The Battle of Neighborhoods -

Social Media Popularity Mapping

Romain Bocher

December 18, 2019

Paris, France – bocher_romain@yahoo.fr

Abstract

In this project, we build a recommendation map of restaurants in Paris, France, in order to help a Millennial person to quickly find relevant areas without being familiar with the city. Mapping is based on geographical clustering and popularity analysis assuming social media to properly capture Millennials' preferences.

1. Introduction

1.1 Background

The digital revolution has dramatically changed the way people decide and plan their leisure activities, such as going to a restaurant or a bar. Today, numerous applications enable individuals to find the right venue (e.g. TripAdvisor, Booking.com, Airbnb, Facebook, Google, etc.). The game changer has been the massive adoption of such applications by so-called Millennials for checking and booking [1]. But every person who has visited an unknown city knows that online reviews might not be fully reliable. Moreover, Millennials tends to check multiple sites, including blogs, in order to find the right place to visit [2]. Indeed, existing reviews platforms fail to fully address people's expectations and there are at least two reasons for that. First, reviews are heterogeneous in terms of demographic cohorts since they can also be published by older or younger people. Second, if you place the same person in two different contexts (e.g. a romantic evening vs a bachelor party), then he/she is likely to have distinct expectations.

1.2 Problem

Assuming Foursquare and Instagram to be popular platforms among Millennials [3][4], this project aims to compare neighborhoods of an unknown city based on the social media popularity of their respective restaurants. The goal is to recommend relevant areas to visit without having to spend time on reviews sites or blogs. The focus was set on Paris, France.

1.3 Interest

This work might gain interest from social travel and booking applications since their current search and recommendation algorithms do not target any specific demographic cohort. It could also prove useful for social media platforms. In the offline world, it could make sense for businesses such as restaurants, bars and nightclubs, or event planners looking for the right location. Last, although it is not the initial purpose of this paper, network analysis could lead to interesting findings about urban organization, online interactions and Millennials preferences.

2. Data Acquisition and Cleaning

2.1 Data Sources

Paris is divided into 20 administrative areas known as *arrondissements*, and each arrondissement is divided into 4 *quartiers* (i.e. neighborhoods). Therefore, there were 80 neighborhoods to analyze. Data such as latitudes, longitudes or perimeters can be imported for free from Paris official web platform. Restaurants search was processed calling Foursquare API, and social media popularity metrics were retrieved calling Foursquare API and scrapping Instagram web pages.

2.2 Data Selection

We were using a free personal account on Foursquare. Thus, there were two main limits to consider:

1. We could only get up to 50 venues per location.
2. We could only make up to 500 premium calls per day.

Premium calls were a critical part of the process since they enabled us to retrieve important details on restaurants like Foursquare rating or Instagram username if one. To overcome this issue, data on restaurants were selected as follows:

Step 1: We imported 50 recommended restaurants for each neighborhood calling Foursquare API. Please note that for each neighborhood we had to set a radius that was consistent with the neighborhood's surface. Moreover, we had to limit the request to the neighborhood administrative limits using 'near' option.

Step 2: We selected 400 restaurants displaying the highest eigenvector centralities. This method is closed to the famous Google PageRank algorithm: restaurants with high eigenvector centrality are geographically closed to restaurants which are also closed to other restaurants [5]. This technique enabled us to detect relevant areas for our study and to focus on a limited number of venues.

Step 3: We retrieved venues details such as Foursquare rating, Foursquare likes count and Instagram followers.

2.3 Data Cleaning and Features Selection

Neighborhoods data we downloaded from one single source with few changes to make to get the following table:

	Neighborhood	PostalCode	Latitude	Longitude	Radius
0	Javel	75015	48.839060	2.278076	1669.482927
1	Saint-Thomas-d'Aquin	75007	48.855263	2.325588	791.864176
2	Faubourg-Montmartre	75009	48.873935	2.343253	576.539498
3	Palais-Royal	75001	48.864660	2.336309	448.322129
4	Clignancourt	75018	48.891668	2.345979	1242.550732
5	Saint-Germain-des-Prés	75006	48.855289	2.333657	530.888347
6	Porte-Saint-Martin	75010	48.871245	2.361504	671.579562
7	Maison-Blanche	75013	48.823128	2.352433	1409.903551
8	Halles	75001	48.862289	2.344899	539.271421
9	Chaillot	75016	48.868434	2.291679	1077.345335

Table 1

For each neighborhood, coordinates and radius were used as variables for an 'explore' request (Step 1), leading to a table of 4000 restaurants. Then, we built a distance matrix using restaurants geographical coordinates, and generated a graph importing Networkx package, in order to calculate eigenvector centralities and select the top 400 restaurants (Step 2). For each selected venue, we retrieved the Foursquare rating, the likes count and the Instagram username (Step 3).

There were several restaurants with no Instagram username, so we had to handle such exception setting the followers count to zero.

Besides, some Instagram accounts referred to the chef and not to the venue itself (e.g. Cyril Lignac). However, a chef can be treated as a brand from a business perspective, meaning that his/her official account is relevant to capture the popularity of his/her restaurants.

Last, we imported Instagram followers count for each username using a simple web scrapping algorithm. Note that we could also have used the Instagram API for that – and it is actually a cleaner solution – but the registration process is not straightforward.

3. Methodology and Data Exploration

3.1 Eigenvector Centrality Sampling

Before sampling, the 4000 restaurants were mapped as follows:

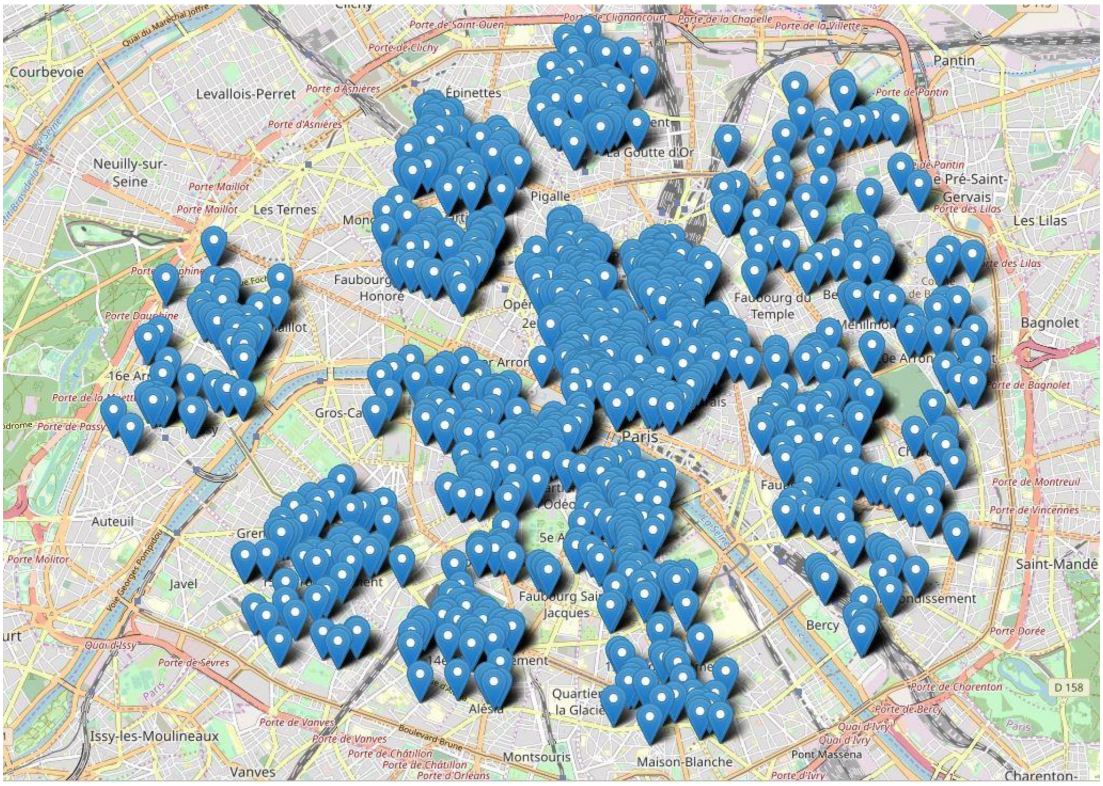


Figure 1

The restaurants dataset was reduced from 4000 venues (Figure 1) to 400 (Figure 2). To achieve such a result, we built an undirected graph connecting each restaurant to one another. In other words, nodes were venues and the weight of each edge was set as the reciprocal of the geographical distance between two restaurants. Thus, we were able estimated eigenvector centrality for all the restaurants of our dataset. As already mentioned before, the remaining 400 restaurants were those with higher centrality measures, leading to the following map:

Figure 2

This method had two advantages. First, geographical outliers (i.e. isolated venues) were excluded, meaning that clustering should be more efficient. Second, it helped to visualize clusters, enabling us to set an accurate estimate for the number of clusters k.

3.2 K-Means Clustering

Thanks to eigenvector centrality sampling, K-means clustering was quite straightforward importing Scikit-Learn library (Figure 3). Moreover, data preprocessing was not required since we based the analysis on venues geographical coordinates.

We set $k = 17$, using the previous map to help us estimate a relevant value for the number of clusters.

However, one could argue that such a discretionary choice might limit the automation process.

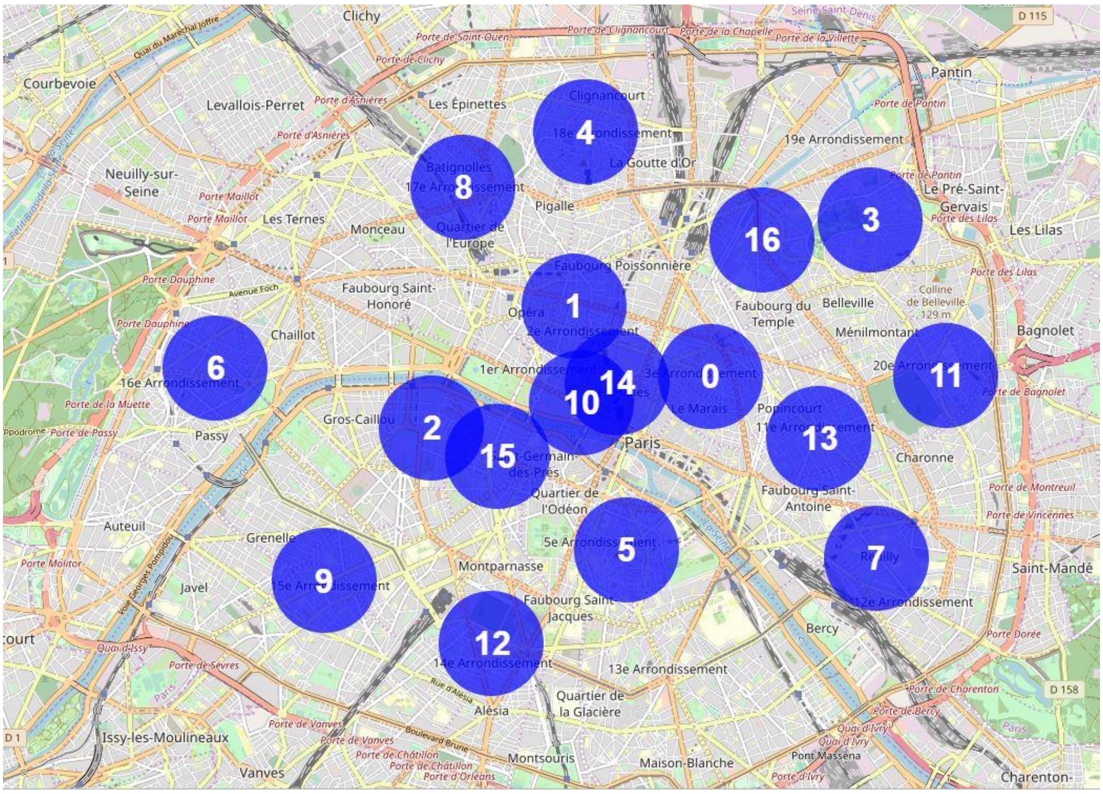


Figure 3

To solve this issue, we could search a hypothetical relationship between a city's size and the number of such restaurant areas, working on other big cities. Nevertheless, we would have to properly define the concept of 'size' of for city (surface, population, etc.). Besides, we might have to make key assumptions about what make two cities similar or dissimilar. Though this might of interest, such a research work is clearly beyond the scope of this project.

The 17 clusters should look consistent for any person familiar with Paris. Instead of using the generic index number, we decided to rename them for more clarity (note that this part is optional):

0. Marais
1. Grands Boulevards

2. Solférino
3. Ourcq
4. Pigalle - Monrtmartre
5. Mouffetard
6. La Muette
7. Reuilly-Diderot
8. Batignolles
9. Pasteur – Convention
10. Chatelet – Les Halles
11. Gaîté – Alésia
12. Voltaire
13. Sentier – Montorgueil
14. Saint Germain des Prés
15. Canal Saint Martin - Jaurès

3.3 Social Media Ranking

At this stage, we were able to recommend geographical areas based on the closeness between restaurants within the area. Thus, these areas became the actual neighborhoods to compare with respect to social media popularity metrics.

Since there might be biases or other statistical problems, we decided to consider three different metrics: Foursquare rating, Foursquare Likes count, Instagram followers count.

3.3.1 Foursquare Rating

To understand how such a metric could be used, it was necessary to plot a frequency distribution chart:

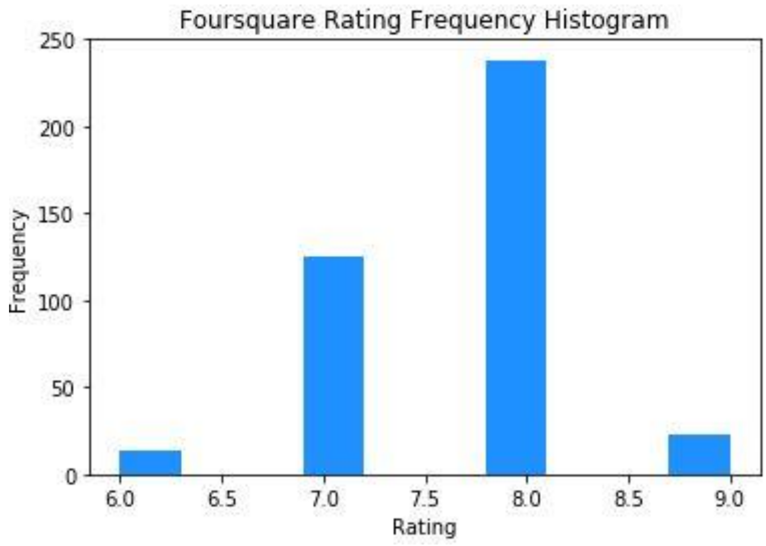


Figure 4

Ratings distribution shape is closed to a normal distribution. Therefore, mean is a consistent estimator and we calculated the average rating of each of neighborhoods based on the rating of each venue within the area:

3.3.2 Foursquare Likes Count

Contrary to the previous metric, the frequency distribution of 'likes' count displays a non-normal distribution, closer to a power-tailed function (Figure 5).

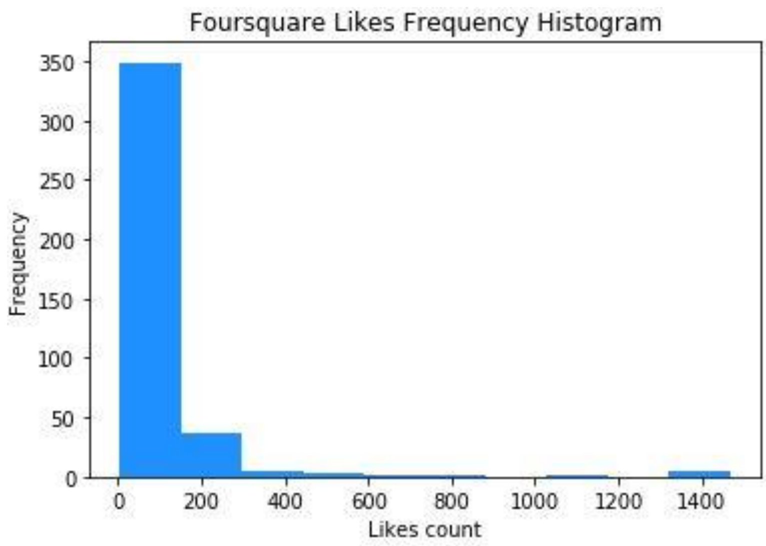


Figure 5

This should not come as a surprise since social media are said to self-organize by forming so-called scale-free networks [6]. Thus, metrics like mean or standard deviation were not consistent to compare the 17 neighborhoods.

Given this power-tailed distribution, we assumed that only the top 30 restaurants in terms of 'likes' count to differentiate from the others (please see Appendix A, Table 2).

3.3.3 Instagram Followers Count

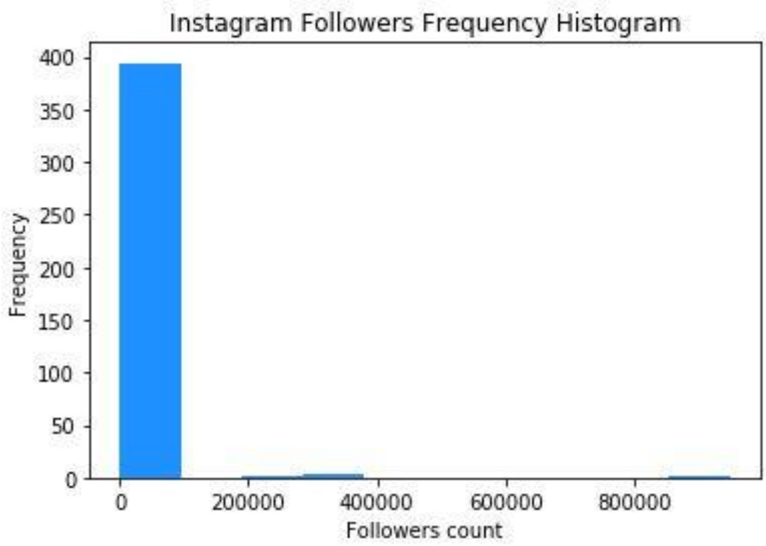


Figure 6

The frequency distribution of 'followers' count is also power-tailed (Figure 6). Therefore, we assumed again that only the top 30 restaurants in terms of 'followers' count could enable us to differentiate neighborhoods (please see Appendix A, Table 3).

4. Recommendation Map

The primary goal of this paper was to recommend neighborhoods to a Millennial without having to navigate on reviews sites or blogs. Thus, we chose to build an interactive map indicating all recommended areas and/or venues (please see Appendix B for more visibility):

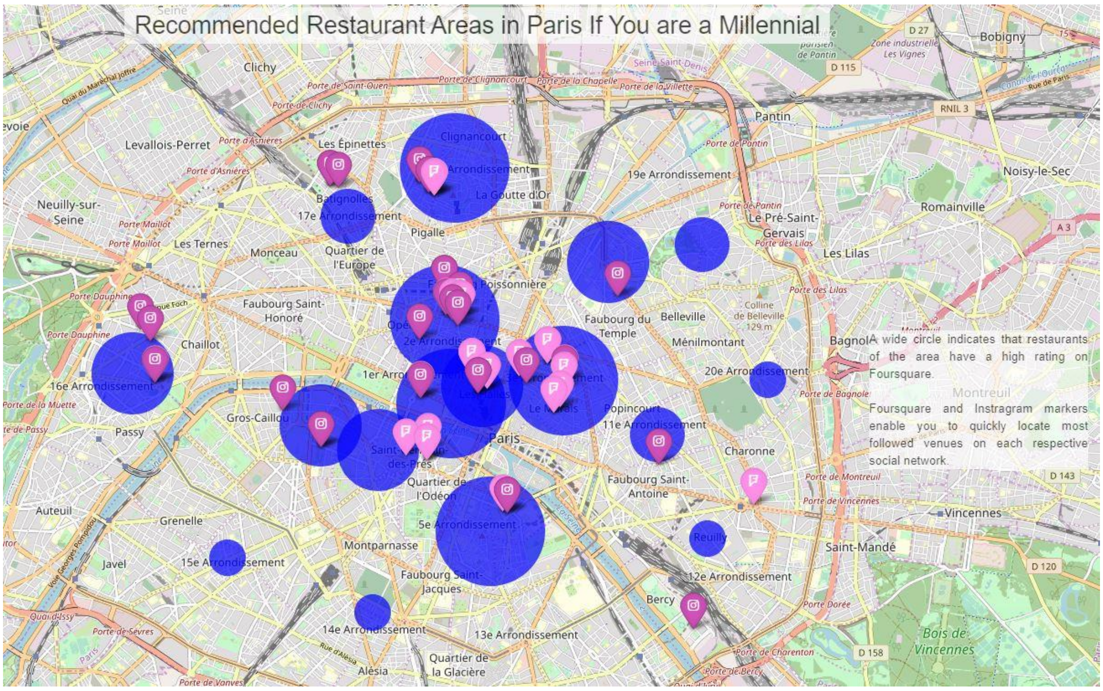


Figure 7

According to the map, the most recommended areas in Paris for a Millennial are:

- Grands Boulevards (1)
- Sentier – Montorgueil (13)
- Marais (0)
- Chatelet – Les Halles (10)
- Pigalle – Montmartre (4)
- Mouffetard (5)
- Canal Saint Martin – Jaurès (16)
- Voltaire (13)
- La Muette (6)

It is important to note that the most followed restaurants on Foursquare and Instagram are located near or within those areas.

To improve clarity, clusters index numbers are not displayed on the recommendation map, but names are available by clicking on the respective circle.

5. Discussion

Once again, the recommendation map should not come as a surprise for any person familiar with Paris. Nevertheless, it is worth noting that it was fully generated with machine learning algorithms. Thus, we might able to apply the method to other comparable cities without being familiar with them.

Moreover, it could also be applied to other venue types such as bars, night clubs, museums, etc.

The map could be easily embedded into a mobile or web application in order to provide a simple and visual tool that would help persons searching for restaurants in an unknown city.

Of course, this method is biased towards the Millennials cohort and should not be relevant for older or younger persons, although one could try to adapt it by focusing on popular social media among other demographic cohorts (e.g. Facebook, Tit Tok), especially generation Z. Besides, even some Millennials might disagree with the recommendation map. However, the map might still make sense for most of them.

Finally, one could wonder whether some Instagram usernames were not missing in the Foursquare database, leading to an incomplete 'followers' count set. Although such a question makes sense, there was no consistent alternative to calling Foursquare API to retrieve Instagram usernames. Indeed, location endpoints have been disabled by Facebook since 2018 due to privacy concerns. Therefore, finding the most followed venues using Instagram API cannot be straightforward and our method was a good approximation for that.

6. Conclusion

This project sought to build a recommendation map of restaurants in an unknown city matching Millennials expectation. From this perspective, we achieved to detect relevant neighborhoods and compare them using social media popularity metrics on Foursquare and Instagram platforms. The method could be applied to other cities and other venue types. Further work could be accomplished to intent to adapt it to at least generation Z or older cohorts. As already mentioned before, we made a discretionary decision by setting the number of clusters relying on data visualization. This process could be improved, but it would require research on a hypothetical relationship between cities and their respective number of restaurant areas. Despite this limitation and although the recommendation model does not fully capture individuals' preferences, this work is a first step towards an algorithmic method to understand the complexity behind human restaurants selection.

7. Acknowledgment

I thank Justine Arnoux for her continuous support and Romain Faure for his detailed comments. I also thank IBM instructors, Joseph Santarcangelo, Alex Aklson, Rav Ahuja, Saeed Aghabozorgi and Polong Lin, for the quality of the courses.

Appendix A

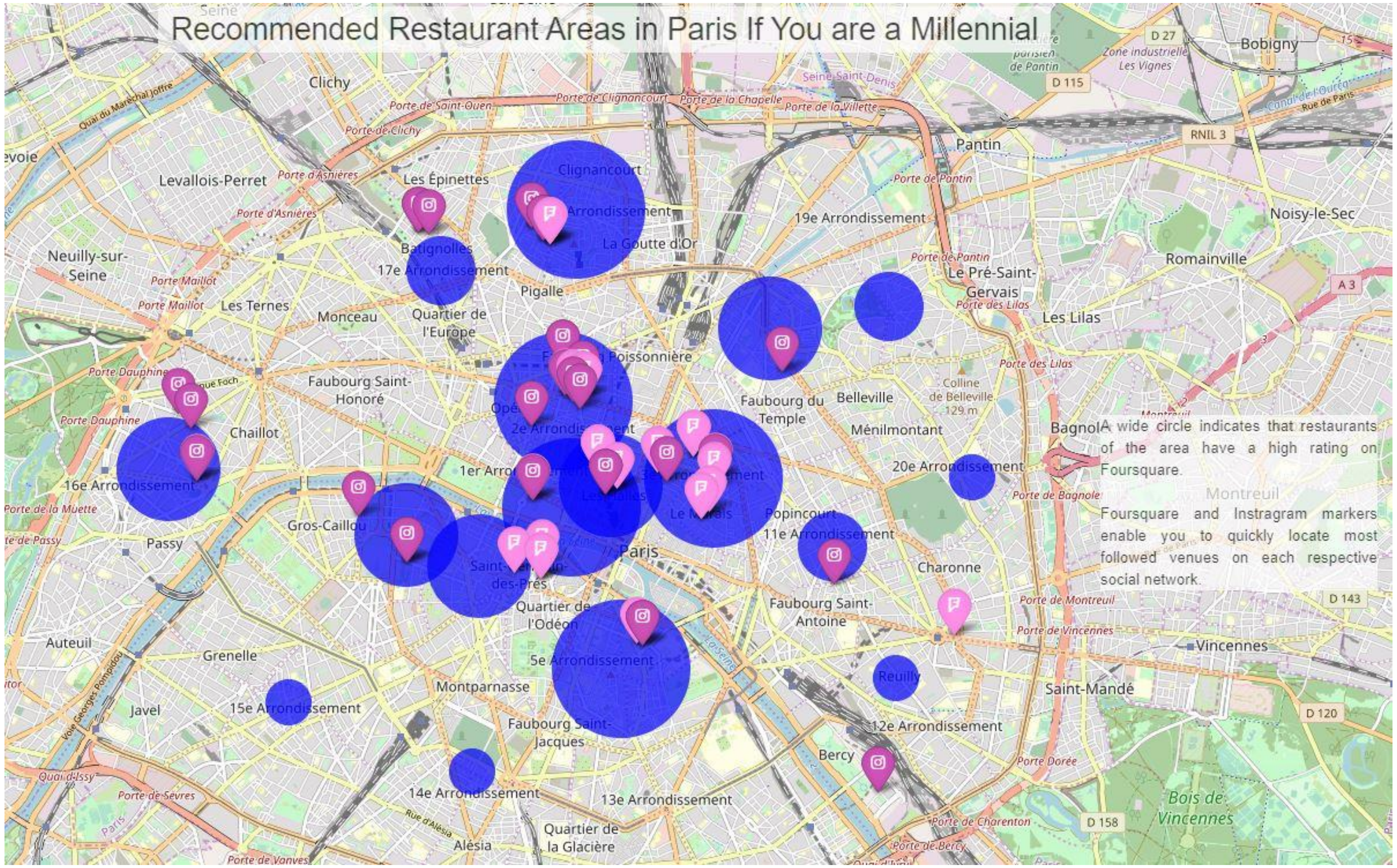
	Neighborhood	Latitude	Longitude	Venue	Id	Eigenvector Centrality	Foursquare_Likes	Foursquare_Rating	Instagram_Account	Instagram_Followers	Centroid
370	Faubourg-Montmartre	48.871944	2.341604	Hard Rock Cafe	4b1d4f79f964a520980e24e3	0.015811	1468	8		0.0	1
17	Chaussée-d'Antin	48.871944	2.341604	Hard Rock Cafe	4b1d4f79f964a520980e24e3	0.015811	1468	8		0.0	1
181	Rochechouart	48.871944	2.341604	Hard Rock Cafe	4b1d4f79f964a520980e24e3	0.015811	1468	8		0.0	1
314	Saint-Georges	48.871944	2.341604	Hard Rock Cafe	4b1d4f79f964a520980e24e3	0.015811	1468	8		0.0	1
60	Rochechouart	48.871861	2.342989	Bouillon Chartier	4b335de3f964a520e51825e3	0.015811	1120	8		0.0	1
364	Saint-Thomas-d'Aquin	48.854678	2.332847	Le Relais de l'Entrecôte	4b3764d6f964a520d94025e3	0.015811	856	8		0.0	15
203	Sainte-Avoie	48.860613	2.361804	Breizh Café	4b003268f964a520673b22e3	0.015811	641	8		0.0	0
200	Sainte-Avoie	48.862940	2.362580	Café Charlot	4b026607f964a520894822e3	0.015811	524	8		0.0	0
322	Arts-et-Métiers	48.864470	2.354178	Le Derrière	4b391111f964a5206c5525e3	0.015811	495	9		0.0	14
246	Place-Vendôme	48.861740	2.335646	Le Café Marly	4adcda04f964a520503221e3	0.015811	446	8	beaumarly_paris	4525.0	10
313	Saint-Georges	48.872090	2.340420	Chipotle Mexican Grill	4eb85eadbe7bfc284b8c1931	0.015811	407	8	chipotle	950114.0	1
320	Arts-et-Métiers	48.859985	2.360735	La Perle	4af77f16f964a520970922e3	0.015811	362	7		0.0	0
172	Sainte-Avoie	48.859985	2.360735	La Perle	4af77f16f964a520970922e3	0.015811	362	7		0.0	0
135	Porte-Dauphine	48.863702	2.285601	Schwartz's Deli	4d59643f296d5481838b58b1	0.015811	340	8	schwartzsdeli	9148.0	6
24	Saint-Germain-l'Auxerrois	48.855370	2.336929	La Palette	4b9bd9e7f964a5201f2d36e3	0.015811	307	8		0.0	10
273	Sorbonne	48.847802	2.351117	Breakfast in America	4adcda14f964a520383721e3	0.015811	289	8		0.0	5
123	Goutte-d'Or	48.886841	2.338186	La Boîte aux Lettres	51c5f55d498e59e2ebcc7742	0.015811	274	8		0.0	4
18	La Chapelle	48.886841	2.338186	La Boîte aux Lettres	51c5f55d498e59e2ebcc7742	0.015811	274	8		0.0	4
294	Clignancourt	48.886841	2.338186	La Boîte aux Lettres	51c5f55d498e59e2ebcc7742	0.015811	274	8		0.0	4
393	Arts-et-Métiers	48.863843	2.362661	Paris New York	551ab159498e0f0eb50a5a1f	0.015811	265	9	pnyburger	29724.0	0
383	Arts-et-Métiers	48.866098	2.359550	La Massara	5083046de4b01d45a22621c0	0.015811	261	8		0.0	0
202	Sainte-Avoie	48.866098	2.359550	La Massara	5083046de4b01d45a22621c0	0.015811	261	8		0.0	0
16	Picpus	48.848576	2.398235	Chez Prosper	4b4b2d20f964a520c39326e3	0.015811	260	8		0.0	7
113	Rochechouart	48.871721	2.343310	La Crème de Paris	55aa3cec498e57df008a0066	0.015811	236	8		0.0	1
374	Mail	48.864555	2.345259	Le Comptoir de la Gastronomie	4adcda14f964a520023721e3	0.015811	225	8		0.0	14
218	Palais-Royal	48.864555	2.345259	Le Comptoir de la Gastronomie	4adcda14f964a520023721e3	0.015811	225	8		0.0	14
1	Saint-Germain-l'Auxerrois	48.864555	2.345259	Le Comptoir de la Gastronomie	4adcda14f964a520023721e3	0.015811	225	8		0.0	14
49	Saint-Germain-l'Auxerrois	48.854067	2.336857	Semilla	4f6e2f86e4b08a0bbae41776	0.015811	224	8		0.0	10
225	Palais-Royal	48.854067	2.336857	Semilla	4f6e2f86e4b08a0bbae41776	0.015811	224	8		0.0	10
182	Gaillon	48.862876	2.348135	Pirouette	5046533ee4b031556a97f2c5	0.015811	222	8		0.0	14

Table 2

	Neighborhood	Latitude	Longitude	Venue	Id	Eigenvector Centrality	Foursquare_Likes	Foursquare_Rating	Instagram_Account	Instagram_Followers	Centroid
313	Saint-Georges	48.872090	2.340420	Chipotle Mexican Grill	4eb85eadbe7bfc284b8c1931	0.015811	407	8	chipotle	950114.0	1
214	Palais-Royal	48.862295	2.346522	Champeaux	5708bc2bcd100e461669a3a4	0.015811	64	7	alainducasse	363918.0	14
57	Saint-Germain-l'Auxerrois	48.862295	2.346522	Champeaux	5708bc2bcd100e461669a3a4	0.015811	64	7	alainducasse	363918.0	14
149	Gros-Caillou	48.855712	2.316920	Arpège	4adcda13f964a520af3621e3	0.015811	166	8	alain_passard	356454.0	2
329	Saint-Thomas-d'Aquin	48.855712	2.316920	Arpège	4adcda13f964a520af3621e3	0.015811	166	8	alain_passard	356454.0	2
71	Picpus	48.833206	2.386972	Five Guys	570fa5dbcd10a1d949b374ef	0.015811	98	6	fiveguys	277446.0	7
362	Saint-Thomas-d'Aquin	48.860195	2.309622	David Toutain	52af6d4a498e6d8ddfc6caa4	0.015811	88	9	david_toutain	50985.0	2
393	Arts-et-Métiers	48.863843	2.362661	Paris New York	551ab159498e0f0eb50a5a1f	0.015811	265	9	pnyburger	29724.0	0
293	Plaine de Monceaux	48.887718	2.320194	Melt	5a60de91ee628b4830a34175	0.015811	26	8	melt.paris	27706.0	8
171	Sainte-Avoie	48.863614	2.355505	Love Juice Bar	58946a922a198259b0ea45a0	0.015811	20	8	lovejuicebarparis	12110.0	0
308	La Chapelle	48.887172	2.337146	Le Coq Rico	4f244789e4b0dc27b94a105c	0.015811	102	8	lecoqrico_paris	10976.0	4
118	Goutte-d'Or	48.887172	2.337146	Le Coq Rico	4f244789e4b0dc27b94a105c	0.015811	102	8	lecoqrico_paris	10976.0	4
135	Porte-Dauphine	48.863702	2.285601	Schwartz's Deli	4d59643f296d5481838b58b1	0.015811	340	8	schwartzsdeli	9148.0	6
132	Porte-Dauphine	48.870227	2.282893	Les Tablettes de Jean-Louis Nomicos	4da89d186e81162ae7a9dd0e	0.015811	46	8	jeanlouisnomicos	8117.0	6
367	Faubourg-Montmartre	48.871363	2.341702	Canard & Champagne	56d9e28fcd103af3d7014386	0.015811	50	9	canardetchampagne	7087.0	1
15	La Chapelle	48.888471	2.335482	Marcel	4d0e110d5c46a09364cd11b4	0.015811	126	8	restaurantmarcel	7063.0	4
117	Goutte-d'Or	48.888471	2.335482	Marcel	4d0e110d5c46a09364cd11b4	0.015811	126	8	restaurantmarcel	7063.0	4
246	Place-Vendôme	48.861740	2.335646	Le Café Marly	4adcda04f964a520503221e3	0.015811	446	8	beaumarly_paris	4525.0	10
47	Rochechouart	48.875007	2.340100	Bioburger	5229b11111d26d734ce3498d	0.015811	136	8	bioburgerfrance	4335.0	1
353	Saint-Georges	48.875007	2.340100	Bioburger	5229b11111d26d734ce3498d	0.015811	136	8	bioburgerfrance	4335.0	1
55	Rochechouart	48.873013	2.342930	Mamie Burger Faubourg Montmartre	55f01057498eeae9cf6c0d7f	0.015811	86	8	mamieburger	4142.0	1
373	Faubourg-Montmartre	48.870752	2.342632	Papadoom Kitchen	5aa038484febd53ed9dfcbcc	0.015811	17	8	papadoomkitchen	3941.0	1
260	Pont-de-Flandre	48.874267	2.372722	Le Galopin	4e7b81e6aeb7a3fb8df5fcfc	0.015811	47	7	le_galopin_	3743.0	16
277	Sorbonne	48.847551	2.351969	Bonvivant	5530cc1f498ea6903c33a3da	0.015811	41	8	bonvivantparis	2570.0	5
205	Saint-Victor	48.847551	2.351969	Bonvivant	5530cc1f498ea6903c33a3da	0.015811	41	8	bonvivantparis	2570.0	5
22	Folie-Méricourt	48.853581	2.380434	Louie Louie	564f964b498e03abff0be2ab	0.015811	80	9	louielouieparis	2326.0	13
48	Rochechouart	48.869037	2.335486	L'Entente, Le British Brasserie	5a229ed7ca18ea73eb85db62	0.015811	24	8	lentente.paris	2167.0	1
120	Porte-Dauphine	48.868771	2.284863	Le Petit Rétro	4bf54e03ff90c9b68b2d5628	0.015811	37	8	lepetitretroparis	1248.0	6
292	Plaine de Monceaux	48.888006	2.318568	Cucuzza	543d1c33498ea4de49ae21d7	0.015811	97	9	cucuzza_ristorante	1146.0	8
332	Batignolles	48.888006	2.318568	Cucuzza	543d1c33498ea4de49ae21d7	0.015811	97	9	cucuzza_ristorante	1146.0	8

Table 3

Appendix B


Recommended Restaurant Areas in Paris If You are a Millennial
A wide circle indicates that restaurants of the area have a high rating on Foursquare.
Foursquare and Instragram markers enable you to quickly locate most followed venues on each respective social network.

References

[1] Gadik, T. (2019). ‘Millennials Choose Hotels Based on Tech, Social Media’. MediaPost. Retrieved from https://www.mediapost.com/publications/article/330264/millennials-choose-hotels-based-on-tech-social-me.html

[2] Anonymous (2016). ‘Millennial Travel Trends: A Look at the Largets Generatios’ Habits’. Internat Marketing Inc. Rerieved from https://www.internetmarketinginc.com/blog/millennial-travel-trends/

[3] Anonymous (2019). YouGov. Retrieved from https://today.yougov.com/topics/media/explore/mobile_application/Foursquare

[4] Janicke-Bowles, S.H., et al. (2018). ‘Social Media for Good? A Survey on Millennials’ Inspirational Social Media Use’. *The Journal of Social Media in Society Fall 2018*. 7(2): 120-140

[5] Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

[6] Albert, R., Barabási, AL. (2002). ‘Statistical Mechanics of Complex Networks’. *Reviews of Modern Physics*. 74(1): 47-97