

# Capstone project: The battle for neighborhoods (week 2) By TETSA TATA Darius

## Introduction/Business Problem

### Context:

Mrs Suzanne is a business woman having a series of restaurants in the United States. She recently decided to extend her business to France and decided to open a restaurant in Paris. Given she has never done such a business in Paris, she came to us asking us to tell her if Paris could be a good destination for her business. If yes, where in Paris should she open the restaurant?

### Problem:

Using foursquare datasets to help choose the best location to start a business.

### Problem background:

The amount of data collected every day can be highly useful when it comes to taking decisions. Whether we want to segment the neighborhoods of Toronto or New York City or we want to choose a location suitable for a restaurant in the city of Paris. Foursquare has built an important dataset on the neighborhoods and their venues. Exploring some of the neighborhoods and analyzing some of their trending venues can permit us to get some important information that can help us in the process of selecting a place to start a restaurant in a city.

### Target Audience:

Business men who are looking for a new place to start either a new business or an existing one

## Data Description

Foursquare has an important quantity of data on the neighborhoods for a good number of cities worldwide. We intend to explore the city of Paris over a radius of 60km, Get all of its trending restaurants and evaluate their ratings and their foot traffics (statistics). We then intend to do some neighborhood segmentation on the data and try to analyze the results. We will recommend to Mrs. Suzanne the neighborhood with the highest foot traffic to the restaurant, and probably the highest density in restaurants. We will adopt the following methodology:

- Data importation from foursquare
- Data cleansing
- Data transformation
- Data analysis
- Result presentation.

## Foursquare Data presentation

Foursquare has a series of variables for each venue amongst which the following.

- id : the Id of the venue
- name: the name of the venue
- location: the location of the venue which is a composite variable and contains,
  - address
  - crossStreet
  - lat
  - lng
  - cc
  - city
  - state
  - country
  - Postal codes

- Categories: the category of the venue, which is also a composite variable.
  - id
  - name
  - vpluralName
  - shortName
  - icon

For each venue, we will get the statistics. The fields are the following

- totalCheckins: Number of total checkins at the venue during the time period.
- newCheckins: Number of new visitors to the venue during the time period.
- uniqueVisitors: Number of unique visitors at the venue during the time period.
- Sharing: A subobject containing fields twitter and facebook indicating the number of check-ins at the venue pushed to twitter and facebook.
- genderBreakdown: A subobject containing fields female and male indicating the number of checkins at the venue by men and women.
- ageBreakdown: An array of subobjects with fields age and checkins indicating the number of check-ins at the venue by people of different age ranges.
- hourBreakdown: A 24-element array of subobjects with fields hour and checkins indicating the number of checkins at the venue during each hour of the day.
- visitCountHistogram: A list of subobjects with fields checkins and users indicating how many users checked in a given number of times.
- topVisitors: An array of the top 10 users by check-in count during the time period. Array elements are objects with fields user, and checkins, the number of checkins by - that user. Does not include users who have opted out of sharing their check-ins with venue managers in their settings.
- recentVisitors: An array of the 10 most recent users. Array elements are objects with fields user and lastCheckin, the timestamp of the user's last check-in. This field is only present if startAt is specified and endAt is NOT specified (i.e., now). Does not include users who have opted out of sharing their check-ins with venue managers in their settings.

## Foursquare Data collection

We used foursquare API to collect the list of restaurants of Paris in a radius of 60km. To this list of venues, we associated the list of their rating and tips to get an idea of the statistics. We finally associated the list of the neighborhoods names from postal codes for easy identification and understanding.

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id	Neighbourhood	Neighbourhood_name	rating	tips
0	Restaurant 58 Tour Eiffel	French Restaurant	Tour Eiffel	FR	Paris	France	1er étage	3007	[Tour Eiffel (1er étage), 75007 Paris, France]	[[{"label": "display", "lat": 48.85836582402155, "lng": 2.294249}]]	48.858366	2.294249	NaN	75007	Île-de-France	4e078807e4cdefcf65dce4f6	7eme arrondissement de paris	Palais-Bourbon	9.2	154
1	Le Restaurant du Musée d'Orsay	French Restaurant	Musée d'Orsay	FR	Paris	France	1 rue de la Légion d'Honneur	916	[Musée d'Orsay (1 rue de la Légion d'Honneur), ...]	[[{"label": "display", "lat": 48.86029986248957, "lng": 2.325392}]]	48.860300	2.325392	NaN	75007	Île-de-France	4cbb05054352a1cd0d4396f5	7eme arrondissement de paris	Palais-Bourbon	7.2	16
2	Restaurant Le Meurice Alain Ducasse	French Restaurant	228 rue de Rivoli	FR	Paris	France	NaN	420	[228 rue de Rivoli, 75001 Paris, France]	[[{"label": "display", "lat": 48.86516, "lng": 2.328070}]]	48.865160	2.328070	NaN	75001	Île-de-France	4adcda13f964a520e53621e3	1er arrondissement de paris	Louvre	8.5	37
3	Restaurant d'Entreprise	Cafeteria	NaN	FR	Paris	France	NaN	142	[75002 Paris, France]	[[{"label": "display", "lat": 48.86791550563465, "lng": 2.333772}]]	48.867916	2.333772	NaN	75002	Île-de-France	4e57797d7dbf55c172826d9	2eme arrondissement de paris	Bourse	0.0	0
4	Restaurant Mon Paris I	French Restaurant	6 rue Édouard VII	FR	Paris	France	NaN	555	[6 rue Édouard VII, 75009 Paris, France]	[[{"label": "display", "lat": 48.870959, "lng": 2.329453}]]	48.870959	2.329453	NaN	75009	Île-de-France	5640db04cd10809c53939c69	9eme arrondissement de paris	Opera	7.6	9
	L'Indica	French					Place		[Palais Garnier]	[[{"label": "display", "lat": 48.873234, "lng": 2.331487}]]	48.873234	2.331487	NaN	75002	Île-de-France	4e57797d7dbf55c172826d9	9eme arrondissement de paris			

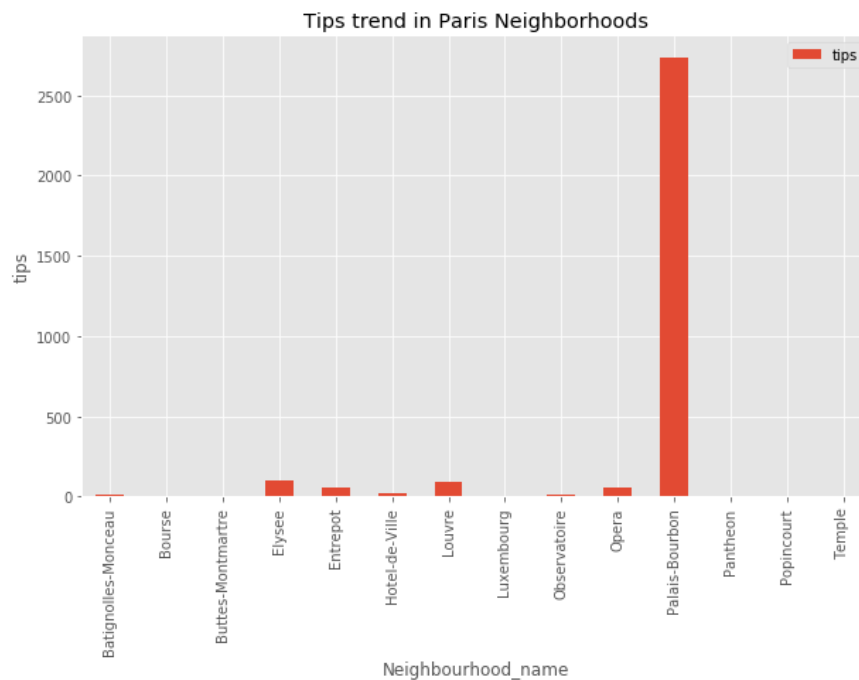
## Methodology

### Exploratory data analysis:

For an exploratory analysis, we started by getting the number of restaurants and grouping them per neighborhood. Out of the 46 restaurants found on foursquare, the Elysee neighborhood comes first with 8 restaurants in its neighborhood as shown in the figure below.

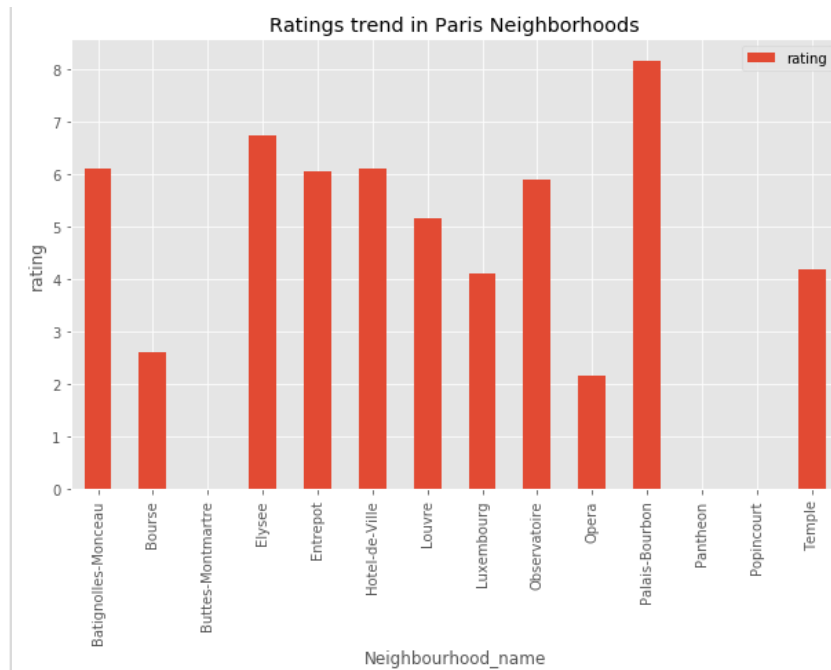
```
Out[17]: Elysee          8
         Opera          7
         Louvre         7
         Palais-Bourbon  5
         Entrepot       5
         Bourse         3
         Temple        2
         Luxembourg    2
         Pantheon       2
         Popincourt    1
         Hotel-de-Ville 1
         Buttes-Montmartre 1
         Batignolles-Monceau 1
         Observatoire  1
         Name: Neighbourhood_name, dtype: int64
```

We then did a plot of the tips of the restaurants per neighborhood.



We can see from this graph that the neighborhood of Palais-Bourbon is outstanding. This could already be an indicator that a new restaurant in this neighborhood could be a success. The next favorable neighborhood for a restaurant could be Elysee. This can be due to high population density in those areas.

We then proceeded by doing another plotting of the ratings of the restaurants per neighborhood.



From the graph, we can once more observe that the restaurants of Palais-Bourbon had a better rating compared to the others. Followed by Elysee. This could be due to business competition in those areas.

We evaluated the correlation between rating and tips. We noticed that those two variables were not correlated. We then did some one-hot analysis on the neighborhoods, some Min-Max normalization and 5 classes K-Means clustering on the venues.

## Results

After carrying out a 5 classes K-Means clustering on our processed data, we obtained the following clusters.

Cluster No 1:

	name	Neighbourhood_name	rating	tips	t
3	Restaurant d'Entreprise	Bourse	0.0	0	
7	Restaurant Aux Tours de Notre-Dame	Hotel-de-Ville	6.1	21	
9	Le Grand Amour Restaurant	Entrepot	8.5	25	
13	Le Restaurant de L'Hôtel	Luxembourg	8.2	8	
18	Le Restaurant des Poètes	Temple	8.4	6	
21	Terrass" Restaurant	Buttes-Montmartre	0.0	0	
23	Restaurant Le Robinet d'Or	Entrepot	7.9	19	
28	Elise Restaurant	Entrepot	7.1	10	
29	Restaurant Buffet Monte Carlo	Batignolles-Monceau	6.1	10	
30	Restaurant Hotel Little Palace	Temple	0.0	0	
32	Restaurant Deluxe	Entrepot	0.0	0	
34	Restaurant (V)ivre	Bourse	7.8	4	
35	Restaurant Mamma Mia	Pantheon	0.0	0	
36	Le Paris Montparnasse Café Restaurant	Observatoire	5.9	16	
37	Restaurant Le Gastronom halal	Popincourt	0.0	0	
38	Restaurant de l'Institut du Monde Arabe.	Pantheon	0.0	2	
42	Restaurant Royal	Bourse	0.0	0	
44	Le Restaurant de L'Hôtel	Luxembourg	0.0	0	
45	A Hot Pot Restaurant 一家	Entrepot	6.7	4	

### Cluster No 2:

	name	Neighbourhood_name	rating	tips
2	Restaurant Le Meurice Alain Ducasse	Louvre	8.5	37
6	Restaurant Le Mona Lisa	Louvre	0.0	0
15	Restaurant Champeaux	Louvre	7.9	24
25	Restaurant de la Cordonnerie	Louvre	6.9	5
31	Restaurants du Monde	Louvre	5.6	5
40	Macéo Restaurant & Bar	Louvre	7.3	19
43	Restaurant du Théâtre	Louvre	0.0	1

### Cluster No3:

	name	Neighbourhood_name	rating	tips
10	Jet Set Restaurant	Elysee	6.6	41
12	Le Restaurant	Elysee	8.6	5
14	Restaurant Au 35	Elysee	7.1	1
17	Restaurant Alain Ducasse	Elysee	8.9	31
20	Le Grand Restaurant - Jean-François Piège	Elysee	8.1	8
24	BLOSSOM Restaurant	Elysee	0.0	7
27	Restaurant & Lounge Bar Le W	Elysee	7.9	6
33	Restaurant du Rond-Point	Elysee	6.7	5

### Cluster No 4:

	name	Neighbourhood_name	rating	tips
4	Restaurant Mon Paris !	Opera	7.6	9
5	L'Opéra Restaurant	Opera	7.5	48
8	Restaurant Sichuan	Opera	0.0	0
11	Restaurant Ayào 阿耀面点	Opera	0.0	1
16	Restaurant Vietnam   Banh Mi	Opera	0.0	0
19	La Cantone - Restaurant d'Entreprise	Opera	0.0	0
41	B.B. Le Restaurant	Opera	0.0	0

### Cluster No 5:

	name	Neighbourhood_name	rating	tips
0	Restaurant 58 Tour Eiffel	Palais-Bourbon	9.2	154
1	Le Restaurant du Musée d'Orsay	Palais-Bourbon	7.2	16
22	Restaurant du Montalembert	Palais-Bourbon	6.9	2
26	Le 144 - Restaurant Petrossian	Palais-Bourbon	8.0	9
39	Tour Eiffel	Palais-Bourbon	9.5	2556

## Discussion

After looking kindling into the different clusters, we observe that Palais-Bourbon, Elysee, Louvre and opera are alone in their clusters. With the individual variable analysis, we could observe that Palais-Bourbon was a good candidate neighborhood to start a new restaurant. It had an outstanding number of tips and good ratings. The high number of tips could indicate a high foot traffic in this neighborhood which is a good criteria when selecting a place to start a restaurant. On the other hand, good ratings could indicate a good level of commercial competition. Where there is commercial competition, there is market share. Finally, from the cluster analysis of Palais-Bourbon, we observe for all of its existing restaurants good ratings and good tips which is an indicator of a good cluster.

Elysee could be considered the next candidate neighborhood to start a restaurant, followed by Louvre, following the same analysis. Opera is not a good candidate neighborhood to start a restaurant.

**Conclusion:**

Our aim was to use foursquare data to recommend to Mrs. Suzanne a neighborhood in Paris to launch her new restaurant.

Based on the data collected from foursquare on the Paris neighborhoods and with a radius of 60km, we carried out the above analysis and we recommended to Mrs. Suzanne 03 candidate neighborhoods to start her new restaurant: Palais-Bourbon, Elysee and Louvre respectively. Palais-Bourbon being the best candidate.