Finding locations to open a breakfasting spot serving eggs in Paris

– IBM Capstone Project Report –

Lara Simonova

March 30, 2020

1. Introduction

1.1. Background

Paris is one of world's most important haute cuisine capitals. However French food consuming culture has its specificity which reflects on restaurant's menu. For example the majority of french people tend to have only an espresso with croissant for breakfast. Even though such diet fits locals, tourists might want to have something more substantial for their breakfast. So a breakfasting spot serving different kinds of egg plates might be quite popular.

1.2. Problem

Data on
– percentage of existing breakfasting spots serving eggs among all breakfasting spots;
– their ratings to determine populairty
– and analysis of their neighborhoods
might give an insight whether there is a niche for a new breakfasting spot serving eggs and on where it might potentially be located

1.3. Interest

The result of this research would be interesting for restorators searching to open a new food serving point.

2. Data Acquisition and Cleaning

2.1. Data Sources

Forsquare API [1] would be the main source of data about breakfasting places and their locations.

Aditional source of already exitsting egg breakfasting places would be a list created by Foursquare: The 15 Best Places for Eggs in Paris [2]. And total number of restaurants is taken from World Cities Culture Forum [3].

Paris Open Data [4] would be used to retrieve Paris neighborhoods (= arrondisements) coordinates

Coordinates grid wold be set up using data from epsg.io [5]

2.2. Data Cleaning and Transformation

Since data about venues was taken from a single source (Foursquare), it is uniform and happen to be of a good quality. The main restriction influencing the project was limited number of requests of different kinds which could be performed in a free version of Foursquare API. Thus only
— 30 venues per one set of coordinates
— 1 tip per venue
could be fetched. This resulted in the necessity to google some numbers (total number of

restaurants in Paris) and do manual checks of tips and photos for each of selected egg-breakfasting places on Foursquare app.

To describe egg-serving breakfasting spots' neighborhoods I planned to select neaby venues categories, calculate their occurences and use them as features for clustering. But my resulting list of egg-venues contained only 17 points, and they have been surrounded by as much as 112 categories of nearby venues. This could result into low accuracy of clustering. To deal with it I decided to reduce number of categories by grouping less frequent of them under their parent terms (either top level parent, or immediate level parent). Final list contains 25 categories. Further reduction could be done by finding correlations between independent variables, but that was not done.

After clustring I created a chart of top 10 most frequently occuring nearby venue categories for each location. It happend that for 3 of 17 egg-breakfasting spots had less then 10 categories of places nearby. Whereas algorithm which selected these categories output some categories with zero occurence as 7-th to 10-th "most frequent places". For correct charts display I replaced those categories by Nan.

2.3. Data and Feature Selection
There have been 3 main questions to get a data based answer on within this research
a. What are existing egg-serving venues in Paris?
b. What is the percentage of egg-serving venues among breakfastng venues and total number of food points?
c. Are they popular?
d. What are characteristics of neighborhoods they are located in?

To answer them mainly following Foursquare data was used: *Venue title, Venue address, Venue categories, Venue tips text, Venue coordinates, Venue rating*

Venue title, categories and *tips text* were used to find initial list breakfasting spots serving eggs. Since Foursquare API has reduced functionality in its free version, the final list was obtained by additional manual analysis of initial list's venues on Foursquare app and appending it with some more venues from The 15 Best Places for Eggs in Paris list [2].

To get high level statistics on egg-serving venues (percentage within all breakfasting places and all restaurants) I also used external sources, mentioned in section 2.1 due to the limitations of free Foursquare API version. Otherwise I'd analyzed an extent at which selected egg places are present within "Breakfast spot" category and "Food" category.

Venue rating has been used to get an idea whether egg places in Paris are popular. To evaluate popularity properly several parameters of egg breakfasting places (rating, number of likes, number of checkins, tips likes and dislikes etc) should be taken into account and compared for other breakfasting places. But it was impossible to get this information due to mentioned limitations. Thus only rating has been taken and no normalization against other food points has been made.

Mean occurences of *venue categories* have been transformed as described in section 2.2 and used as features for clustering analysis to create profiles for egg-serving venues neighborhoods. Same logic has been applied to regular coordinate grid points which would be tested to belong to any of egg-serving venues neighborhoods' profiles.

Venue coordinates (and generated grid points coordinates) have been used to locate egg-venues and potential spots for a new venue on map.

Venue title and address served for better readability and filtering

"Get similar venues" functionality of Forsquare was used to search for yet another set of candidates places serving eggs based on places already found (similarity was defined by Foursquare algorithms)

3. Methodology

3.1. Getting a list of egg-serving breakfasting spots in Paris
The main goal of this section is to obtain a clean list of egg breakfasting places in Paris with all information needed for further analysis. There was several approches on how to tackle that:

a. Search by "egg" string in title (in query) with "Food" category ID. That gave 3 places. After manual check of photos and tips on Foursquare 2 of them have been dropped as irrelevant: one was a creperie, another one was closed.

b. Search by a list of keywords in tips
This was not a particularly good approach since there are at least 2 problems:
– There is no possibility to get a full list of keywords used by Foursquare for venue tagging.
– There is no possibility to get a list of tips containing a key word directly and analyze them: we can only either query a specific tip by its id, or get 1 tip per specified place.

Thus the approach was following:
– Get all venues within "breakfast spot" category
– Fetch one tip for each of them (since no more are available)
– Manually check top comments for several results of "egg breakfast" + "Paris" query in Foursquare app and create a list of keywords
– Select only those breakfast spots that contain at least one of keywords

That gave 3 places one of which has already been found via first approach

c. Find "similar" places for those 3 venues and analyze resulting list manually by relevance.
I used "get similar venues" Foursquare query and got a list of 13 venues 2 of which have been present among previously found spots. I manually checked this list in Foursquare app for keywords, photos and comments and removed 5 of them because of their irrelevance. Thus I totally found 9 egg-serving breakfasting spots.

d. Manual adding of missing places
As far as I found "The 15 Best Places for Eggs in Paris" list [2] created by Foursquare, I decided to check it for more places I potentially missed due to free Foursquare API version restrictions. That list gave me another 8 places abscent in my list. This gave a resulting list of 17 egg-serving breakfasting spots (Table 1)

name	categories	rating	tips count	latitude	longitude	postal code	address	tip text
Benedict	French Restaurant	9.0	172	48.85820815365001	2.3560811411196494	75004	19 rue Sainte-Croix-de-la-Bretonnerie	Brunch fo
Le Saint-Régis	Bistro	8.6	190	48.852930295842626	2.353724241256714	75004	6 rue Jean du Bellay	Must have
Holybelly 19	Breakfast Spot	8.9	234	48.87236651589251	2.360927357451203	75010	19 rue Lucien Sampaix	Si vous a'
Carette	Tea Room	8.9	317	48.86358902223995	2.287205457687378	75016	4 place du Trocadéro	Typicall P
Le Mary Céleste	Cocktail Bar	9.1	166	48.86174155463238	2.3650123178958893	75004	1 rue Commines	Don't thin
Café de Flore	Café	8.4	558	48.85399681424528	2.3326457751586753	75006	172 boulevard Saint-Germain	Sit outsid
Angelina	Tea Room	8.8	605	48.865089750224186	2.3284433919743606	75001	226 Rue de Rivoli	Surprising
Ladurée	Pastry Shop	8.9	996	48.870780615282726	2.3030948638916016	75008	75 Avenue des Champs Elysées	Gorgeous
Les Bonnes Sœurs	French Restaurant	7.4	37	48.85600439835367	2.366941119545712	75003	8 rue du Pas de la Mule	Get the "F
Biglove Caffè	Italian Restaurant	9.0	91	48.86206260694734	2.363556952325989	75003	30 rue Debelleyme	AMAZINC
Eggs & Co	French Restaurant	8.8	148	48.85311560765672	2.331547737121582	75006	11 rue Bernard Palissy	Super cut
Café Marlette	Breakfast Spot	8.2	69	48.88021167483201	2.340392007241914	75009	51 rue des Martyrs	Brunch pa
Claus - La table du petit-déjeuner	Breakfast Spot	8.3	150	48.862457	2.34062	75001	14 rue Jean-Jacques Rousseau	Amazing
Le Pain Quotidien	Breakfast Spot	7.3	21	48.880029714349604	2.340559959411621	75009	54 rue des Martyrs	Même si l
Hardware Société	Breakfast Spot	9.3	125	48.886901473803164	2.3444633609475147	75018	10 rue Lamarck	This has l
Twinkie Breakfasts	Breakfast Spot	7.8	110	48.865297558872626	2.350472361762968	75002	167 rue Saint-Denis	Très fréqu
Paperboy	Breakfast Spot	8.7	123	48.864665	2.366582	75011	137 rue Amelot	Was here

Table 1: Final list of egg-serving breakfasting spots in Paris

3.2. Occurence of egg-serving breakfasting spots within total number of breakfasting spots and total number of food points in Paris
The goal of this section was to understand whether there is a demand on egg-serving breakfasting spots in Paris by determinig:
– the percentage of egg-serving breakfasting spots within "breakfast spot" category venues and within "food" category venues (which is basically a parent category for all restaurants and fastfoods)
– popularity of egg-serving breakfasting spots among visitors

But since Foursquare API limits result list by 30 venues,
– the total number of breakfasting spots I put equal to the maximum number of resuls displayed by Foursquare app when a manually entering query "breakfast spot" + "Paris". It is 120.
– the total number of resaurants I found it on World Cities Culture Forum [3]. It is 44.896 for the year 2017. By "restaurants" different food point types are meant here (caffes, bistros, etc)

The percentage is shown in Table 2.

	total number	% in breakfast spots	% in restaurants
Restaurants	44896	NaN	100.000000
Breakfast spots	120	100.000000	0.267284
Egg serving Breakfast spots	17	14.166667	0.037865

Table 2: Percentage of egg-serving breakfasting spots within "breakfast spot" category venues and within "food" category venues

Venue rating has been used to get an idea whether egg places in Paris are popular. To evaluate popularity properly several parameters of egg breakfasting places (rating, number of likes, number of checkins, tips likes and dislikes etc) should be taken into account and compared for other breakfasting places. But it was impossible to get this information due to mentioned limitations. Thus only rating has been taken to calculate mean rating (= *8.5523*) and no normalization against other food points has been made.

Conclusion:
– Only 0.27 % of total number of food points in Paris are mentioned as serving breakfasts
– Around 14% of them serve eggs (0.04% from total food point number)
– Those spots are popular: mean rating is ~8.5

Which gives an impression that it could make sense to open yet another egg-serving breakfasting place.

3.3. Finding egg-serving breakfasting spots' profiles based on nearby venues categories mean occurence
Next question is where should such new egg-serving breakfasting place be located. To address this I decided to collect profiles of selected existing egg-serving spots neighborhoods and cluster them under a number of neighborhood categories based on feature distribution similarity.

The method of features selection for such profiling was discussed in section 2.3. Some other properties of egg-serving breakfasting spots might also impact the accuracy of clustering. For example, not only presence of certain venues categories, but abscence of some other venues categories nearby. But for this research only mean occurence of categories of present nearby venues was taken.

For each egg-serving breakfasting spot from the list obtained in section 3.1. I retreived all nearby venues in radius of 200 m. That gave 411 nearby venues of 112 categories. Number of nearby of venues for each spot is shown in Table 3.

For future analysis it is important to notice, that some spots have very few nearby venues.

venue id	name	number of nearby venues
5293ae7d11d2fba382d9f652	Benedict	29
4b1411c6f964a520c09c23e3	Le Saint-Régis	27
53f32591498e1cd3c3ec2555	Holybelly 19	17
4adcda14f964a5203a3721e3	Carette	29
5116b70ce4b0d096ad258d22	Le Mary Céleste	29
4adcda04f964a520323221e3	Café de Flore	29
4adcda12f964a520543621e3	Angelina	29
4bc5e23151b376b0ce8e1a6f	Ladurée	29
4b8a4680f964a520a76632e3	Les Bonnes Sœurs	27
583025d07ff1e43c19cd8599	Biglove Caffè	29
4cc9f623b878a093404b799a	Eggs & Co	29
53037fbb498e6f8b7ada68a3	Café Marlette	7
4de77728e4cdfedb8a9dad41	Claus - La table du petit-déjeuner	16
524fef8411d29554626d9a1a	Le Pain Quotidien	8
5710c77a498e3021c0641aa9	Hardware Société	21
4b8a5057f964a520146832e3	Twinkie Breakfasts	27
531499ec11d2a01b87e9e3a3	Paperboy	29

Table 3: Number of venues of different categories nearby to egg-serving breakfasting spots

The breakdown of venues by categories was also performed. But since the accuracy of clustering 17 datapoints (= egg venues) according 112 features (= nearby venues categories) would be low, I decided to reduce number of features by replacing some of them by their parent terms according to the logic described in section 2.1. A part of mapping is shown in Table 4, full version is available via link under reference [6].

main category id	venue main category	final category tag id	final category tag
4bf58dd8d48988d1e7931735	Jazz Club	4d4b7104d754a06370d81259	Arts & Entertainment
4bf58dd8d48988d137941735	Theater	4d4b7104d754a06370d81259	Arts & Entertainment
4bf58dd8d48988d1e2931735	Art Gallery	4d4b7104d754a06370d81259	Arts & Entertainment
52e81612bcbc57f1066b79e7	Circus	4d4b7104d754a06370d81259	Arts & Entertainment
4deefb944765f83613cdba6e	Historic Site	4d4b7104d754a06370d81259	Arts & Entertainment

Table 4: First rows of a table for mapping of initial categories to their parent categories.
Full version is available via link under reference [7]

The final list contains 25 categories, which are displayed in Table 5 along with corresponding number of venues nearby egg-serving breakfasting spots.

The next step was to find how often a venue of each category occures near each of egg-serving breakfasting spot. To do this I:
– One hot encoded categories
– Calculated mean frequency of their occurence and used those as features for further clustering (a part of features is shown in Table 6, full version is available via link under reference [7])
– To make results more visual we'll create a chart of top 10 nearby venues categories and arrange them by descending occurence. In cases there would be less nearby venues' categories than 10, well put nan to remaining cells. In Table 7 see that 3 spots are surrounded by 7 and 9 categories of other venues.

selected category id	selected category	occurence
4d4b7104d754a06370d81259	Arts & Entertainment	18
4bf58dd8d48988d142941735	Asian Restaurant	11
4f4528bc4b90abdf24c9de85	Athletics & Sports	6
4bf58dd8d48988d16a941735	Bakery	14
4bf58dd8d48988d116941735	Bar	28
4bf58dd8d48988d143941735	Breakfast Spot	4
4bf58dd8d48988d16d941735	Café	12
4bf58dd8d48988d145941735	Chinese Restaurant	7
4bf58dd8d48988d103951735	Clothing Store	30
4bf58dd8d48988d1e0931735	Coffee Shop	19
52e81612bcbc57f1066b79f2	Creperie	7
4bf58dd8d48988d1f6941735	Department Store	1
4bf58dd8d48988d1d0941735	Dessert Shop	21
4d4b7105d754a06374d81259	Food	33
4bf58dd8d48988d1f9941735	Food & Drink Shop	11
4bf58dd8d48988d10c941735	French Restaurant	55
4bf58dd8d48988d1fa931735	Hotel	23
4bf58dd8d48988d110941735	Italian Restaurant	13
4bf58dd8d48988d111941735	Japanese Restaurant	11
4d4b7105d754a06376d81259	Nightlife Spot	2
4d4b7105d754a06377d81259	Outdoors & Recreation	7
4bf58dd8d48988d10f951735	Pharmacy	2
4bf58dd8d48988d164941735	Plaza	11
4d4b7105d754a06378d81259	Shop & Service	39
4bf58dd8d48988d1c4941735	Restaurant	26

Table 5: Final list of categories used as features and corresponding number of venues located nearby egg-serving breakfast spots

egg place venue name	egg place venue id	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Bakery	Bar	Breakfast Spot	Café	Chinese Restaurant	...	French Restaurant	Hotel
Angelina	4adcda12f964a520543621e3	0.000000	0.000000	0.000000	0.000000	0.068966	0.000000	0.000000	0.034483	...	0.172414	0.172414
Benedict	5293ae7d11d2fba382d9f652	0.068966	0.000000	0.000000	0.000000	0.103448	0.000000	0.034483	0.000000	...	0.137931	0.000000
Biglove Caffè	583025d07ff1e43c19cd8599	0.068966	0.000000	0.000000	0.000000	0.137931	0.000000	0.034483	0.000000	...	0.034483	0.000000
Café Marlette	53037fbb498e6f8b7ada68a3	0.000000	0.000000	0.000000	0.142857	0.000000	0.142857	0.000000	0.000000	...	0.142857	0.142857
Café de Flore	4adcda04f964a520323221e3	0.034483	0.034483	0.000000	0.000000	0.000000	0.000000	0.034483	0.034483	...	0.103448	0.068966
Carette	4adcda14f964a5203a3721e3	0.137931	0.068966	0.000000	0.000000	0.034483	0.000000	0.034483	0.000000	...	0.206897	0.103448
Claus - La table du petit-déjeuner	4de77728e4cdfedb8a9dad41	0.062500	0.000000	0.000000	0.062500	0.125000	0.000000	0.000000	0.062500	...	0.312500	0.000000
Eggs & Co	4cc9f623b878a093404b799a	0.034483	0.034483	0.000000	0.000000	0.034483	0.000000	0.068966	0.034483	...	0.103448	0.000000
Hardware Société	5710c77a498e3021c0641aa9	0.047619	0.047619	0.000000	0.047619	0.095238	0.000000	0.000000	0.000000	...	0.190476	0.000000
Holybelly 19	53f32591498e1cd3c3ec2555	0.000000	0.117647	0.058824	0.117647	0.000000	0.058824	0.000000	0.058824	...	0.117647	0.058824
Ladurée	4bc5e23151b376b0ce8e1a6f	0.000000	0.000000	0.034483	0.034483	0.000000	0.034483	0.034483	0.034483	...	0.137931	0.172414
Le Mary Céleste	5116b70ce4b0d096ad258d22	0.103448	0.000000	0.034483	0.034483	0.068966	0.000000	0.068966	0.000000	...	0.000000	0.000000
Le Pain Quotidien	524fef8411d29554626d9a1a	0.000000	0.000000	0.000000	0.125000	0.000000	0.125000	0.000000	0.000000	...	0.250000	0.125000
Le Saint-Régis	4b1411c6f964a520c09c23e3	0.000000	0.000000	0.000000	0.037037	0.000000	0.000000	0.000000	0.000000	...	0.259259	0.037037
Les Bonnes Sœurs	4b8a4680f964a520a76632e3	0.037037	0.000000	0.000000	0.037037	0.000000	0.000000	0.074074	0.000000	...	0.185185	0.037037
Paperboy	531499ec11d2a01b87e9e3a3	0.068966	0.103448	0.034483	0.103448	0.172414	0.000000	0.000000	0.000000	...	0.034483	0.068966
Twinkie Breakfasts	4b8a5057f964a520146832e3	0.000000	0.037037	0.074074	0.037037	0.222222	0.000000	0.037037	0.037037	...	0.074074	0.037037

Table 6: Mean venue categories occurences for egg-serving breakfasting spots

selected category id	selected category	occurence
4d4b7104d754a06370d81259	Arts & Entertainment	18
4bf58dd8d48988d142941735	Asian Restaurant	11
4f4528bc4b90abdf24c9de85	Athletics & Sports	6
4bf58dd8d48988d16a941735	Bakery	14
4bf58dd8d48988d116941735	Bar	28
4bf58dd8d48988d143941735	Breakfast Spot	4
4bf58dd8d48988d16d941735	Café	12
4bf58dd8d48988d145941735	Chinese Restaurant	7
4bf58dd8d48988d103951735	Clothing Store	30
4bf58dd8d48988d1e0931735	Coffee Shop	19
52e81612bcbc57f1066b79f2	Creperie	7
4bf58dd8d48988d1f6941735	Department Store	1
4bf58dd8d48988d1d0941735	Dessert Shop	21
4d4b7105d754a06374d81259	Food	33
4bf58dd8d48988d1f9941735	Food & Drink Shop	11
4bf58dd8d48988d10c941735	French Restaurant	55
4bf58dd8d48988d1fa931735	Hotel	23
4bf58dd8d48988d110941735	Italian Restaurant	13
4bf58dd8d48988d111941735	Japanese Restaurant	11
4d4b7105d754a06376d81259	Nightlife Spot	2
4d4b7105d754a06377d81259	Outdoors & Recreation	7
4bf58dd8d48988d10f951735	Pharmacy	2
4bf58dd8d48988d164941735	Plaza	11
4d4b7105d754a06378d81259	Shop & Service	39
4bf58dd8d48988d1c4941735	Restaurant	26

Table 5: Final list of categories used as features and corresponding number of venues located nearby egg-serving breakfast spots

egg place venue name	egg place venue id	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Bakery	Bar	Breakfast Spot	Café	Chinese Restaurant	...	French Restaurant	Hotel
Angelina	4adcda12f964a520543621e3	0.000000	0.000000	0.000000	0.000000	0.068966	0.000000	0.000000	0.034483	...	0.172414	0.172414
Benedict	5293ae7d11d2fba382d9f652	0.068966	0.000000	0.000000	0.000000	0.103448	0.000000	0.034483	0.000000	...	0.137931	0.000000
Biglove Caffè	583025d07ff1e43c19cd8599	0.068966	0.000000	0.000000	0.000000	0.137931	0.000000	0.034483	0.000000	...	0.034483	0.000000
Café Marlette	53037fbb498e6f8b7ada68a3	0.000000	0.000000	0.000000	0.142857	0.000000	0.142857	0.000000	0.000000	...	0.142857	0.142857
Café de Flore	4adcda04f964a520323221e3	0.034483	0.034483	0.000000	0.000000	0.000000	0.000000	0.034483	0.034483	...	0.103448	0.068966
Carette	4adcda14f964a5203a3721e3	0.137931	0.068966	0.000000	0.000000	0.034483	0.000000	0.034483	0.000000	...	0.206897	0.103448
Claus - La table du petit-déjeuner	4de77728e4cdfedb8a9dad41	0.062500	0.000000	0.000000	0.062500	0.125000	0.000000	0.000000	0.062500	...	0.312500	0.000000
Eggs & Co	4cc9f623b878a093404b799a	0.034483	0.034483	0.000000	0.000000	0.034483	0.000000	0.068966	0.034483	...	0.103448	0.000000
Hardware Société	5710c77a498e3021c0641aa9	0.047619	0.047619	0.000000	0.047619	0.095238	0.000000	0.000000	0.000000	...	0.190476	0.000000
Holybelly 19	53f32591498e1cd3c3ec2555	0.000000	0.117647	0.058824	0.117647	0.000000	0.058824	0.000000	0.058824	...	0.117647	0.058824
Ladurée	4bc5e23151b376b0ce8e1a6f	0.000000	0.000000	0.034483	0.034483	0.000000	0.034483	0.034483	0.034483	...	0.137931	0.172414
Le Mary Céleste	5116b70ce4b0d096ad258d22	0.103448	0.000000	0.034483	0.034483	0.068966	0.000000	0.068966	0.000000	...	0.000000	0.000000
Le Pain Quotidien	524fef8411d29554626d9a1a	0.000000	0.000000	0.000000	0.125000	0.000000	0.125000	0.000000	0.000000	...	0.250000	0.125000
Le Saint-Régis	4b1411c6f964a520c09c23e3	0.000000	0.000000	0.000000	0.037037	0.000000	0.000000	0.000000	0.000000	...	0.259259	0.037037
Les Bonnes Sœurs	4b8a4680f964a520a76632e3	0.037037	0.000000	0.000000	0.037037	0.000000	0.000000	0.074074	0.000000	...	0.185185	0.037037
Paperboy	531499ec11d2a01b87e9e3a3	0.068966	0.103448	0.034483	0.103448	0.172414	0.000000	0.000000	0.000000	...	0.034483	0.068966
Twinkie Breakfasts	4b8a5057f964a520146832e3	0.000000	0.037037	0.074074	0.037037	0.222222	0.000000	0.037037	0.037037	...	0.074074	0.037037

Table 6: Mean venue categories occurences for egg-serving breakfasting spots

egg place venue id	egg place venue name	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
2f964a520543621e3	Angelina	Clothing Store	French Restaurant	Hotel	Shop & Service	Japanese Restaurant	Bar	Dessert Shop	Food	Restaurant	Coffee Shop
'd11d2fba382d9f652	Benedict	Dessert Shop	French Restaurant	Shop & Service	Clothing Store	Bar	Restaurant	Arts & Entertainment	Nightlife Spot	Italian Restaurant	Food
d07ff1e43c19cd8599	Biglove Caffè	Bar	Food	Shop & Service	Food & Drink Shop	Coffee Shop	Clothing Store	Restaurant	Dessert Shop	Arts & Entertainment	Japanese Restaurant
b498e6f8b7ada68a3	Café Marlette	Dessert Shop	Creperie	Bakery	Breakfast Spot	Hotel	French Restaurant	Coffee Shop	NaN	NaN	NaN
4f964a520323221e3	Café de Flore	Clothing Store	Shop & Service	Italian Restaurant	French Restaurant	Food	Restaurant	Plaza	Japanese Restaurant	Hotel	Asian Restaurant
4f964a5203a3721e3	Carette	French Restaurant	Food	Arts & Entertainment	Hotel	Plaza	Asian Restaurant	Bar	Café	Clothing Store	Coffee Shop
8e4cdfedb8a9dad41	Claus - La table du petit-déjeuner	French Restaurant	Food	Bar	Shop & Service	Bakery	Chinese Restaurant	Clothing Store	Food & Drink Shop	Arts & Entertainment	NaN
3b878a093404b799a	Eggs & Co	Italian Restaurant	Clothing Store	French Restaurant	Shop & Service	Café	Restaurant	Dessert Shop	Japanese Restaurant	Plaza	Chinese Restaurant
a498e3021c0641aa9	Hardware Société	French Restaurant	Restaurant	Food	Outdoors & Recreation	Bar	Dessert Shop	Asian Restaurant	Bakery	Creperie	Arts & Entertainment
1498e1cd3c3ec2555	Holybelly 19	Coffee Shop	Asian Restaurant	Bakery	French Restaurant	Food	Restaurant	Athletics & Sports	Breakfast Spot	Hotel	Chinese Restaurant
151b376b0ce8e1a6f	Ladurée	Shop & Service	Hotel	French Restaurant	Clothing Store	Restaurant	Athletics & Sports	Bakery	Breakfast Spot	Café	Chinese Restaurant
ce4b0d096ad258d22	Le Mary Céleste	Clothing Store	Shop & Service	Coffee Shop	Dessert Shop	Arts & Entertainment	Bar	Café	Italian Restaurant	Athletics & Sports	Bakery
411d29554626d9a1a	Le Pain Quotidien	French Restaurant	Dessert Shop	Creperie	Bakery	Breakfast Spot	Hotel	Coffee Shop	NaN	NaN	NaN
6f964a520c09c23e3	Le Saint-Régis	French Restaurant	Outdoors & Recreation	Creperie	Shop & Service	Restaurant	Food & Drink Shop	Dessert Shop	Italian Restaurant	Food	Bakery
0f964a520a76632e3	Les Bonnes Sœurs	French Restaurant	Coffee Shop	Food	Shop & Service	Café	Restaurant	Food & Drink Shop	Bakery	Arts & Entertainment	Hotel
c11d2a01b87e9e3a3	Paperboy	Bar	Shop & Service	Asian Restaurant	Bakery	Restaurant	Hotel	Food & Drink Shop	Arts & Entertainment	Food	Athletics & Sports
7f964a520146832e3	Twinkie Breakfasts	Bar	French Restaurant	Plaza	Athletics & Sports	Japanese Restaurant	Food	Bakery	Clothing Store	Coffee Shop	Restaurant

Table 7: Top 10 nearby venues’ categories for each egg-serving breakfasting spot

3.4. Clustering egg-serving breakfasting spots’ profiles and locating them on Paris map

The goal of this section is to understand whether egg-serving breakfasting spots’ neigbourhoods have any similarities (= same frofile of nearby venues' categorie's occurece). To do this I desided to perform k-means clustering.

The preliminary step was to deside, how much clusters we should take. For this I calculated Average Within Cluster Sum of Squares for each number of clusters to select a number which gives most dense clusters but having adequate (= not too small) number of points within (Figure 1).

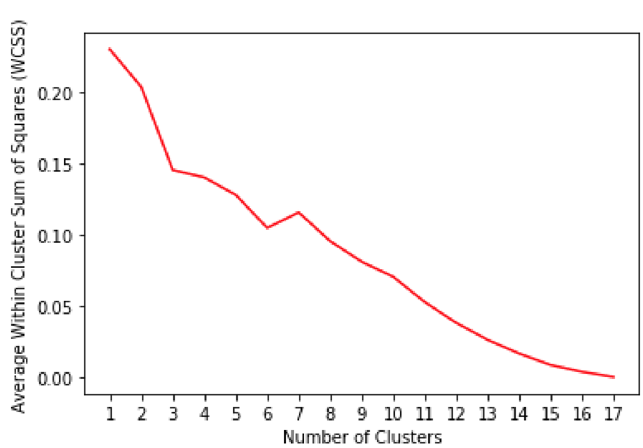


Figure 1: Average WCSS for different number of clusters

I chose 3 clusters as an optimal number because there is a significant drop of the Av. WCSS, and relatively big number of datapoints within clusters. Given that I run k-means clustering analysis, obtained 3 clusters of 2, 9 and 6 points (Table 8) and located them on Paris map using Folium library (Figure 2).

egg place venue id	egg place venue name	categories	cluster label
5293ae7d11d2fba382d9f652	Benedict	French Restaurant	1
4b1411c6f964a520c09c23e3	Le Saint-Régis	Bistro	2
53f32591498e1cd3c3ec2555	Holybelly 19	Breakfast Spot	2
4adcda14f964a5203a3721e3	Carette	Tea Room	2
5116b70ce4b0d096ad258d22	Le Mary Céleste	Cocktail Bar	1
4adcda04f964a520323221e3	Café de Flore	Café	1
4adcda12f964a520543621e3	Angelina	Tea Room	1
4bc5e23151b376b0ce8e1a6f	Ladurée	Pastry Shop	1
4b8a4680f964a520a76632e3	Les Bonnes Sœurs	French Restaurant	2
583025d07ff1e43c19cd8599	Biglove Caffè	Italian Restaurant	1
4cc9f623b878a093404b799a	Eggs & Co	French Restaurant	1
53037fbb498e6f8b7ada68a3	Café Marlette	Breakfast Spot	0
4de77728e4cdfedb8a9dad41	Claus - La table du petit-d	Breakfast Spot	2
524fef8411d29554626d9a1a	Le Pain Quotidien	Breakfast Spot	0
5710c77a498e3021c0641aa9	Hardware Société	Breakfast Spot	2
4b8a5057f964a520146832e3	Twinkie Breakfasts	Breakfast Spot	1
531499ec11d2a01b87e9e3a3	Paperboy	Breakfast Spot	1

Table 8: Egg-serving breakfasting spots clusters

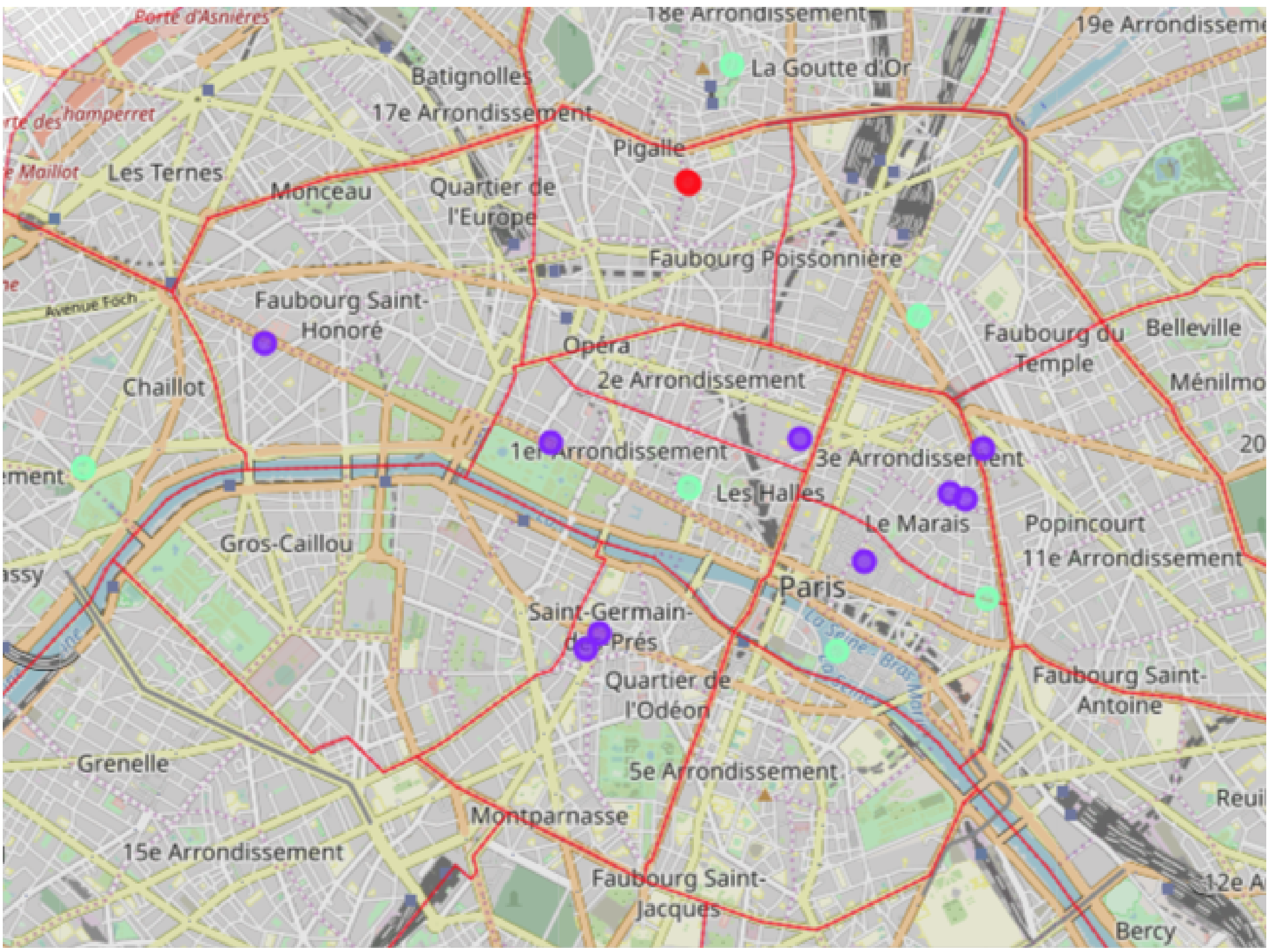


Figure 2: Map of existing egg-serving breakfasting points in Paris clustered according their neighborhood profiles

In order to understand what are main characteristics of clusters I analyzed top 10 most frequent nearby venue categories for each egg-serving breakfasting spot in each cluster (Table 9). The result aggregated result is displayed in Table 10.

egg place venue name	categories	cluster label	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
Café Marlette	Breakfast Spot	0	Dessert Shop	Creperie	Bakery	Breakfast Spot	Hotel	French Restaurant	Coffee Shop	NaN	NaN	NaN
Le Pain Quotidien	Breakfast Spot	0	French Restaurant	Dessert Shop	Creperie	Bakery	Breakfast Spot	Hotel	Coffee Shop	NaN	NaN	NaN
egg place venue name	**categories**	**cluster label**	**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**
Benedict	French Restaurant	1	Dessert Shop	French Restaurant	Shop & Service	Clothing Store	Bar	Restaurant	Arts & Entertainment	Nightlife Spot	Italian Restaurant	Food
Le Mary Céleste	Cocktail Bar	1	Clothing Store	Shop & Service	Coffee Shop	Dessert Shop	Arts & Entertainment	Bar	Café	Italian Restaurant	Athletics & Sports	Bakery
Café de Flore	Café	1	Clothing Store	Shop & Service	Italian Restaurant	French Restaurant	Food	Restaurant	Plaza	Japanese Restaurant	Hotel	Asian Restaurant
Angelina	Tea Room	1	Clothing Store	French Restaurant	Hotel	Shop & Service	Japanese Restaurant	Bar	Dessert Shop	Food	Restaurant	Coffee Shop
Ladurée	Pastry Shop	1	Shop & Service	Hotel	French Restaurant	Clothing Store	Restaurant	Athletics & Sports	Bakery	Breakfast Spot	Café	Chinese Restaurant
Biglove Caffè	Italian Restaurant	1	Bar	Food	Shop & Service	Food & Drink Shop	Coffee Shop	Clothing Store	Restaurant	Dessert Shop	Arts & Entertainment	Japanese Restaurant
Eggs & Co	French Restaurant	1	Italian Restaurant	Clothing Store	French Restaurant	Shop & Service	Café	Restaurant	Dessert Shop	Japanese Restaurant	Plaza	Chinese Restaurant
Twinkie Breakfasts	Breakfast Spot	1	Bar	French Restaurant	Plaza	Athletics & Sports	Japanese Restaurant	Food	Bakery	Clothing Store	Coffee Shop	Restaurant
Paperboy	Breakfast Spot	1	Bar	Shop & Service	Asian Restaurant	Bakery	Restaurant	Hotel	Food & Drink Shop	Arts & Entertainment	Food	Athletics & Sports
egg place venue name	**categories**	**cluster label**	**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**
Le Saint-Régis	Bistro	2	French Restaurant	Outdoors & Recreation	Creperie	Shop & Service	Restaurant	Food & Drink Shop	Dessert Shop	Italian Restaurant	Food	Bakery
Holybelly 19	Breakfast Spot	2	Coffee Shop	Asian Restaurant	Bakery	French Restaurant	Food	Restaurant	Athletics & Sports	Breakfast Spot	Hotel	Chinese Restaurant
Carette	Tea Room	2	French Restaurant	Food	Arts & Entertainment	Hotel	Plaza	Asian Restaurant	Bar	Café	Clothing Store	Coffee Shop
Les Bonnes Sœurs	French Restaurant	2	French Restaurant	Coffee Shop	Food	Shop & Service	Café	Restaurant	Food & Drink Shop	Bakery	Arts & Entertainment	Hotel
Claus - La table du petit-déjeuner	Breakfast Spot	2	French Restaurant	Food	Bar	Shop & Service	Bakery	Chinese Restaurant	Clothing Store	Food & Drink Shop	Arts & Entertainment	NaN
Hardware Société	Breakfast Spot	2	French Restaurant	Restaurant	Food	Outdoors & Recreation	Bar	Dessert Shop	Asian Restaurant	Bakery	Creperie	Arts & Entertainment

Table 9: Top 10 most frequent nearby venues categories for egg-serving breakfasting places

cluster label	number of datapoints	descriptive characteristics
0	2	relatively fast eating spots
1	9	clothing stores, bars, shops
2	6	french restaurants, other food points

Table 10: Egg-serving breakfasting places neighborhood clusters descriptions

It is important to point out that cluster 0 consists of only 2 datapoints, which, taking into account the number of features (= 25) could give a lot of "false positive" results when trying to attribute coordinate grid points to this cluster (see section 3.7. further)

3.5. Getting a coordinate grid of points which would be tested to belong to one of egg-serving breakfasting spots' clusters

Further sections are dedicated to finding locations falling into same clusters, which would be proposed as candidates to open a new egg-serving breakfast spot. And the first step was to cover Paris with coordinates grid. The distance between grid point was chosen to be 300 m – thus all venues would belong to a radius of 200 m of at least one grid point.

Following approach to cover Paris was used:
– A pair of points have been chosen in a way that they defined South-West and North-East angles of a rectangle, which would include Paris
– This rectangle has been filled in with 300 m coordinates grid (= 3.111 points)
– Each of this grid has been checked on being located within Paris coordinates polygone

As far as Foursquare and Folium work with degree coordinates and to set up the grid we needed metric coordinates, an EPSG:3035 Mercator projection for Europe has been used [5].
– Paris degree coordinates polygone and a pair of rectangle defining poinst have been broadcasted into metric coordinates
– Metric grid points mask for Paris has been obtained according to the logic described above,
– And the result has been broadcasted back to obtain final 300 m grid within Paris boardes (= 2.345 points) in degrees to make it usable with Foursquare queries (Figure 3).
I used shapely and pyproj libraries to work on geographic shapes and projections

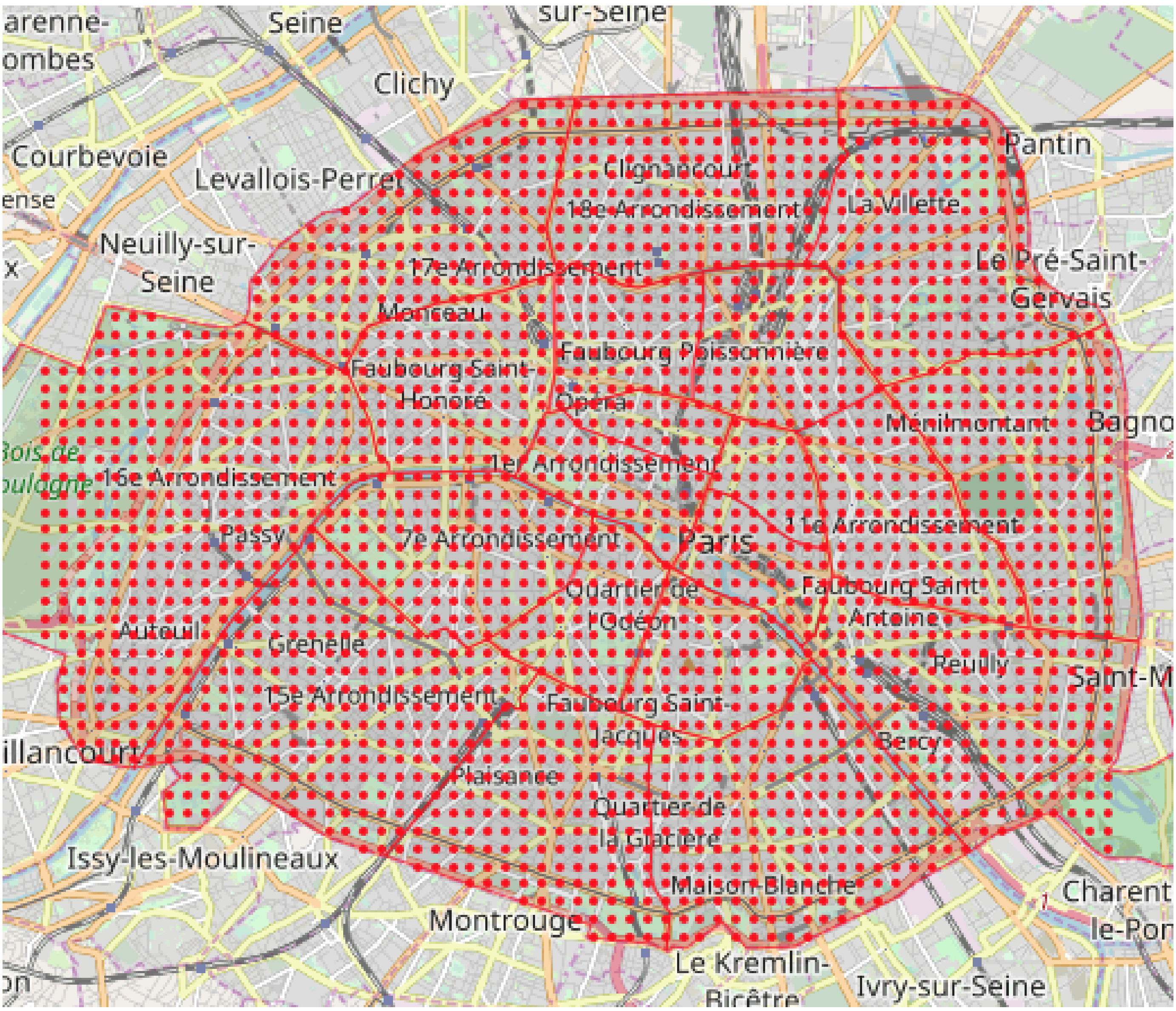


Figure 3: 300 m coordinate grid of 2.345 points displayed on Paris map

3.6. Find profiles for each of grid points based on nearby venues categories mean occurence
The next step is to get the profile of each grid point based on occurence of neighborhood venues categories within 200 m radius.

Even though for egg-serving breakfasting points we retreived up to 30 nearby venues, here this number was limited to 24, since this significantly reduce the query time. But this decision might impact the accuracy of grid points attributions to clusters defined earlier. The first version of this list contained 25.566 venues.

Since initial egg-serving spots have nearby places of only 112 raw categories, I removed from the list all places which do not fall into the list of such categories, because they won't have any significance for clustering. Thus the number of venues for all grid points decreased to 20.521 and some points which have been surrounded by only irrelevant categories also disappeared from the list, reducing number of candidate points to 2.177.

The further logic of neighborhood profiling is exactly the same as described for egg-serving breakfasting spots in section 3.3. and includes:
– Replacement of initial 112 categories by 25 final categories according to the mapping table (Table 4)
– One hot encoding of 25 categories for each of 2.177 grid points
– Calculation of mean occurences of each category for each grid point to obtain a feature matrix, which can be found via link under reference [8].

3.7. Attribute each grid point to one of clusters earlier defined or set it as an outlier
The aim of this step is to reduce initial number of candidate points from 2.177 by finding those which are located within neighborhoods having profiles falling into one of 3 clusters earlier defined.

The potential candidate cluster for attribution of each point have been selected based on the euclidian distance to each cluster centers. The closest cluster has been chosen. Results for several points are shown in Table 11. The breakdown of candidate grid points by clusters is shown in Table 12.

point index	closest cluster index	distance to cluster center
0	2	0.357567
1	1	0.991687
9	2	0.816250
10	2	0.816250
12	2	0.998388

Table 11: Examples of candidate grid points to be attributed to one of 3 clusters

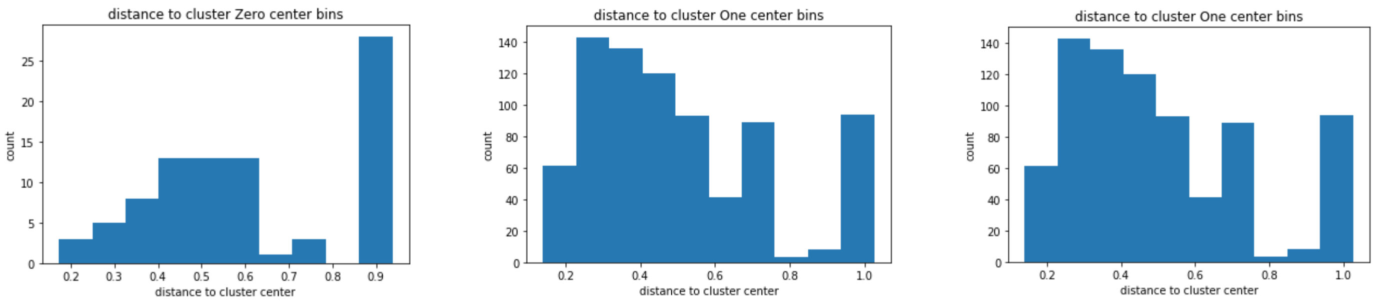
cluster label	number of candidate grid points
0	87
1	788
2	1302

Table 12: The breakdown of candidate grid points by clusters

Next step was to set up a parameter which defined whether a point belong to the closest cluster or is an outlier. For each cluster the distribution of distances from its center to each of its closest points has been built (Figure 4).

The initial idea was to take an X-quantile as a cutoff parameter which gives a possibility to calculate a cutoff distance. But the distribution didn't give any insights on what cutoff X-quantile should be selected. That's why I decided to set a cutoff distance itself based on distances from cluster center of corresponding egg-serving breakfasting spots.

I initially decided to set a cutoff distance equal to the distance from a cluster center to the most distant of its points. But that gave me > 800 points all over the city.



Figures 4 (a, b, c): Distribution of distances to closest cluster center for candidate grid points

egg place venue id	egg place venue name	categories	cluster label	distance to cluster center
524fef8411d29554626d9a1a	Le Pain Quotidien	Breakfast Spot	0	0.324383
5116b70ce4b0d096ad258d22	Le Mary Céleste	Cocktail Bar	1	0.164253
4b8a4680f964a520a76632e3	Les Bonnes Sœurs	French Restaurant	2	0.147337

Table 13: Distances from cluster centers to their respective closest cluster points (= egg-serving breakfasting spots)

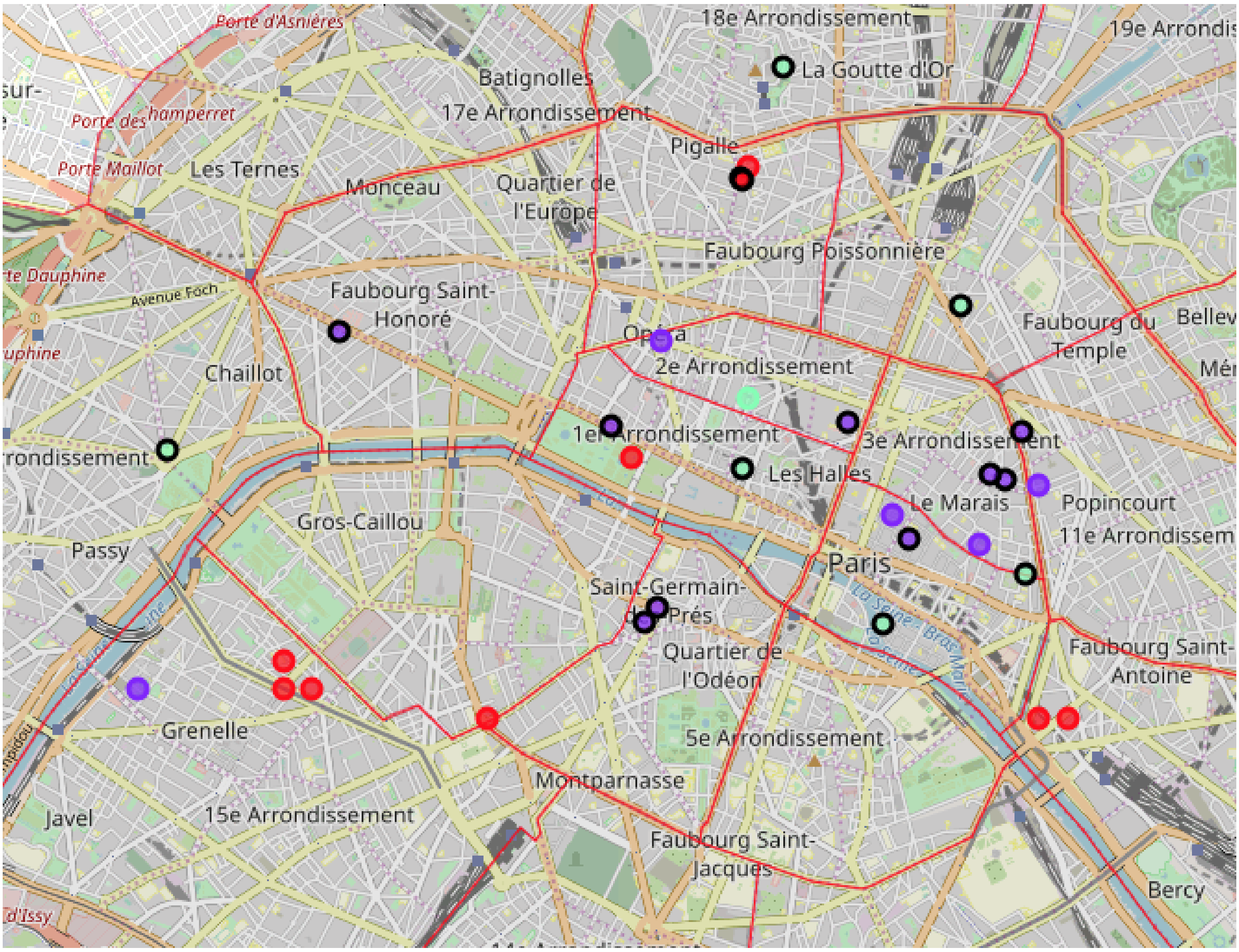
Thus I took distance to closest points as cutoff distance (Table 13). All grid points which are closer to corresponding cluster center than the closest cluster point were considered to belong to the cluster. That gave a list of 14 candidate grid points (Table 14): 8 for cluster 0, 5 for cluster 1 and 1 for cluster 2.

point index	closest cluster index	distance to cluster center	lon	lat
508	0	0.315592	2.297973	48.848990
509	0	0.214822	2.297973	48.850763
545	0	0.243628	2.300668	48.848990
790	0	0.320684	2.316838	48.847216
1027	0	0.318689	2.330312	48.863175
1229	0	0.173273	2.341092	48.880901
1688	0	0.318689	2.368042	48.847216
1735	0	0.313880	2.370737	48.847216
336	1	0.161974	2.284498	48.848990
1079	1	0.150637	2.333007	48.870266
1456	1	0.145813	2.354567	48.859629
1599	1	0.144165	2.362652	48.857856
1696	1	0.140671	2.368042	48.861402
1221	2	0.128865	2.341092	48.866721

Table 14: List of 14 candidate grid points falling into one of 3 clusters

Relatively big number of points 8 points belonging to cluster 0 was not surprizing since this cluster has only 2 initial egg-serving reference points which, taking into account the number of features (= 25) could give a lot of "false positive" grid points.

I used follium to locate selected candidate points on Paris map (Figure 5).



Figures 5: Distribution of 14 candidate grid points on Paris map (existing egg-spots have black border)

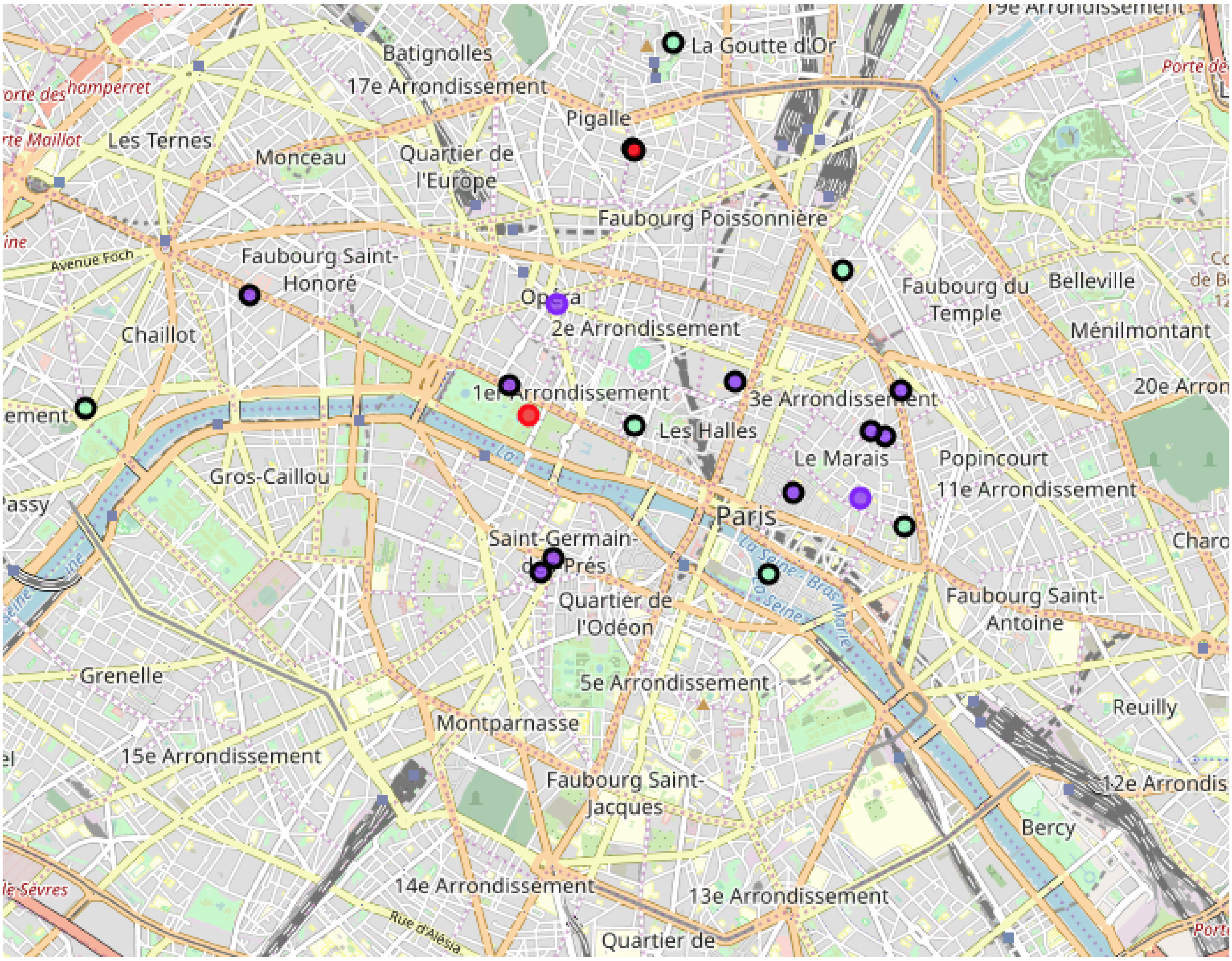
3.8. Select final candidate points based on their distance from city center and initial egg-serving breakfasting spots and locate them on map

After examining Figure 5, I decided that 2 more restrictions would make the list of candidate points more relevant. Best fit would be defined by:
– their proximity to city center: they should be located in less than 3 km from Louvre
– absense of other egg-serving breakfasting spots nearby: since Paris is a very dense city, the distance of 300 m should be fine

To impose those restrictions I performed the same set of coordinates transformations from degrees to meters and, after calculating distances, the reverse. That gave me the final list of 4 candidate points (Table 15) which I ploted on Paris map using Folium (Figure 6).

point index	closest cluster index	lon	lat	distance from louvre (m)	distance from nearest existing egg place (m)
1027	0	2.330312	48.863175	875.0	385.0
1079	1	2.333007	48.870266	1596.0	1013.0
1599	1	2.362652	48.857856	2843.0	571.0
1221	2	2.341092	48.866721	986.0	723.0

Table 15: Final list of candidate points to exmine their neighborhoods for opening a new egg-serving breakfasting spot



Figures 6: Distribution of final 4 candidate grid points on Paris map (existing egg-spots have black border)

This concludes my analysis. I have found 4 points which neighborhoods should be checked by business owners as good options to locate a new egg-breakfasting venue in Paris, since they are:
– close to city center (< 3000 m from Louvre)
– far enough from already existing egg breakfasting venues (> 300 m)
– in neighborhoods which are similar to those of already existing egg-breakfasting places

4. Results

a. Althoug Paris has a great number of restaurants, and a lot of breakfasting spots (> 120), only few of them serve eggs (= 17), which is around 14% of all breakfasting places and 0.04% of all food points. And egg spots are popular with mean rating is ~8.5. So it seems that to open another one might be a good idea.

b. After analyzing categories of each egg breakfasting point's nearby venues which have been supposed to be used as features for clustering, it has been decided to reduce their initial number (= 112) by replacing less frequent of them by their parent categories. Otherwise number of features would be too large compared to the number of datapoints (= 17) That gave 25 features instead of 112.

c. Using new categories of nearby venues, initial egg-points have been clustered in 3 groups. Number of clusters have been selected by analyzing Average Within Cluster Sum of Squares: a balance between its minimal values and relatively big number of datapoints within clusters should have been found.

– Cluster 0: characterized by relatively fast eating spots (only 2 spots fell into this cluster, which would potentially reduce
– Cluster 1: characterized by clothing stores, bars, shops (9 points)
– Cluster 2: characterized by french restaurants + other food points (6 points)

d. The next step was to find points in Paris having same caracteristics that egg-places clusters have
– After dropping a 300 m coordinates grid over Paris we analyzed each grid point nearby venues' categories, replacing them by reduced list of parent categories the same way we did for egg-spots nearby venues
– We dropped all nearby venues with categories which were not in the list of egg-spots nearby venues categories (the reduced one). Around 2000 grid points left for analysis
– Each point has been attributed either to one of 3 clusters or marked as outlier based on a cutoff distance from nearet cluster center. Cutoff distance has been selected as minimal distance from an egg place belonging to a cluster to this cluster's center (this approach could be challenged)
14 points have been found.

* 8 belonging to cluster 0
* 5 belonging to cluster 1
* 1 belonging to cluster 2

e. Among selected points we did last round of filtering by distance to city center (not far than 3000 m from Louvre) and distance to nearest existing egg breakfasting spot (not less than 300 m). That gave us 4 final candidate points
– Two of them (1599, 1079) are in Marais and Opera neighborhoods which are touristic and shopping intensive (and belong to Cluter 1 characterized by clothing stores, bars, shops
– One in Bourse area (1221) which is a typical location for restaurants (and it is confirmed by the fact that it belongs to Cluster 2 => french resto + other food points)
– And the last one (1027) is located in Tuileries, which even though very touristic is not really adapted to open a breakfasting spot. This poor match can be explained since the point belongs to Cluster 0 having only 2 breakfasting points which is not enough for accurate clustering

5. Discussion

Although proper results have been obtained, there are several descisions that could be challenged to improve the model in future.

a. Selection of egg serving breakfasting places could be improved by acessing a paid version of Foursquare API which gives possibility to do proper tips text analysis and get full number of nearby places.

b. Popularity of egg places was not a subject of a proper analysis in this project. To evaluate popularity properly several parameters of egg breakfasting places (rating, number of likes, number of checkins, tips likes and dislikes etc) should be taken into account and compared for other breakfasting places.

c. Clustering by k-means algorithm of 17 egg breakfasting points based on 25 features is also an approach that might be challenged (too much features for such small number of points)

d. Further reduction of the number of features could be done for example by finding correlations between independent variables, but that was not done

e. Some other properties of egg breakfasting spots might be also significant to perform accurate clustering. For example, not only presence of certain venues categories, but abscence of some other venues categories nearby.

f. For egg-serving breakfasting venues we retreived up to 30 nearby venues, but for grid points this number was limited to 24, since this significantly reduce the query time. But this decision might impact the accuracy of further grid points attributions to clusters defined earlier

g. Selection of cutoff distance which defines whether a grid point belong to a cluster is also a subect of discussion. Initially I thought that distance from a cluster center to its farest point should be taken. But that gave too much points for analysis (more than 400). Thus I decided to take the distance to closest cluster point.

h. Attribution of grid points (= location candidates to open new egg venue) to cluster 0 is supposed not to be accurate enough since there are only 2 points in this cluster (with 25 features!)

i. For final points selection I took certain distances from city center and from existing egg breakfasting spots as cutoff, but Paris is a dense and in a way decentralized city. It might occur that good location to open a new egg-breakfasting-venue would not meet those two creteria. Final descision should be made by business owners.

j. It would be good to provide addresses for these points, not their coordinates, but it can be mabe only by google API, which is paid. Or manually, which does not make sense within this project.

6. Conclusions

The purpose of this project was to understand whether there is a niche in Paris for an egg-breakfasting point, and, if yes, to identify points in Paris which would potentially be good to open one. Those points should have:
– similar neighborhoods as existing egg-breakfasting points (in terms of mean occurences of nearby venues categories)
– be close to city center
– be far enough from already existing egg-breakfasting points

By fetching from Foursquare an analyzing existing egg-breakfasting points in Paris and their properties we understood that there is a niche for another one. Clustering analysis of those points taking occurences of their nearby venues categories as feature gave us 3 profiles (= clusters) other potential locations should match to be a good candidate. Attribution of coordinates grid points to those clusters reduced number of candidates to 14. Additional restrictions on distances from city center and from existing egg breakfasting spots left 4 final points.

Final decission on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods around those points, taking into consideration additional factors like levels of noise, transportation infrastructure, real estate availability, prices, social and economic dynamics of every neighborhood etc.

7. References

[1] Foursquare API: https://developer.foursquare.com/docs/api-reference/venues/listed/
[2] The 15 Best Places for Eggs in Paris: https://foursquare.com/top-places/paris/best-places-eggs
[3] World Cities Culture Forum: http://worldcitiescultureforum.com/data/number-of-restaurants
[4] Paris Open Data: https://opendata.paris.fr/explore/dataset/arrondissements/table/
[5] Mercator projections: https://epsg.io/
[6] https://github.com/CarexNigra/coursera_ibm_data_science/blob/master/ibm_capstone_project/data/04_unique_categories_mapping_to_final_categories_list.csv
[7] https://github.com/CarexNigra/coursera_ibm_data_science/blob/master/ibm_capstone_project/data/07_egg_places_nearby_venues_replaced_categories_frequency_of_occurence.csv
[8] https://github.com/CarexNigra/coursera_ibm_data_science/blob/master/ibm_capstone_project/data/13_venues_nearby_to_grid_points_replaced_categories_frequency_of_occurence.csv