

The Battle of the Neighborhoods

Identification of appropriate location for new Italian restaurant in Paris

May 2, 2020

1. Introduction

1.1. Background

The study area of the project is Paris city, capital and most populous city in France. Paris is one of the main business cities in Europe, and with Brexit there are some possibilities that different headquarters will move from London to Paris. Developed economy and business activities with strong tourism, all of these makes city very attractive for small and medium-sized enterprises.

This also means that market is very competitive. The high cost of living also increases the cost of business and makes it challenging. As a result, any business decision must be analysed carefully. Proper analytical work provides good understanding of the business environment and helps to identify targeting audience. This analysis also helps to reduce risks and find new opportunities for expansion and diversification.

1.2. Business Problem

This project will examine possible locations for setting up Italian restaurant in Paris. Restaurant is a business that prepares and serves food and drinks to customers. Meals are generally served and eaten on the premises, but many restaurants also offer take-out and food delivery services. Paris is famous for its diverse cuisines from different cultures. So it is significant to create accurate strategic plan to survive in such a high competitive market.

One of the main parameters in restaurant launching is location. Once it is launched the most difficult to change is location, so this business plan step should be done carefully. If this step is successful, there will be new opportunities for other ventures in different locations. So, choosing location is very important step.

1.3. Target Audience

- Investors
- Businessmen interested in launching new restaurant
- Businessmen interested in expansion of existing restaurant chain

Target audience is restaurant management team, which hired me as Data Scientist to propose effective location for restaurant venture in Paris. As part of analytical work, data scientist should prepare argumentative report with final recommendations for the management team.

2. Data acquisition and cleaning

2.1. Data Sources

In this project two main data sources will be used.

Data 1. In order to divide the region into smaller clusters, it was decided to use metro stations as centers for neighborhood analysis. The region of Il-de-France has 25,338 metro stations. For every metro station the data for location, including latitude and longitude is required. This information is gained from RATP, official information website for Paris metropolitan users.

- This dataset exists for free on the web. Link to the dataset is :

<https://dataratp2.opendatasoft.com/>

Data 2. Il-de-France (Paris city) geographical coordinates data will be used as input for the Foursquare API, that will be leveraged to provision venues information for every communes. We will use the Foursquare API to explore different business activities with detailed information such as name, category, location, responses and others.

- This dataset exists for free on the web. Link to the dataset is : <https://foursquare.com>

2.2. Data Cleaning

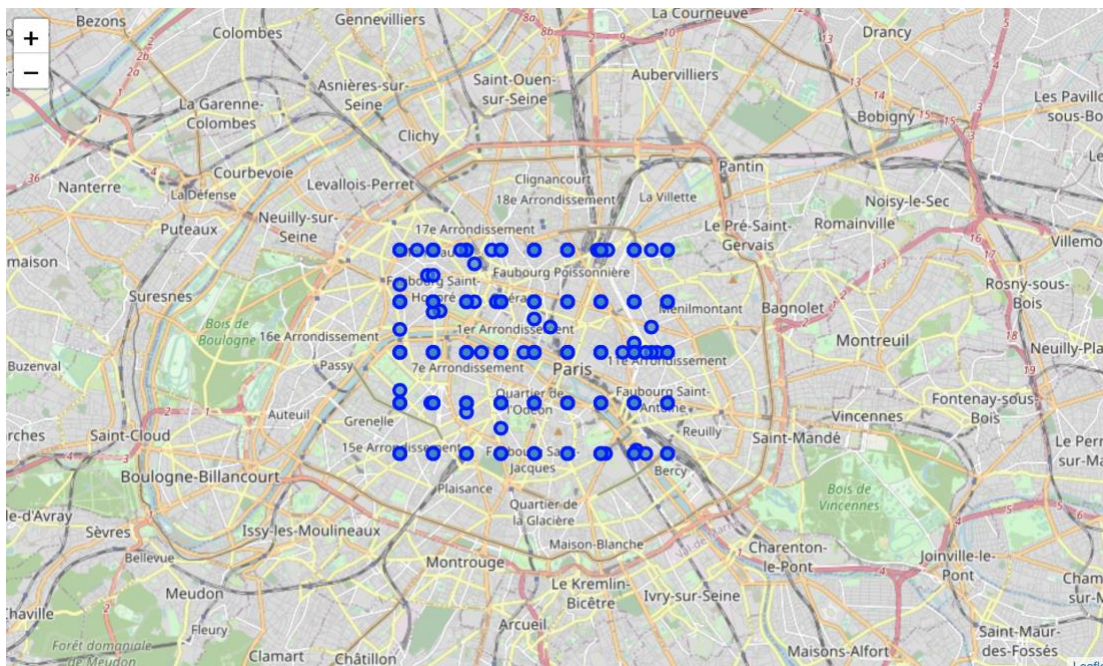
Downloaded dataset for metro station location had some problems, it was not usable for further analysis by pandas. Firstly, coordinates for each station, latitude and longitude, were given in one cell in string. So, these data were split and put in new columns called 'Latitude' and 'Longitude', and values were changed into float.

Secondly, description of the stations included some text and post code of the location. So, this text was removed, and only post code remained. Using these post codes, the dataset was decreased, and only metro stations located in Paris city remained.

After that, stations were grouped by names using mean. This helped to combine stations with similar names but different directions and locate them between each other. This resulted in 1140 stations.

Such size of dataset was too large for further exploration using Foursquare API. So, to decrease the number of stations for future analysis outliers were removed as shown below.

```
df=df[(df['Latitude']<=48.88) & (df['Longitude']<=2.38)]
df=df[(df['Latitude']>=48.84) & (df['Longitude']>=2.30)]
df
```



Now, in order to explore the neighborhood of the stations Foursquare API was utilized using following parameters:

- CLIENT_ID = Foursquare ID
- CLIENT_SECRET = Foursquare Secret
- VERSION = '20200310'
- LIMIT = 50
- Radius = 150
- categoryId = '4d4b7105d754a06374d81259'

Since stations are located close to each other, the radius of neighborhood venues was installed as 150 meters with limit in number of venues 50. Also, to decrease the number of output venues and to target competitive venues 'Food' category was added.

Table 1. Output table from Foursquare API call

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ABBE GROULT	48.84	2.30	Je thé...me	48.838866	2.300375	French Restaurant
1	ABBE GROULT	48.84	2.30	Les Vignes du Liban	48.839402	2.300697	Lebanese Restaurant
2	ABBE GROULT	48.84	2.30	L'Inattendu	48.840485	2.299322	French Restaurant
3	ABBE GROULT	48.84	2.30	Le Cap	48.840937	2.300072	French Restaurant
4	ABBE GROULT	48.84	2.30	Le Quartier du Pain	48.839282	2.300139	Bakery
...
3856	YVONNE LE TAC	48.88	2.34	Le Clou	48.880905	2.340473	French Restaurant
3857	YVONNE LE TAC	48.88	2.34	Sushi Shop	48.878660	2.339924	Sushi Restaurant
3858	YVONNE LE TAC	48.88	2.34	Le Pain Quotidien	48.880030	2.340560	Breakfast Spot
3859	YVONNE LE TAC	48.88	2.34	L'Oriental	48.880744	2.341271	Moroccan Restaurant
3860	YVONNE LE TAC	48.88	2.34	1ndix Café	48.880764	2.341361	Café

One problem with that table is that some venues are shown multiple times for different stations. So, checking 'Venue', 'Venue Latitude' and 'Venue Longitude' columns, duplications were removed.

After that pivot table with columns representing 'Venue Category' and rows representing 'Stations' was constructed. For better understanding of each station this pivot table was grouped by 'Station' using mean.

Table 2. Pivot table ‘Station’ vs ‘Venue Category’ grouped using mean

	Station	African Restaurant	Alsatian Restaurant	American Restaurant	Asian Restaurant	Auvergne Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	...	Souvlaki Shop	Spanish Restaurant	Steakhouse	Rest
0	ABBE GROULT	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.166667	...	0.0	0.0	0.0	0.0
1	ABBESSES	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.000000	...	0.0	0.0	0.0	0.0
2	ALIBERT	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0
3	ALLERAY	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0
4	ALLERAY-LA QUINTINIE	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.166667	...	0.0	0.0	0.0	0.0
...
457	VOLTAIRE - LEON BLUM	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.250000	...	0.0	0.0	0.0	0.0
458	WAGRAM	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0
459	WAGRAM - COURCELLES	0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0
460	WAGRAM - PRONY	0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	...	0.0	0.0	0.0	0.0
461	YVONNE LE TAC	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.000000	...	0.0	0.0	0.0	0.0

From pivot table demonstrated above, for each station only 10 most common venue categorized were chosen using coefficients (values) in that pivot table.

Table 3. Pivot table with Most Common Venues for Stations

	Station	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
0	ABBE GROULT	French Restaurant	Bistro	Lebanese Restaurant	Bakery	Falafel Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Ethiopian Restaurant
1	ABBESSES	French Restaurant	Bakery	Breakfast Spot	Café	Sushi Restaurant	Moroccan Restaurant	Creperie	Eastern European Restaurant	Dim Sum Restaurant	Diner
2	ALIBERT	French Restaurant	Vietnamese Restaurant	Diner	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Eastern European Restaurant	Restaurant	Gastropub	Bagel Shop
3	ALLERAY	French Restaurant	Italian Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Fast Food Restaurant

4. Clustering

On the table above it can be seen, that there are stations that have similar categories. As a result, one of the most popular unsupervised learning K-means algorithm was used to cluster stations. In our case, k-means method was used to divide the stations into 5 clusters. The stations in each cluster are similar to each other in terms of venue types located nearby.

Table 4. Stations with Cluster Labels

	Station	Latitude	Longitude	ClusterLabels	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
0	ABBE GROULT	48.840000	2.300000	0	French Restaurant	Bistro	Lebanese Restaurant	Bakery	Falafel Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant
1	ABBESSES	48.880000	2.340000	0	French Restaurant	Bakery	Breakfast Spot	Café	Sushi Restaurant	Moroccan Restaurant	Creperie	Eastern European Restaurant	Diner
2	ALIBERT	48.870000	2.370000	0	French Restaurant	Vietnamese Restaurant	Diner	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Eastern European Restaurant	Restaurant	Gas
3	ALLERAY	48.840000	2.310000	2	French Restaurant	Italian Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant

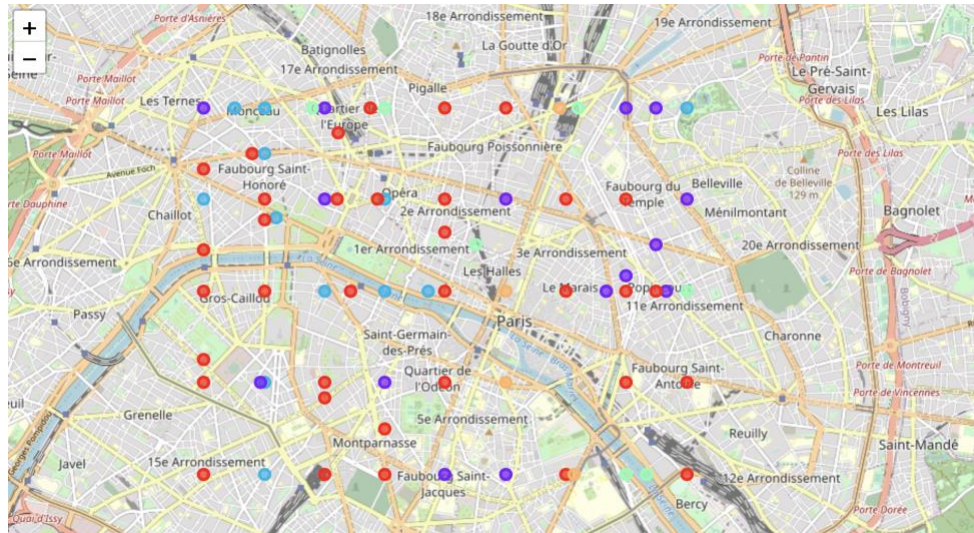


Figure 3. Stations after clustering

Red color – Cluster 0

Violet color – Cluster 1

Blue color – Cluster 2

Green color – Cluster 3

Orange color – Cluster 4

5. Discussion

After gaining clusters, each cluster can be examined deeper and some cluster categorizing venues can be identified.

- Cluster 0: Total number – 35

Table 5. Cluster 0

	Station	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
0	ABBE GROULT	French Restaurant	Bistro	Lebanese Restaurant	Bakery	Falafel Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Ethiopian Restaurant
1	ABBESSES	French Restaurant	Bakery	Breakfast Spot	Café	Sushi Restaurant	Moroccan Restaurant	Creperie	Eastern European Restaurant	Dim Sum Restaurant	Diner
2	ALIBERT	French Restaurant	Vietnamese Restaurant	Diner	Vegetarian / Vegan Restaurant	Italian Restaurant	Japanese Restaurant	Eastern European Restaurant	Restaurant	Gastropub	Bagel Shop
5	ALMA - MARCEAU	French Restaurant	Seafood Restaurant	Brasserie	Café	Sandwich Place	Falafel Restaurant	Mediterranean Restaurant	Vietnamese Restaurant	Dim Sum Restaurant	Diner
6	ALMA-MARCEAU	French Restaurant	Italian Restaurant	Café	Bakery	Indian Restaurant	Asian Restaurant	Chinese Restaurant	Pizza Place	Vietnamese Restaurant	Eastern European Restaurant

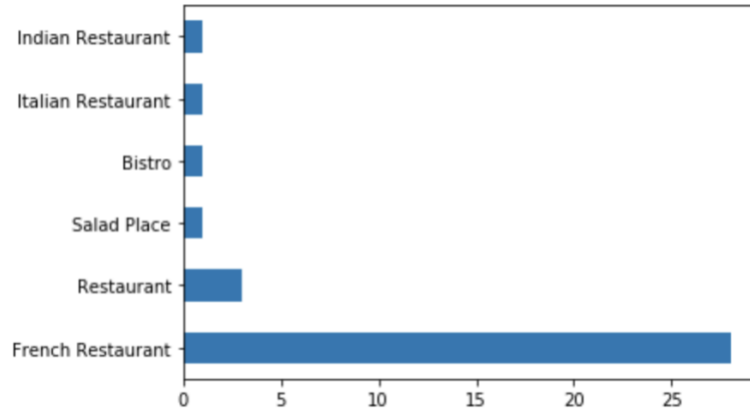


Figure 4. Venue categories count in Cluster 0

As it can be seen from the table and figure above, this cluster is mainly associated with French Restaurant.

- Cluster 1: Total number – 15

Table 6. Cluster 1

	Station	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
7	ANJOU - CHAUVEAU LAGARDE	Café	Sushi Restaurant	Cafeteria	Salad Place	Sandwich Place	Spanish Restaurant	Pizza Place	Asian Restaurant	French Restaurant	Fast Food Restaurant
17	ASSAS - DUGUAY TROUIN	Pizza Place	Café	Sandwich Place	Bakery	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Vietnamese Restaurant
35	BELLEVILLE	Restaurant	BBQ Joint	Diner	Spanish Restaurant	Pizza Place	Fondue Restaurant	Food Truck	Fast Food Restaurant	French Restaurant	Corsican Restaurant
37	BERTHOLLET - VAUQUELIN	Café	Greek Restaurant	French Restaurant	Tapas Restaurant	Japanese Restaurant	Pizza Place	Moroccan Restaurant	Mexican Restaurant	Donut Shop	Deli / Bodega
40	BOLIVAR	Food Truck	Vietnamese Restaurant	Corsican Restaurant	Greek Restaurant	Gastropub	French Restaurant	Fondue Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
41	BONNE NOUVELLE	Vietnamese Restaurant	Bistro	Mediterranean Restaurant	Pizza Place	Poke Place	Burger Joint	Seafood Restaurant	Indian Restaurant	Turkish Restaurant	Bagel Shop

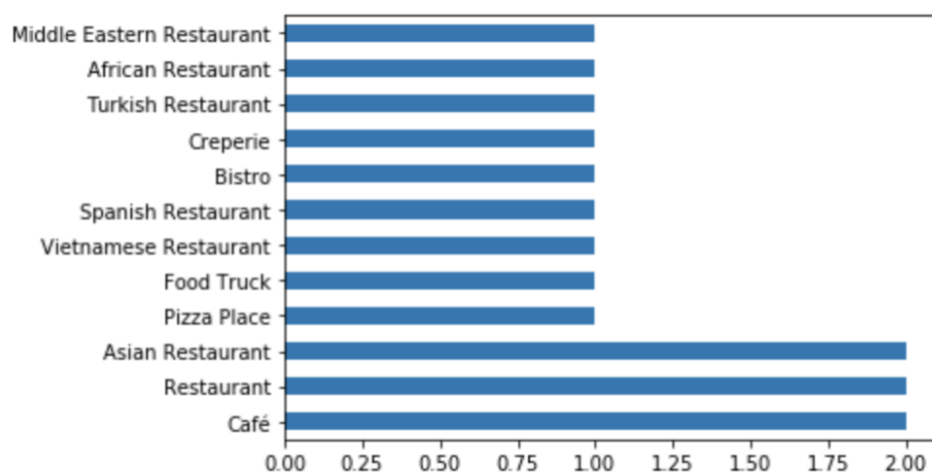


Figure 5. Venue categories count in Cluster 1

In this cluster, there are no specific venue category. Different cuisines are presented in this cluster, however some pattern for eastern and African restaurants can be noticed.

- Cluster 2: Total number – 12

Table 7. Cluster 2

	Station	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
3	ALLERAY	French Restaurant	Italian Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Fast Food Restaurant
12	ARMAND CARREL - MAIRIE DU 19E	French Restaurant	Vietnamese Restaurant	Corsican Restaurant	Greek Restaurant	Gastropub	Food Truck	Fondue Restaurant	Fast Food Restaurant	Falafel Restaurant	Ethiopian Restaurant
18	ASSEMBLEE NATIONALE	French Restaurant	Bistro	Italian Restaurant	Café	Falafel Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Ethiopian Restaurant
20	AUBER	French Restaurant	Restaurant	Breakfast Spot	Auvergne Restaurant	Salad Place	Bagel Shop	Café	Dim Sum Restaurant	Diner	Donut Shop
25	BALZAC	French Restaurant	Japanese Restaurant	Sandwich Place	Mediterranean Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop

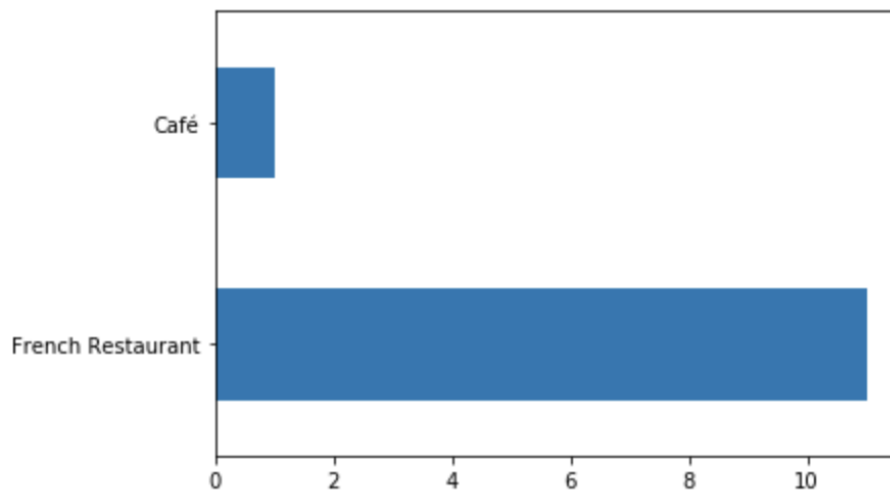


Figure 6. Venue categories count in Cluster 2

Here we can see that French restaurants are common in this cluster as in cluster 0. So, to get better understanding of difference, cluster 2 was deeper examined. Counting 2nd and 3rd most common venues demonstrated that along with French restaurants, Vietnamese restaurants are also popular in this cluster.


```
cluster2[ '2nd Most Common Venue'].value_counts()
```

```
Café 4
Vietnamese Restaurant 2
Restaurant 1
Italian Restaurant 1
Bistro 1
French Restaurant 1
Japanese Restaurant 1
Bakery 1
Name: 2nd Most Common Venue, dtype: int64
```

```
cluster2['3rd Most Common Venue'].value_counts()
```

```
Vietnamese Restaurant 5
Corsican Restaurant 2
Sandwich Place 1
Falafel Restaurant 1
Snack Place 1
Italian Restaurant 1
Breakfast Spot 1
Name: 3rd Most Common Venue, dtype: int64
```

Figure 7. 2nd and 3rd most common venue categories count in Cluster 2

- Cluster 3: Total number – 7

Table 8. Cluster 3

	Station	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
28	BASFROI	Bistro	Italian Restaurant	Café	Ethiopian Restaurant	Falafel Restaurant	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Vietnamese Restaurant
39	BLANCHE	Italian Restaurant	Bistro	Steakhouse	Indian Restaurant	Donut Shop	Pizza Place	Bagel Shop	Creperie	Fondue Restaurant	Fast Food Restaurant
82	CHATEAU LONDON	Italian Restaurant	Restaurant	Snack Place	Japanese Restaurant	Fast Food Restaurant	Burger Joint	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop
95	CITE DE LA MODE ET DU DESIGN	Russian Restaurant	Italian Restaurant	Ethiopian Restaurant	Creperie	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Vietnamese Restaurant
123	ETIENNE MARCEL - MONTMARTRE	Italian Restaurant	Bakery	Ramen Restaurant	Souvlaki Shop	Peruvian Restaurant	Café	French Restaurant	Restaurant	Seafood Restaurant	Noodle House

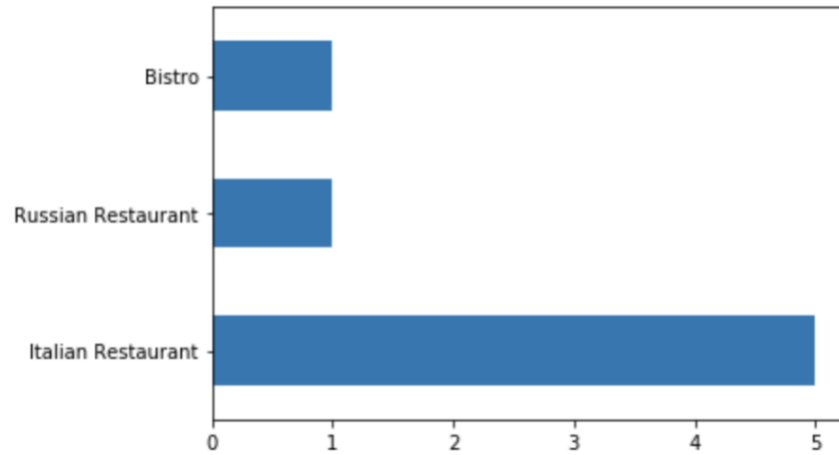


Figure 8. Venue categories count in Cluster 3

Cluster 3 is place where Italian restaurants are densely concentrated. As shown on the table and figure above stations in this cluster, mostly have 1st most common venue – Italian Restaurant. Even only two exceptions have Italian restaurants as 2nd most common venue.

- Cluster 4: Total number - 4

Table 9. Cluster 4

	Station	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
63	CARDINAL LEMOINE	Bakery	Japanese Restaurant	Burger Joint	Seafood Restaurant	Mexican Restaurant	Vietnamese Restaurant	Ethiopian Restaurant	Dim Sum Restaurant	Diner	Donut Shop
70	CENTRE GEORGES POMPIDOU	Bakery	Falafel Restaurant	Deli / Bodega	Dim Sum Restaurant	Diner	Donut Shop	Eastern European Restaurant	Ethiopian Restaurant	Vietnamese Restaurant	Corsican Restaurant
139	GARE D'AUSTERLITZ	Corsican Restaurant	Korean Restaurant	Indian Restaurant	American Restaurant	Deli / Bodega	Greek Restaurant	Gastropub	French Restaurant	Alsatian Restaurant	Food Truck
149	GARE DU NORD	Indian Restaurant	Bakery	Greek Restaurant	Gastropub	French Restaurant	Food Truck	Fondue Restaurant	Fast Food Restaurant	Corsican Restaurant	Falafel Restaurant

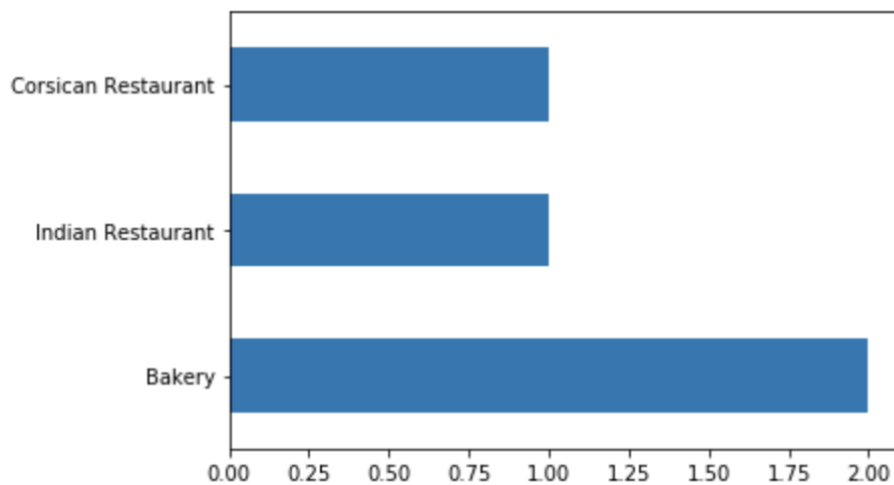


Figure 9. Venue categories count in Cluster 4

6. Conclusion

Based on the analysis information above, we can give some recommendations for possible Italian restaurant set up location. Firstly, cluster 3 station's location should be deeply examined. May be such concentration if Italian restaurants can be explained by Italian district, which is actually a perfect location for our restaurant in case if there are still uncovered demand for such cuisine.

Another suggestion is to choose location far from cluster 3, which is related to Italian restaurant. As it can be seen from the map above, south-west part of the city doesn't have Italian restaurant clusters, so in that part of the city it will be easier to locate Italian restaurant successfully.