Analysing cultural activity of the neighbourhoods of Paris in order to chose the best place for starting a new restaurant

Denis LALLEE

Coursera – IBM

Capstone Project - The Battle of Neighborhoods-

Introduction

- ✓ Paris is the most populous municipality and the capital of France.
- ✓ The city of Paris has 2.187 million inhabitants called Parisians. The Parisian agglomeration has 10.73 million inhabitants and its urban area has 12.78 million inhabitants. The Parisian agglomeration is thus the most populous in France, the third in Europe and the 32nd in the world.
- ✓ Paris symbolizes French culture and is nicknamed the City of Light, it is home to world famous monuments and is one of the most visited cities in the world.
- ✓ Paris has three opera houses, hundreds of theaters and cabarets, as well as a large number of music halls and performance venues .

Source: https://fr.wikipedia.org/wiki/Paris

Business problem and background

- ✓ Find the best place to open a new restaurant
- ✓ For this, we have to do a neighbourhood choice which would be probably the most likely to give a profitable business.
- ✓ Neighbourhood is one of choice inducteurs, like number of competitors and type of proposed services for instance, but probably than the level of cultural activity available in the neighbourhood , with the presence of several different venues close from the futur business, will help to develop it.
- ✓ Starting hypothesis: the more the cultural activity is available, the more competitors there are.

Description of the data

- Source for neighbourhood of Paris
  - Wikipedia page via data scraping: https://fr.wikipedia.org/wiki/Liste_des_quartiers_administratifs_de_Paris
- Geographical coordinates of the neighbourhoods
  - Use of GeoPy library: https://geopy.readthedocs.io/en/stable/
  - And ArcGIS Online Geocoding Service: https://geocode.arcgis.com/arcgis/
- “Density” of cultural places and competing restaurants by neighbourhood
  - Use of Foursquare API: https://developer.foursquare.com/docs/places-api/

Methodology (1/2)

- ✓ Selection of all venues by neighborhood / categories by neighborhood

Entrée [8]:

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):
"""
Function for getting all venues given in parameter "names"
Return a DF with venue name, venue latitude & longitude, and its category (for each location name passed in parameters)
"""
    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        # create 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)

        # 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)
```

Entrée [ ]:

```
paris_venues = getNearbyVenues(names=df['Neighborhood'], latitudes=df['Latitude'],longitudes=df['Longitude'])
```

Methodology (2/2)

- ✓ One hot encoding

Entrée [53]:

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

# add neighborhood column back to dataframe
paris_onehot['Neighborhood'] = paris_venues['Neighborhood']

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

- ✓ K-Means Clustering

Entrée [74]:

```
# set number of clusters
kclusters = 8

paris_grouped_clustering = paris_grouped.drop('Neighborhood', 1)

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

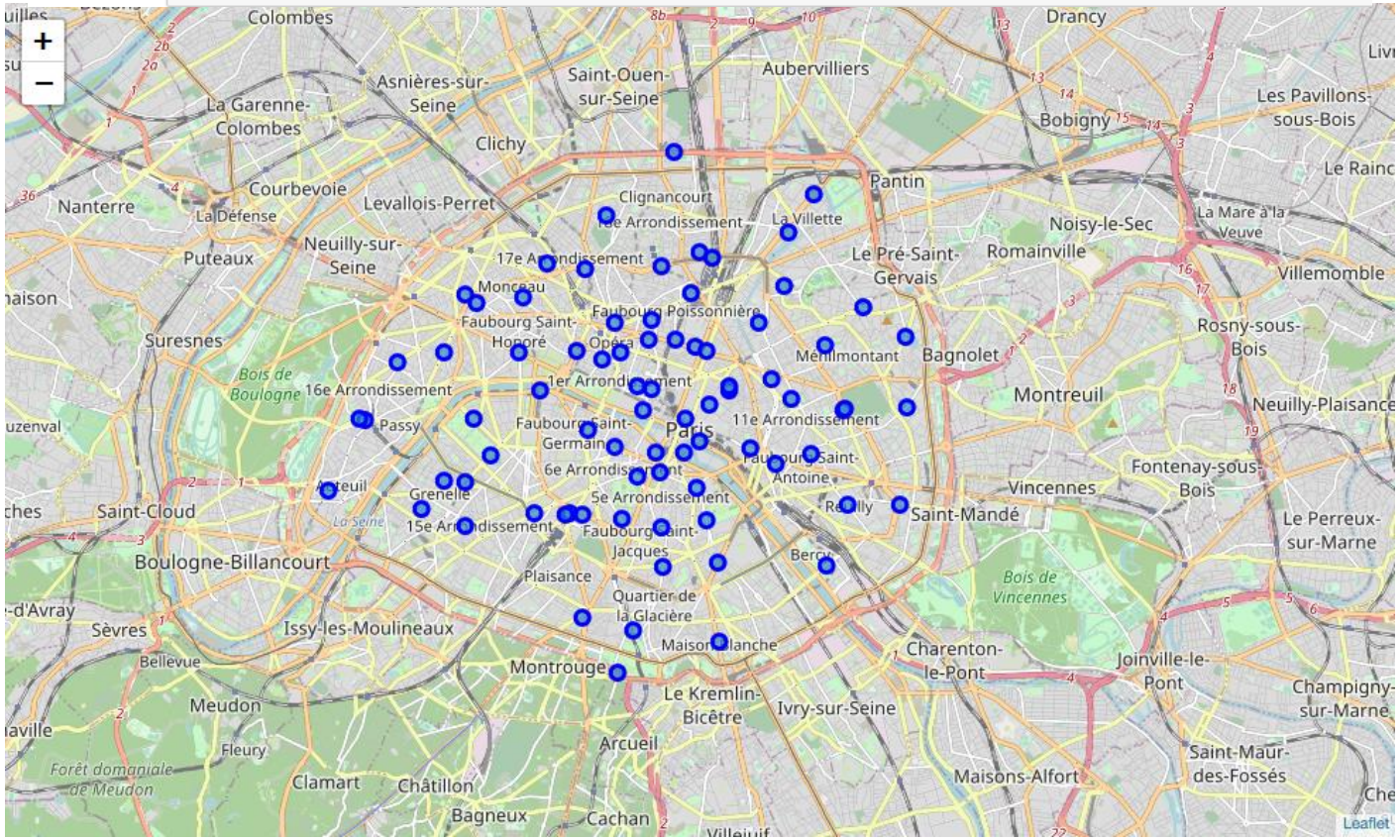
Plotting on map with neighborhood coordinates

Entrée [7]:

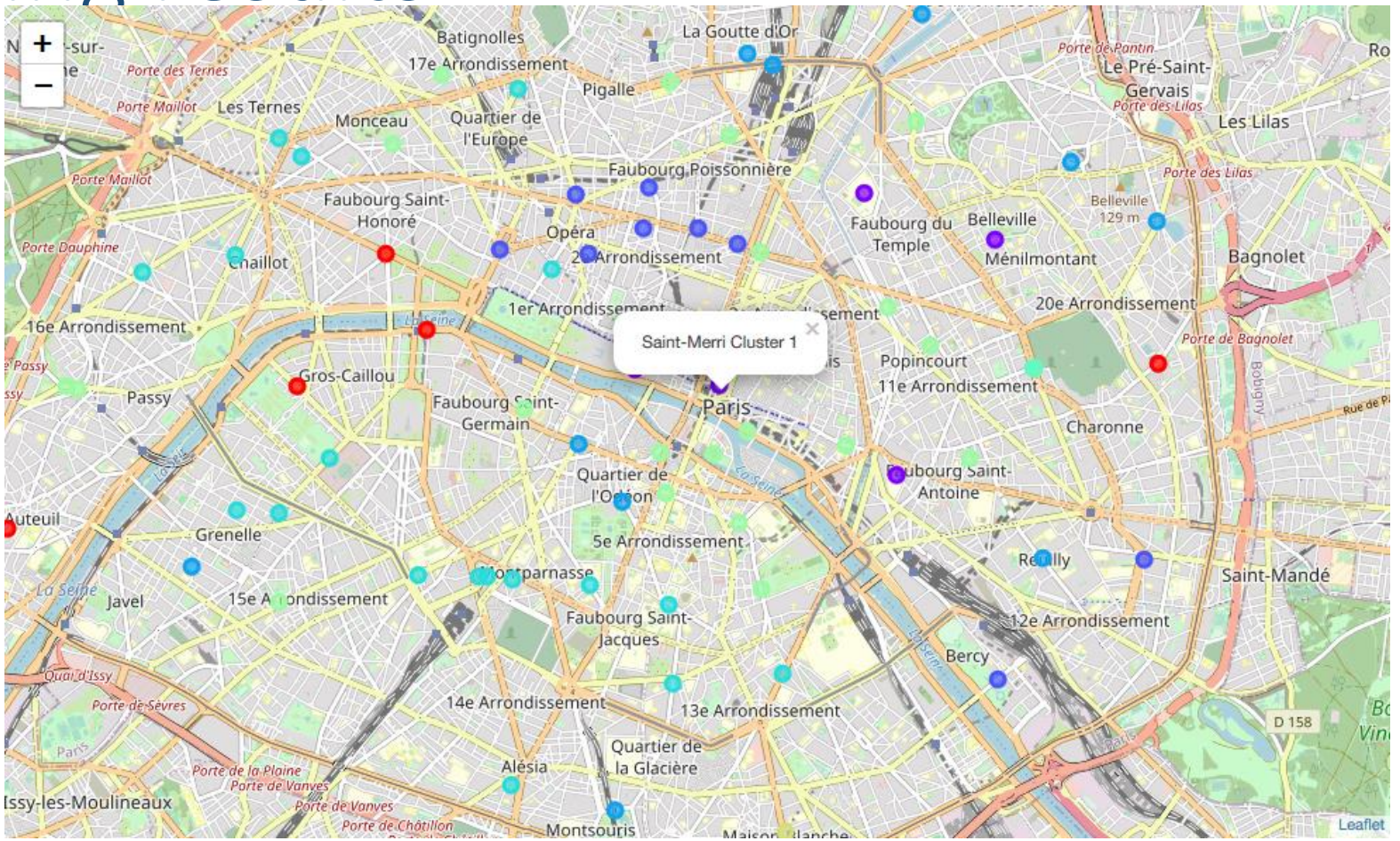
```
# create map of Paris using latitude and longitude values
map_paris = folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, bor, neigh in zip(df['Latitude'], df['Longitude'], df['Borough'], df['Neighborhood']):
    label = '{} {}'.format(bor, neigh)
    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_paris)

map_paris
```



Clustering results



	Neighborhood	Longitude	Cluster Labels	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
12	Saint-Merri	2.350530	1	Art Gallery	French Restaurant	Plaza	Restaurant	Ice Cream Shop	Art Museum	Park	Lebanese Restaurant	Pub	Burger Joint

Discussion

- ✓ The most suitable neighborhood with cultural activities are mainly in cluster 2, 3, 4 et 5.
- ✓ The cluster 5, with Theater venues, could a good choice for starting an evening restaurant.
- ✓ Clusters 3 and 4 are place for Museums and Historic sites, so more for day restaurant, but the competitors are already numerous.
- ✓ The cluster 2 has a large variety of cultural activities (Museums, Art galleries, Historic sites, Exhibits, Theatres) so a possibility to start a restaurant with spread opened hours on all day. And concurrence seems to be less important.
- ✓ The best neighborhood could be “St Merri” to open a new restaurant ou food business.

Conclusion

- ✓ BeautifulSoup, GeoPy and Folium are very useful Python libraries which are quite easy to use and give a real avantage to analyse geographical data.
- ✓ Data analysis and machine learning can be a help to deals with some complex business problems.
- ✓ The K-means algorithm, for the unsupervised learning, has help in our case to cluster the similar Paris neighborhoods.
- ✓ Finally, we can give a recommendation and advice (and the better places) for who want to launch a business such a new restaurant.