

The Battle of Neighborhoods in Europe: Choosing A City For Relocation

By Sergei Perfilyev
June 2021

INTRODUCTION

This project aims to solve the following problem:

A hypothetical family from Russia is planning to relocate to some place in the European Union (EU). They are aiming at broadening their career opportunities (as for the parents) and, most importantly, getting access to high-quality education for the children, so that the kids could study at top ranked European universities. In addition, to feel less uncomfortable while adjusting to living abroad and getting accustomed to foreign culture as well as potential changes of work/life balance, they would prefer to live in a city that resembles their hometown.

The family currently resides in Novosibirsk, which is the third-largest city in Russia with a population of over 1.5 million people. The city is one of the major business and transport hubs. A substantial number of large companies' headquarters are located here as well as an international airport, a river port and a railway station. Novosibirsk has many parks, open spaces, and squares throughout the city. Attractions and entertainment facilities include museums, theaters, concert halls, a water park and several large movie theaters. Novosibirsk is known for its shopping malls, numerous restaurants, bars and coffee shops. So, our friends are willing to know which of European cities tend to be similar to their own hometown in everyday life.

Certainly, it would be a very complex task for them to gather all the information manually and perform such a comparison for each city. Therefore, we will apply data science techniques to analyze data on European cities and help our friends compare all available options in order to choose the most appropriate and comfortable city for their relocation.

DATA

We intend to use the following data in order to implement this project as well as for learning purpose:

1. Eurostat open data website <https://ec.europa.eu/eurostat>

Eurostat is the statistical office of the EU whose mission is to provide high quality statistics and data on Europe. For example, we can obtain current, up-to-date information on member countries of the EU from this page:

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Country_codes

Member States of the [European Union \(EU\)](#) and other countries have been assigned a two-letter **country code**, always written in capital letters, and often used as an abbreviation in statistical analyses, tables, figures or maps.

The **protocol order** in which countries are often listed is based on the alphabetical list of countries in their national language for EU and [EFTA](#) Member States and for [candidate countries](#); for [potential candidates](#), it is based on the alphabetical order of their country code.

EU Member States come first, followed by [European Free Trade Association \(EFTA\)](#) Member States, candidate countries for EU membership, potential candidates and, finally, other countries. The order in the tables below is first column down, then second column down, etc..

European Union (EU)

Belgium	(BE)	Greece	(EL)	Lithuania	(LT)	Portugal	(PT)
Bulgaria	(BG)	Spain	(ES)	Luxembourg	(LU)	Romania	(RO)
Czechia	(CZ)	France	(FR)	Hungary	(HU)	Slovenia	(SI)
Denmark	(DK)	Croatia	(HR)	Malta	(MT)	Slovakia	(SK)
Germany	(DE)	Italy	(IT)	Netherlands	(NL)	Finland	(FI)
Estonia	(EE)	Cyprus	(CY)	Austria	(AT)	Sweden	(SE)
Ireland	(IE)	Latvia	(LV)	Poland	(PL)		

2. Times Higher Education website <https://www.timeshighereducation.com>

This site contains a lot of statistics, articles and other useful data on higher education in the world including various university rankings. In particular, we will be using the 'Best Universities in Europe' ranking:

<https://www.timeshighereducation.com/student/best-universities/best-universities-europe>

The page contains summary in a tabular format, and hyperlinks to more detailed page for each institution, where we can find and parse all information necessary for the project, including location data with complete address of the university. Below is an excerpt from the webpage:

Best universities in Europe 2021: the results in full

Click on each institution to view its [World University Rankings 2021](#) result

World University Rank 2021	Europe Rank 2021	University	Country
1	1	University of Oxford	United Kingdom
6	2	University of Cambridge	United Kingdom
11	3	Imperial College London	United Kingdom
14	4	ETH Zurich	Switzerland
16	5	UCL	United Kingdom
27	6	London School of Economics and Political Science	United Kingdom
30	7	University of Edinburgh	United Kingdom
32	8	LMU Munich	Germany
35	9	King's College London	United Kingdom
=36	10	Karolinska Institute	Sweden

3. Wikipedia

In order to analyze cities in the EU countries, we will parse the Wikipedia page `List of cities in the European Union by population within city limits`:

https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_population_within_city_limits

The tabular data from the webpage will help us create a list of potential places to relocate, and then rank them by population and link to the previously obtained data on educational institutions.

Cities by population within the city boundary [edit]

Cities in bold are capital cities of their respective countries.

Rank ↕	City ↕	Member State ↕	Official population ↕	Date of census ↕	Reference	Photography
1	Berlin	Germany	3,669,495	31 December 2019	[1]	
2	Madrid	Spain	3,348,536	1 February 2020	[2]	
3	Rome	Italy	2,856,133	31 December 2018	[3]	
4	Bucharest	Romania	2,155,240	1 July 2020	[4]	

4. Foursquare API

We have previously created a developer account on <https://developer.foursquare.com> and will be using the API to collect data on recommended venues in the cities of interest. Then we will analyze distribution of the venues in order to cluster the cities into groups and find the most suitable ones for relocation.

5. HERE.com API

In order to obtain geographical coordinates of various objects (cities, venues, etc.) we have created a developer account on <https://developer.here.com> and received credentials for using HERE Location Services REST APIs.

In particular, we will be using HERE Geocoding service to get latitude/longitude coordinates, since this service has proved to be robust and accurate (in contrast to OSM). We will use it from Python via GeoPy library.