Comparison of Paris and Berlin Neighbourhoods

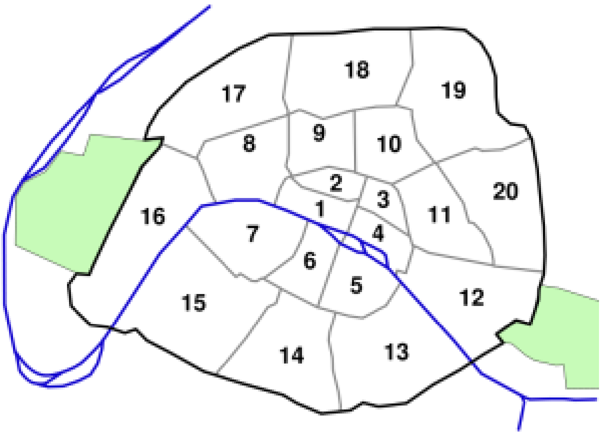
1 Introduction

1.1 Background

One of the important steps in planning travel or relocation to a new city is to choose which area to visit, stay, or live. It would be useful to have insight to the character of the neighbourhoods in order to find the places most suited to your personal preference and situation.

The character of a neighbourhood is defined by its demographic and the kinds of commerce and landmarks it contains. For example, the type and quantity of establishments may reveal whether an area is animated with bars and nightlife as opposed to a residential area occupied predominantly by schools and supermarkets. The type of restaurants in an area may reveal information about its demographic, and other features such as proximity to parks or shopping may of interest.

In addition, a way to make this information more relatable could be to draw parallels between the neighbourhoods of an unfamiliar city, to those of a city with which one is already familiar.



Preliminary research shows that Paris (left) is made up of 20 districts, each containing 4 neighbourhoods to to form a total of 80 neighbourhoods, while Berlin (right) is made up of 12 boroughs containing a total of 96 neighbourhoods. The full list of neighbourhoods and boroughs can be found in Appendix A.

The following figures from August 2017 show that although Paris has around 60% of Berlin's population, it occupies less than 12% its area, making the population density of Paris about 5 fold that of Berlin.

City	Population	Area (km^2)	Density (/km^2)
Paris	2203817	105.4	20909
Berlin	3711930	891.7	4162

Source : https://en.wikipedia.org/wiki/List_of_European_Union_cities_proper_by_population_density

1.2 Objective

This project will cluster the neighbourhoods of Paris, France and Berlin, Germany to try and identify similar neighbourhoods between the two cities, and produce a list of the most abundant establishments which characterise each cluster.

1.3 Hypothesis

As Paris and Berlin are both multicultural European cities, the expectation is that simliar neighbourhoods between the two cities should be readily identifiable. However, if there is little overlap and such parallels cannot be drawn, it would be a valuable conlcusion as well to see how the cities differ.

2 Data Requirements

Requirement	Purpose	Source
A list of neighbourhoods for each city.	The neighbourhoods are the objects to be clustered.	Wikipedia [i]
The top 100 venues for each neighbourhood.	Neighbourhood venues form the feature set for input to K-means clustering algorithm.	Foursquare
Each neighbourhood's geospatial coordinates.	Input to Foursquare venue search.	GeoPy package
Each city's average neighbourhood radius [ii]	Input to Foursquare venue search.	Calculated from the neighbourhoods' geospatial coordinates.

[i] https://en.wikipedia.org/wiki/Quarters_of_Paris
https://en.wikipedia.org/wiki/Boroughs_and_neighborhoods_of_Berlin

[ii] The average neighbourhood radius per city will be used for simplicity rather than the radius of each individual neighbourhood as it is the relative density between the cities that needs to be corrected for.

3 Methodology

3.1 Data Collection

Neighbourhood Radius

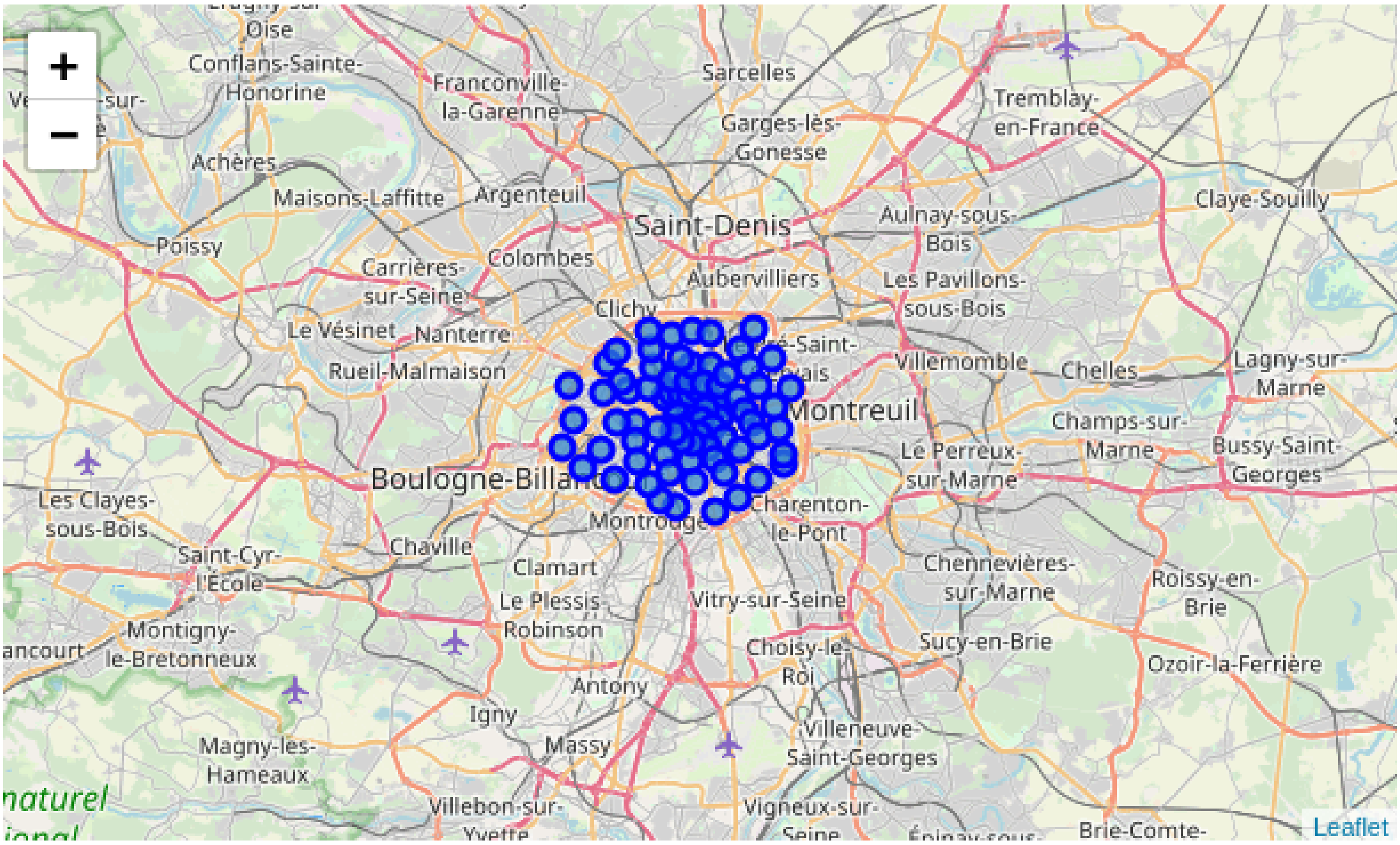
The following maps show the 80 neighbourhoods of Paris, and 96 neighbourhoods of Berlin plotted on the same scale. It is evident that Paris neighbourhoods are much smaller compared to Berlin neighbourhoods.

In order to account for this difference in size, KD Tree algorithm was used to find the distance of each neighbourhood to its closest neighbour. The average distance was then calculated for each city and divided by 2 to find the average neighbourhood radius.

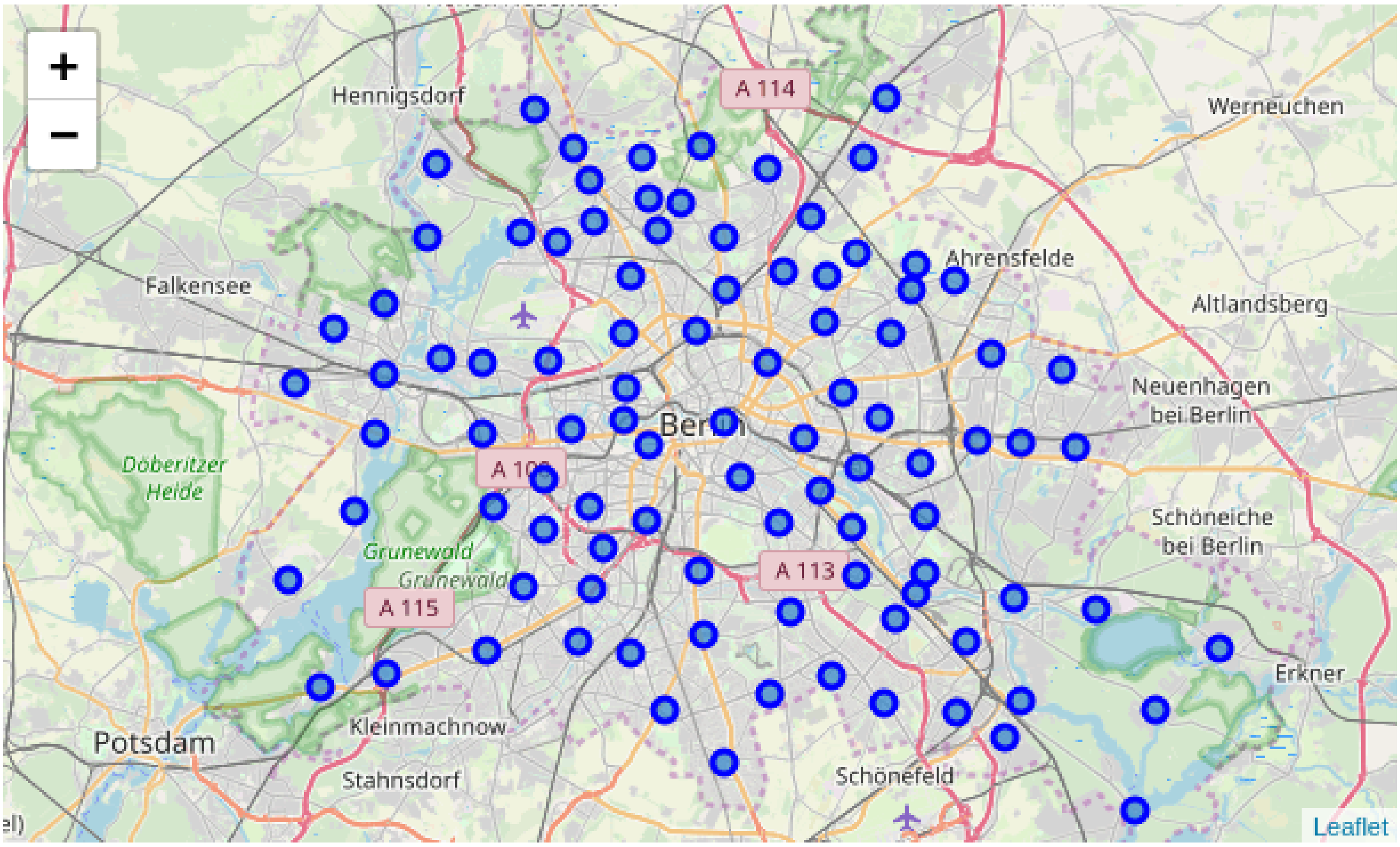
The top 100 venues were requested from Foursquare

- for each Paris neighbourhood within a radius of 521 metres
- for each Berlin neighbourhood within a radius of 1573 metres

The coordinates of Paris, France are 48.8566969, 2.3514616.



The coordinates of Berlin, Deutschland are 52.5170365, 13.3888599.



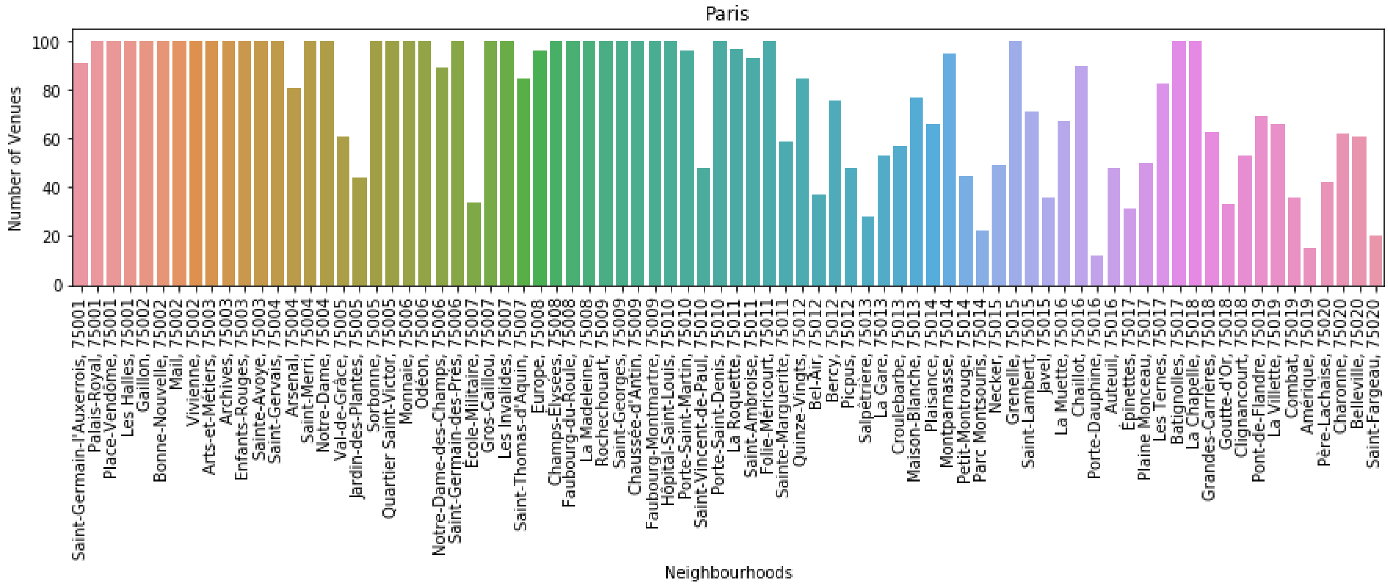
3.2 Exploratory Data Analysis

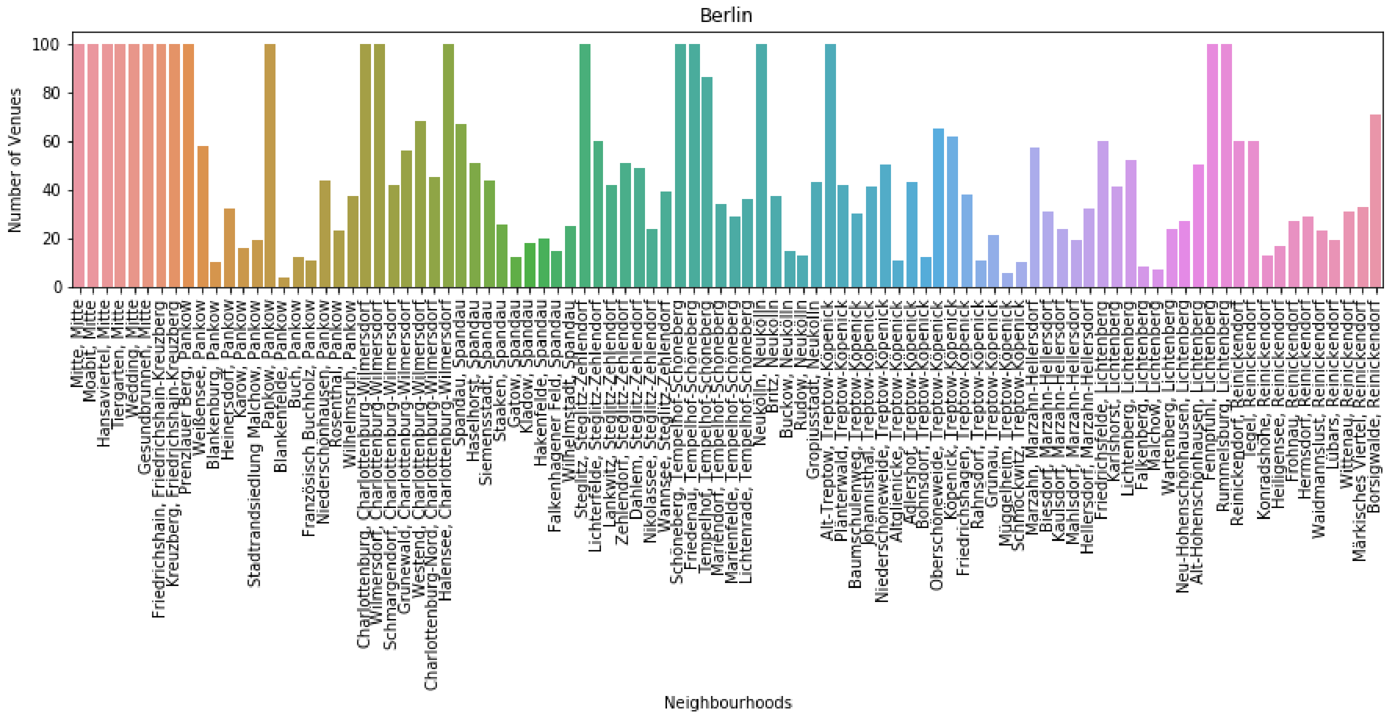
Distribution of venues

Foursquare venue search results :

	Paris	Berlin
Total venues	6120	4570
Average venues per neighbourhood	77	48

Number of venues per neighbourhoods :



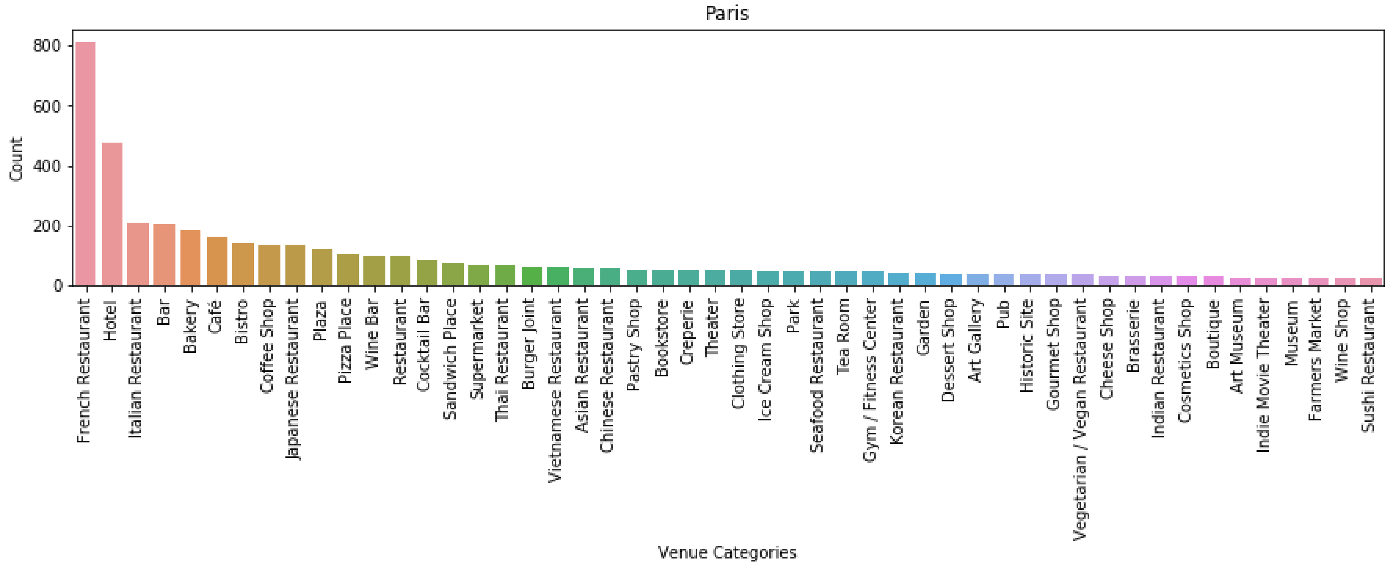


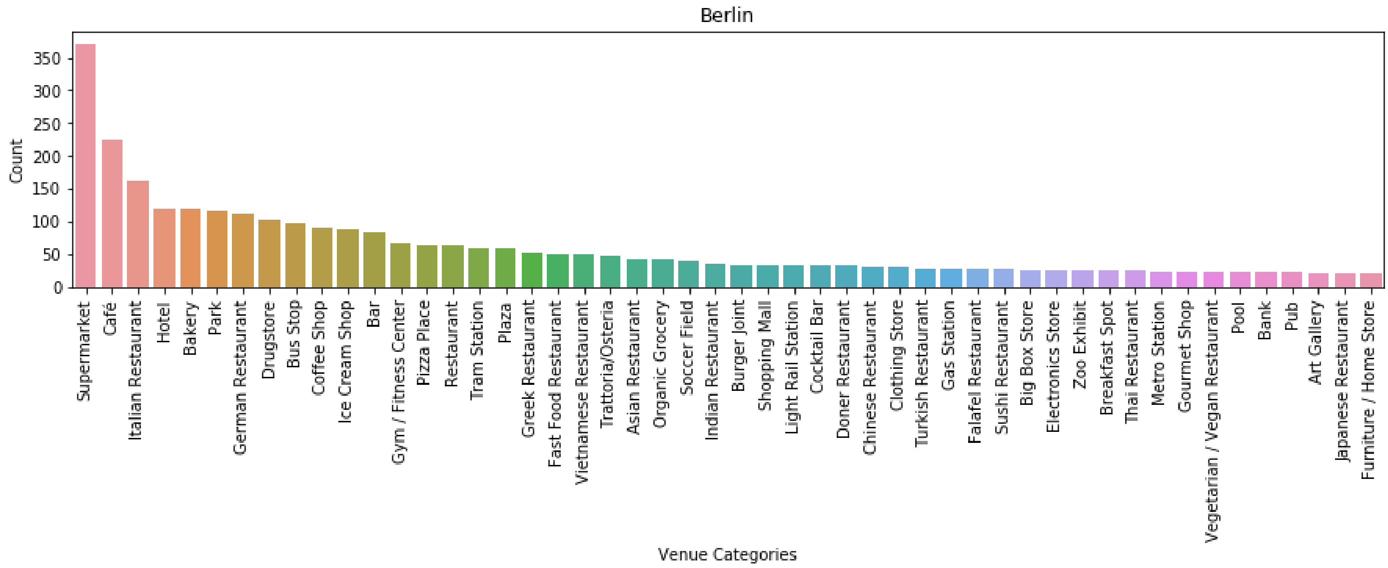
On average, each neighbourhoods in Paris contains more venues than in Berlin.

In addition, 43% of Paris neighbourhoods returned the maximum number of 100 venues, compared to only 21% for Berlin.

Venue Categories

Some cleaning is required to make the venue categories fit for purpose.





Problems with the category labels include :

- Too specific, e.g. 'French Restaurant', 'Bistro', 'Creperie', 'Brasserie' are all types of French restaurants; 'Sushi Restaurant' could be grouped with other types of Japanese restaurants; it is not necessary to differentiate between 'Museum', 'Art Museum' and 'Art Gallery'. Such categories will be generalised.
- Too general, e.g. 'Tree', 'Building', 'Platform'. Venues in such categories will be dropped.
- Inconsistent naming – the same type of venue under different labels, e.g. 'Tram Station' and 'Light Rail Station', 'Café' and 'Coffee Shop'. Such category labels will be aggregated.
- Location specific labels – 'French Restaurant' in Paris and 'German Restsaurant' in Berlin will cause unnecassary segregation when clustering when both types of restaurants are the local cuisine. These will be classified to simply 'Restaurant'.
- Misleading bias whether due to the presence of unuseful data, or deficiency in one of the cities will be removed, e.g. Berlin has categories 'ATM', 'Hardware Store' which also exist in Paris in reality but are not present in the data. Venues in such categories will be dropped.
- Lastly, any venues categories with too few occurrences will be dropped. The minimum threshold for inclusion of a category is set at 20 total occurrences.

Data Cleaning

Number of venue categories :

	Total	Shared	Paris only	Berlin only
Before cleaning	422	231	76	115
After cleaning	32	31	0	1

The full lists of changes is detailed in Appendix B.

Final venue categories :

Shared categories
['American Restaurant', 'Attraction', 'Auto', 'Bar / Pub', 'Beauty', 'Boat / Ferry', 'Café', 'Convenience Store', 'East Asian Restaurant', 'European Restaurant', 'Fast Food', 'Fitness / Sports', 'Hotel / Tourism', 'Italian Restaurant', 'Liquor Store', 'M.E. / African Restaurant', 'Market', 'Museum / Exhibit', 'Nightclub', 'Park', 'Performing Arts', 'Plaza', 'Restaurant', 'Service', 'Shopping', 'South Asian Restaurant', 'Specialty Food', 'Stadium', 'Supermarket', 'Urban Entertainment', 'Waterfront']

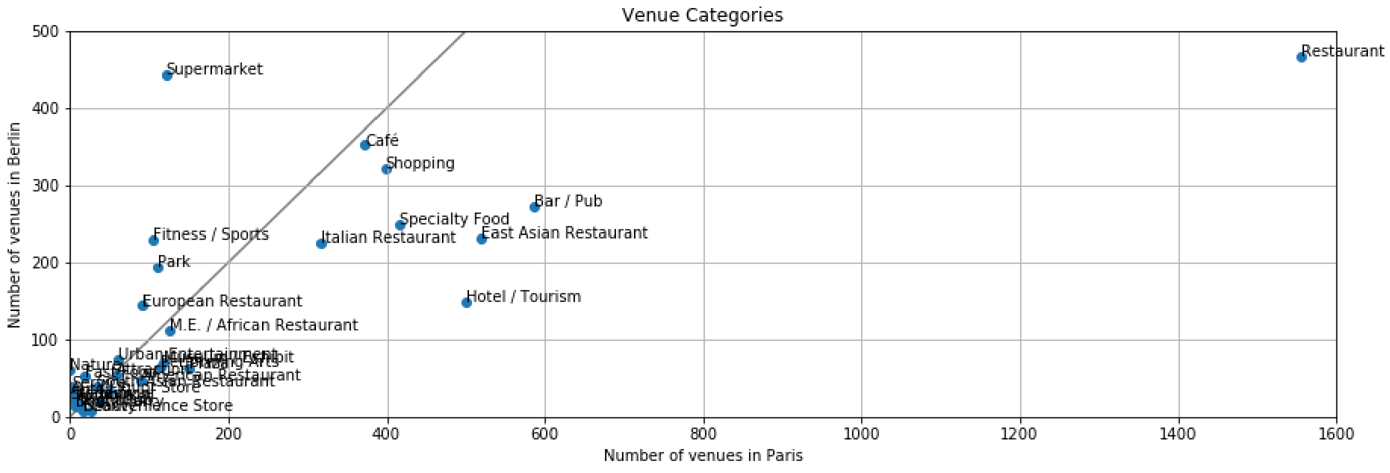
Paris only categories
[]

Berlin only categories
['Nature']

Data Analysis

Now that the categories have been cleaned, it is possible to compare venues for the two cities by plotting the occurrences of each category in one city against the other.

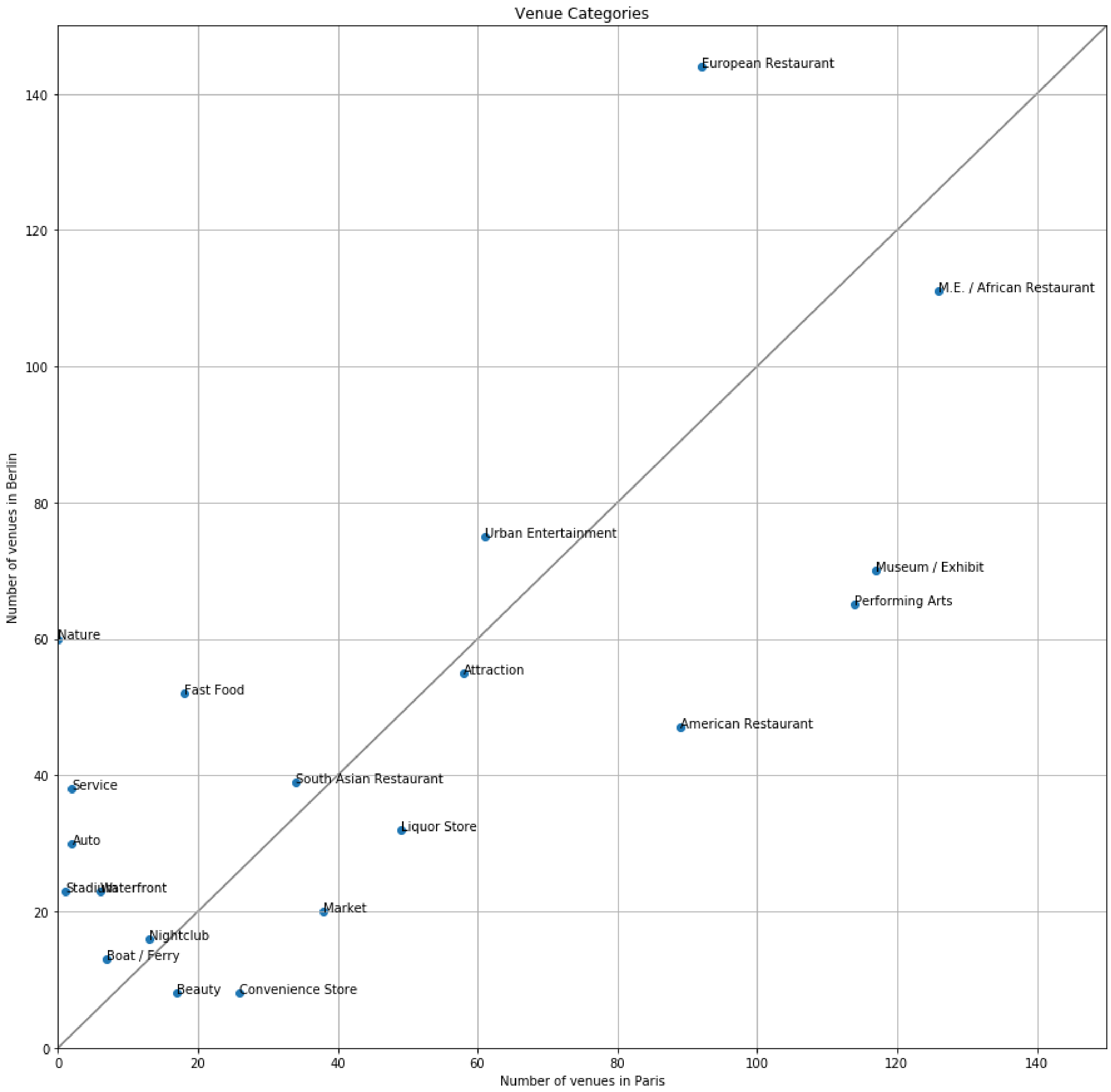
Full view :



Notable observations :

- The number of restaurants in Paris far surpasses any other type of venue, whereas in Berlin the number of restaurants and supermarkets are on par.
- There is an abundance and about equal number of cafés in both cities.
- Fitness and sport facilities are more readily available in Berlin, whereas Paris is dominated by restaurants, bars and shopping for recreation.
- Paris is more geared for tourism.
- Berlin has significantly more parks and nature spots than Paris.

Zoomed-in view :



3.3 Feature Set

To build the feature set, begin by onehot coding the venue list so that each venue is mapped to the category to which it belongs. Next, the venues are grouped by neighbourhood and the means of venues in the same category are calculated for each neighbourhood. This forms a *neighbourhood profile* which corresponds to one row in the feature set.

The resulting feature set is a 2-dimensional array where

- each row index is a neighbourhood,
- each column header is a venue category, and
- each cell value is the proportion that the venue category contributes to the neighbourhood.

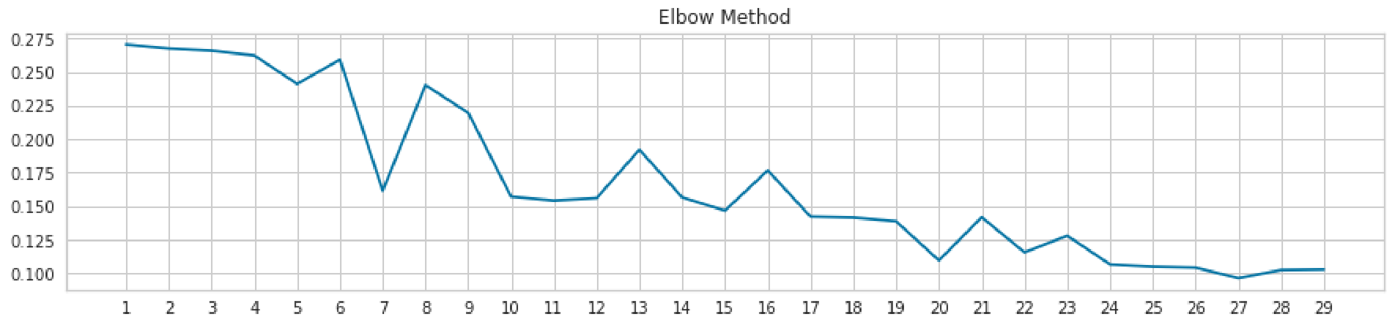
3.4 Modeling

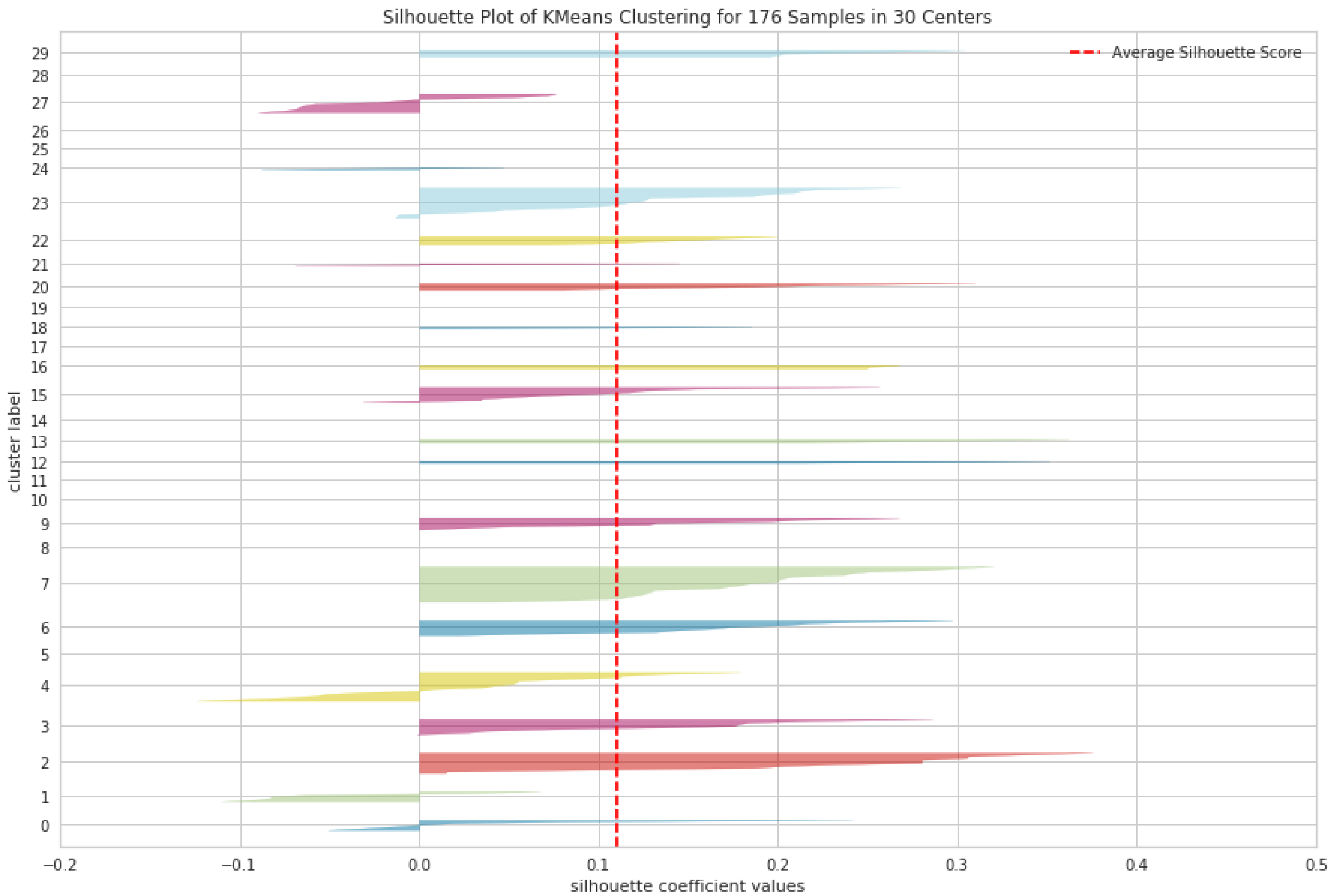
K-means clustering algorithm will be run on this feature set to group neighbourhoods with similar profiles.

Determining Optimal K

The Elbow method and the Silhouette Scores are used in complement to determine the optimal number of clusters for running K-means.

K=7 appears to be a reasonable choice given the results below. It indicates the potential start of the elbow 'bend' (albeit not a very clear one) and its silhouette coefficient is one of the higher ones.



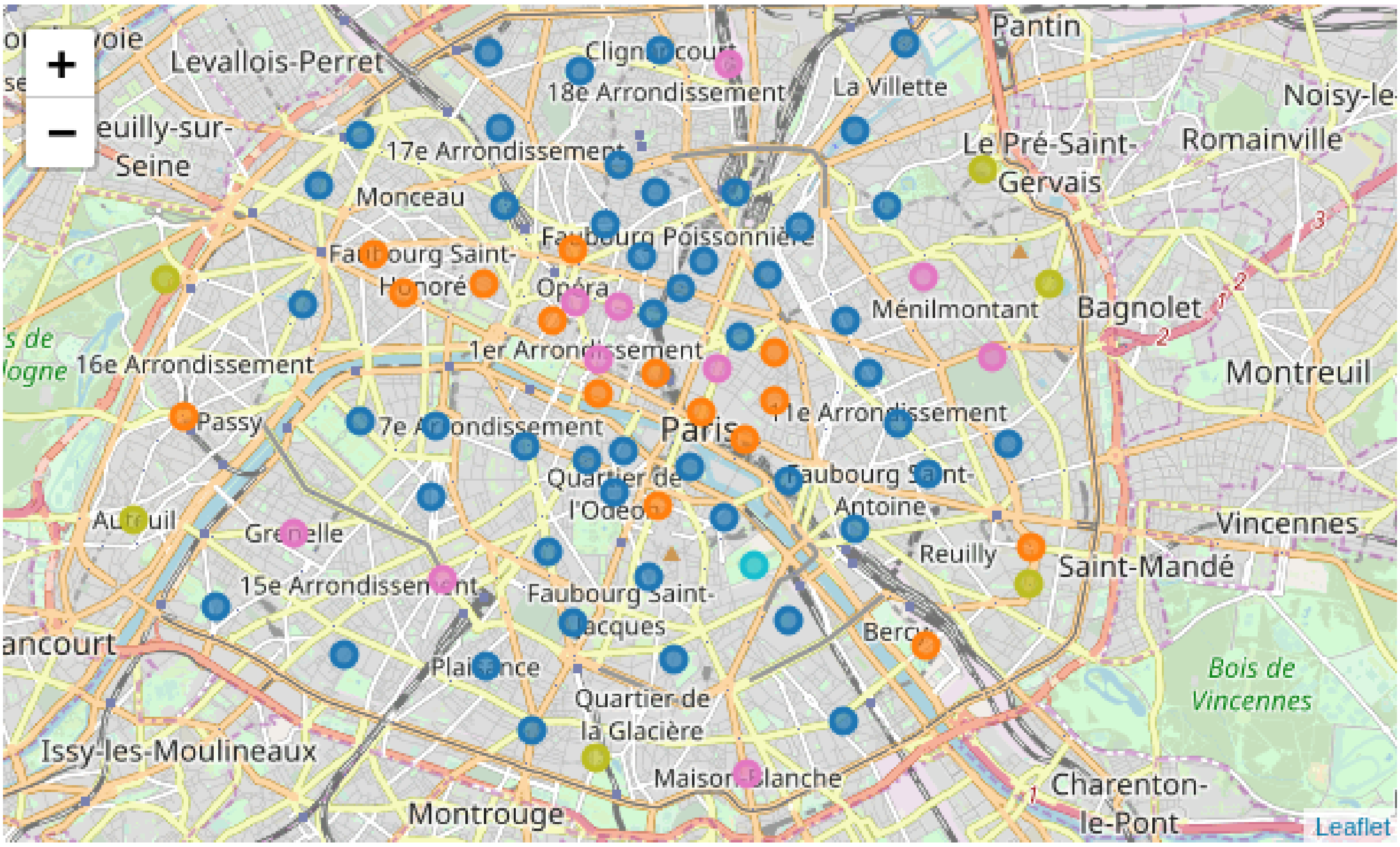


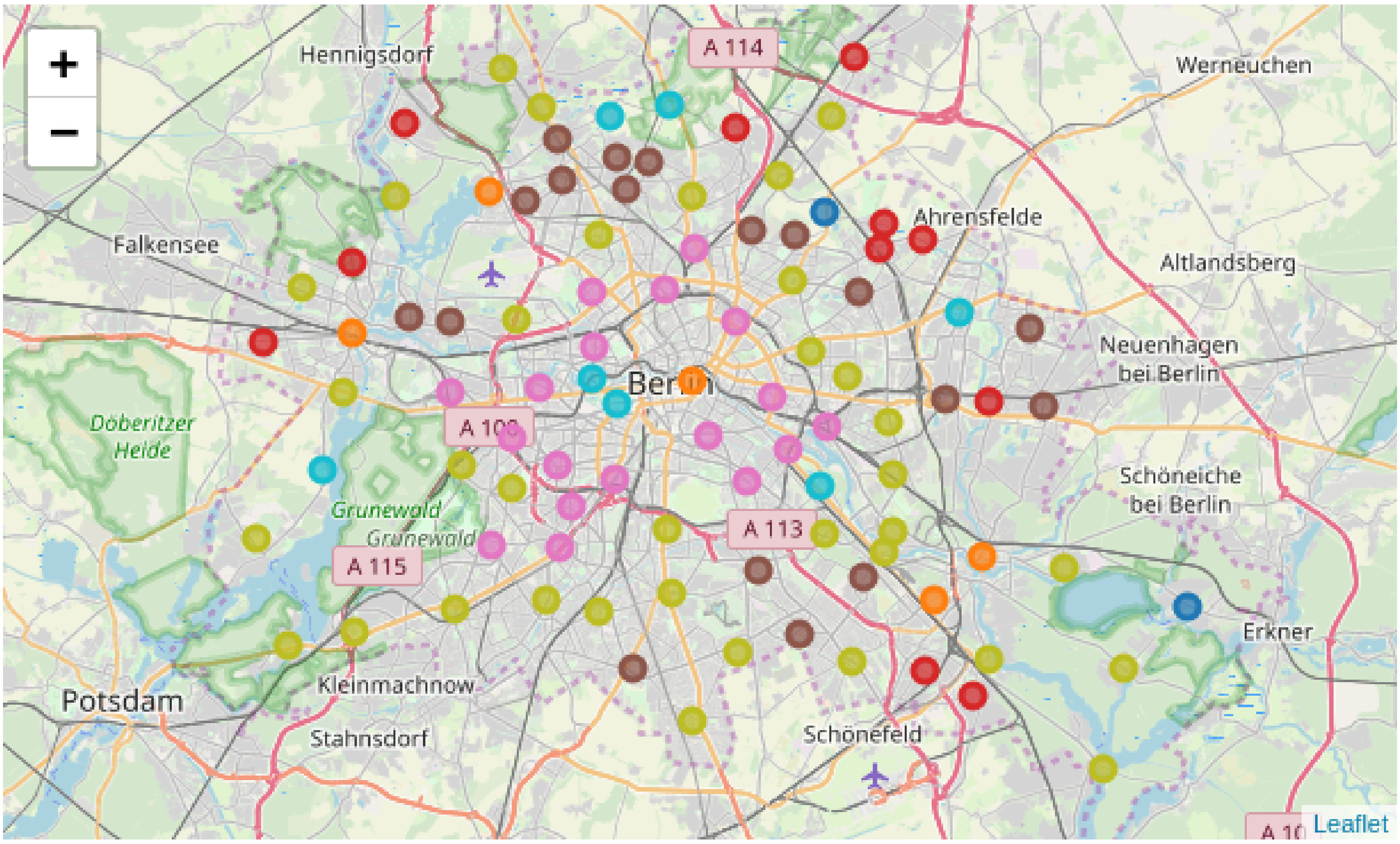
4 Results

From a total of 7 clusters,

- 5 clusters contain neighbourhoods from both cities [Clusters 0, 1, 4, 5, 6]
- 2 clusters contain only Berlin neighbourhoods [Clusters 2, 3]

The colour of the clusters correspond to the bar graphs below.

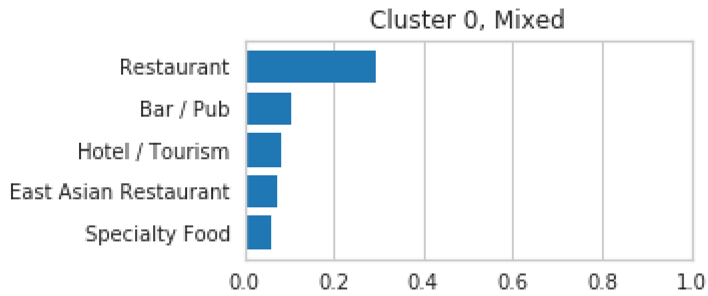




Cluster Details

---Cluster 0---

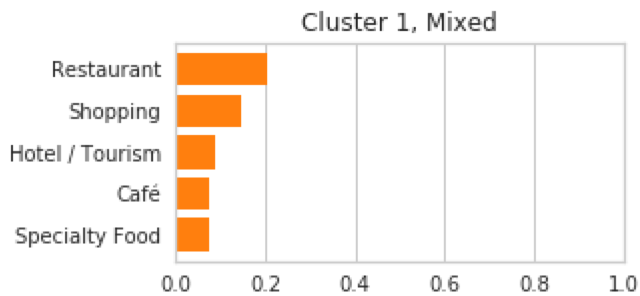
6	Mail, Paris
7	Bonne-Nouvelle, Paris
8	Arts-et-Métiers, Paris
14	Arsenal, Paris
15	Notre-Dame, Paris
16	Quartier Saint-Victor, Paris
18	Val-de-Grâce, Paris
20	Monnaie, Paris
21	Odéon, Paris
22	Notre-Dame-des-Champs, Paris
23	Saint-Germain-des-Prés, Paris
24	Saint-Thomas-d'Aquin, Paris
25	Les Invalides, Paris
26	École-Militaire, Paris
27	Gros-Caillou, Paris
31	Europe, Paris
32	Saint-Georges, Paris
34	Faubourg-Montmartre, Paris
35	Rochechouart, Paris
36	Saint-Vincent-de-Paul, Paris
37	Porte-Saint-Denis, Paris
38	Porte-Saint-Martin, Paris
39	Hôpital-Saint-Louis, Paris
40	Folie-Méricourt, Paris
41	Saint-Ambroise, Paris
42	La Roquette, Paris
43	Sainte-Marguerite, Paris
47	Quinze-Vingts, Paris
48	Salpêtrière, Paris
49	La Gare, Paris
51	Croulebarbe, Paris
52	Montparnasse, Paris
54	Petit-Montrouge, Paris
55	Plaisance, Paris
56	Saint-Lambert, Paris
59	Javel, Paris
63	Chaillot, Paris
64	Les Ternes, Paris
65	Plaine Monceau, Paris
66	Batignolles, Paris
67	Épinettes, Paris
68	Grandes-Carrières, Paris
69	Clignancourt, Paris
71	La Chapelle, Paris
72	La Villette, Paris
73	Pont-de-Flandre, Paris
75	Combat, Paris
79	Charonne, Paris
146	Rahnsdorf, Berlin
159	Malchow, Berlin



Neighbourhoods : Total 50, Paris 48, Berlin 2

---Cluster 1---

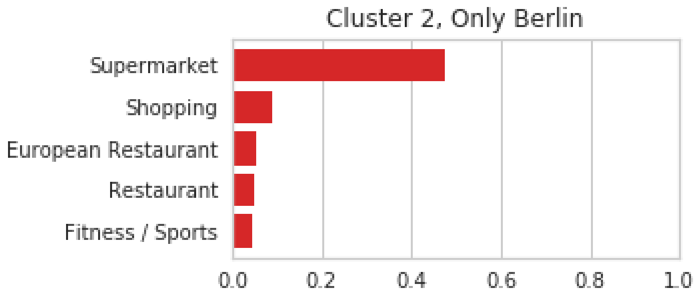
0	Saint-Germain-l'Auxerrois, Paris
1	Les Halles, Paris
3	Place-Vendôme, Paris
9	Enfants-Rouges, Paris
10	Archives, Paris
12	Saint-Merri, Paris
13	Saint-Gervais, Paris
19	Sorbonne, Paris
28	Champs-Élysées, Paris
29	Faubourg-du-Roule, Paris
30	La Madeleine, Paris



33 Chaussée-d'Antin, Paris
45 Picpus, Paris
46 Bercy, Paris
61 La Muette, Paris
80 Mitte, Berlin
108 Spandau, Berlin
141 Adlershof, Berlin
144 Köpenick, Berlin
166 Tegel, Berlin

Neighbourhoods : Total 20, Paris 15, Berlin 5

---Cluster 2---
96 Buch, Berlin
97 Französisch Buchholz, Berlin
111 Staaken, Berlin
114 Hakenfelde, Berlin
140 Altglienicke, Berlin
142 Bohnsdorf, Berlin
152 Kaulsdorf, Berlin
158 Falkenberg, Berlin
160 Wartenberg, Berlin
161 Neu-Hohenschönhausen, Berlin
168 Heiligensee, Berlin

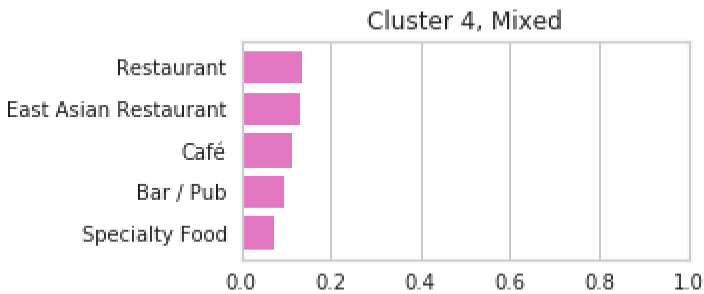


Neighbourhoods : Total 11, Paris 0, Berlin 11

---Cluster 3---
91 Heinersdorf, Berlin
93 Stadtrandsiedlung Malchow, Berlin
99 Rosenthal, Berlin
100 Wilhelmsruh, Berlin
109 Haselhorst, Berlin
110 Siemensstadt, Berlin
128 Marienfelde, Berlin
131 Britz, Berlin
134 Gropiusstadt, Berlin
138 Johannisthal, Berlin
151 Biesdorf, Berlin
153 Mahlsdorf, Berlin
154 Hellersdorf, Berlin
162 Alt-Hohenschönhausen, Berlin
171 Waidmannslust, Berlin
173 Wittenau, Berlin
174 Märkisches Viertel, Berlin
175 Borsigwalde, Berlin

Neighbourhoods : Total 18, Paris 0, Berlin 18

---Cluster 4---
2 Palais-Royal, Paris
4 Gaillon, Paris
5 Vivienne, Paris
11 Sainte-Avoye, Paris
50 Maison-Blanche, Paris
57 Necker, Paris
58 Grenelle, Paris
70 Goutte-d'Or, Paris
76 Belleville, Paris
78 Père-Lachaise, Paris
81 Moabit, Berlin
84 Wedding, Berlin
85 Gesundbrunnen, Berlin
86 Friedrichshain, Berlin
87 Kreuzberg, Berlin
88 Prenzlauer Berg, Berlin
94 Pankow, Berlin
101 Charlottenburg, Berlin
102 Wilmersdorf, Berlin
105 Westend, Berlin
107 Halensee, Berlin
117 Steglitz, Berlin



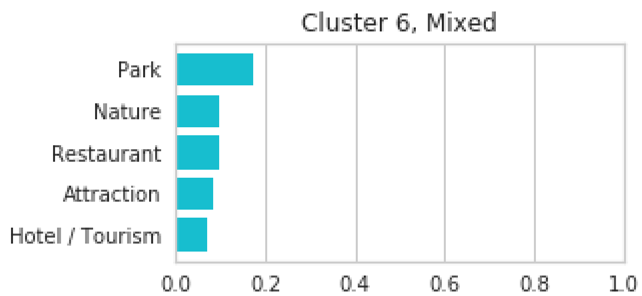
Neighbourhoods : Total 28, Paris 10, Berlin 18

121 Dahlem, Berlin
124 Schöneberg, Berlin
125 Friedenau, Berlin
130 Neukölln, Berlin
135 Alt-Treptow, Berlin
164 Rummelsburg, Berlin

---Cluster 5---
44 Bel-Air, Paris
53 Parc Montsouris, Paris
60 Auteuil, Paris
62 Porte-Dauphine, Paris
74 Amérique, Paris
77 Saint-Fargeau, Paris
89 Weißensee, Berlin
90 Blankenburg, Berlin
92 Karow, Berlin
98 Niederschönhausen, Berlin
103 Schmargendorf, Berlin
104 Grunewald, Berlin
106 Charlottenburg-Nord, Berlin
113 Kladow, Berlin
115 Falkenhagener Feld, Berlin
116 Wilhelmstadt, Berlin
118 Lichterfelde, Berlin
119 Lankwitz, Berlin
120 Zehlendorf, Berlin
122 Nikolassee, Berlin
123 Wannsee, Berlin
126 Tempelhof, Berlin
127 Mariendorf, Berlin
129 Lichtenrade, Berlin
132 Buckow, Berlin
133 Rudow, Berlin
137 Baumschulenweg, Berlin
139 Niederschöneweide, Berlin
143 Oberschöneweide, Berlin
145 Friedrichshagen, Berlin
147 Grünau, Berlin
148 Müggelheim, Berlin
149 Schmöckwitz, Berlin
155 Friedrichsfelde, Berlin
156 Karlshorst, Berlin
157 Lichtenberg, Berlin
163 Fennpfuhl, Berlin
165 Reinickendorf, Berlin
167 Konradshöhe, Berlin
169 Frohnau, Berlin
170 Hermsdorf, Berlin

Neighbourhoods : Total 41, Paris 6, Berlin 35

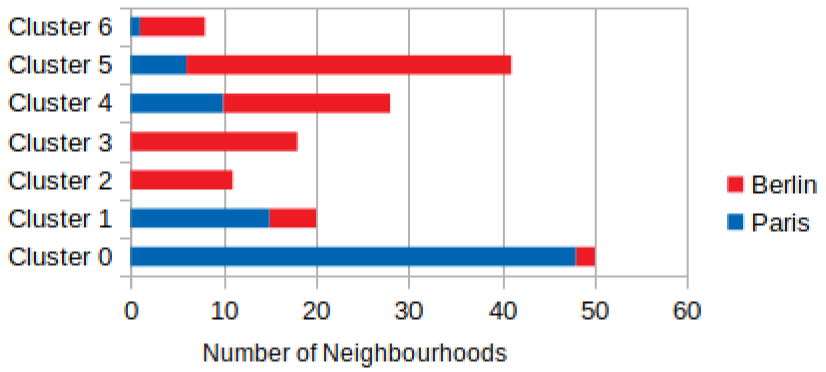
---Cluster 6---
17 Jardin-des-Plantes, Paris
82 Hansaviertel, Berlin
83 Tiergarten, Berlin
95 Blankenfelde, Berlin
112 Gatow, Berlin
136 Plänterwald, Berlin
150 Marzahn, Berlin
172 Lübars, Berlin



Neighbourhoods : Total 8, Paris 1, Berlin 7

Neighbourhood Distribution

	Paris	Berlin	Total
Cluster 6	1	7	8
Cluster 5	6	35	41
Cluster 4	10	18	28
Cluster 3	0	18	18
Cluster 2	0	11	11
Cluster 1	15	5	20
Cluster 0	48	2	50



5 Discussion

Paris is dominated by cluster 0 throughout and at its centre a mix of 1 and 4.

Berlin can be described as the clusters 4, 5, 3, 2 forming loosely concentric circles from inside out, spotted with cluster 6 for parkland.

- Cluster 0 which is spread all across Paris is characterised restaurants and bars / pubs. This is the most ‘Parisien’ cluster as it covers most areas of Paris and only includes 2 Berlin neighbourhoods – Rahnsdort and Malchow.
- Cluster 1 which covers central Paris along the River Seine and the Champs Elysées is primarily shopping and cafés in addition to restaurants. Mitte of central Berlin falls into this cluster along with the centres of other animated Berlin boroughs.
- Hotels and tourism services are abundant all over Paris as they appear in both Clusters 0 and 1.
- Specialty food stores (bakery, patisserie, cheese shops, butchers) are also everywhere in Paris, whereas Berlin relies more on supermarkets.
- Cluster 2 sits towards the outskirts of Berlin and appears to be mainly residential with almost half of the venues being supermarkets.
- Cluster 3 is also on the periphery of Berlin, a little closer to the centre than cluster 2 and with more restaurants and shopping.
- Cluster 4 has many restaurants, cafés and bars but not as abundant as for clusters 0 and 1. Most neighbourhoods from the inner-most circle of Berlin fall into this cluster. It is further differentiated from clusters 0 and 1 by its relatively high number of East Asian restaurants (mainly Chinese and Japanese).
- Cluster 5 has a fairly similar profile to cluster 3 in terms of supermarkets, restaurants and sport facilities, but with cafés instead of shopping. This cluster makes up a good part of Berlin and includes a handful of pheripheric neighbourhoods of Paris.
- Clusters 2, 3 and 5 all of which are made up either exclusively or predominantly of Berlin neighbourhoods, have a notable number of sport facilities.
- Cluster 6 is the most green of the clusters, representing the neighbourhoods abundant in parks and nature.

6 Conclusion

City Features

The data paints an overview of the cities characterised by the following features :

Paris

- dense with restaurants, bars and tourism in all parts of the city
- specialty stores for grocery shopping
- urban forms of recreation such as shopping, cafés and dining out

Berlin

- dense in central boroughs Mitte and Friedrichshain-Kreuzberg
- supermarkets for groceries
- accessibility to nature, parks and sporting facilities for recreation

Clusters Summary

Cluster	Description	Paris	Berlin	**Defining features** (top contributors to 40% of venues)
0	Parisian	96%	4%	Restaurants, bars
1	City centre	75%	25%	Restaurants, shopping, tourism
2	Berlin residential outskirts	0%	100%	Supermarkets Presence of (non-German/French) European cuisines.
3	Berlin residential with commerce	0%	100%	Shopping, supermarkets, sporting facilities
4	Midtown	36%	64%	Restaurants, East-Asian cuisine, cafés
5	Residential	15%	85%	Supermarkets, restaurants, Italian cuisine
6	Green spaces	13%	88%	Parks, nature, restaurants, attractions

Drawbacks

The data obtained from Foursquare was limited to 100 results (even when the limit was set to higher). In addition, the search results were extremely restaurant oriented especially for Paris which may have obscured other types of venues from being retrieved.

Future Studies

It could be interesting to filter for particular types of venues, such as daycares, schools, universities or places of worship to answer more specific questions regarding the demographic of neighbourhoods; or transport facilities and landscape to assess the adequacy or infrastructure.

7 Appendix

A) List of neighbourhoods

	neighbourhood	postcode	city
0	Saint-Germain-l'Auxerrois	75001	Paris
1	Les Halles	75001	Paris
2	Palais-Royal	75001	Paris
3	Place-Vendôme	75001	Paris
4	Gaillon	75002	Paris
5	Vivienne	75002	Paris
6	Mail	75002	Paris
7	Bonne-Nouvelle	75002	Paris
8	Arts-et-Métiers	75003	Paris
9	Enfants-Rouges	75003	Paris
10	Archives	75003	Paris
11	Sainte-Avoye	75003	Paris
12	Saint-Merri	75004	Paris
13	Saint-Gervais	75004	Paris
14	Arsenal	75004	Paris
15	Notre-Dame	75004	Paris
16	Quartier Saint-Victor	75005	Paris
17	Jardin-des-Plantes	75005	Paris
18	Val-de-Grâce	75005	Paris
19	Sorbonne	75005	Paris
20	Monnaie	75006	Paris
21	Odéon	75006	Paris
22	Notre-Dame-des-Champs	75006	Paris
23	Saint-Germain-des-Prés	75006	Paris
24	Saint-Thomas-d'Aquin	75007	Paris
25	Les Invalides	75007	Paris
26	École-Militaire	75007	Paris
27	Gros-Caillou	75007	Paris
28	Champs-Élysées	75008	Paris
29	Faubourg-du-Roule	75008	Paris
30	La Madeleine	75008	Paris
31	Europe	75008	Paris
32	Saint-Georges	75009	Paris
33	Chaussée-d'Antin	75009	Paris
34	Faubourg-Montmartre	75009	Paris
35	Rochechouart	75009	Paris
36	Saint-Vincent-de-Paul	75010	Paris
37	Porte-Saint-Denis	75010	Paris
38	Porte-Saint-Martin	75010	Paris
39	Hôpital-Saint-Louis	75010	Paris
40	Folie-Méricourt	75011	Paris
41	Saint-Ambroise	75011	Paris
42	La Roquette	75011	Paris
43	Sainte-Marguerite	75011	Paris
44	Bel-Air	75012	Paris
45	Picpus	75012	Paris
46	Bercy	75012	Paris

	neighbourhood	postcode	city
47	Quinze-Vingts	75012	Paris
48	Salpêtrière	75013	Paris
49	La Gare	75013	Paris
50	Maison-Blanche	75013	Paris
51	Croulebarbe	75013	Paris
52	Montparnasse	75014	Paris
53	Parc Montsouris	75014	Paris
54	Petit-Montrouge	75014	Paris
55	Plaisance	75014	Paris
56	Saint-Lambert	75015	Paris
57	Necker	75015	Paris
58	Grenelle	75015	Paris
59	Javel	75015	Paris
60	Auteuil	75016	Paris
61	La Muette	75016	Paris
62	Porte-Dauphine	75016	Paris
63	Chaillot	75016	Paris
64	Les Ternes	75017	Paris
65	Plaine Monceau	75017	Paris
66	Batignolles	75017	Paris
67	Épinettes	75017	Paris
68	Grandes-Carrières	75018	Paris
69	Clignancourt	75018	Paris
70	Goutte-d'Or	75018	Paris
71	La Chapelle	75018	Paris
72	La Villette	75019	Paris
73	Pont-de-Flandre	75019	Paris
74	Amérique	75019	Paris
75	Combat	75019	Paris
76	Belleville	75020	Paris
77	Saint-Fargeau	75020	Paris
78	Père-Lachaise	75020	Paris
79	Charonne	75020	Paris

	neighbourhood_id	neighbourhood	borough	city
0	0101	Mitte	Mitte	Berlin
1	0102	Moabit	Mitte	Berlin
2	0103	Hansaviertel	Mitte	Berlin
3	0104	Tiergarten	Mitte	Berlin
4	0105	Wedding	Mitte	Berlin
5	0106	Gesundbrunnen	Mitte	Berlin
6	0201	Friedrichshain	Friedrichshain-Kreuzberg	Berlin
7	0202	Kreuzberg	Friedrichshain-Kreuzberg	Berlin
8	0301	Prenzlauer Berg	Pankow	Berlin
9	0302	Weißensee	Pankow	Berlin
10	0303	Blankenburg	Pankow	Berlin
11	0304	Heinersdorf	Pankow	Berlin
12	0305	Karow	Pankow	Berlin
13	0306	Stadtrandsiedlung Malchow	Pankow	Berlin
14	0307	Pankow	Pankow	Berlin
15	0308	Blankenfelde	Pankow	Berlin

	neighbourhood_id	neighbourhood	borough	city
16	0309	Buch	Pankow	Berlin
17	0310	Französisch Buchholz	Pankow	Berlin
18	0311	Niederschönhausen	Pankow	Berlin
19	0312	Rosenthal	Pankow	Berlin
20	0313	Wilhelmsruh	Pankow	Berlin
21	0401	Charlottenburg	Charlottenburg-Wilmersdorf	Berlin
22	0402	Wilmersdorf	Charlottenburg-Wilmersdorf	Berlin
23	0403	Schmargendorf	Charlottenburg-Wilmersdorf	Berlin
24	0404	Grunewald	Charlottenburg-Wilmersdorf	Berlin
25	0405	Westend	Charlottenburg-Wilmersdorf	Berlin
26	0406	Charlottenburg-Nord	Charlottenburg-Wilmersdorf	Berlin
27	0407	Halensee	Charlottenburg-Wilmersdorf	Berlin
28	0501	Spandau	Spandau	Berlin
29	0502	Haselhorst	Spandau	Berlin
30	0503	Siemensstadt	Spandau	Berlin
31	0504	Staaken	Spandau	Berlin
32	0505	Gatow	Spandau	Berlin
33	0506	Kladow	Spandau	Berlin
34	0507	Hakenfelde	Spandau	Berlin
35	0508	Falkenhagener Feld	Spandau	Berlin
36	0509	Wilhelmstadt	Spandau	Berlin
37	0601	Steglitz	Steglitz-Zehlendorf	Berlin
38	0602	Lichterfelde	Steglitz-Zehlendorf	Berlin
39	0603	Lankwitz	Steglitz-Zehlendorf	Berlin
40	0604	Zehlendorf	Steglitz-Zehlendorf	Berlin
41	0605	Dahlem	Steglitz-Zehlendorf	Berlin
42	0606	Nikolassee	Steglitz-Zehlendorf	Berlin
43	0607	Wannsee	Steglitz-Zehlendorf	Berlin
44	0701	Schöneberg	Tempelhof-Schöneberg	Berlin
45	0702	Friedenau	Tempelhof-Schöneberg	Berlin
46	0703	Tempelhof	Tempelhof-Schöneberg	Berlin
47	0704	Mariendorf	Tempelhof-Schöneberg	Berlin
48	0705	Marienfelde	Tempelhof-Schöneberg	Berlin
49	0706	Lichtenrade	Tempelhof-Schöneberg	Berlin
50	0801	Neukölln	Neukölln	Berlin
51	0802	Britz	Neukölln	Berlin
52	0803	Buckow	Neukölln	Berlin
53	0804	Rudow	Neukölln	Berlin
54	0805	Gropiusstadt	Neukölln	Berlin
55	0901	Alt-Treptow	Treptow-Köpenick	Berlin
56	0902	Plänterwald	Treptow-Köpenick	Berlin
57	0903	Baumschulenweg	Treptow-Köpenick	Berlin
58	0904	Johannisthal	Treptow-Köpenick	Berlin
59	0905	Niederschöneweide	Treptow-Köpenick	Berlin
60	0906	Altglienicke	Treptow-Köpenick	Berlin
61	0907	Adlershof	Treptow-Köpenick	Berlin
62	0908	Bohnsdorf	Treptow-Köpenick	Berlin
63	0909	Oberschöneweide	Treptow-Köpenick	Berlin
64	0910	Köpenick	Treptow-Köpenick	Berlin
65	0911	Friedrichshagen	Treptow-Köpenick	Berlin
66	0912	Rahnsdorf	Treptow-Köpenick	Berlin
67	0913	Grünau	Treptow-Köpenick	Berlin

	neighbourhood_id	neighbourhood	borough	city
68	0914	Müggelheim	Treptow-Köpenick	Berlin
69	0915	Schmöckwitz	Treptow-Köpenick	Berlin
70	1001	Marzahn	Marzahn-Hellersdorf	Berlin
71	1002	Biesdorf	Marzahn-Hellersdorf	Berlin
72	1003	Kaulsdorf	Marzahn-Hellersdorf	Berlin
73	1004	Mahlsdorf	Marzahn-Hellersdorf	Berlin
74	1005	Hellersdorf	Marzahn-Hellersdorf	Berlin
75	1101	Friedrichsfelde	Lichtenberg	Berlin
76	1102	Karlshorst	Lichtenberg	Berlin
77	1103	Lichtenberg	Lichtenberg	Berlin
78	1104	Falkenberg	Lichtenberg	Berlin
79	1106	Malchow	Lichtenberg	Berlin
80	1107	Wartenberg	Lichtenberg	Berlin
81	1109	Neu-Hohenschönhausen	Lichtenberg	Berlin
82	1110	Alt-Hohenschönhausen	Lichtenberg	Berlin
83	1111	Fennpfuhl	Lichtenberg	Berlin
84	1112	Rummelsburg	Lichtenberg	Berlin
85	1201	Reinickendorf	Reinickendorf	Berlin
86	1202	Tegel	Reinickendorf	Berlin
87	1203	Konradshöhe	Reinickendorf	Berlin
88	1204	Heiligensee	Reinickendorf	Berlin
89	1205	Frohnau	Reinickendorf	Berlin
90	1206	Hermsdorf	Reinickendorf	Berlin
91	1207	Waidmannslust	Reinickendorf	Berlin
92	1208	Lübars	Reinickendorf	Berlin
93	1209	Wittenau	Reinickendorf	Berlin
94	1210	Märkisches Viertel	Reinickendorf	Berlin
95	1211	Borsigwalde	Reinickendorf	Berlin

B) Category labels cleaning

Misclassified venues

venue_name Lidl, venue_category changed from Discount Store to Supermarket : 2 converted
venue_name La Cave Prenzlauer Berg, venue_category changed from Winery to Wine Bar : 1 converted
venue_name Pfennigpfeiffer, venue_category changed from Shop & Service to Paper / Office Supplies Store : 1 converted
venue_name Le Midnight, venue_category changed from Other Nightlife to Hookah Bar : 1 converted

Generalised or aggregated categories

venue_category changed from ['Asian Restaurant', 'Noodle House', 'Chinese Restaurant', 'Cantonese Restaurant', 'Dim Sum Restaurant', 'Dumpling Restaurant', 'Jiangxi Restaurant', 'Shanxi Restaurant', 'Szechuan Restaurant', 'Taiwanese Restaurant', 'Japanese Restaurant', 'Soba Restaurant', 'Okonomiyaki Restaurant', 'Ramen Restaurant', 'Sushi Restaurant', 'Udon Restaurant', 'Korean Restaurant', 'Cambodian Restaurant', 'Thai Restaurant', 'Vietnamese Restaurant', 'Indonesian Restaurant'] to East Asian Restaurant : 751 converted
venue_category changed from ['Afghan Restaurant', 'Indian Restaurant', 'Pakistani Restaurant', 'Tibetan Restaurant'] to South Asian Restaurant : 73 converted
venue_category changed from ['American Restaurant', 'Cajun / Creole Restaurant', 'Southern / Soul Food Restaurant', 'New American Restaurant'] to American Restaurant : 16 converted
venue_category changed from ['Mexican Restaurant', 'Burrito Place', 'Taco Place', 'Central American', 'Caribbean Restaurant', 'Cuban Restaurant', 'South American Restaurant', 'Arepa Restaurant', 'Argentinian Restaurant', 'Brazilian Restaurant', 'Colombian Restaurant', 'Empanada Restaurant', 'Peruvian Restaurant', 'Venezuelan Restaurant', 'Latin American Restaurant'] to American Restaurant : 104 converted
venue_category changed from ['Hawaiian Restaurant', 'Poke Place'] to Restaurant : 5 converted
venue_category changed from ['English Restaurant', 'Scandinavian Restaurant', 'Belgian Restaurant', 'Modern European Restaurant', 'Swiss Restaurant', 'Austrian Restaurant', 'Schnitzel Restaurant', 'Eastern

European Restaurant', 'Russian Restaurant', 'Romanian Restaurant', 'Mediterranean Restaurant', 'Turkish Restaurant', 'Greek Restaurant', 'Taverna', 'Kumpir Restaurant', 'Portuguese Restaurant', 'Spanish Restaurant', 'Tapas Restaurant'] to European Restaurant : 236 converted
venue_category changed from ['Italian Restaurant', 'Pizza Place'] to Italian Restaurant : 171 converted
venue_category changed from ['Bavarian Restaurant', 'German Restaurant', 'Alsatian Restaurant', 'Auvergne Restaurant', 'Basque Restaurant', 'Breton Restaurant', 'Burgundian Restaurant', 'Corsican Restaurant', 'Fondue Restaurant', 'Provençal Restaurant', 'Savoyard Restaurant', 'Southwestern French Restaurant', 'French Restaurant', 'Molecular Gastronomy Restaurant', 'Bistro', 'Creperie', 'Lyonese Bouchon'] to Restaurant : 1193 converted
venue_category changed from ['Middle Eastern Restaurant', 'Caucasian Restaurant', 'Kurdish Restaurant', 'Lebanese Restaurant', 'Persian Restaurant', 'Syrian Restaurant', 'Yemeni Restaurant', 'Israeli Restaurant', 'Jewish Restaurant', 'Doner Restaurant', 'Falafel Restaurant', 'Halal Restaurant', 'Kebab Restaurant', 'Souvlaki Shop', 'Kofte Place', 'Pide Place', 'Shawarma Place'] to M.E. / African Restaurant : 181 converted
venue_category changed from ['Moroccan Restaurant', 'African Restaurant', 'Ethiopian Restaurant'] to M.E. / African Restaurant : 56 converted
venue_category changed from ['Restaurant', 'Comfort Food Restaurant', 'Diner', 'BBQ Joint', 'Steakhouse', 'Burger Joint', 'Seafood Restaurant', 'Fish & Chips Shop', 'Snack Place', 'Fried Chicken Joint', 'Food', 'Mac & Cheese Joint', 'Hot Dog Joint', 'Ice Cream Shop', 'Donut Shop', 'Bubble Tea Shop', 'Bagel Shop', 'Pie Shop', 'Sandwich Place', 'Soup Place', 'Food Truck', 'Street Food Gathering', 'Salad Place', 'Juice Bar', 'Currywurst Joint'] to Restaurant : 593 converted
venue_category changed from ['Gluten-free Restaurant', 'Vegetarian / Vegan Restaurant'] to Restaurant : 69 converted
venue_category changed from ['Fast Food Restaurant', 'Food Court', 'Snack Place', 'Fried Chicken Joint', 'Food', 'Mac & Cheese Joint', 'Hot Dog Joint', 'Ice Cream Shop', 'Donut Shop', 'Bubble Tea Shop', 'Bagel Shop', 'Pie Shop', 'Sandwich Place', 'Soup Place', 'Food Truck', 'Street Food Gathering', 'Salad Place', 'Juice Bar', 'Currywurst Joint'] to Fast Food : 70 converted
venue_category changed from ['Café', 'Coffee Roaster', 'Coffee Shop', 'Pet Café', 'Tea Room', 'Cafeteria', 'Breakfast Spot', 'College Cafeteria'] to Café : 340 converted
venue_category changed from ['Bar', 'Wine Bar', 'Hotel Bar', 'Beer Bar', 'Sports Bar', 'Dive Bar', 'Beach Bar', 'Brasserie', 'Gay Bar', 'Hookah Bar', 'Cocktail Bar', 'Piano Bar', 'Champagne Bar', 'Whisky Bar', 'Roof Deck', 'Speakeasy', 'Lounge', 'Gastropub', 'Irish Pub', 'Beer Garden', 'Brewery', 'Beer Bar', 'Pub'] to Bar / Pub : 859 converted
venue_category changed from ['Nightclub'] to Nightclub : 0 converted
venue_category changed from ['Supermarket', 'Food & Drink Shop', 'Grocery Store', 'Discount Store', 'Organic Grocery', 'Health Food Store'] to Supermarket : 123 converted
venue_category changed from ['Market', 'Farmers Market', 'Flea Market'] to Market : 41 converted
venue_category changed from ['Bakery', 'Pastry Shop', 'Cheese Shop', 'Dessert Shop', 'Chocolate Shop', 'Cupcake Shop', 'Gourmet Shop', 'Candy Store', 'Butcher', 'Deli / Bodega', 'Fish Market', 'Trattoria/Osteria'] to Specialty Food : 665 converted
venue_category changed from ['Liquor Store', 'Beer Store', 'Wine Shop'] to Liquor Store : 55 converted
venue_category changed from ['Drugstore', 'Pharmacy', 'Optical Shop'] to Pharmacy / Healthcare : 123 converted
venue_category changed from ['Convenience Store', 'Herbs & Spices Store'] to Convenience Store : 1 converted
venue_category changed from ['Salon / Barbershop', 'Spa', 'Health & Beauty Service', 'Massage Studio'] to Beauty : 25 converted
venue_category changed from ['Bookstore', 'Boutique', 'Used Bookstore', 'Shoe Store', 'Furniture / Home Store', 'Clothing Store', 'Arts & Crafts Store', 'Toy / Game Store', "Women's Store", "Men's Store", 'Jewelry Store', 'Accessories Store', 'Leather Goods Store', 'Electronics Store', 'Paper / Office Supplies Store', 'Kids Store', 'Pet Store', 'Lingerie Store', 'Camera Store', 'Music Store', 'Thrift / Vintage Store', 'Kitchen Supply Store', 'Stationery Store', 'Baby Store', 'Carpet Store', 'Outdoor Supply Store', 'Gift Shop', 'Cosmetics Shop', 'Smoke Shop', 'Perfume Shop', 'Miscellaneous Shop', 'Sporting Goods Shop', 'Record Shop', 'Flower Shop', 'Comic Shop', 'Costume Shop', 'Tailor Shop', 'Antique Shop', 'Fabric Shop', 'Pop-Up Shop', 'Mobile Phone Shop', 'Hobby Shop', 'Bike Shop', 'Furniture / Home Store', 'Kitchen Supply Store', 'Carpet Store', 'Garden Center', 'Hardware Store', 'Big Box Store', 'Shopping Mall', 'Shopping Plaza', 'Department Store', 'Automotive Shop', 'Motorcycle Shop', 'Auto Dealership', 'Fishing Store', 'Adult Boutique', 'Newsstand', 'Gun Shop'] to Shopping : 720 converted
venue_category changed from ['Auto Garage', 'Auto Workshop', 'Gas Station'] to Auto : 32 converted
venue_category changed from ['Yoga Studio', 'Sports Club', 'Tennis Court', 'Gym', 'Gym / Fitness Center', 'Pool', 'Athletics & Sports', 'Fitness Center', 'Gym Pool', 'Cycle Studio', 'Recreation Center', 'Dance Studio', 'Squash Court', 'Hockey Rink', 'Soccer Field', 'Baseball Field', 'Boxing Gym', 'Volleyball Court', 'Climbing Gym', 'Rock Climbing Spot', 'Basketball Court', 'Hockey Field', 'College Gym', 'Golf Course', 'Martial Arts Dojo', 'College Rec Center'] to Fitness / Sports : 335 converted

venue_category changed from ['Bowling Alley', 'Gaming Cafe', 'General Entertainment', 'Laser Tag', 'Indie Movie Theater', 'Movie Theater', 'Multiplex', 'Pool Hall', 'Arcade', 'Skating Rink', 'Karaoke Bar', 'Indoor Play Area'] to Urban Entertainment : 136 converted
venue_category changed from ['Theme Park Ride / Attraction', 'Paintball Field', 'Go Kart Track', 'Racetrack', 'Fair', 'Theme Park'] to Outdoor Entertainment : 14 converted
venue_category changed from ['Bed & Breakfast', 'Hostel', 'Hotel', 'Rental Car Location', 'Souvenir Shop', 'Motel', 'Resort', 'Tourist Information Center', 'Tour Provider'] to Hotel / Tourism : 649 converted
venue_category changed from ['Recording Studio', 'Design Studio', 'Photography Studio', 'Photography Lab', 'Lawyer', 'Food Service', 'Event Service', 'Office', 'Entertainment Service', 'Industrial Estate', 'Government Building', 'Tech Startup', 'IT Services', 'Construction & Landscaping', 'Coworking Space', 'Insurance Office', 'Business Service'] to Workplace : 31 converted
venue_category changed from ['Bank', 'Post Office', 'Laundry Service', 'Laundromat', 'Credit Union', 'Dry Cleaner', 'Shipping Store', 'Print Shop', 'Tattoo Parlor'] to Service : 40 converted
venue_category changed from ['Church'] to Place of Worship : 5 converted
venue_category changed from ['Stadium', 'Rugby Stadium', 'Soccer Stadium', 'Track', 'Track Stadium', 'Football Stadium', 'Tennis Stadium'] to Stadium : 17 converted
venue_category changed from ['Circus', 'Comedy Club', 'Concert Hall', 'Jazz Club', 'Opera House', 'Performing Arts Venue', 'Indie Theater', 'Music Venue', 'Theater', 'Rock Club', 'Amphitheater'] to Performing Arts : 179 converted
venue_category changed from ['Museum', 'Art Gallery', 'Art Museum', 'History Museum', 'Science Museum', 'Planetarium', 'Exhibit', 'Aquarium'] to Museum / Exhibit : 187 converted
venue_category changed from ['Library', 'Community Center', 'Cultural Center', 'Auditorium', 'Event Space', 'Arts & Entertainment', 'Outdoor Event Space'] to Community Venue : 16 converted
venue_category changed from ['Botanical Garden', 'Cemetery', 'Garden', 'Park', 'Sculpture Garden', 'Skate Park', 'Playground', 'Zoo', 'Zoo Exhibit', 'Dog Run'] to Park : 137 converted
venue_category changed from ['Plaza', 'Pedestrian Plaza'] to Plaza : 29 converted
venue_category changed from ['Harbor / Marina', 'Waterfront', 'Canal'] to Waterfront : 19 converted
venue_category changed from ['Historic Site', 'Memorial Site', 'Monument / Landmark', 'Fountain', 'Outdoor Sculpture', 'Palace', 'Scenic Lookout', 'Castle', 'Windmill', 'Cable Car', 'Street Art'] to Attraction : 113 converted
venue_category changed from ['Nature Preserve', 'Campground', 'Mountain', 'Forest', 'Beach', 'Lake', 'River'] to Nature : 60 converted
venue_category changed from ['Stables', 'Farm', 'Distillery'] to Agriculture : 14 converted
venue_category changed from ['Bus Station', 'Bus Stop', 'Bike Rental / Bike Share'] to Bus / Bike : 125 converted
venue_category changed from ['Train Station', 'Metro Station', 'Light Rail Station', 'Tram Station'] to Rail / Tram : 151 converted
venue_category changed from ['Boat or Ferry', 'Pier', 'Boat Rental'] to Boat / Ferry : 20 converted
venue_category changed from ['Airport Lounge', 'Airport Service'] to Airport : 6 converted

Too few in category
venue_category Community Venue : 16 venues dropped
venue_category Outdoor Entertainment : 14 venues dropped
venue_category Agriculture : 14 venues dropped
venue_category Trail : 10 venues dropped
venue_category Platform : 9 venues dropped
venue_category Airport : 6 venues dropped
venue_category Place of Worship : 5 venues dropped
venue_category Bridge : 5 venues dropped
venue_category ATM : 5 venues dropped
venue_category Building : 4 venues dropped
venue_category Intersection : 3 venues dropped
venue_category Neighborhood : 3 venues dropped
venue_category Tree : 3 venues dropped
venue_category Rest Area : 2 venues dropped
venue_category Tunnel : 2 venues dropped
venue_category Storage Facility : 1 venues dropped
venue_category Animal Shelter : 1 venues dropped
venue_category Canal Lock : 1 venues dropped
venue_category Field : 1 venues dropped
venue_category School : 1 venues dropped
venue_category Boarding House : 1 venues dropped
venue_category Taxi Stand : 1 venues dropped
venue_category Residential Building (Apartment / Condo) : 1 venues dropped
venue_category Road : 1 venues dropped
venue_category Warehouse Store : 1 venues dropped

111 venues dropped

Data biased or not useful
venue_category Rail / Tram : 151 venues dropped
venue_category Bus / Bike : 125 venues dropped
venue_category Pharmacy / Healthcare : 123 venues dropped
venue_category Workplace : 31 venues dropped
430 venues dropped

C) References used for venue category grouping

https://en.wikipedia.org/wiki/Geography_of_Asia#Regions
https://en.wikipedia.org/wiki/Kurdistan
https://en.wikipedia.org/wiki/Caucasus
https://en.wikipedia.org/wiki/List_of_regions_of_Latin_America#Cultural_regions
https://en.wikipedia.org/wiki/List_of_European_cuisines
https://en.wikipedia.org/wiki/
Middle_East#Territories_and_regions_usually_considered_within_the_Middle_East

D) Jupyter notebook

https://gist.github.com/jeyel496/7d56ff1bbb9cf7e486de87f1045738d4