**EU member states capitals comparison**

Victor Silva

January 11, 2020

**1. Introduction**

**1.1 Background**

For someone who have EU citizenship it is easy to move between all of member states, but with 28 countries to choose one can be confused on where to live or visit. The number of options is so high that, if we consider locations with more than 300,000 inhabitants, one would have to choose between 120 alternatives, according to the list of cities in the European Union by population within city limits on Wikipedia, and if all villages, towns and cities were considered this number surpasses thousands of places.

**1.2 Problem**

Data might help those who want to move to other countries to choose where to live, based on what venues each city and neighbourhood has to offer. This project aims to group similar capitals of EU countries into clusters and to compare similar neighbourhoods on those clusters to help people who want to live abroad or just meet places similar of those they love.

**1.3 Interest**

This project will be helpful for anyone who wants to move to an EU member state or want to know places there but don’t know where to start.

**2. Data acquisition and cleaning**

**2.1 Data sources**

To achieve our objective the model will consider the type of venues existing on each neighbourhood of each place, which will be purchased using Foursquare API, and the cost of living on those cities, based on NUMBEO Cost of Living Index, which is explained below:

*These indices are relative to New York City (NYC). Which means that for New York City, each index should be 100(%). If another city has, for example, rent index of 120, it means that on an average in that city rents are 20% more expensive than in New York City. If a city has rent index of 70, that means on an average in that city rents are 30% less expensive than in New York City. (NUMBEO)*

Additionally, a CSV dataset was created manually to map all neighbourhoods of the UE members capitals, their names, number of inhabitants and areas.

Those data will be used in order to cluster cities and, later, neighbourhoods inside each city cluster.

**2.2 Data cleaning**

**2.3 Feature selection**

**3. Exploratory analysis**

**4. Modelling**

**5. Conclusions**

**6.Future directions**

**7. References**

ABOUT COST OF LIVING INDICES AT THIS SITE. NUMBEO. Source: <https://www.numbeo.com/cost-of-living/cpi\_explained.jsp>. Accessed January 11, 2020.

CURRENT COST OF LIVING INDEX. Source: <https://www.numbeo.com/cost-of-living/rankings\_current.jsp>. Accessed January 11, 2020.