2. Data

To answer this question, we will use a series of data combining several themes. These data relate 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

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

These three criteria were normalized and then combined using the arithmetic mean to calculate an index for each city. From this index, we selected the 100 highest ranked cities to study them further by adding to the study what they offer as infrastructure which will be extracted from the Foursquare location data.

- Infrastructure of candidate cities

The infrastructure of the selected cities is extracted using data provided by the Foursquare API.

The main data considered regarding our subject are made up of the following categories of venues :

Category	Subcategory	Sub Subcategory
Arts & Entertainment	Stadium	

College & University	General College & University	
Outdoors & Recreation	Athletics & Sports	Gym / Fitness Center
Garden		
Professional & Other Places	Business Center Cultural Center	
Government Building	Embassy / Consulate	
Medical Center	Hospital Maternity Clinic	
Office	Recruiting Agency Tech Startup	
Social Club		
Shop & Service	Bank	
Travel & Transport	Airport Train Station	

For each city, we calculated the number of the considered venue's categories that were retrieved from foursquare location data. therefore, we add these data to the three initial city's characteristics (population, standard of living and willingness to receive immigrants) to constitute the attributes of each city.

Based on these features, we conducted a clustering analysis based on K-Means algorithm to divide cities into homogeneous categories which will be proposed to African applicants for immigration as the most suitable candidate cities where to go.