The battle of neighborhoods
Best place to start a restaurant in Paris

Zakaria BOUZIANE

November 23, 2019

1. Introduction

1.1 Background

 Paris is the capital of France and the most populous city in this country. The City of Paris is part of *île-de-France* region and it is considered as one of economic centers in Europe. It is multicultural city and provides many business opportunities. It is also known by its history with presence of historical buildings and museum like museum of Louvre. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in Paris.

Attractivity brings competitivity. Paris is the most competitive market in France. For starting a business there, business analysis should be provided to better understanding business environment, reduce risk and have reasonable return of investment.

1.2 Problem description

A restaurant is a business which prepares and serves food and drink to customers for certain profit. Paris is famous by its culinary art and *haute cuisine* and presence of star-studded Chef. The city is home of the finest and most diverse haute cuisine restaurants in the world. In addition to French cuisine, food culture includes also an array of international cuisines influenced by multiculturality of the town.

According to the guardian, there is international cuisines that is exists in Paris are:

1. North African cuisine: Couscous, Tagines
2. Asian cuisine especially Vietnamian cuisine: Pho Ta, sushi
3. Africa south of the Sahara cuisine: Poulet Yassa
4. Italian cuisine: Pizzas and pastas
5. Lebanese cuisine: m'tabal, kibbeh nayyeh, falafel
6. Turkish cuisine: kebab, baklava... etc

To survive in such competitive market, it is very important to strategically plan. Various factors need to be studied in order to decide on the location such as:

1. Paris Population
2. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
3. Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
4. Who are the competitors in that location?
5. Cuisine served / Menu of the competitors
6. Segmentation of the Borough
7. Untapped markets
8. Saturated markets etc

Even though well-funded a restauration company named XYZ Company. This company needs to choose the correct location to start. If this is successful, so it can replicate the same in other locations. First move is very important, thereby choice of location is very important.

1.3 Target Audience

Objective of this study is to locate and recommend to **anyone who wants to open a new restaurant in Paris** which neighborhood will be best choice to start.

1.4 Success criteria

The success criteria of the project will be a good recommendation of neighborhood choice to XYZ Company based on lake of restaurant specialized in certain type of restaurants based on location and nearest suppliers of ingredients.

Objective of this study is to locate and recommend to **anyone who wants to open a new restaurant in Paris** which neighborhood will be best choice to start.

2. Data:

This project will analyze neighborhoods of the city of Paris.

The datasets below will be used for this analysis.

2.1 Data1: Boroughs and neighborhoods in Paris

Paris has in total 20 boroughs (called *arrondissements* in French) and 80 neighborhoods. To explore, analyze and segment neighborhoods, longitude and latitude of each neighbor and borough will be added.

This dataset exists for free on the web.
Links to the dataset are:

- https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
- https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin

2.2 Data 2: Paris population

Number of populations in each borough will be analyzed to find most populous borough in Paris (more populous Borough = Higher gain).

The dataset that exists in link below will be used: https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e

	num_arr	nom_dept	nom_region	population	postal_code	statut
0	1	PARIS	ILE-DE-FRANCE	17.6	75001	Capitale d'état
1	2	PARIS	ILE-DE-FRANCE	22.4	75002	Chef-lieu canton
2	3	PARIS	ILE-DE-FRANCE	35.7	75003	Chef-lieu canton
3	4	PARIS	ILE-DE-FRANCE	28.2	75004	Chef-lieu canton
4	5	PARIS	ILE-DE-FRANCE	61.5	75005	Chef-lieu canton
5	6	PARIS	ILE-DE-FRANCE	43.1	75006	Chef-lieu canton
6	7	PARIS	ILE-DE-FRANCE	57.4	75007	Chef-lieu canton
7	8	PARIS	ILE-DE-FRANCE	40.3	75008	Chef-lieu canton
8	9	PARIS	ILE-DE-FRANCE	60.3	75009	Chef-lieu canton
9	10	PARIS	ILE-DE-FRANCE	95.9	75010	Chef-lieu canton
10	11	PARIS	ILE-DE-FRANCE	152.7	75011	Chef-lieu canton

The table above shows 11 first borough in Paris. Population column considers number of people that live in each borough per 1000 person.

2.3 Data3: Restaurants and markets in Paris

This data will cover different **restaurants** that already exists in Paris and different **food markets**. Analysis of localization of food markets permits to have an idea of location that will be near to fresh ingredients suppliers. This data will be provided by **Foursquare API**. **Foursquare API** permits to provide venues information for each neighborhood in Paris.

Foursquare API logo

Table below gives some restaurant in Paris:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Restaurant	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	Bonne-Nouvelle	48.867150	2.350080	Les Amis des Messina	48.867166	2.348420	Italian
1	Bonne-Nouvelle	48.867150	2.350080	Kapunka	48.866226	2.347763	Thai
2	Bonne-Nouvelle	48.867150	2.350080	Dionysos	48.866942	2.347266	Greek
3	Bonne-Nouvelle	48.867150	2.350080	Frenchie	48.867739	2.347993	French
4	Bonne-Nouvelle	48.867150	2.350080	Rice & Fish	48.865150	2.351435	Sushi
5	Bonne-Nouvelle	48.867150	2.350080	Pupetta	48.864719	2.350271	Italian
6	Bonne-Nouvelle	48.867150	2.350080	Le Pas Sage	48.864646	2.350066	French
7	Bonne-Nouvelle	48.867150	2.350080	La Maison du Saké	48.864521	2.349226	Japanese
8	Bonne-Nouvelle	48.867150	2.350080	Tir-Bouchon	48.864648	2.348728	French
9	Bonne-Nouvelle	48.867150	2.350080	Ristorante National - Cucina Italiana	48.865983	2.353209	Italian

To have fresh ingredient for dishes, it is preferable for business of restauration to be close to farmer market. In this section, farmer markets through each borough in Paris will be discovered. We should find here borough that contains higher number of farmer markets to ensure quality and competitivity (Borough where there is a big number of farmer market, it will be more competitive in order of products that are related to restauration business, so cost will be smaller). Table below give some food markets in Paris and characteristics of them as localization and opening time. This table is provided by following link: https://opendata.paris.fr/explore/dataset/marches-decouverts/table/?disjunctive.produit&disjunctive.ardt&disjunctive.jours_tenue&disjunctive.gestionnaire&basemap=jawg.dark&location=12,48.86175,2.33563

id_marche	jeudi	jours_tenue	lineaire	localisation	lundi	mardi	mercredi	nom_court	nom_long	produit
12.0	0.0	vendredi apm	176.0	place d'Anvers le long du square d'Anvers en v...	0.0	0.0	0.0	ANVERS	MARCHÉ ANVERS	Alimentaire
17.0	0.0	mardi, vendredi	293.0	terre-plein boulevard de Ménilmontant, entre l...	0.0	1.0	0.0	PERE LACHAISE	MARCHÉ PERE LACHAISE	Alimentaire
39.0	0.0	mercredi, dimanche	259.0	avenue Villemain sur le terre-plein, entre cet...	0.0	0.0	1.0	VILLEMAIN	MARCHÉ VILLEMAIN	Alimentaire
31.0	0.0	mardi, vendredi	60.0	boulevard de l'Hôpital, le long du square Mari...	0.0	1.0	0.0	SALPETRIERE	MARCHÉ SALPETRIERE	Alimentaire
42.0	0.0	mercredi, dimanche	843.0	terre-plein du boulevard de Grenelle, entre la...	0.0	0.0	1.0	GRENELLE	MARCHÉ GRENELLE	Alimentaire

3. Methodology:

3.1 Business Understanding:

Main goal of this project is to get locations that will be suitable for opening new restaurant business in Paris.

3.2 Analytic Approach:

Paris has a total of 20 boroughs and 80 neighborhoods (4 neighborhoods for each borough). In this project neighborhoods will be clustered following exploratory data that will be discovered in the next part.

3.3 Analytic Approach:

3.3.1 Data 1: population of Paris

This data is contained in a JSON file that is downloaded from link:
https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e
This data is also cleaned to extract the necessary of information and processed like this:

- Transform the data of nested python dictionaries into a data frame.
- This data frame contains geographical localization and population by 1000 of each borough in Paris.

	num_arr	nom_dept	nom_region	population	postal_code	statut	geo_point_2d
0	1	PARIS	ILE-DE-FRANCE	17.6	75001	Capitale d'état	[48.8626304851685, 2.336293446550539]
1	2	PARIS	ILE-DE-FRANCE	22.4	75002	Chef-lieu canton	[48.86790337886785, 2.344107166658533]
2	3	PARIS	ILE-DE-FRANCE	35.7	75003	Chef-lieu canton	[48.86305413181178, 2.359361058970589]
3	4	PARIS	ILE-DE-FRANCE	28.2	75004	Chef-lieu canton	[48.854228281954754, 2.357361938142205]
4	5	PARIS	ILE-DE-FRANCE	61.5	75005	Chef-lieu canton	[48.844508659617546, 2.349859385560182]

This table show data frame that contains different element for each borough in Paris. Each borough in Paris is called in French *Arrondissement* and it is determined by a number (from 1 to 20). You will find in the notebook the definition of each column in the table.

- This data will be used to get venues data from Foursquare.
- Geopy and Folium is two libraries that will be used to create a map of Paris with marks for each borough where number of borough and number of population are superimposed on top.

- Same thing is done for finding neighborhoods in Paris. Data is containing in JSON file that is downloaded from: https://opendata.paris.fr/explore/dataset/quartier_paris/download/?format=json&timezone=Europe/Berlin

- Geopy and Folium is two libraries that will be used to create a map of Paris with marks for each neighborhood where number of borough where neighborhood is located, name of neighborhood and postal code are superimposed on top.

3.3.2 Data 2: Cuisine of Paris:

Data about cuisine in Paris is generated by Foursquare API by providing different restaurants that are near to a neighborhood or located in a certain borough.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Restaurant	Restaurant Latitude	Restaurant Longitude	Restaurant Category
0	Bonne-Nouvelle	48.867150	2.350080	Les Amis des Messina	48.867166	2.348420	Italian
1	Bonne-Nouvelle	48.867150	2.350080	Kapunka	48.866226	2.347763	Thai
2	Bonne-Nouvelle	48.867150	2.350080	Dionysos	48.866942	2.347266	Greek
3	Bonne-Nouvelle	48.867150	2.350080	Frenchie	48.867739	2.347993	French
4	Bonne-Nouvelle	48.867150	2.350080	Rice & Fish	48.865150	2.351435	Sushi
5	Bonne-Nouvelle	48.867150	2.350080	Pupetta	48.864719	2.350271	Italian
6	Bonne-Nouvelle	48.867150	2.350080	Le Pas Sage	48.864646	2.350066	French
7	Bonne-Nouvelle	48.867150	2.350080	La Maison du Saké	48.864521	2.349226	Japanese
8	Bonne-Nouvelle	48.867150	2.350080	Tir-Bouchon	48.864648	2.348728	French
9	Bonne-Nouvelle	48.867150	2.350080	Ristorante National - Cucina Italiana	48.865983	2.353209	Italian

For example, that table above represent some restaurants that are near to Bonne-Nouvelle neighborhood in Paris. This table contains information about name and localization of restaurant and restaurant category (type of restaurant e.g. Italian).

This data will help us to understand what is the type of restauration that have the domination on each borough. For this task, we will use word cloud. Word cloud is a good visual tool that permit to have a global idea of domination of a category over others. Word cloud is an image composed of words used in a particular text or subject, in which the size of each word indicates its frequency or importance.

For example, we built word cloud for all categories of restaurants in Paris and it gives:

We seen that French, Italian and Japanese restaurant are restaurants that dominate Paris market. We see a category named Restaurant. It is a category of restaurant that is not specialized in certain national dishes.

For our project, word clouds are built for each borough in Paris. For example, for borough 10:

Arrondissement10

We see here that Indian restauration is also dominating in borough 10.

3.3.3 Data 3: Farmer market:

This data is contained in a JSON file that is downloaded from link:
https://opendata.paris.fr/explore/dataset/marches-
decouverts/table/?disjunctive.produit&disjunctive.ardt&disjunctive.jours_tenue&disjunctive.gestionnair
e&basemap=jawg.dark&location=12,48.86175,2.33563

This data is also cleaned to extract the necessary of information and processed like this:

- Transform the data of nested python dictionaries into a data frame.

- This data frame contains geographical localization of famer market and information as opening time, company that is managed the market and type of products that are bought in the market.

- Geopy and Folium is two libraries that will be used to create a map of Paris with marks for each market where name of market and number of borough where it is located are superimposed on top.
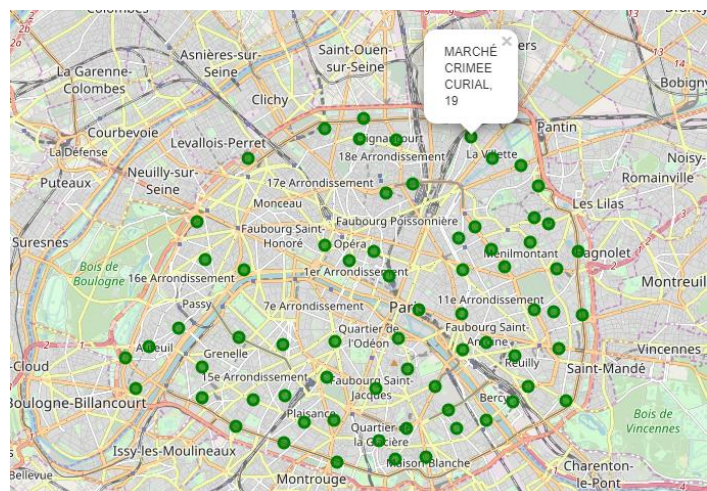
There is a total of 68 farmers market in Paris. Highest numbers are for boroughs 13, 12 and 16. Lowest numbers are for 7 boroughs (2, 4, 6, 7, 8, 9 and 10).


Borough
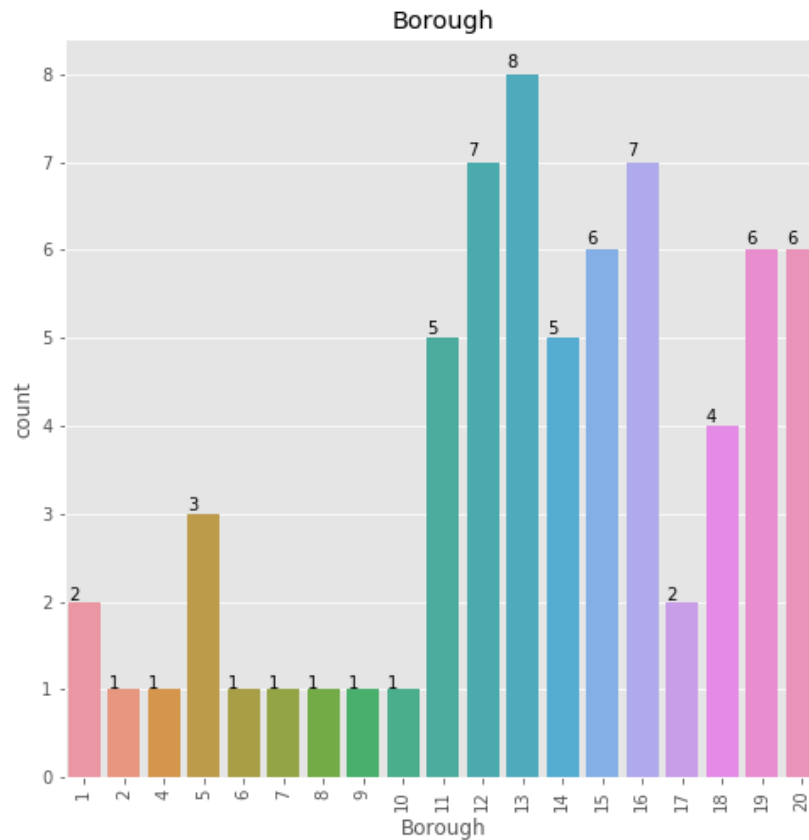
3.4 Machine learning uses:

For segmenting neighborhood in Paris based on existed restaurant, population and localization of farmer markets, and finding recommendation for a good start of restaurant business, **clustering algorithm** will be used especially **K-mean method**. This method is used as unsupervised algorithm. It doesn't need previous recommendations to build a model. **K-mean method** is good for segmentation. It divides the data into clusters without any cluster-internal structures or label

4. Results:

From venues data, dataset is filtered and used only data where venues are restaurant as we focused only on restaurant business.

4.1 Neighborhood K-Means clustering based on mean occurrence of venue category:

To cluster the neighborhoods into two clusters we used K-Means clustering Algorithm. K-Mean clustering aims to partition n observation into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach. In this project, 2 clusters are chosen. The map below, that is constructed using Geopy and Folium libraries, shows clustering of neighborhoods in Paris into two clusters.
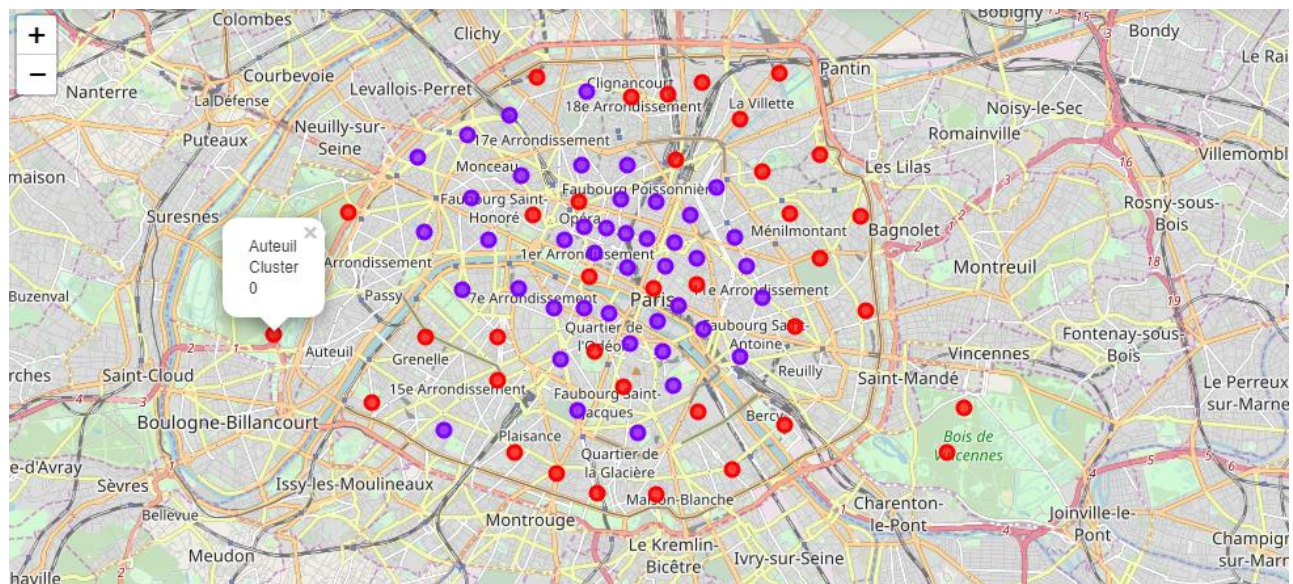
Cluster 0 is marked by red circle markers. It is about neighborhoods where restauration market is not saturated and we can start here a restaurant business. (37 neighborhoods)

Cluster 1 is marked by red purple markers. It is about neighborhoods where restauration market is saturated and it is not recommended here to start a restaurant business. (43 neighborhoods)

5. Conclusion:

As we can see from data and map of clustered neighborhoods, it is more preferable to start restaurant business neighborhoods in periphery of Paris, but in general we can start it in neighborhoods where restauration market is not saturated.

In this project, we have used data available in the net about population of boroughs and already existed restaurant in Paris. The fissionability of farmer markets in each borough of Paris for fresh ingredient is also analyzed. The K-mean algorithm, that is a clustering algorithm for unsupervised learning, is used to show us what are neighborhoods that are good places where we can start restaurant business.

6. Future directions:

For refining our model, another type of data can be added as attractivity and events that are present in each borough or neighborhood in Paris to have an idea about the place that are probably the most profitable. In addition, other methods of unsupervised learning can be used to solve the problem as agglomerative algorithm or density-based clustering.

Thank you for reading this report.

Sources:

[1] : https://en.wikipedia.org/wiki/Paris

[2]: https://www.theguardian.com/travel/2011/may/06/top-10-best-paris-ethnic-restaurants

[3]: https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwirs_yQpPvlAhUQqxoKHZSgDaMQjRx6BAgBEAQ&url=https%3A%2F%2Ftwitter.com%2Ffoursquareapi&psig=AOvVaw2VPQomGJpQuIYzK4sWDZ9w&ust=1574424948931233