IBM Data Science

Applied Data Science Capstone

**Optimum relocation**

**Xavier Martínez**

January 2020

**TABLE OF CONTENT**

[1. Introduction - 2 -](#_Toc29767972)

[1.1 Background - 2 -](#_Toc29767973)

[1.2 Objective - 2 -](#_Toc29767974)

[2. Data - 3 -](#_Toc29767975)

[2.1 Data sources - 3 -](#_Toc29767976)

[2.1.1 Foursquare data - 3 -](#_Toc29767977)

[2.1.2 Numbeo data - 3 -](#_Toc29767978)

[2.1.3 Google maps data - 3 -](#_Toc29767979)

[2.1.4 Folium library - 3 -](#_Toc29767980)

[2.2 Data preparation - 4 -](#_Toc29767981)

[2.2.1 Foursquare data - 4 -](#_Toc29767982)

[2.2.2 Numbeo data - 4 -](#_Toc29767983)

[2.3 Example implemented - 4 -](#_Toc29767984)

[3. Methodology - 6 -](#_Toc29767985)

[4. Results - 7 -](#_Toc29767986)

[5. Discussion - 8 -](#_Toc29767987)

[6. Conclusions - 9 -](#_Toc29767988)

[7. Future directions - 10 -](#_Toc29767989)

Appendix 1.

# Introduction

## Background

Several people nowadays relocate for different reasons. While undertaking a quick research to try to put a figure to the scale of this phenomenon, I found on a piece of information that, even it only relates to the USA, is quite indicative:

*“Each year, roughly 40 million Americans, or about 14 percent of the U.S. population, move at least once”.* From the article *Population migration patterns: US cities we are flocking to* by Michael B. Sauter, published Oct 4, 2018.

The truth is that in many occasions the person relocating may not know the city in which he will be moving into; for example, because the relocation relates to the needs of the company where this person works for.

It could be the case that a person wants or needs to move, but that the neighbourhood or even the city, where this person would relocate is not determined.

The circumstances for relocating may be very diverse, but an element that probably is common in any case is that every individual relocating will have an idea of what kind of place it should be.

An easy way to describe the neighbourhood where one would like to relocate is by referring to a known one. This known neighbourhood could be the one where you live or the one where you like to live.

## Objective

The objective of this project is to develop a tool that would allow identifying the area or neighbourhood in a city or a group of cities that best matches the requirements of an individual that will be relocating.

The requirements would be expressed as an area or neighbourhood that the person relocating would indicate.

The criteria to be considered in comparing areas would include the presence of a wide range of venues such as social, cultural, entertainment, schools or natural environment; and other factors such as quality of life, cost of living, safety, health care, climate and pollution.

# Data

As indicated in the previous section, the criteria in comparing areas would include the presence of a wide range of venues as well as other factors such as quality of life, cost of living, safety, health care, climate and pollution. Therefore data related to all these factors will be required.

Also, along the process of finding venues, the conversion from an address to latitude and longitude will be needed.

And to present the locations considered, mapping them will be the best option. Therefore, map information will be required as well.

## Data sources

### Foursquare data

The data related to venues will be obtained from Foursquare (), using their API. The data obtained will be for all, the neighbourhood that would represent the requirements of the person who relocates, and all the areas that will be compared to the former.

### Numbeo data

The data related to the quality of life, cost of living, safety, health care, climate and pollution, will be obtained from Numbeo (), using their API. Again, the data obtained will be for the neighbourhood that represents the requirements of the person who relocates and all the areas that will be compared to the former.

It needs to be noted here that the granularity of the information from Numbeo is different from the granularity provided by Foursquare. Numbeo provides information representative of the average of the city.

### Google maps data

The data related to the transformation from address to geographical coordinates will be gathered from Google Maps API geocoding.

### Folium library

Folium will be the python library used in generating the maps that will show the areas considered.

## Data preparation

Once the data is obtained from the different sources, it will be checked for adequacy and completeness.

There are some particular considerations for the Foursquare and Numbeo data exposed below.

### Foursquare data

In some occasions, the data obtained from Foursquare may present a very large number of categories and subcategories. And given that different countries will be compared, it may happen that similar venue concepts are categorised differently (i.e. a Pub in the UK versus a Bar in Spain). The avoid this potential issue the information obtained for all the areas to be compared will be analysed and grouped in meaningful categories, such as restaurants, schools, natural environment, health centres, or sports.

In other occasions, the amount of information received may the scarce. That could lead to biased categorisation. In those situations, the particular area will be removed from the potential candidates.

### Numbeo data

The data that can be obtained in Numbeo for some of the fields, such as the cost of living, can be either in absolute terms (i.e. cost of living expressed in monthly GPB, USD or EUR), or indexes. To facilitate the comparison between different cities, the use of indices is more convenient. The indices that will be considered are:

* Quality of life
* Cost of living
* Safety
* Health care
* Climate
* Pollution

## Example implemented

The code developed will be presenting a particular case. And this case will be:

* Area to take as requirements: Richmond, Surrey, United Kingdom
* Cities where relocation will be considered:
  + Madrid, Spain
  + Frankfurt, Germany
  + Singapore, Singapore
  + New York, USA

# Methodology

# Results

# Discussion

# Conclusions

# Future directions

Appendix 1.