Capstone Project
The Battle of Neighborhoods

OBJECTIVE

¿It's possible to compare neighborhoods of 5 multicultural cities (London, Sydney, Paris, Amsterdan and San Francisco) in order to find groups of neighborhoods that are similar according to similarities in the categories of businesses that are in them?

Data

The data used for resolve the problem is:

Foursquare dataset
- Bussiness geo location
- Category
- Name

Neigbourhoods geo location data from http://insideairbnb.com/get-the-data.html

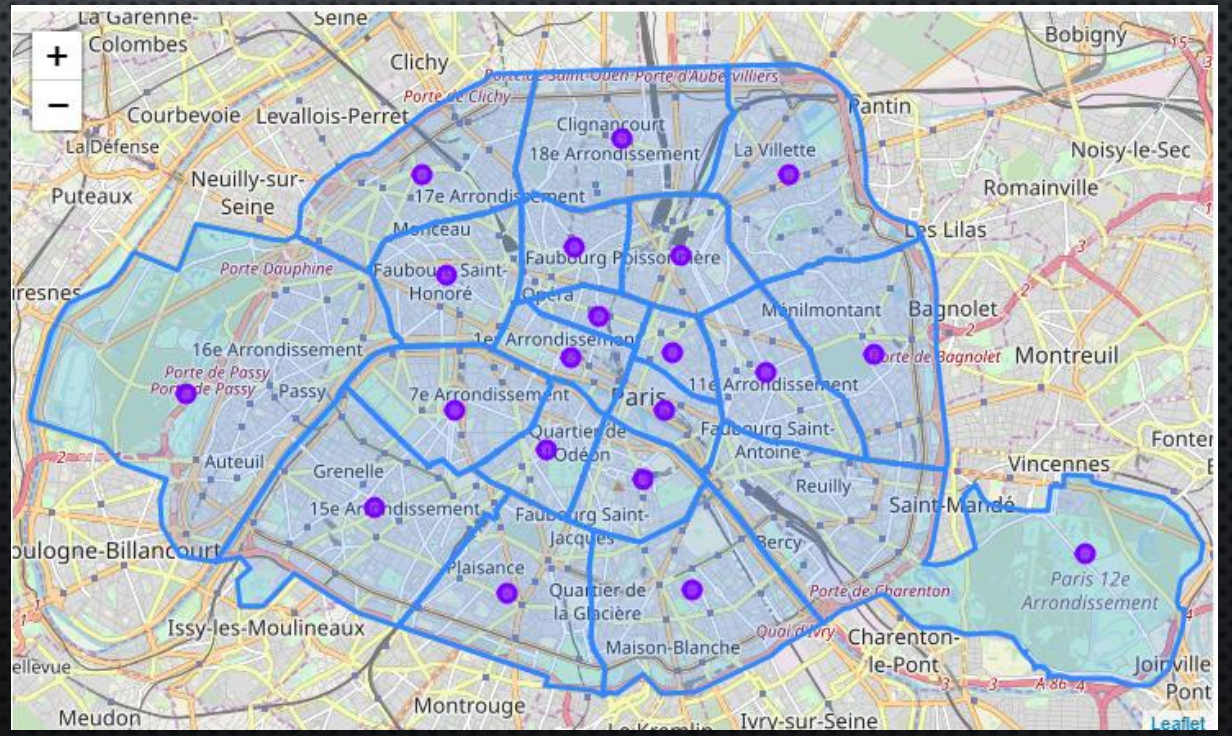
The features that were used from the data are:

- Neigbourhood Geo Location
- Neigbourhood Name

Methodology

For modeling, the k-means algorithm were used in order to identify groups of neighborhoods similar to each other in the distribution of categories of places that are in their surroundings.

The calculation of the center of each neighborhood was made using the shapely library and a test of the data was conducted with the city of Paris



Results

Proportion of venue categories for each cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Bakery	0.224165	0.0	0.056003	0.072941
Café	0.175256	0.0	0.680374	0.100392
Coffee Shop	0.195146	0.0	0.055889	0.720196
Italian Restaurant	0.269075	0.0	0.066020	0.030098
Park	0.136358	1.0	0.141713	0.076373

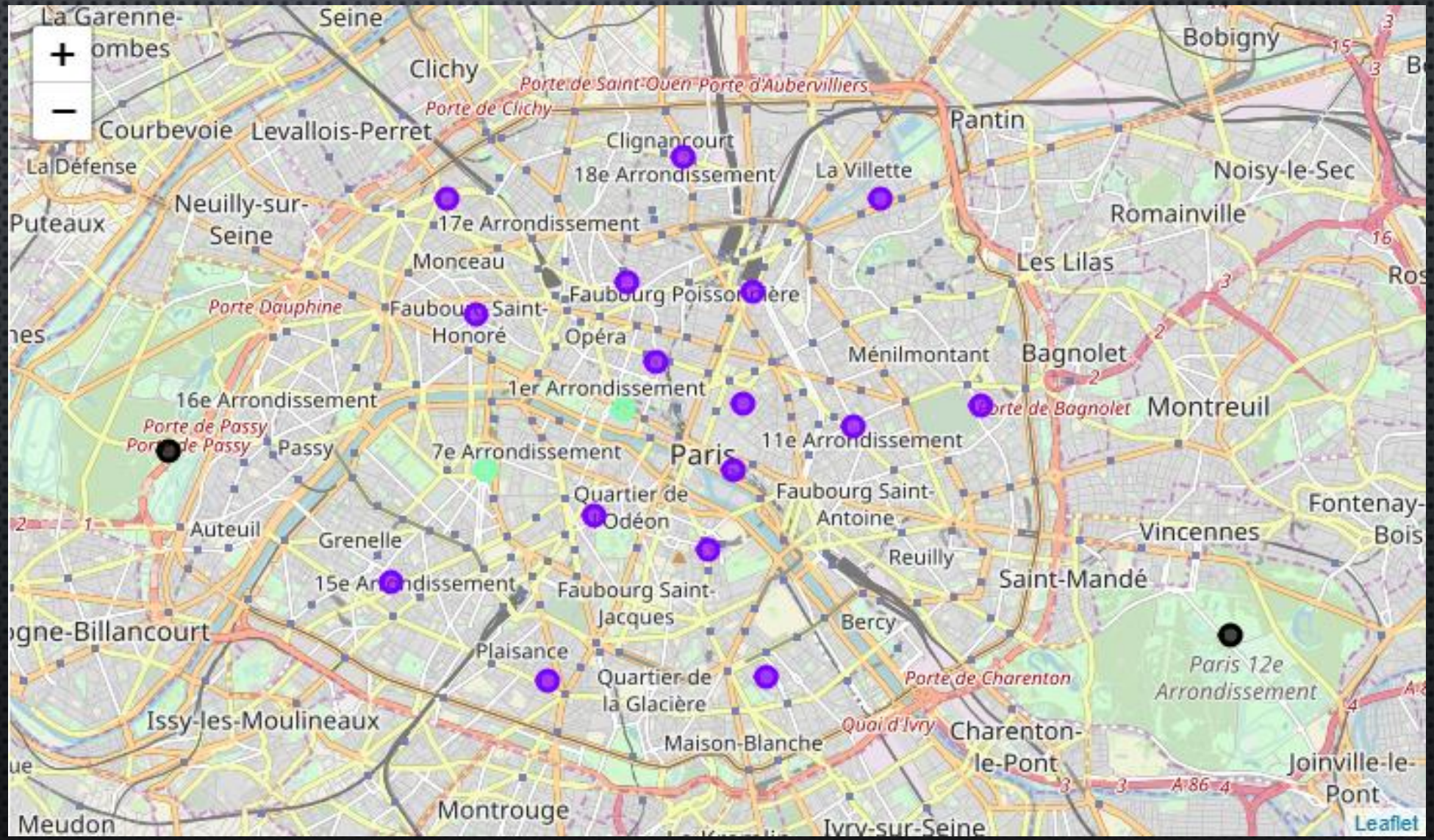
Number of neighborhoods for city for cluster

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Amsterdam	9	3	3	1
London	8	4	4	7
Paris	16	0	2	0
San Francisco	15	8	4	8
Sydney	3	3	11	1

Cluster 1 has a great variety of categories of venues. Cluster 2 corresponds to neighborhoods with 100% of parks. Cluster 3 is more represented by neighborhoods that own cafes and parks, while Cluster 4 by coffee shops and cafes.

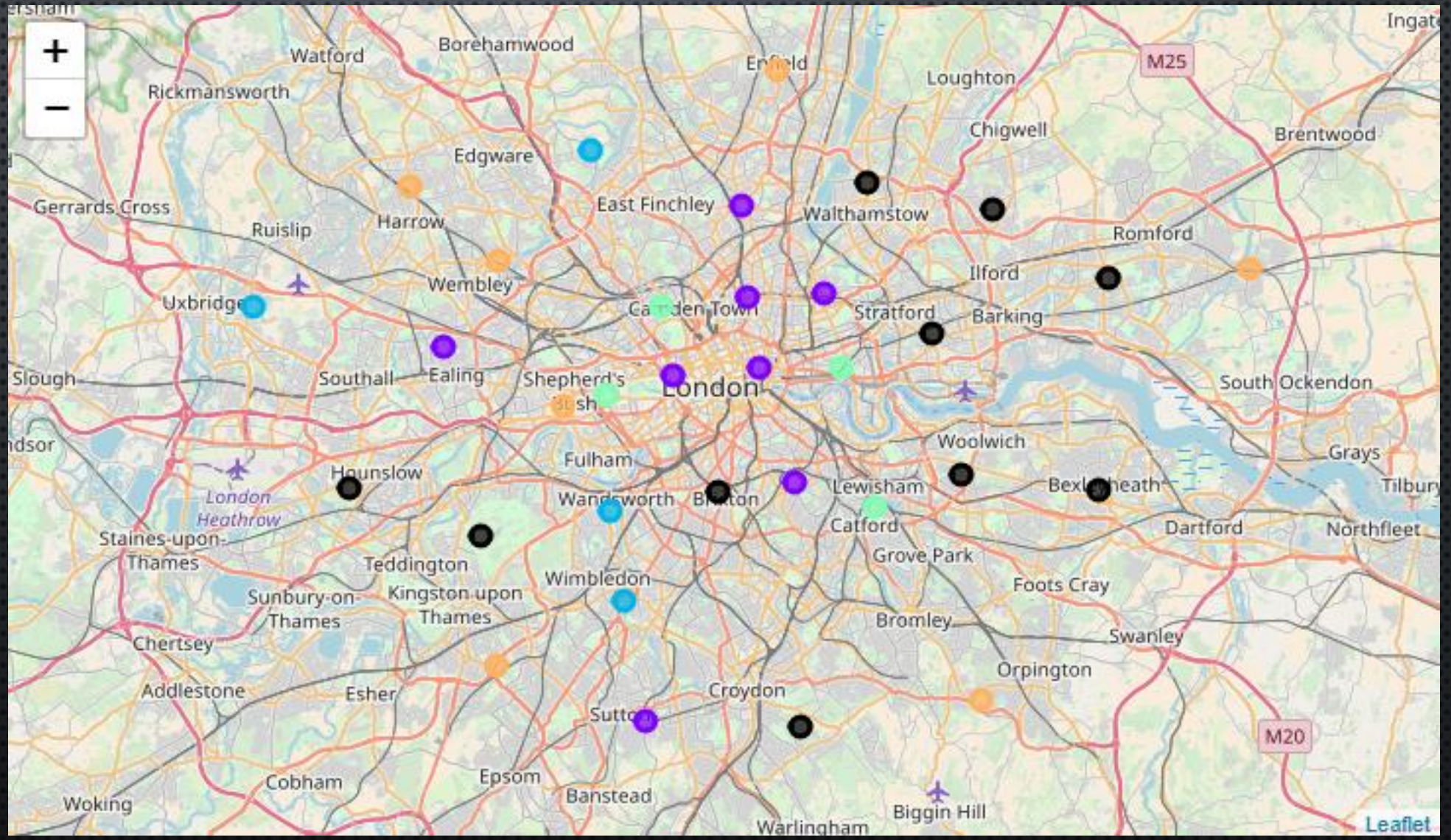
Results: Paris

Clusters distribution:



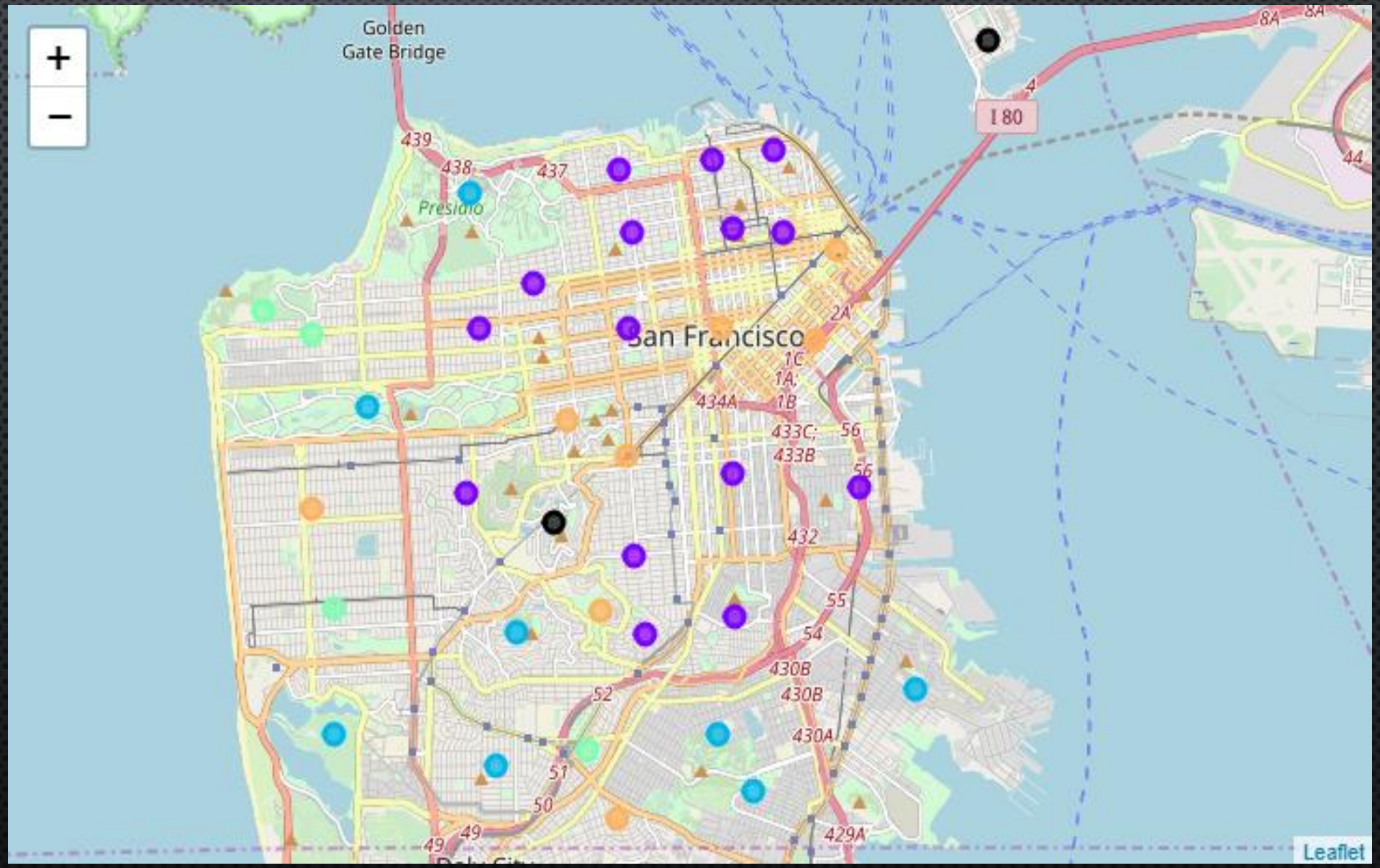
Results: London

Clusters distribution:



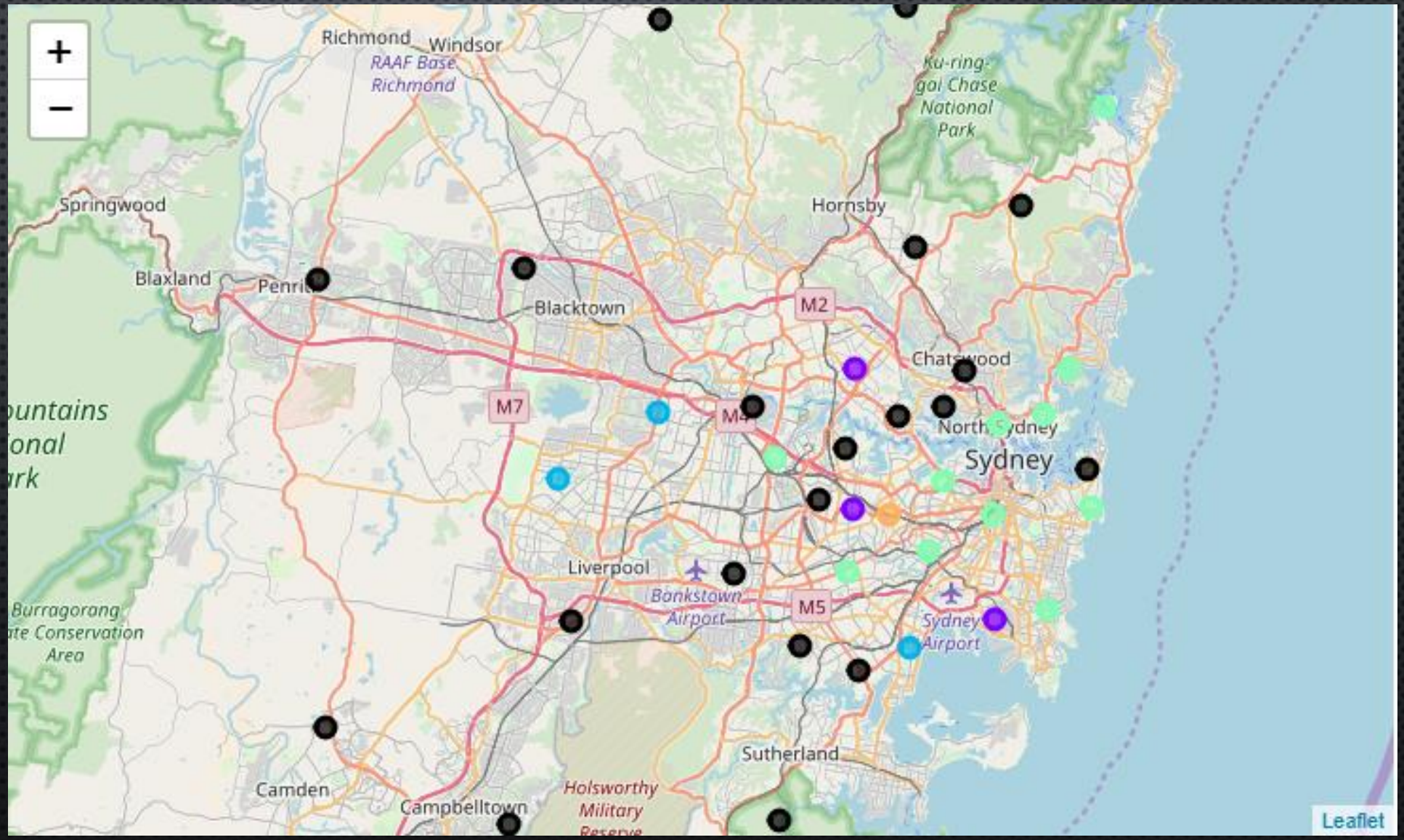
Results: San Francisco

Clusters distribution:



Results: Sidney

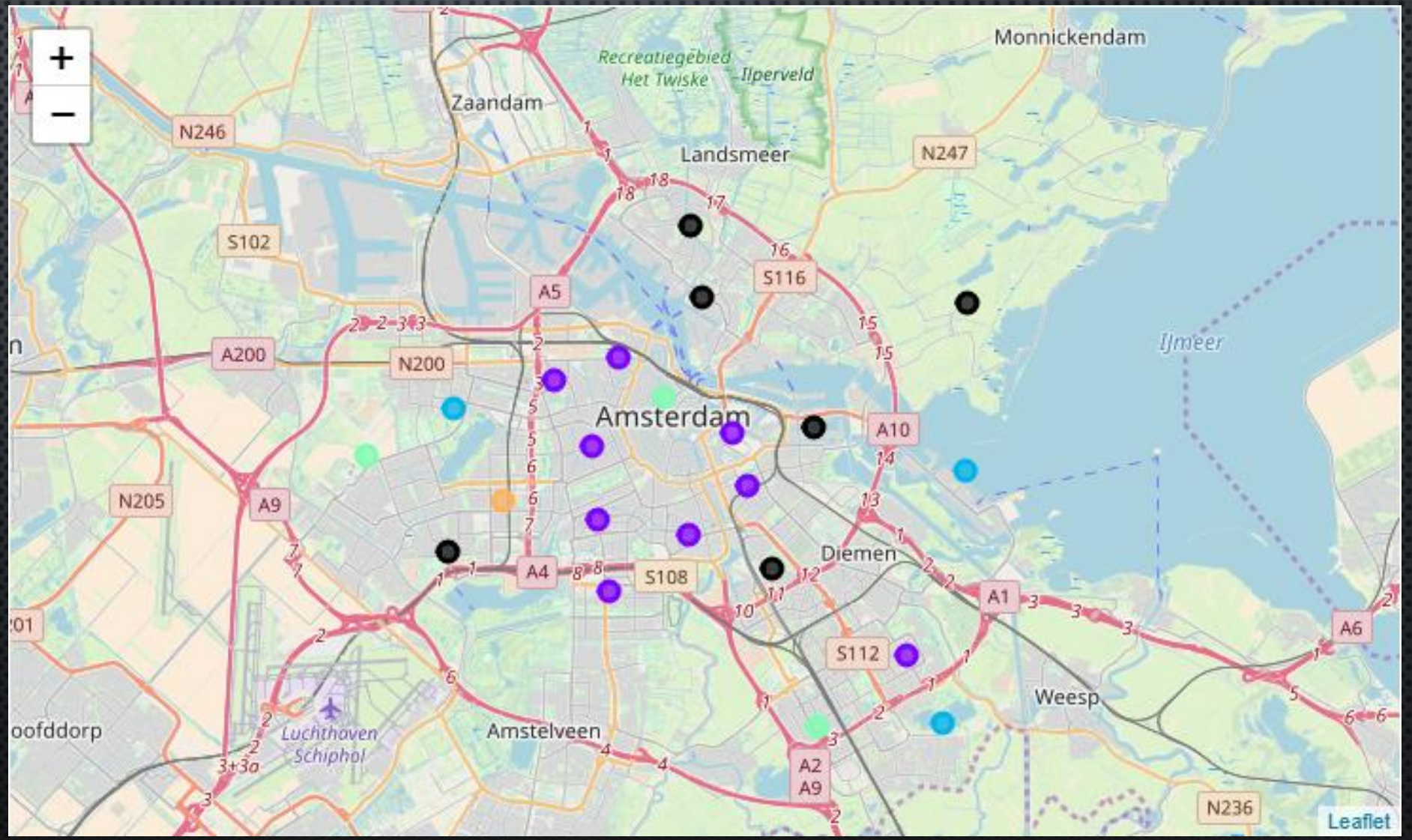
Clusters distribution:



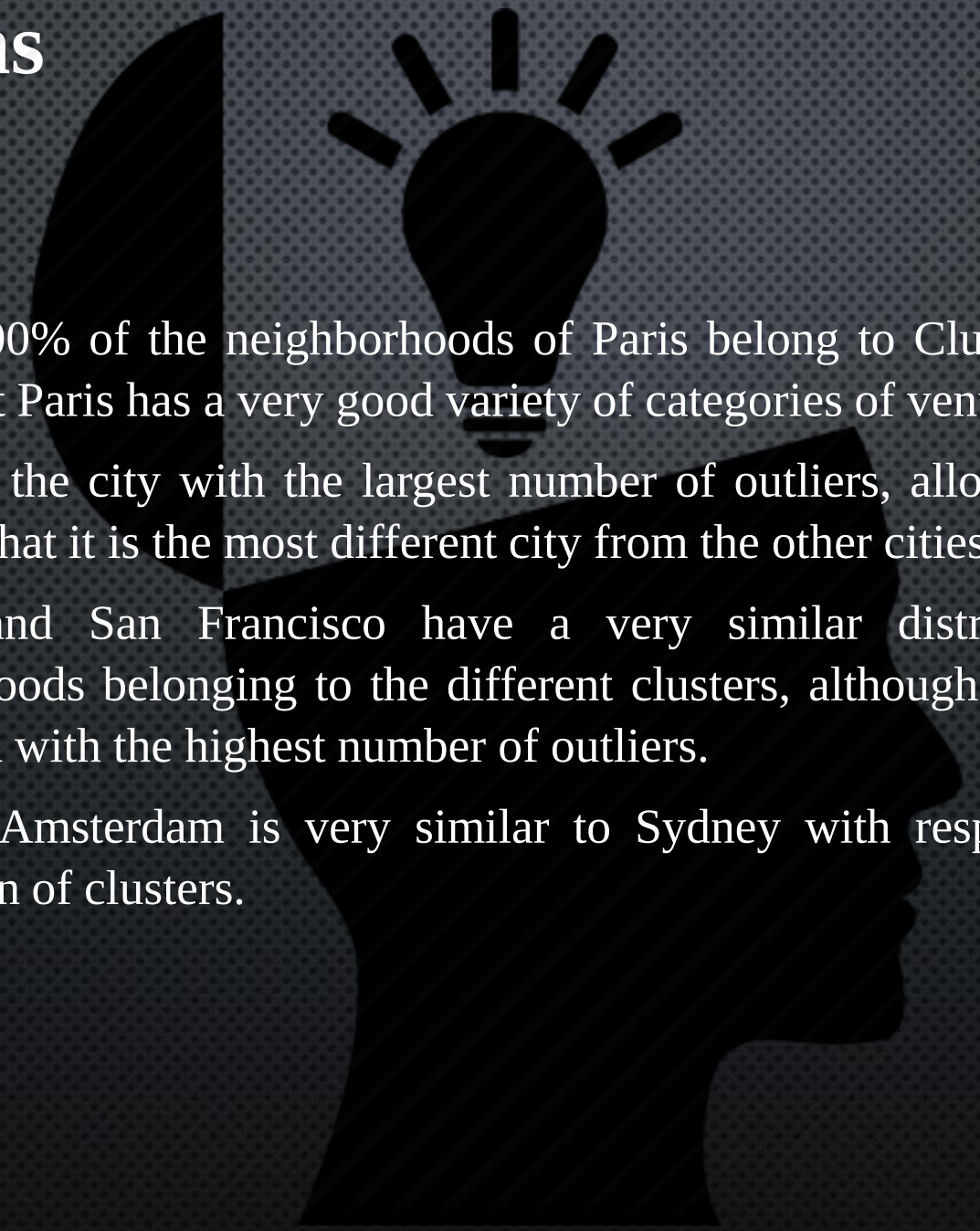
Results: Amsterdam

Clusters distribution:

- Cluster 1
- Cluster 2
- Cluster 3
- Cluster 4
- Outliers



Conclusions



Almost 100% of the neighborhoods of Paris belong to Cluster 1, this means that Paris has a very good variety of categories of venues.

Sydney is the city with the largest number of outliers, allowing us to conclude that it is the most different city from the other cities evaluated.

London and San Francisco have a very similar distribution of neighborhoods belonging to the different clusters, although London is the second with the highest number of outliers.

Likewise Amsterdam is very similar to Sydney with respect to the distribution of clusters.