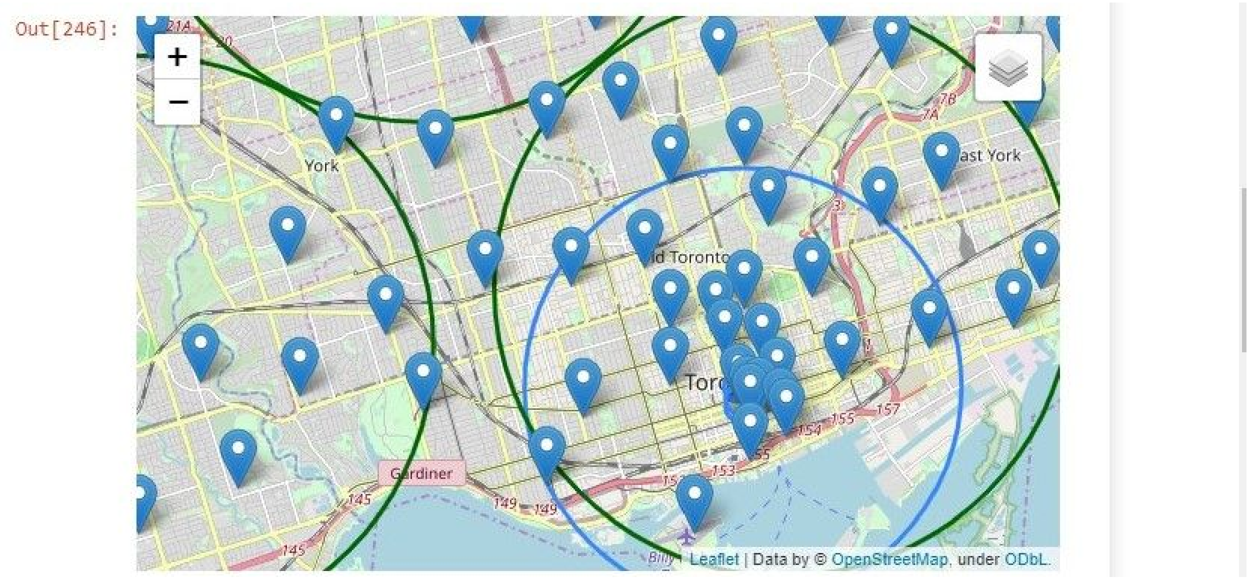
Capstone Project Report:by Nyam Ochir BOLD

Goal:

In this project I try to some interactive map jupyter notebook that can offer some nice features like showing venues around your neighborhood and listing events and will use some machine learning algorithms for real life implementation. Also, I will use several website's api for this to finish it



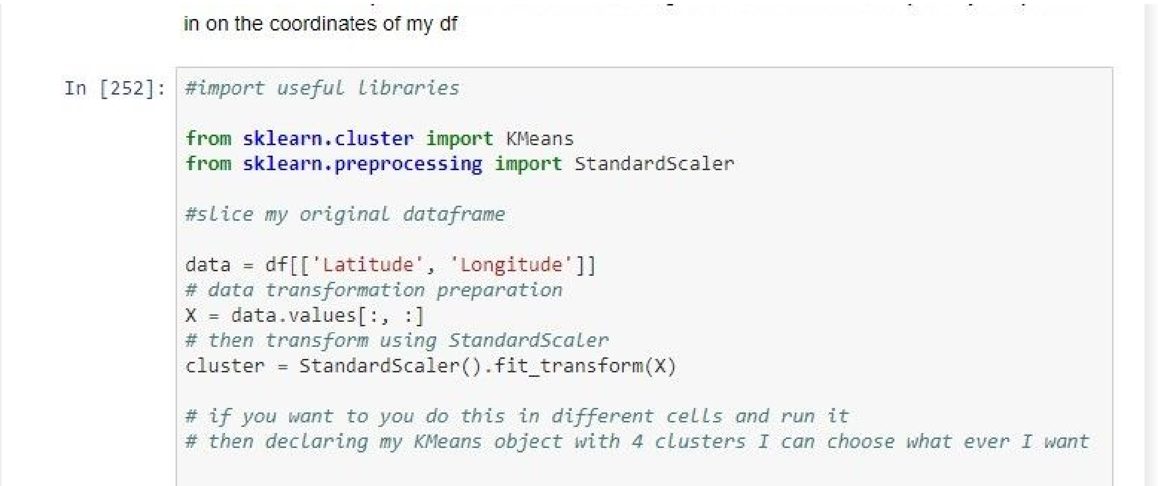
Out[246]:

It is very important to cluster the things into desired segments. This practice practice very important for many sectors of the world, for example, in The business environment, companies want to target certain sectors of their customers but how can we accomplish this task is

Problem to solve by doing this project is to make it easier for any person who have project prepared set up for running the notebook can find a good restaurant to go and good events that they can attend every outsourcing information are very high quality that is providing very better chance to have futuristic project.

Features

It starts from clustering the zipcodes, It is possible to have larger datasets, I will try that out on my android application soon enough. But currently I will show how to use KMeans cluster for very optimized clustering of the location



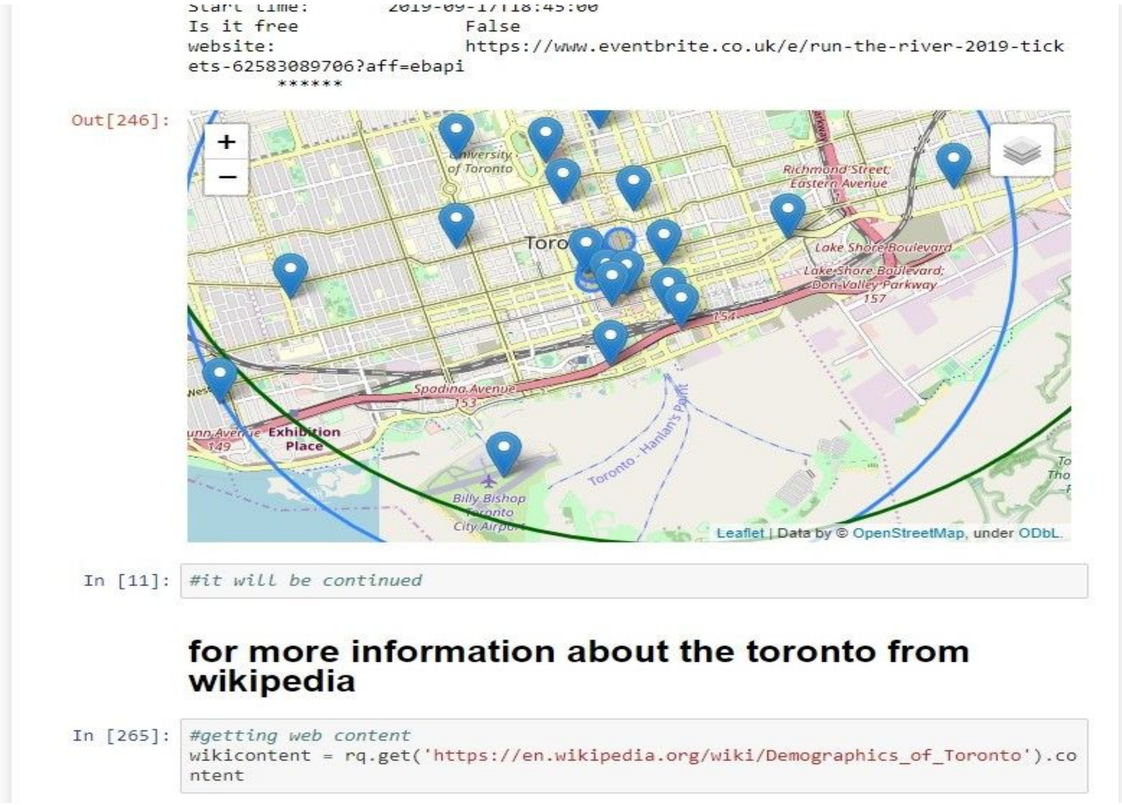
in on the coordinates of my df

In [252]:

```
#import useful libraries

from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

#slice my original dataframe

data = df[['Latitude', 'Longitude']]
# data transformation preparation
X = data.values[:, :]
# then transform using StandardScaler
cluster = StandardScaler().fit_transform(X)

# if you want to you do this in different cells and run it
# then declaring my KMeans object with 4 clusters I can choose what ever I want
```

I am putting just a part of code since it is not meaningful.

Second thing is using API extensively is really makes the world easy for example I can get the exact longitude and latitude of any certain query of the location using google maps api, this is very cool part. So using this feature I can provide very high end application with many external sources that connects together build for that certain customer that needed to have important event information in his or her neighborhood.



```
Start time:       2019-09-17T18:45:00
Is it free                False
website:                  https://www.eventbrite.co.uk/e/run-the-river-2019-tick
ets-62583089706?aff=ebapi
        ******
```

Out[246]:

+
−
University of Toronto
Richmond Street; Eastern Avenue
Toro
Lake Shore Boulevard
Lake Shore Boulevard; Don Valley Parkway
157
154
Spadina Avenue
153
Exhibition Place
149
Toronto - Hanlan's Point
Billy Bishop Toronto City Airport
Leaflet | Data by © OpenStreetMap, under ODbL.

In [11]:

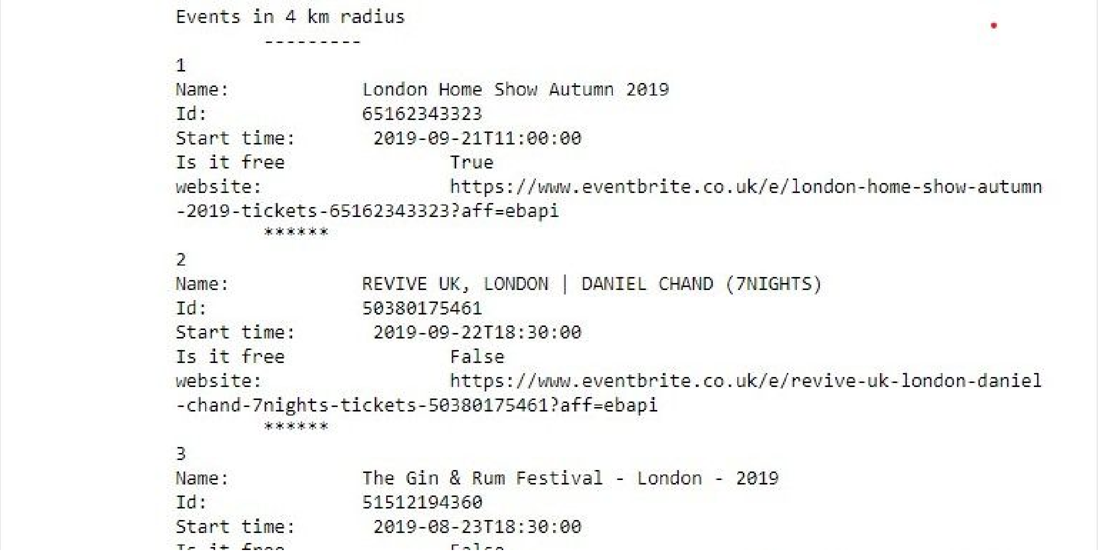
```
#it will be continued
```

for more information about the toronto from wikipedia

In [265]:

```
#getting web content
wikicontent = rq.get('https://en.wikipedia.org/wiki/Demographics_of_Toronto').co
ntent
```

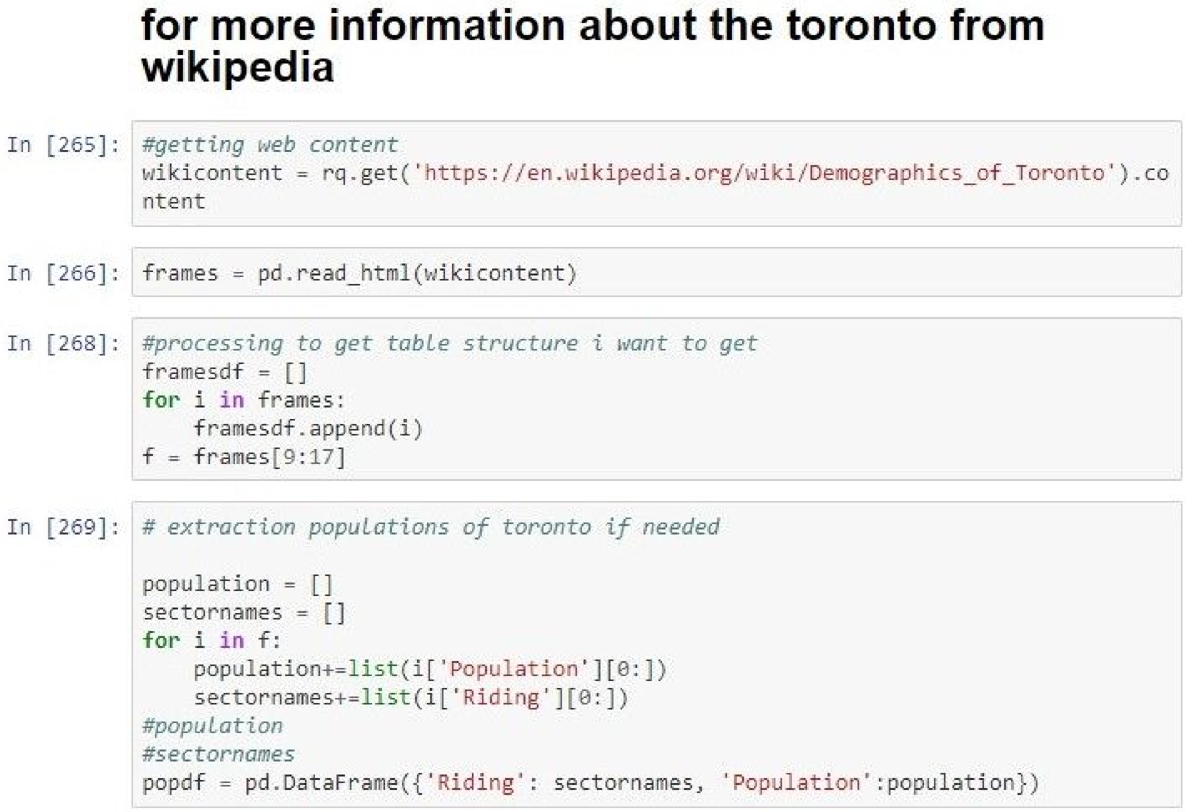
As I mention picture above there places that is political geographic units of toronto, drawn as pinned points and gray and black points are the points for place I searched and the application I build recommended to go. And also



```
Events in 4 km radius
        ---------
1
Name:            London Home Show Autumn 2019
Id:              65162343323
Start time:       2019-09-21T11:00:00
Is it free                True
website:                  https://www.eventbrite.co.uk/e/london-home-show-autumn
-2019-tickets-65162343323?aff=ebapi
        ******
2
Name:            REVIVE UK, LONDON | DANIEL CHAND (7NIGHTS)
Id:              50380175461
Start time:       2019-09-22T18:30:00
Is it free                False
website:                  https://www.eventbrite.co.uk/e/revive-uk-london-daniel
-chand-7nights-tickets-50380175461?aff=ebapi
        ******
3
Name:            The Gin & Rum Festival - London - 2019
Id:              51512194360
Start time:       2019-08-23T18:30:00
Is it free                False
```

let’s see picture this data is extracted on my application if I search for London in my query then it will print events in a list that’s in 4km radius that is very cool that a person can walk there for attending the event. There are map features that mentioned in the blogpost. There is also static data scraping from the website, which mean very unlikely to change soon in time.

For example, extracted data from wikipedia population of Toronto, by divisions



for more information about the toronto from wikipedia

In [265]:

```
#getting web content
wikicontent = rq.get('https://en.wikipedia.org/wiki/Demographics_of_Toronto').co
ntent
```

In [266]:

```
frames = pd.read_html(wikicontent)
```

In [268]:

```
#processing to get table structure i want to get
framesdf = []
for i in frames:
    framesdf.append(i)
f = frames[9:17]
```

In [269]:

```
# extraction populations of toronto if needed

population = []
sectornames = []
for i in f:
    population+=list(i['Population'][0:])
    sectornames+=list(i['Riding'][0:])
#population
#sectornames
popdf = pd.DataFrame({'Riding': sectornames, 'Population':population})
```

Data is clear but needs some understanding of json data structure or html rendering extract all the table but in this it very easy to pandas read_html method, that will return a lot of tables that I can observe one by one do lot of slicing and can get the desired output I want. By using this, I can obtain very good datasets. For my analysis that I can simply display the populations in graph. Find out from the notebook.

Insights:

The also there last graph that build that is about static data manipulation from the website.
For example statistical data on wikipedia can be considered as very static because of its nature of being not changed daily or weekly even monthly on the other, If i want to make analysis on the price of the cryptocurrencies or stocks then I need real time or very recent time data provider that I can even do predictive analysis. So the thing that I want to say is if you know the data you want, then it will be very easy for you to choose where to extract the data. There are lots of graphical application libraries in python, but in the project I have just done best suited with Folium library in python since I can represent the information very understandable.Connecting it modern GUI applications for example in android platform, direct end user or public or mass user usable tools are very possible when using those tool kits development especially machine learning algorithms with RESTFUL api would make the very nice featured application. For further information, I published an article on my blogspot it is very welcome to see and visit.

Conclusion, Using extensive tools and algorithms from python libraries I can do things but If I know how to connect them to use for bigger applications like realtime reviews and number of customers in restaurants are very possible with several platform connections.

Further readings:
https://github.com/nyamochir/Coursera_Capstone/blob/master/nb_003.html
https://medium.com/@chrieke/essential-geospatial-python-libraries-5d82fcc38731
https://developer.foursquare.com/docs/api/venues/search
https://www.eventbrite.com/platform/api
https://en.wikipedia.org/wiki/Api
https://developers.google.com/maps/documentation/
https://python-visualization.github.io/folium/
https://pandas.pydata.org/