The Battle of Neighborhoods | Business Proposal

Pooriya Ghanbari Merdasi

13th April 2021

Introduction

Paris is a quite the popular tourist and vacation destination for people all around the world. The city is diverse and multicultural and offers a wide variety of experiences that are widely sought after. We try to group the neighbourhoods of Paris and draw insights to what they look like.

Business Problem

The aim is to help tourists choose their destinations depending on the experiences that the neighbourhoods have to offer and what they would want to have. This also helps people make decisions if they are thinking about migrating to Paris or even if they want to relocate neighbourhoods within the city. Our findings will help stakeholders make informed decisions and address any concerns they have including the different kinds of cuisines, provision stores and what the city has to offer.

Data Description:

We require geolocation data for Paris. Postal codes in each city serve as a starting point. Using Postal codes we use can find out the neighbourhoods, boroughs, venues and their most popular venue categories.

To derive our solution, We leverage JSON data available at https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
The JSON file has data about all the neighbourhoods in France, we limit it to Paris.

postal_code : Postal codes for France
nom_comm : Name of Neighbourhoods in France
nom_dept : Name of the boroughs, equivalent to towns in France
geo_point_2d : Tuple containing the latitude and longitude of the Neighbourhoods.

Foursquare API Data

We will need data about different venues in different neighbourhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighbourhood. For each neighbourhood, we have chosen the radius to be 500 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

Neighbourhood : Name of the Neighbourhood
Neighbourhood Latitude : Latitude of the Neighbourhood
Neighbourhood Longitude : Longitude of the Neighbourhood
Venue : Name of the Venue
Venue Latitude : Latitude of Venue
Venue Longitude : Longitude of Venue
Venue Category : Category of Venue

Methodology

We will be creating our model with the help of Python so we start off by importing all the required packages.

```
import pandas as pd
import requests
import numpy as np
import matplotlib.cm as cm
import matplotlib.colors as colors
import folium

# import k-means for the clustering stage
from sklearn.cluster import KMeans
```

Neighbourhoods of Paris

Data Collection

We read the json data with pandas.

```
!wget -q -O 'france-data.json' https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
print("Data Downloaded!")
```

```
Data Downloaded!
```

```
paris_raw = pd.read_json('france-data.json')
paris_raw.head()
```

	datasetid	recordid	fields	geometry	record_timestamp
0	correspondances-code-insee-code-postal	21e809b1d4480333c8b6fe7addd8f3b06f343e2c	{'code_comm': '003', 'nom_dept': 'VAL-DE-MARNE...	{'type': 'Point', 'coordinates': [2.3335102498...	2016-09-21T00:29:06.175+02:00
1	correspondances-code-insee-code-postal	c38925e974a8875071da3eb1391a6935d9c97e07	{'code_comm': '430', 'nom_dept': 'SEINE-ET-MAR...	{'type': 'Point', 'coordinates': [2.7879422114...	2016-09-21T00:29:06.175+02:00
2	correspondances-code-insee-code-postal	7c0aa8ba7a7b4320a9cf5abf12288eb76e3eead8	{'code_comm': '412', 'nom_dept': 'SEINE-ET-MAR...	{'type': 'Point', 'coordinates': [2.5107818983...	2016-09-21T00:29:06.175+02:00
3	correspondances-code-insee-code-postal	b123405b4d069c33725418aab20ca0b741f8a5d8	{'code_comm': '598', 'nom_dept': 'VAL-D'OISE',...	{'type': 'Point', 'coordinates': [2.3004997834...	2016-09-21T00:29:06.175+02:00
4	correspondances-code-insee-code-postal	33dea89ab43606076200134a51f2b9d2d7d62256	{'code_comm': '040', 'nom_dept': 'SEINE-ET-MAR...	{'type': 'Point', 'coordinates': [2.5699190953...	2016-09-21T00:29:06.175+02:00

Data Preprocessing

We break down each of the nested fields and create the dataframe that we need

```
paris_field_data = pd.DataFrame()
for f in paris_raw.fields:
    dict_new = f
    paris_field_data = paris_field_data.append(dict_new, ignore_index=True)

paris_field_data.head()
```

	code_arr	code_cant	code_comm	code_dept	code_reg	geo_point_2d	geo_shape	id_geofla	insee_com	nom_comm	nom_dept	nom_regic
0	3	34	003	94	11	[48.80588035965699, 2.333510249842654]	{'type': 'Polygon', 'coordinates': [[[2.343851...	32123	94003	ARCUEIL	VAL-DE-MARNE	ILE-DE-FRANCE
1	1	09	430	77	11	[49.07375705850074, 2.787942211427862]	{'type': 'Polygon', 'coordinates': [[[2.779756...	33986	77430	SAINT-PATHUS	SEINE-ET-MARNE	ILE-DE-FRANCE
2	2	32	412	77	11	[48.47430724856196, 2.510781898359126]	{'type': 'Polygon', 'coordinates': [[[2.505938...	18225	77412	SAINT-GERMAIN-SUR-ECOLE	SEINE-ET-MARNE	ILE-DE-FRANCE
3	2	24	598	95	11	[48.98840778751709, 2.300499783419528]	{'type': 'Polygon', 'coordinates': [[[2.297857...	32145	95598	SOISY-SOUS-MONTMORENCY	VAL-D'OISE	ILE-DE-FRANCE
4	2	32	040	77	11	[48.51197791258124, 2.569919095331142]	{'type': 'Polygon', 'coordinates': [[[2.558809...	28430	77040	BOISSISE-LE-ROI	SEINE-ET-MARNE	ILE-DE-FRANCE

Feature Selection

We take the columns that we require, in case of paris it would be the geo_point_2d, nom_dept, nom_comm and postal_code

```
df_2 = paris_field_data[['postal_code','nom_comm','nom_dept','geo_point_2d']]
df_2
```

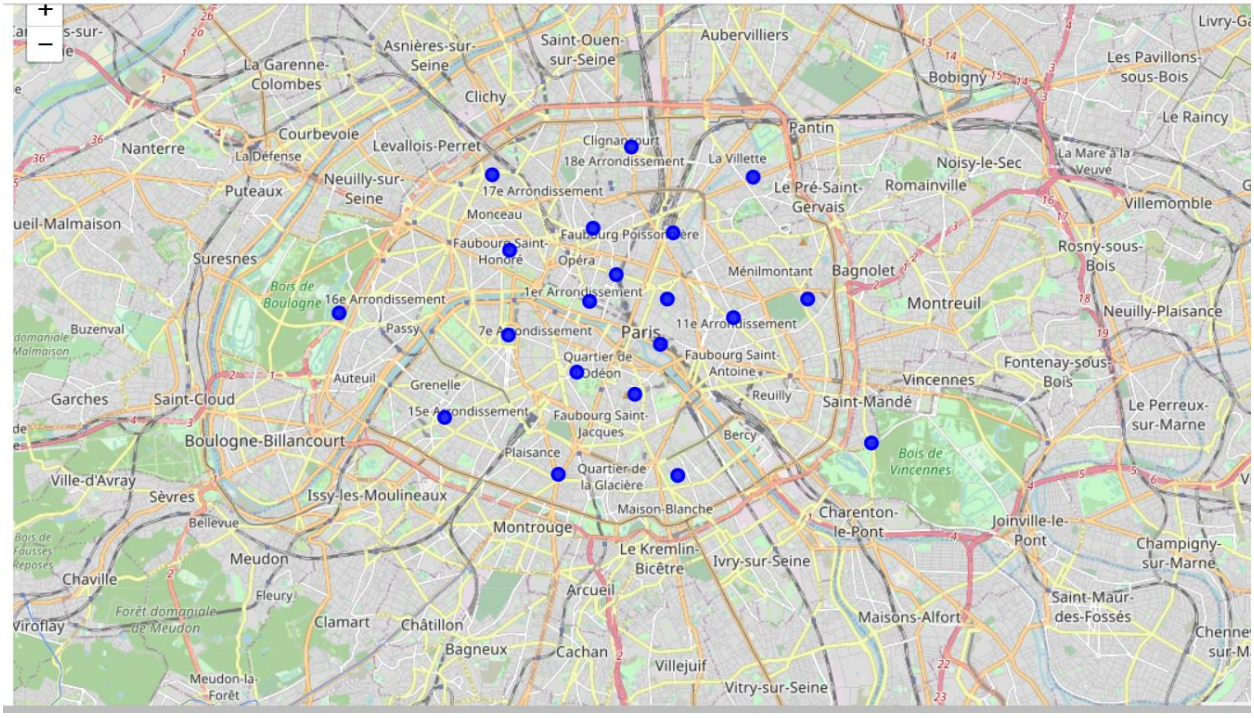
	postal_code	nom_comm	nom_dept	geo_point_2d
0	94110	ARCUEIL	VAL-DE-MARNE	[48.80588035965699, 2.333510249842654]
1	77178	SAINT-PATHUS	SEINE-ET-MARNE	[49.07375705850074, 2.787942211427862]
2	77930	SAINT-GERMAIN-SUR-ECOLE	SEINE-ET-MARNE	[48.47430724856196, 2.510781898359126]
3	95230	SOISY-SOUS-MONTMORENCY	VAL-D'OISE	[48.98840778751709, 2.300499783419528]
4	77310	BOISSISE-LE-ROI	SEINE-ET-MARNE	[48.51197791258124, 2.569919095331142]
...	...	...	...	...
1295	77230	VINANTES	SEINE-ET-MARNE	[49.00336689019137, 2.739932368905603]
1296	91350	GRIGNY	ESSONNE	[48.6568896115566, 2.387896478580817]
1297	77440	JAIGNES	SEINE-ET-MARNE	[48.98705524224735, 3.074558248708759]
1298	77220	GRETZ-ARMAINVILLIERS	SEINE-ET-MARNE	[48.74953288926905, 2.7262357852056303]
1299	77580	VOULANGIS	SEINE-ET-MARNE	[48.84031218010945, 2.889716285735505]

1300 rows × 4 columns

We have managed to collect 1300 records.

We localize to only include Paris

We gather the Gelocations of the Neighbourhoods of Paris and visualize the map of Paris



We then get the nearby venues in each neighborhood and then group venues by category.

Grouping by Venue Categories

We need to now see how many Venue Categories are there for further processing

```
venues_in_Paris.groupby('Venue Category').max()
```

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue
Venue Category				
Afghan Restaurant	PARIS-11E-ARRONDISSEMENT	48.859415	2.378741	Afghanistan
African Restaurant	PARIS-9E-ARRONDISSEMENT	48.876896	2.361113	Wally Le Saharien
American Restaurant	PARIS-19E-ARRONDISSEMENT	48.892735	2.384694	Harper's
Antique Shop	PARIS-9E-ARRONDISSEMENT	48.876896	2.337460	Hôtel des Ventes Drouot
Argentinian Restaurant	PARIS-3E-ARRONDISSEMENT	48.863054	2.359361	Anahi
...	...	...	...	...
Wine Bar	PARIS-9E-ARRONDISSEMENT	48.892735	2.400820	Vingt Vins d'Art
Wine Shop	PARIS-3E-ARRONDISSEMENT	48.892735	2.361113	Trois Fois Vin
Women's Store	PARIS-2E-ARRONDISSEMENT	48.867903	2.344107	& Other Stories
Zoo	PARIS-12E-ARRONDISSEMENT	48.835156	2.419807	Parc zoologique de Paris
Zoo Exhibit	PARIS-12E-ARRONDISSEMENT	48.835156	2.419807	Grande Serre du Parc Zoologique de Paris

198 rows × 4 columns

We cluster the city of Paris to roughly 5 to make it easier to analyze. We use the K Means clustering technique to do so. And visualize the clustered data.

```
map_clusters_paris = folium.Map(location=[paris_lat_coords, paris_lng_coords], zoom_start=12)

# set color scheme for the clusters
x = np.arange(k_num_clusters)
ys = [i + x + (i*x)**2 for i in range(k_num_clusters)]
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(paris_data_nonan['Latitude'], paris_data_nonan['Longitude'], paris_data_nonan['nom_comm'], paris_data_n
onan['Cluster Labels']):
    label = folium.Popup('Cluster ' + str(int(cluster) +1) + ' ' + str(poi) , parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[int(cluster-1)],
        fill=True,
        fill_color=rainbow[int(cluster-1)],
        fill_opacity=0.8
        ).add_to(map_clusters_paris)

map_clusters_paris
```

Results and Discussion

Paris is small in size geographically. It has a wide variety of cusines and eateries including French, Thai, Cambodian, Asian, Chinese etc. There are a lot of hangout spots including many Restaurants and Bars. Paris has a lot of Bistro's. Different means of public transport in Paris which includes buses, bikes, boats or ferries. For leisure and sight seeing, there are a lot of Plazas, Trails, Parks, Historic sites, clothing shops, Art galleries and Museums. Overall, Paris seems like the relaxing vacation spot with a mix of lakes, historic spots and a wide variety of cusines to try out.

Conclusion

The purpose of this project was to explore Paris and see how attractive it is to potential tourists and migrants. We explored the city based on the postal codes and then extrapolated the common venues present in each of the neighbourhoods. We could see that each of the neighbourhoods have a wide variety of experiences to offer which is unique in it's own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion. Paris seems to offer a vacation stay or a romantic gateaway with a lot of places to explore, beautiful landscapes and a wide variety of culture. Overall, it's upto the stakeholders to decide which experience they would prefer more and which would more to their liking.