**Background**

Europe with its Eurozone is a perfect place for people who likes work travelling. My friend, Bart, currently is living in Warsaw, Poland and looking to move to some other city in Europe that is as comfortable as Warsaw but with higher income. I think Bart is not alone with such idea in mind, hence categorizing European cities by similarities is great topic that I might use myself as well.

**Idea**

The goal here is to build a clustering model to group European cities based on certain parameters. Leveraging Foursquare API I gathered top cities venues, because variety of the restaurants and places you can go is one of the important factors feeling comfortable in the city. Also, the important factor is weather, even though there is plenty of restaurants out there if its showring outdoor. I know Bart likes the weather in Warsaw, hence I gathered some main weather statistics for each city. And of course, Bart is looking for city with higher income, GDP per capita as one of the main indicators of higher salaries is important indicator.

**Data**

Data for the research comes from various places:

- European data bank: https://ec.europa.eu/eurostat - provided population and GDP figures
- Geolocation will mainly be pulled from Foursquare (better precision) and from geopy library
- City venues – Foursquare
- Annual weather report – to be scraped from https://weatherspark.com

**Data Cleaning**

The data for GDP and population was not of the best quality, so I gathered data from 2017-2019 year and picked 2019 year filling the gaps for the cities from previous years. This allowed me cover more cities, but even though more then half was skipped due to lack of data (600 missing out of 800), so I arrived at 200 European cities for analysis.

To my regret the “geopy” library did not managed to recognize some of the cities correctly and provided a bunch of false coordinates. Thus, I used a Foursquare API and **venues** endpoint and option “nearby” to get coordinates for cities – worked well!

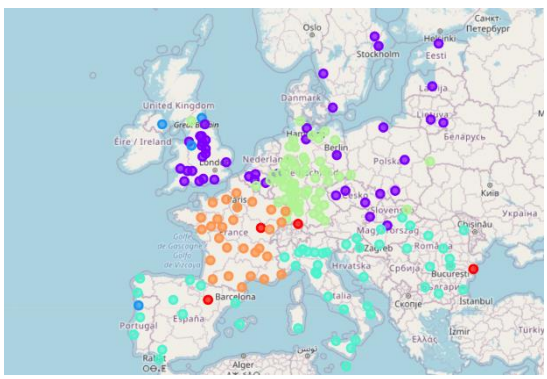
The weather report was scraped from weatherspark.com. The website has a standardized report generated for location; hence it was easy to build a simple scraper to gather required details: temperature, cloudiness, and daylight statistics.

**Exploratory Data Analysis**

After completion of data collection, categorized data was encoded. Venues end up having more than 500 different categories. Weather report produced 10+ variable on temperature, cloudiness, and daylight stats. GDP per Capita was derived from cities GDP and population.

**Clustering**

After testing different number of clusters, I end up with 6 of them. And from the map view I would say that the clustering has weather statistics and GDP per capita. Some groups are located almost strictly on the territory by certain countries. However, I think venues categories specific to region might played a big role.




Nonetheless, a great result. More then 200 cities categorized. Choose your criteria and pack your bags. Do you want more sun or higher salary?!

To conclude, I think I managed to complete the task and I will be able to suggest Bart to move to one of these cities: Manchester, Cambridge, Oxford, Exeter, Ipswich, Malmö, Stockholm, Göteborg, Düsseldorf, Leicester, Bruxelles, Gent, Coventry, Antwerpen, Middlesbrough, Bristol, Portsmouth, Kiel, Hamburg, Cardiff, Leeds, Uppsala, Southampton, Colchester, Ostrava, Doncaster, Bratislava, Swansea, Plzen, Liège, Praha, Brno, Berlin