

To which European cities can Africans go?

1. Introduction/Business Problem

Each year, a thousand of African executives leave the continent to settle elsewhere and particularly in Europe. This migration of the African elite has been amplified in the last decade due the technological boom which has led to a very high demand for skilled labor, particularly in the fields related to digital (IT, data science, artificial intelligence, etc.). and health services (doctors, nurses, etc.).

Is this a good thing or a bad thing for Africa? two points of view are generally cited in relation to this phenomenon.

The first, which is the most known and the most widespread, considers this migration as a loss for Africa and an obstacle of the development of African countries. The main argument cited at this level is that these elites were well trained in the continent with taxpayers' money and that Europe recovers these ready-made skills, without any investment or consideration, to develop its economic fabric and advance its scientific and technological ecosystem. Africa then loses the return on investment made in the training of these executives which costs it quite expensive. Therefore, this only increases the gap in development and standard of living between the continent and the other continents receiving these elites.

The second point of view, considers this migration as a positive element for Africa since these executives will further develop their skills in the host countries and may return thereafter to their original countries to invest and build up there some projects they cannot carry out if they stay on the continent. They support their position by the fact that statistics show that most of these emigrants participate in the development of their countries from their resident countries by sending money to their families and helping to improve the standard of living of their relatives. In addition, these persons can constitute a vector of development and influence of African countries in various fields as is the case of India in the field of IT where most of the IT giants' CEO are originally from India.

So, far from the controversy on this question of migration of the African elite towards Western countries and particularly towards Europe, the objective of this study is to examine the cities where would be more judicious for an African that is preparing to migrate to go there.

Migration generally takes place in large cities which provide more employment opportunities and better conditions in terms of standard of living and social mix and particularly in terms of the acceptance of foreigners.

For this, the choice of cities to study was determined using an indicator combining the population of the city, the standard of living and the disposition to receive emigrants. Also, because most people migrate to countries that have occupied their countries for the language reason. This criterion was added to adjust the score of cities. A 10% bonus has been allocated to the colonizing countries. 100 candidate cities were then selected and they were then classified into categories using the clustering algorithms based on their socio-economic properties and what they also offer as infrastructure extracted using Foursquare location data.

2. Data

To answer this question, we used a series of data combining several themes. These data are related to the following elements:

- The population of European cities

The list of cities with their populations in 2020 was extracted from the website:

<https://worldpopulationreview.com/continents/cities-in-europe/>

This site lists the 500 most populous cities in Europe. The population of these cities varies between 10,381,222 for Moscow in Russia and 156,238 for the city Ruse in Bulgaria.

These data are provided by the website in CSV or Json format.

- Standard of living

The standard of living was determined at the country level from the Gross Domestic Product (GDP) of the countries per capita. These data estimated for 2019 from the International Monetary Fund (IMF), using GDP based on purchasing power parity (PPP) per capita, are extracted from Wikipedia using the bellow link:

[https://en.wikipedia.org/wiki/List_of_sovereign_states_in_Europe_by_GDP_\(PPP\)_per_capita](https://en.wikipedia.org/wiki/List_of_sovereign_states_in_Europe_by_GDP_(PPP)_per_capita)

The data of this website was extracted directly by python program to a pandas dataframe.

- willingness to receive immigrants

The willingness of cities to receive immigrants are estimated based on the statistics of immigrant population in European countries calculated based on the United Nations report Trends in International Migrant Stock for the year of 2013 ([International Migration and Development](#)". *Esa.un.org*. Retrieved 13 May 2016).

These data were retrieved from the following website:

https://en.wikipedia.org/wiki/Immigration_to_Europe#2013_data

- Infrastructure of candidate cities

The infrastructure of the selected cities is extracted using the Foursquare data location. A set of venue categories were chosen to form the features on the basis of which cities will be compared.

The venue categories considered regarding our subject are made up of the following items:

Table 1: List of venue categories used

Category	Subcategory	Category ID
College & University	University	4bf58dd8d48988d1ae941735
Outdoors & Recreation	Gym / Fitness Center	4bf58dd8d48988d175941735
Professional & Other Places	Business Center	56aa371be4b08b9a8d573517
	Cultural Center	52e81612bc57f1066b7a32
Government Building	Embassy / Consulate	4bf58dd8d48988d12c951735
Medical Center	Hospital	4bf58dd8d48988d196941735
Shop & Service	Bank	4bf58dd8d48988d10a951735
Travel & Transport	Train Station	4bf58dd8d48988d129951735

3. Methodology

a. Selecting the candidate cities

We started our analysis using the list of the 500 most populated cities in Europe. We added to this criteria (population), the information about the living standard and the willingness to receive immigrants calculated at the country level. Thereafter, were normalized these three attributes using the MinMaxScaler to give each city a value comprising between 0 and 1 for each criterion. Thereafter, a 10% bonus has been allocated to the colonizing countries of Africa (France, England, Spain, Italy and Belgium) because African people migrate usually to countries that have occupied their countries especially for the language reason.

The city's score was done calculated using the arithmetic mean of these four criteria. Based on this score, 100 highest ranked cities were then selected to form the list of candidate cities.

The table 2 gives the distribution of candidate cities by country.

Table 2: Number of candidate cities by country

country	Austria	Belgium	Cyprus	Denmark	France	Germany	Ireland	Italy	Netherlands	Norway	Russia	Spain	Sweden	Switzerland	United Kingdom
city	6	2	1	1	18	19	3	1	10	2	1	2	3	3	28

The location of the candidate cities is shown in the next figure. These cities went to be studied further in the next section by adding what they offer as infrastructure which will be extracted from the Foursquare location data.

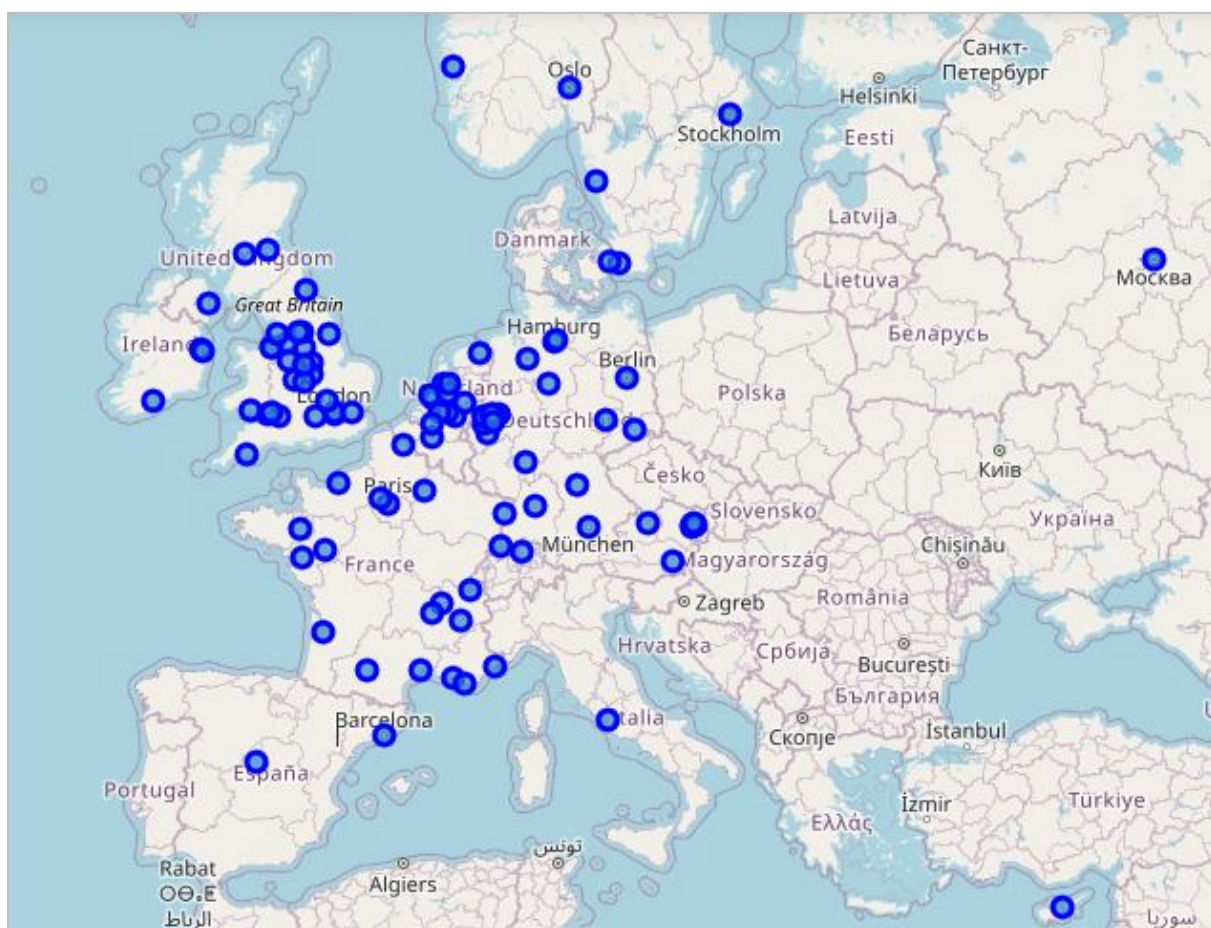


Figure 1: Map of candidate cities

b. Using the Foursquare API to get city's venues

To compare the candidate cities, two set of criteria were used:

- The socioeconomic characteristics that are already calculated in the previous section (population, living standard and willingness to receive immigrants).
- The infrastructure properties of the city that were extracted from the Foursquare location data. 8 main features (venue categories) were selected for this target to characterize the attractiveness of each city in terms of education, health, leisure, business, sports, culture and consular services. These items are presented in table 1 above.

For each city, we calculated the number of the considered venue categories that were retrieved from Foursquare AP using the next function.

```

1  #function to get the number of venues for each city
2  dv=dict()
3  for city_adr in cities_adr:
4      city=city_adr.split(',') [0]
5      dv[city]=[]
6      for category in categ.keys():|
7          dv[city].append((count_categ_city(city,categ[category])))

```

The number of venue categories retrieved for some cities is illustrated in the table 3.

Table 3: number of venue categories for some European Cities

	University	Hospital	Business Center	Cultural Center	Embassy Consulate	Bank	Train Station	Gym Center
Zurich	32	33	4	10	11	66	79	88
Geneve	17	17	11	5	5	58	18	44
Basel	9	12	11	8	8	44	26	26
London	147	122	45	18	97	185	257	217
Dublin	54	15	10	1	13	46	12	79

The two sets of characteristics were merged in one table (dataframe) for the next step.

c. Applying clustering algorithm to candidate cities

Based on the features presented in the previous section, we conducted a clustering analysis based on machine learning algorithm called K-Means. The aim was to divide cities into homogeneous classes which will be proposed to African applicants for immigration as the set of similar cities from which they can choose the city where to go.

Thus, using the K-Means algorithm with k=5, Five classes of homogeneous cities were generated.

4. Results and discussion

The main result of this study is the clustering of the candidate cities into 5 classes of similar cities that have common properties. The figure 2 illustrates the map of resulting clusters.

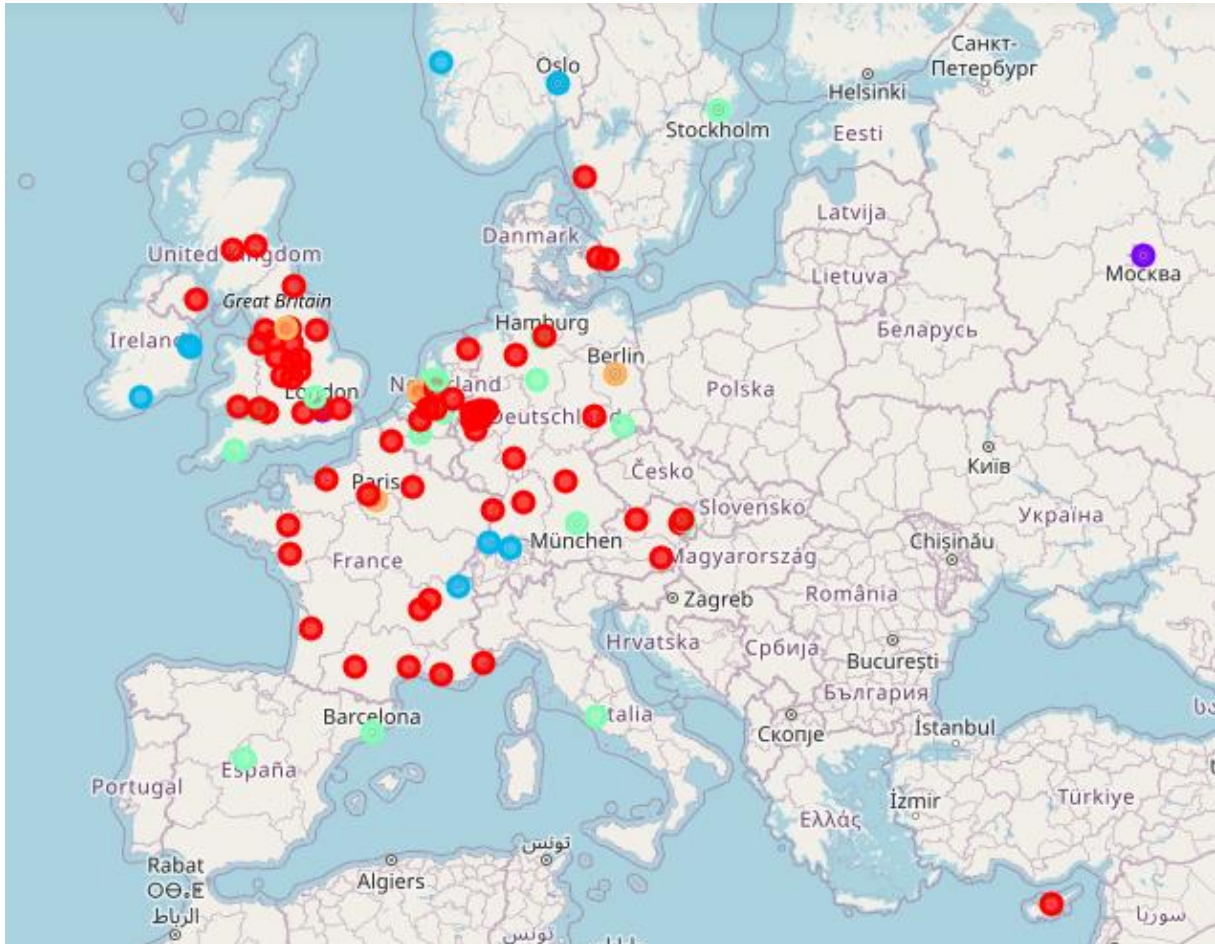


Figure 2: Map of resulting clusters

The First cluster contains 2 cities London and Moscow:

	city	country
3	London	United Kingdom
5	Moscow	Russia

This cluster is made up of mega-cities characterized by a large population and the concentration of business and political establishments

The second cluster is composed of 4 cities (Paris, Berlin, Bradford and The Hague):

	city	country
10	Berlin	Germany
13	Paris	France
29	The Hague	Netherlands
86	Bradford	United Kingdom

This cluster consists mainly of the main European capitals known for their very strong economic activity.

The Third cluster contains the following 8 cities shown in the table below.

	city	country
0	Zurich	Switzerland
1	Geneve	Switzerland
2	Basel	Switzerland
4	Dublin	Ireland
6	Cork	Ireland
7	Dun Laoghaire	Ireland
8	Oslo	Norway
9	Bergen	Norway

This cluster is made up of Swiss, Irish and Norwegian cities characterized by a very high standard of living and good quality equipment.

The fourth cluster contains the following 19 cities:

	city	country			
	11	Vienna	Austria	32	Liverpool United Kingdom
	12	Stockholm	Sweden	36	Eindhoven Netherlands
	14	Madrid	Spain	40	Almere Stad Netherlands
	16	Hamburg	Germany	47	Essen Germany
	21	Donaustadt	Austria	53	Barcelona Spain
	23	Munich	Germany	56	Hannover Germany
	24	Amsterdam	Netherlands	60	Cardiff United Kingdom
	27	Rome	Italy	62	Dresden Germany
	28	Brussels	Belgium	94	Plymouth United Kingdom
				95	Luton United Kingdom

This cluster consists of the cities of western Europe with a dominance of the German and English cities as well as capitals of the other European countries.

The fifth cluster contains 64 cities. It is made up of medium-sized cities characterized by a medium population and activity. The list of these cities are:

['Goeteborg', 'Malmoe', 'Graz', 'Linz', 'Favoriten', 'Floridsdorf', 'Rotterdam', 'Birmingham', 'Koeln', 'Marseille', 'Nottingham', 'Utrecht', 'Sheffield', 'Tilburg', 'Groningen', 'Bristol', 'Breda', 'Nijmegen', 'Frankfurt am Main', 'Glasgow', 'Nicosia', 'Lyon', 'Stuttgart', 'Dortmund', 'Duesseldorf', 'Toulouse', 'Leicester', 'Bremen', 'Edinburgh', 'Leeds', 'Leipzig', 'Duisburg', 'Nuernberg', 'Nice', 'Manchester', 'Stoke-on-Trent', 'Antwerpen', 'Coventry', 'Nantes', 'Wandsbek', 'Strasbourg', 'Sunderland', 'Bochum', 'Montpellier', 'Birkenhead', 'Bochum-Hordel', 'Islington', 'Reading', 'Copenhagen', 'Kingston upon Hull', 'Preston', 'Bordeaux', 'Lille', 'Newport', 'Wuppertal', 'Swansea', 'Southend-on-Sea', 'Rennes', 'Reims', 'Belfast', 'Derby', 'Le Havre', 'Cergy-Pontoise', 'Saint-Etienne']

5. Conclusion

The application of the K-Means algorithm on European cities shows that there is a clear difference between these cities in terms of their attractiveness to immigrants. by referring to the criteria used in this study, two cities in particular stand out for their megalopolis, namely London and Moscow. they are followed by the other European capitals in a second level, namely: Berlin and Paris.

Nordic and Swiss cities are also included in a single cluster, which shows their common characteristics both in terms of standard of living and their facilities.

Finally, the medium-sized cities characterized by a medium population and activity are regrouped to constitute a single cluster.

This work can be further developed by adding other evaluation criteria (venues categories) and may well be extended to other geographical contexts.