

Capstone Project — The Battle of Neighbourhoods

1. Introduction

Lyon is the third-largest city and second-largest urban area of France. It is located at the confluence of the rivers Rhône and Saône, about 470 km south-east of Paris.

The city of Lyon proper had a population of 516,092 in 2017.

Lyon is the capital of the region of Auvergne-Rhône-Alpes as well as the seat of the Rhône departmental council. Lyon is divided into 9 of municipal arrondissements, each of which is identified by a number and has its own council.

Former capital of the Gauls at the time of the Roman Empire, Lyon is the seat of an archbishopric whose holder bears the title of Primate of the Gauls. Lyon became a major economic hub during the Renaissance. The city is recognized for its cuisine and gastronomy, and historical and architectural landmarks; as such, the districts of Old Lyon, the Fourvière hill, the Presqu'île and the slopes of the Croix-Rousse are inscribed on the UNESCO world heritage list. Lyon was historically an important area for the production and weaving of silk. Lyon played a significant role in the history of cinema: it is where Auguste and Louis Lumière invented the cinematograph. It is also known for its light festival, the Fête des Lumières, which begins every 8 December and lasts for four days, earning Lyon the title of Capital of Lights.

2. Business problem

Economically, Lyon is a major centre for banking, as well as for the chemical, pharmaceutical, and biotech industries. The city contains a significant software industry with a particular focus on video games, and in recent years has fostered a growing local start-up sector. Lyon hosts the international headquarters of Interpol, the International Agency for Research on Cancer and Euronews. According to the Globalization and World Rankings Research Institute, Lyon is considered a Beta city, as of 2018. It ranked second in France and 40th globally in Mercer's 2019 liveability rankings. so to help the stackholders when they want to run their own buisness in Lyon, we will search for the best place to open it such as a restaurat or a bakery.

3. Data source

This dataset comes from a certified public service

https://www.data.gouv.fr/fr/datasets/points-de-distribution-du-magazine-de-la-metropole-de-lyon/#_
(https://www.data.gouv.fr/fr/datasets/points-de-distribution-du-magazine-de-la-metropole-de-lyon/#_)

4. Data Exploration

In [2]:

```
import numpy as np
import pandas as pd

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import requests
from bs4 import BeautifulSoup
from sklearn.cluster import KMeans
import matplotlib.cm as cm
import matplotlib.colors as colors
```

In [24]:

```
import types
import pandas as pd
from botocore.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell

df_data_1 = pd.read_csv(body, ";")
#filtered_df = df_data_1[df_data_1['code_postal'].isnull()]
df_data_1.loc[(df_data_1['nom'] == 'Bibliothèque Lyon Sud'), 'code_postal'] = 69310
#filtered_df
df_data_1["code_postal"] = df_data_1["code_postal"].astype(int)
df_data_1["code_postal"] = df_data_1["code_postal"].astype(str)
```

In []:

In [25]:

```
lyon_data = df_data_1[df_data_1["code_postal"].str.contains("6900")]
lyon_data = lyon_data[lyon_data["sous_type"].str.contains("Mairie")]
lyon_data = lyon_data[lyon_data["nom"].str.contains("arrondissement")]
lyon_data.columns = ["nom", "adresse", "adresse_complement", "code_postal", "ville", "type",
"sous_type", "accessibilite", "commentaire", "code_insee", "longitude", "latitude"]
lyon_data.head(9)
```

Out[25]:

	nom	adresse	adresse_complement	code_postal	ville	type	sous_type
275	Mairie de Lyon 1er arrondissement	2 place Sathonay	NaN	69001	Lyon 1er	Public	Mairie
333	Mairie de Lyon 2ème arrondissement	2 rue d'Enghien	NaN	69002	Lyon 2ème	Public	Mairie
418	Mairie de Lyon 3ème arrondissement	215 rue Duguesclin	NaN	69003	Lyon 3ème	Public	Mairie
493	Mairie de Lyon 4ème arrondissement	133 boulevard de la Croix- Rousse	NaN	69004	Lyon 4ème	Public	Mairie
529	Mairie de Lyon 5ème arrondissement	14 rue Dr Edmond Locard	NaN	69005	Lyon 5ème	Public	Mairie
564	Mairie de Lyon 6ème arrondissement	58 rue de Sèze	NaN	69006	Lyon 6ème	Public	Mairie
611	Mairie de Lyon 7ème arrondissement	16 place Jean Macé	NaN	69007	Lyon 7ème	Public	Mairie
675	Mairie de Lyon 8ème arrondissement	12 avenue Jean Mermoz	NaN	69008	Lyon 8ème	Public	Mairie
713	Mairie de Lyon 9ème arrondissement	6 place du marché	NaN	69009	Lyon 9ème	Public	Mairie

In [26]:

```
lyon_data.shape
```

Out[26]:

(9, 12)

In [22]:

```
! pip install folium==0.5.0
import folium
```

Collecting folium==0.5.0

Downloading folium-0.5.0.tar.gz (79 kB)

|██| 79 kB 8.1 MB/s eta 0:00:01

Collecting branca

Downloading branca-0.4.2-py3-none-any.whl (24 kB)

Requirement already satisfied: jinja2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium==0.5.0) (2.11.2)

Requirement already satisfied: requests in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium==0.5.0) (2.24.0)

Requirement already satisfied: six in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from folium==0.5.0) (1.15.0)

Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from jinja2->folium==0.5.0) (1.1.1)

Requirement already satisfied: chardet<4,>=3.0.2 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium==0.5.0) (3.0.4)

Requirement already satisfied: idna<3,>=2.5 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium==0.5.0) (2.9)

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium==0.5.0) (2020.12.5)

Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /opt/conda/envs/Python-3.7-main/lib/python3.7/site-packages (from requests->folium==0.5.0) (1.25.9)

Building wheels for collected packages: folium

Building wheel for folium (setup.py) ... done

Created wheel for folium: filename=folium-0.5.0-py3-none-any.whl size=76240 sha256=fbcfb24d93b4da910473bf40156f6c4a7fdb43f84c560f56965b62339266170ff

Stored in directory: /tmp/wsuser/.cache/pip/wheels/b2/2f/2c/109e446b990d663ea5ce9b078b5e7c1a9c45cca91f377080f8

Successfully built folium

Installing collected packages: branca, folium

Successfully installed branca-0.4.2 folium-0.5.0

In [23]:

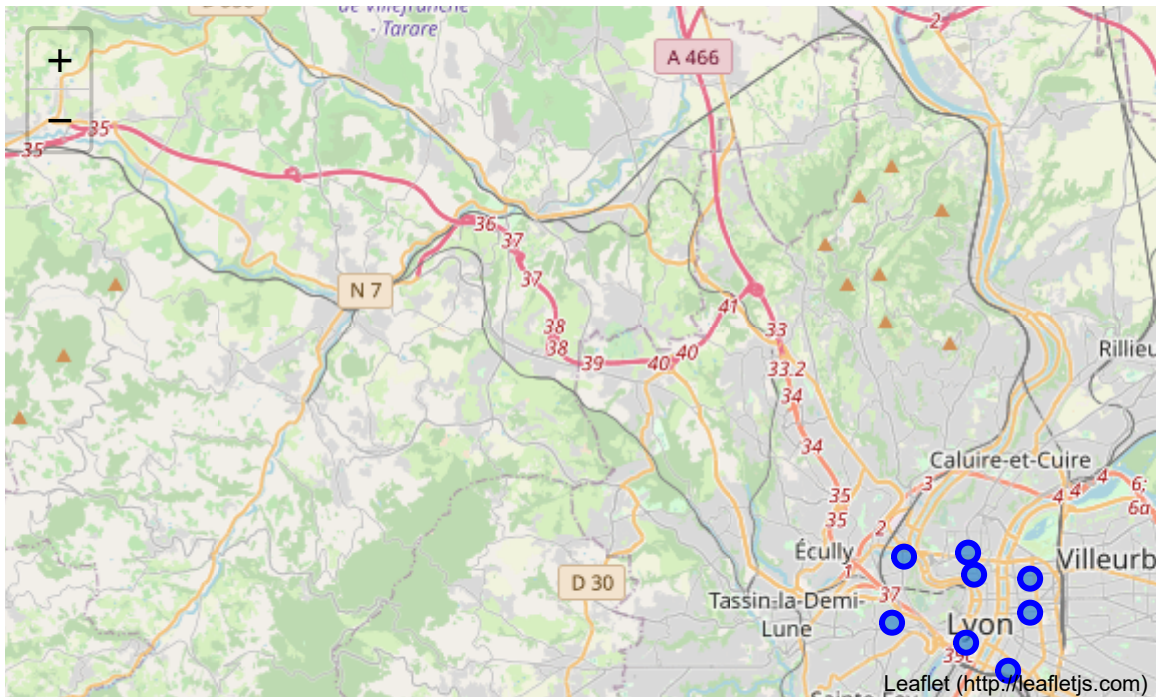
```
latitude =45.764043
longitude =4.835659

# create map of Lyon using Latitude and Longitude values
map_lyon = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, borough, neighborhood in zip(lyon_data['latitude'], lyon_data['longitude'],
    lyon_data['ville'], lyon_data['nom']):
    label = '{} , {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_lyon)

map_lyon
```

Out[23]:



In [27]:

```
# @hidden_cell
```

In [28]:

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]["groups"][0]["items"]

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                            'Neighborhood Latitude',
                            'Neighborhood Longitude',
                            'Venue',
                            'Venue Latitude',
                            'Venue Longitude',
                            'Venue Category']

    return(nearby_venues)
```

In [29]:

```
lyon_venues = getNearbyVenues(names=lyon_data['nom'],  
                               latitudes=lyon_data['latitude'],  
                               longitudes=lyon_data['longitude']  
                               )
```

Mairie de Lyon 1er arrondissement
Mairie de Lyon 2ème arrondissement
Mairie de Lyon 3ème arrondissement
Mairie de Lyon 4ème arrondissement
Mairie de Lyon 5ème arrondissement
Mairie de Lyon 6ème arrondissement
Mairie de Lyon 7ème arrondissement
Mairie de Lyon 8ème arrondissement
Mairie de Lyon 9ème arrondissement

In [31]:

```
print(lyon_venues.shape)  
lyon_venues.head()
```

(259, 7)

Out[31]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mairie de Lyon 1er arrondissement	45.769497	4.830127	Hopstore	45.768267	4.830576	Bar
1	Mairie de Lyon 1er arrondissement	45.769497	4.830127	Fresque des Lyonnais	45.768189	4.828085	Public Art
2	Mairie de Lyon 1er arrondissement	45.769497	4.830127	Le Tasse Livre	45.768711	4.830788	Café
3	Mairie de Lyon 1er arrondissement	45.769497	4.830127	Bangkok royal	45.768773	4.831194	Thai Restaurant
4	Mairie de Lyon 1er arrondissement	45.769497	4.830127	Soda Bar	45.768216	4.829118	Cocktail Bar

In [32]:

```
lyon_venues.groupby('Neighborhood').count()
```

Out[32]:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Mairie de Lyon 1er arrondissement	59	59	59	59	59	59
Mairie de Lyon 2ème arrondissement	57	57	57	57	57	57
Mairie de Lyon 3ème arrondissement	35	35	35	35	35	35
Mairie de Lyon 4ème arrondissement	33	33	33	33	33	33
Mairie de Lyon 5ème arrondissement	8	8	8	8	8	8
Mairie de Lyon 6ème arrondissement	31	31	31	31	31	31
Mairie de Lyon 7ème arrondissement	21	21	21	21	21	21
Mairie de Lyon 8ème arrondissement	4	4	4	4	4	4
Mairie de Lyon 9ème arrondissement	11	11	11	11	11	11

In [33]:

```
print('There are {} uniques categories.'.format(len(lyon_venues['Venue Category'].unique())))
```

There are 83 uniques categories.

In [34]:

```
# one hot encoding
lyon_onehot = pd.get_dummies(lyon_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
lyon_onehot['Neighborhood'] = lyon_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [lyon_onehot.columns[-1]] + list(lyon_onehot.columns[:-1])
lyon_onehot = lyon_onehot[fixed_columns]

lyon_onehot.head()
```

Out[34]:

	Wine Bar	Art Museum	Asian Restaurant	Bagel Shop	Bakery	Bar	Beer Garden	Big Box Store	Bistro	Bookstore	Boutique
0	0	0	0	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0

In [37]:

```
lyon_onehot.shape
```

Out[37]:

(259, 83)

In [36]:

```
lyon_grouped = lyon_onehot.groupby('Neighborhood').mean().reset_index()
lyon_grouped
```

Out[36]:

	Neighborhood	Wine Bar	Art Museum	Asian Restaurant	Bagel Shop	Bakery	Bar	Beer Garden	S
0	Mairie de Lyon 1er arrondissement	0.016949	0.016949	0.000000	0.000000	0.033898	0.135593	0.000000	0
1	Mairie de Lyon 2ème arrondissement	0.017544	0.017544	0.017544	0.017544	0.035088	0.000000	0.000000	0
2	Mairie de Lyon 3ème arrondissement	0.057143	0.000000	0.028571	0.000000	0.000000	0.000000	0.000000	0
3	Mairie de Lyon 4ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.090909	0.060606	0.030303	0
4	Mairie de Lyon 5ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0
5	Mairie de Lyon 6ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.064516	0.000000	0.032258	0
6	Mairie de Lyon 7ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.047619	0.047619	0.000000	0
7	Mairie de Lyon 8ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0
8	Mairie de Lyon 9ème arrondissement	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0

In [38]:

```
lyon_grouped.shape
```

Out[38]:

(9, 83)

In [40]:

```
num_top_venues = 5

for hood in lyon_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = lyon_grouped[lyon_grouped['Neighborhood'] == hood].T.reset_index()
    temp.columns = ['venue', 'freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

----Mairie de Lyon 1er arrondissement ----

	venue	freq
0	French Restaurant	0.15
1	Bar	0.14
2	Plaza	0.07
3	Café	0.05
4	Pub	0.03

----Mairie de Lyon 2ème arrondissement----

	venue	freq
0	French Restaurant	0.14
1	Hotel	0.11
2	Sandwich Place	0.07
3	Café	0.07
4	Pub	0.05

----Mairie de Lyon 3ème arrondissement----

	venue	freq
0	French Restaurant	0.14
1	Wine Bar	0.06
2	Seafood Restaurant	0.06
3	Hotel	0.06
4	Plaza	0.06

----Mairie de Lyon 4ème arrondissement----

	venue	freq
0	French Restaurant	0.12
1	Pub	0.09
2	Bakery	0.09
3	Bar	0.06
4	Plaza	0.06

----Mairie de Lyon 5ème arrondissement----

	venue	freq
0	Farmers Market	0.12
1	Home Service	0.12
2	Plaza	0.12
3	Bakery	0.12
4	Paper / Office Supplies Store	0.12

----Mairie de Lyon 6ème arrondissement----

	venue	freq
0	French Restaurant	0.16
1	Japanese Restaurant	0.10
2	Café	0.06
3	Chinese Restaurant	0.06
4	Plaza	0.06

----Mairie de Lyon 7ème arrondissement----

	venue	freq
0	Restaurant	0.14
1	Japanese Restaurant	0.10
2	Vegetarian / Vegan Restaurant	0.05
3	Smoke Shop	0.05
4	Indie Movie Theater	0.05

----Mairie de Lyon 8ème arrondissement----

	venue	freq
0	Resort	0.25
1	Supermarket	0.25
2	Comedy Club	0.25
3	Hotel	0.25
4	Tram Station	0.00

----Mairie de Lyon 9ème arrondissement----

	venue	freq
0	Pizza Place	0.18
1	Supermarket	0.09
2	Food & Drink Shop	0.09
3	Hotel	0.09
4	Department Store	0.09

In [41]:

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [57]:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = lyon_grouped['Neighborhood']

for ind in np.arange(lyon_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(lyon_grouped.
iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head(9)
```

Out[57]:

	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	7 C
0	Mairie de Lyon 1er arrondissement	French Restaurant	Bar	Plaza	Café	Bakery	Cocktail Bar	
1	Mairie de Lyon 2ème arrondissement	French Restaurant	Hotel	Café	Sandwich Place	Pub	Plaza	Re
2	Mairie de Lyon 3ème arrondissement	French Restaurant	Plaza	Japanese Restaurant	Seafood Restaurant	Hotel	Wine Bar	(
3	Mairie de Lyon 4ème arrondissement	French Restaurant	Bakery	Pub	Bar	Café	Plaza	S
4	Mairie de Lyon 5ème arrondissement	Garden	Paper / Office Supplies Store	Home Service	Bakery	Fast Food Restaurant	Big Box Store	
5	Mairie de Lyon 6ème arrondissement	French Restaurant	Japanese Restaurant	Bakery	Chinese Restaurant	Plaza	Café	
6	Mairie de Lyon 7ème arrondissement	Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Bistro	French Restaurant	Hotel	Indi
7	Mairie de Lyon 8ème arrondissement	Resort	Hotel	Comedy Club	Supermarket	Vegetarian / Vegan Restaurant	Diner	Coffi
8	Mairie de Lyon 9ème arrondissement	Pizza Place	Plaza	Food & Drink Shop	Fast Food Restaurant	Hotel	Italian Restaurant	Supe



5.Cluster Neighborhoods

In [58]:

```
# set number of clusters
kclusters = 5

lyon_grouped_clustering = lyon_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(lyon_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[58]:

```
array([0, 0, 0, 0, 4, 0, 3, 1, 2], dtype=int32)
```


In [59]:

```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

lyon_merged = lyon_data

# merge Lyon_grouped with Lyon_data to add latitude/longitude for each neighborhood
lyon_merged = lyon_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='nom')

lyon_merged.head(9) # check the last columns!
```

Out[59]:

	nom	adresse	adresse_complement	code_postal	ville	type	sous_type
275	Mairie de Lyon 1er arrondissement	2 place Sathonay	NaN	69001	Lyon 1er	Public	Mairie
333	Mairie de Lyon 2ème arrondissement	2 rue d'Enghien	NaN	69002	Lyon 2ème	Public	Mairie
418	Mairie de Lyon 3ème arrondissement	215 rue Duguesclin	NaN	69003	Lyon 3ème	Public	Mairie
493	Mairie de Lyon 4ème arrondissement	133 boulevard de la Croix- Rousse	NaN	69004	Lyon 4ème	Public	Mairie
529	Mairie de Lyon 5ème arrondissement	14 rue Dr Edmond Locard	NaN	69005	Lyon 5ème	Public	Mairie
564	Mairie de Lyon 6ème arrondissement	58 rue de Sèze	NaN	69006	Lyon 6ème	Public	Mairie
611	Mairie de Lyon 7ème arrondissement	16 place Jean Macé	NaN	69007	Lyon 7ème	Public	Mairie
675	Mairie de Lyon 8ème arrondissement	12 avenue Jean Mermoz	NaN	69008	Lyon 8ème	Public	Mairie
713	Mairie de Lyon 9ème arrondissement	6 place du marché	NaN	69009	Lyon 9ème	Public	Mairie

In [53]:

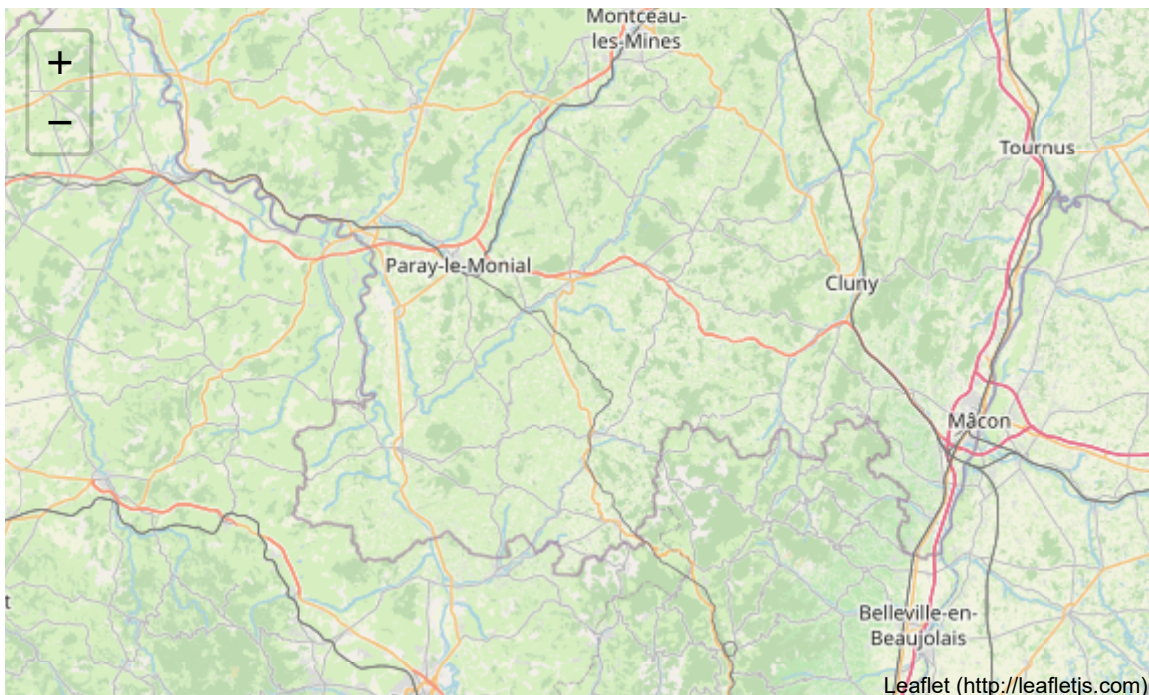
```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(lyon_merged['latitude'], lyon_merged['longitude'], lyon_merged['nom'], lyon_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[53]:



As result

- 1- The most common venues in Lyon are French Restaurant.
- 2- Cluster 2, 3 and 5 don't have an French Restaurant.
- 3- From the geographical representation of the clusters, Lyon 5th 8th 9th seems a good place open an French Restaurant.

6. Conclusion

We did the analysis on small set of data, we can achieve better results by increasing the neighborhood information. in Lyon there are many different types of new restaurant business to offer and I think we have gone through the process of identifying the business problem, specifying the data required, clean the datasets, performing a machine learning algorithm using k-means clustering and providing some useful tips to our stakeholder.