Find your new home

Olesya Tsapenko

June 07, 2020

1. Introduction

1.1 Background

Every year thousands and thousands international students graduate from European Universities. Fair share of graduates want to stay and work in the EU. A lot of them might face the challenge of not knowing the local language. They start to search for cities/countries where they can find a job in English preferably in the EU. There are at least three well-known EU capitals for it - Amsterdam, Paris and Berlin.

1.2 Problem

To find a job purely in English - it is a big problem but to find a new home may be even harder. On one hand - each city is a unique experience. But on the other hand there are many neighborhoods which have pretty similar venues around. At this work I will compare 3 EU capitals' neighborhoods and cluster them to identify some similar places for living based on surrounding venues. In the second part I will add my current neighborhood and show a comparison between it and possible places for living. The script is flexible and assumes that any user can provide their address and look at the personal recommendations.

1.3 Interest

Freshly graduated international students from EU universities would be interested to find a new home for themself which is similar to their living place. Comparison of neighborhoods/cities/countries can also help to find a place which is absolutely different with the current life situation if this is the goal of the user.
Others who are interested in this comparison such as educated english-speaking people may also want to choose a place to move in the EU.

2. Data acquisition and cleaning

2.1 Data sources

The first challenge which I met is to find a list of all neighborhoods of given cities. For this, I parsed Wikipedia web-pages using the beautiful-soup python framework. The next step is to receive coordinates (latitude and longitude) of each neighborhood which I did with the help of a

geopy library (also this library helps me to find coordinates of my current living place). When we already have latitude and longitude of each neighborhood it is easy to find venues in the given radius (500 meters) using API requests to Foursquare service.

2.2 Data cleaning

Unfortunately, I could not receive coordinates of absolutely all neighborhoods using geopy so these places were excluded from future analysis. Also I dropped neighborhoods which have less than 4 venues in the radius of 500 meters. On top of that, I renamed French, German and Dutch restaurants categories to one category "Local cuisine restaurant" to make clustering more unbiased.

As a next step of data cleaning, I calculated all unique venue categories and encoded each of them using one-hot-encoding. Later I group all venues by neighborhood and by taking the mean of the frequency of occurrence of each category. It became a feature for my future clustering and recommendation. I want to mention that I didn't use any geographical-related features such as latitude, longitude, city or country names, neighborhood names to make the clustering more objective for geographically-separated places.

3. Methodology

Because my data is unlabeled I have to choose one of the unsupervised Machine Learning methods. For comparison of different neighborhoods I've picked k-mean clustering.

I set up 10 clusters and repeated k-mean clustering 1000 times in order to find the solution close to global optimum instead of local one.
After receiving results of clustering, I excluded clusters with 3 or less neighborhoods from future analysis.

4. Results

I've plotted clusters individually for each city and for my current location.
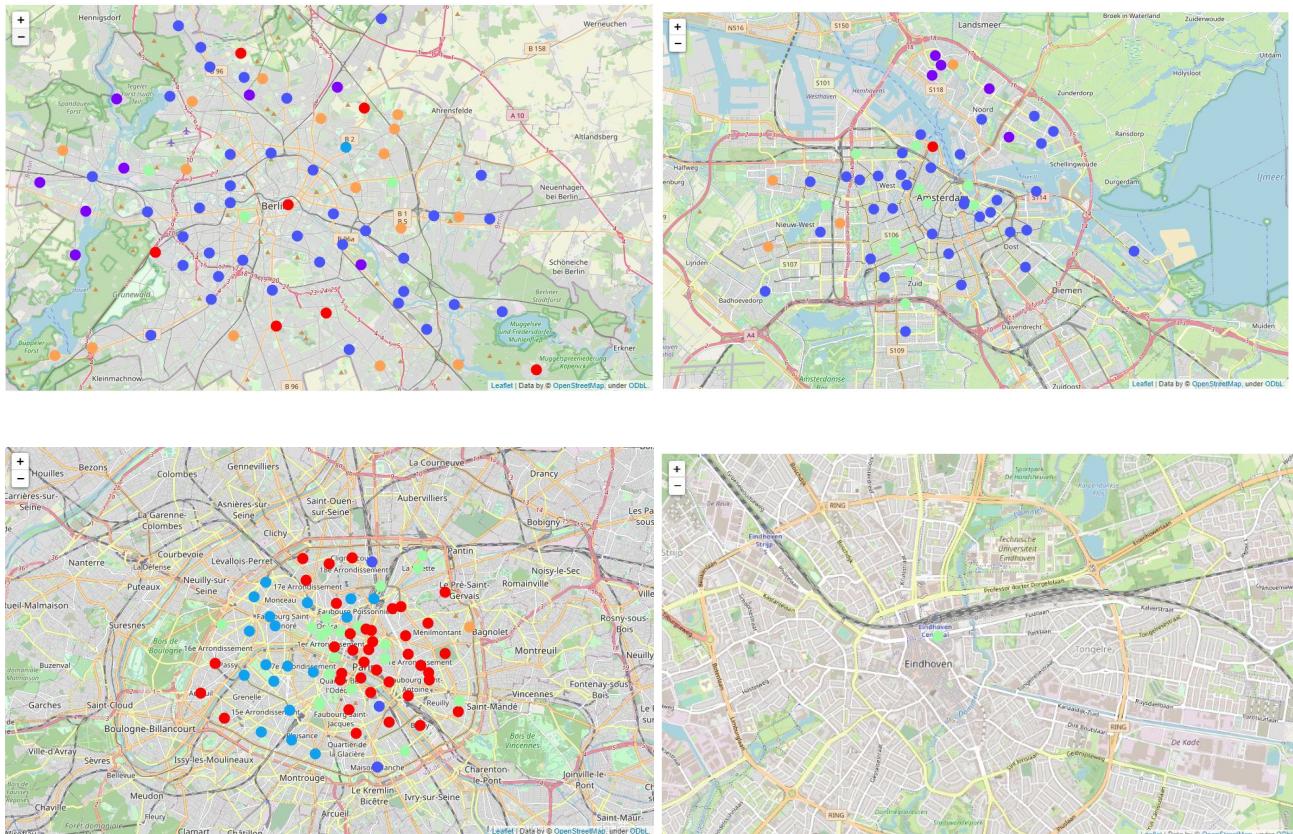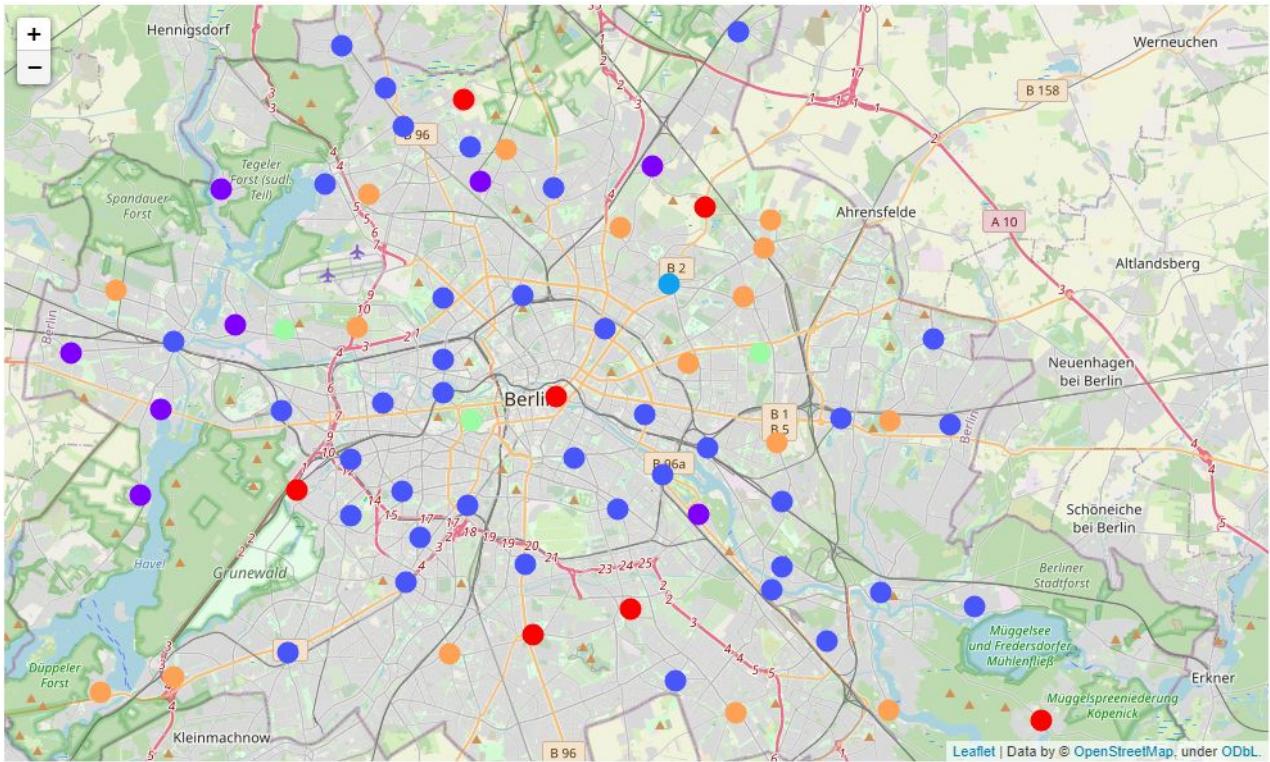
Figure 1. Clusters from left to right from top to bottom:
Berlin, Amsterdam, Paris, Eindhoven (current location)

The full statistics about clusters for all neighborhoods:

Cluster Labels	Neighborhood	
1.0	13	🟣
2.0	76	🔵
3.0	19	🔵
6.0	35	🟢
8.0	21	🟠
10.0	46	🔴

4.1 Berlin

Let's have a look closely at clusters in Berlin:

Neighborhood		
Cluster Labels		
1.0	8	🟣
2.0	38	🔵
3.0	1	🔵
6.0	3	🟢
8.0	16	🟠
10.0	7	🔴
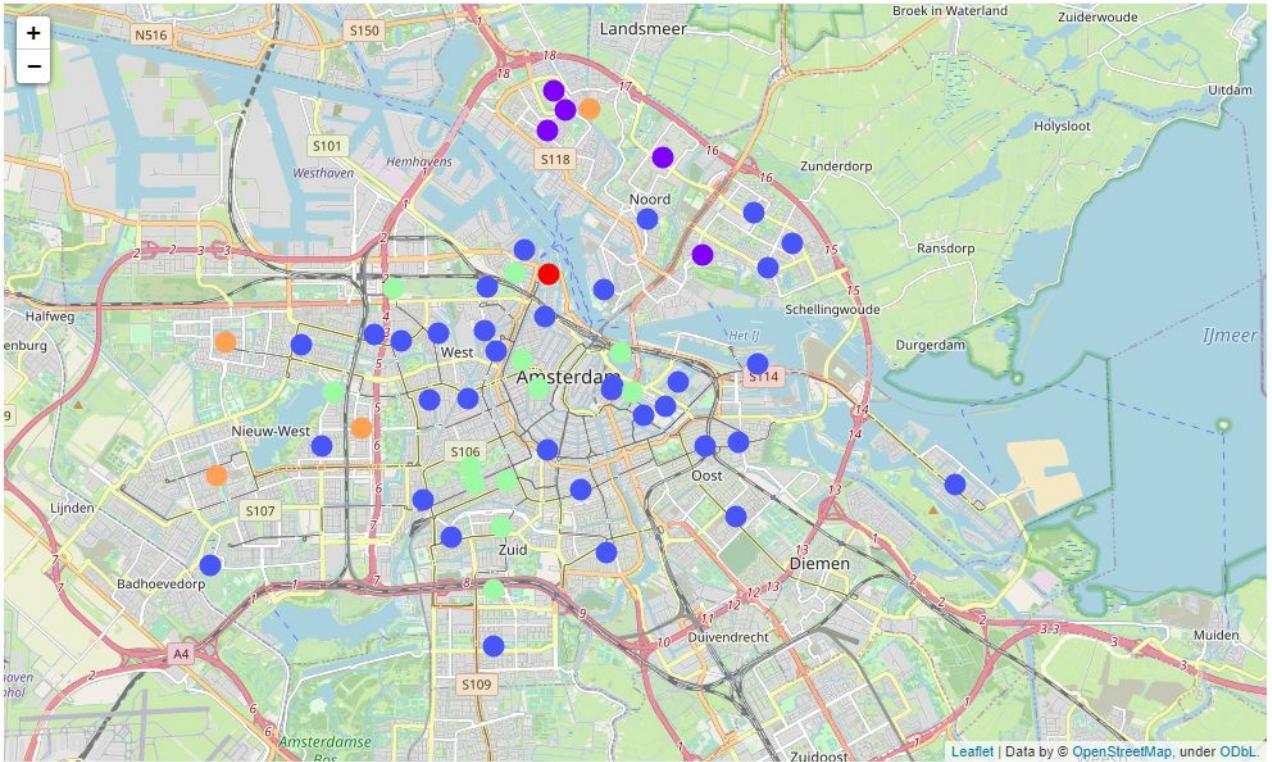
We can see that Berlin is a very diverse city and at the same time it has only 3 neighborhoods which are similar to my current location. Also Berlin has neighborhoods from each cluster which we received.

4.2 Amsterdam

Let's have a look closely at clusters in Amsterdam:

Cluster Labels	Neighborhood	
1.0	5	🟣
2.0	35	🔵
6.0	12	🟢
8.0	4	🟠
10.0	1	🔴

We can see that Amsterdam is less diverse than Berlin and at the same time it has a lot of neighborhoods (12) which are similar to my current location. Also Amsterdam doesn't have any neighborhoods from cluster 3.
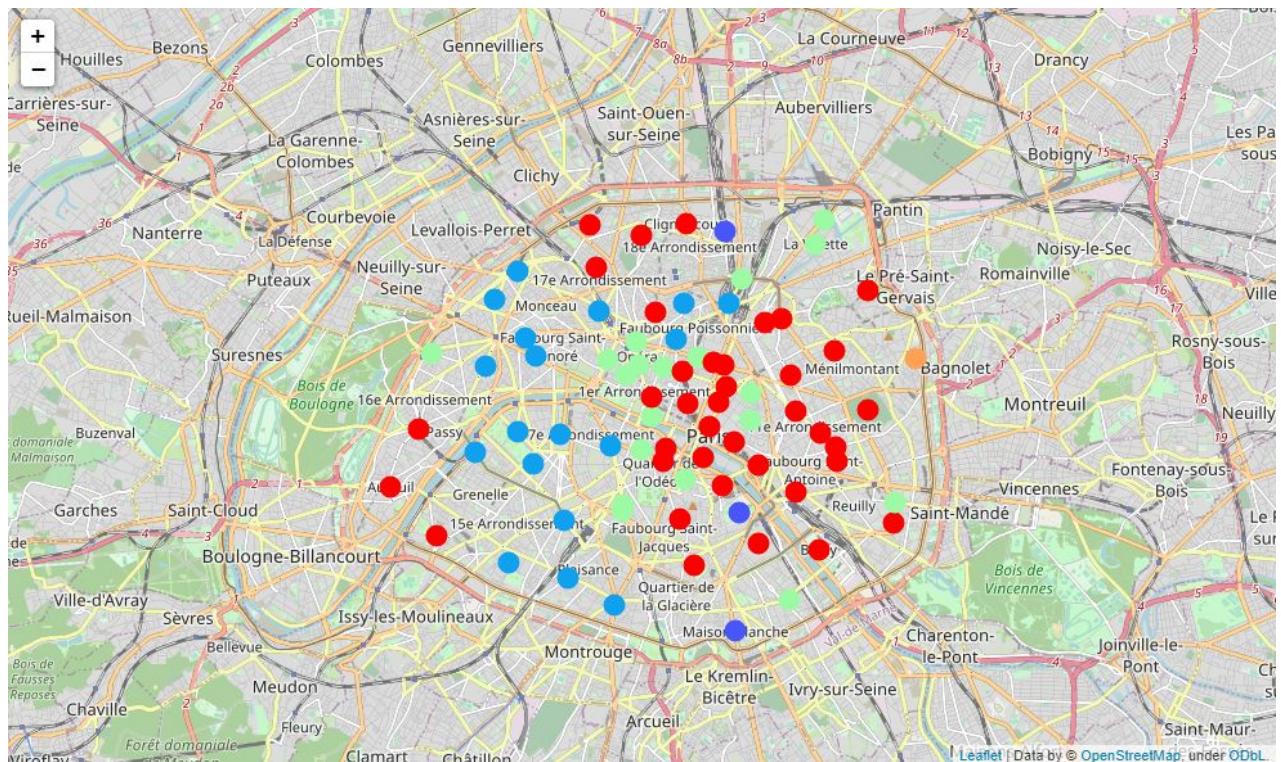
Let's look closely to cluster 3 (part of it):

|	Neighborhood	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
9	(0302) Weißensee, Berlin, Germany	Local Cuisine Restaurant	Hotel	Park	Beach	Vietnamese Restaurant	Café	Fishing Store	Falafel Restaurant	Farm	Farmers Market
183	Saint-Thomas-d'Aquin, Paris, France	Local Cuisine Restaurant	Hotel	Italian Restaurant	Art Gallery	Café	Bakery	American Restaurant	Cheese Shop	Bistro	Tailor Shop
184	Les Invalides, Paris, France	Hotel	Local Cuisine Restaurant	Italian Restaurant	Plaza	Café	History Museum	Cocktail Bar	Japanese Restaurant	Cheese Shop	Coffee Shop
185	École-Militaire, Paris, France	Local Cuisine Restaurant	Hotel	Plaza	Coffee Shop	Japanese Restaurant	Historic Site	Sculpture Garden	Brasserie	Supermarket	Bistro
186	Gros-Caillou, Paris, France	Local Cuisine Restaurant	Italian Restaurant	Hotel	Bistro	Ice Cream Shop	Coffee Shop	Café	Cheese Shop	Bakery	Garden
187	Champs-Élysées, Paris, France	Hotel	Local Cuisine Restaurant	Japanese Restaurant	Italian Restaurant	Tailor Shop	Clothing Store	Art Gallery	Boutique	Cosmetics Shop	Spa
188	Faubourg-du-Roule, Paris, France	Hotel	Local Cuisine Restaurant	Jewelry Store	Italian Restaurant	Japanese Restaurant	Electronics Store	Sporting Goods Shop	Café	Bakery	Scandinavian Restaurant
190	Europe, Paris, France	Local Cuisine Restaurant	Hotel	Thai Restaurant	Restaurant	Bistro	Asian Restaurant	Korean Restaurant	Mediterranean Restaurant	Sushi Restaurant	Sandwich Place
193	Faubourg-Montmartre, Paris, France	Local Cuisine Restaurant	Hotel	Italian Restaurant	Pizza Place	Cocktail Bar	Japanese Restaurant	Chinese Restaurant	Gym / Fitness Center	Burger Joint	Bakery

We can observe that these places are full of local restaurants, Hotels, Cafes, etc. and Amsterdam has a lack of such neighborhoods.

4.3 Paris

Let's have a look closely at clusters in Paris:

Neighborhood

Cluster Labels		
2.0	3	🔵
3.0	18	🔵
6.0	19	🟢
8.0	1	🟠
10.0	38	🔴

We can see that Paris is also less diverse than Berlin and at the same time it has a lot of neighborhoods (19) which are similar to my current location. Also Paris doesn't have any neighborhoods from cluster 1.

Let's look closely to cluster 1:

	Neighborhood	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
10	(0303) Blankenburg, Berlin, Germany	Bus Stop	Supermarket	Greek Restaurant	Café	Flower Shop	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
20	(0313) Wilhelmsruh, Berlin, Germany	Bus Stop	Post Office	Mexican Restaurant	Supermarket	Bakery	Cycle Studio	Dance Studio	Farmers Market	Fast Food Restaurant	Field
29	(0502) Haselhorst, Berlin, Germany	Park	Furniture / Home Store	Supermarket	Automotive Shop	Bus Stop	Metro Station	Flea Market	Falafel Restaurant	Farm	Farmers Market
31	(0504) Staaken, Berlin, Germany	Lake	Supermarket	Restaurant	Bus Stop	Flower Shop	Exhibit	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
32	(0505) Gatow, Berlin, Germany	Italian Restaurant	Supermarket	Trattoria/Osteria	Bus Stop	Harbor / Marina	Hotel	Fishing Store	Exhibit	Falafel Restaurant	Farm
36	(0509) Wilhelmstadt, Berlin, Germany	Bus Stop	Harbor / Marina	Lake	Supermarket	Boat or Ferry	Park	Sporting Goods Shop	Department Store	Fondue Restaurant	Farmers Market
56	(0902) Plänterwald, Berlin, Germany	Bus Stop	Supermarket	Light Rail Station	Flower Shop	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop
87	(1203) Konradshöhe, Berlin, Germany	Bus Stop	Italian Restaurant	Supermarket	Restaurant	River	Diner	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
108	Banne Buiksloot, Amsterdam, Netherlands	Bus Stop	Supermarket	Park	Bakery	Shopping Mall	Restaurant	Turkish Restaurant	Fish Market	Event Space	Exhibit
109	Buiksloot, Amsterdam, Netherlands	Bus Stop	Park	Health & Beauty Service	Campground	Rest Area	Ice Cream Shop	Fishing Store	Falafel Restaurant	Farm	Farmers Market
113	Molenwijk, Amsterdam, Netherlands	Pharmacy	Theater	Shopping Mall	Plaza	Bookstore	Supermarket	Bus Stop	Zoo Exhibit	Farm	Farmers Market
115	Oostzanerwerf, Amsterdam, Netherlands	Bus Stop	Supermarket	Furniture / Home Store	Shopping Mall	Bookstore	Zoo Exhibit	Flower Shop	Farm	Farmers Market	Fast Food Restaurant
118	Tuindorp Oostzaan, Amsterdam, Netherlands	Bus Stop	Plaza	Theater	Snack Place	Business Service	Motorcycle Shop	Baseball Field	Convenience Store	Food Service	Farm

These neighborhoods mostly have a lot of bus stops, farms and farmers markets and Paris has a lack of these places.

5. Conclusion

Personally, I'm searching for a city/neighborhood which is not similar to my current life situation. Because of it I don't choose Amsterdam and Paris as the cities with many neighborhoods similar to mine. I see that Berlin is the most interesting city for me with planty different neighborhoods.