A Foursquared European Trip

1 Business Understanding

"What is the problem that you are trying to solve?"

Problem: There are many people who keep saying that the inner cities in Europe basically all look the same: a big square with a cathedral, nearby mostly the same types of restaurants, bars, shops – which mostly belong to a bigger chain; a castle, a big river, the cool trendy quarter, the not so cool run-down quarter, the train station. Thus, a european trip can be boring for *tourists*.

Idea: Why not using open data to avoid boring holidays? Why not using e.g. *Foursquare data* to determine a trip with the "most diverse" cities or city centers in Europe?

The goals are:
- To **compare** about 70 **cities** or their city centers in Italy, France, Spain, Portugal
- To **select** 3 italian, 3 french, 3 spanish and 1 portugese **cities** as such so that they are as
- **diverse** as possible.

The results of this analysis might be useful for tourists planning to visit European cities.

2 Data

"How can you use data to answer the question? What data do you need to answer the question?"

To conquer this problem we need:
- A **list of cities** from Italy (20 cities), France (20), Spain (20), Portugal (10). Please don't feel discriminated right now if you are Portuguese. The list of cities could e.g. be retrieved from **Wikipedia**
  - City | Country
- The **coordinates** of each city (which should also roughly be the city center) based on **geopy** / **Nomatim** data
  - Latitude Longitude
- Access to **Foursquare** to determine venues of interest
- For visualization purposes we need the **Folium library**

The data will be processed using **Python** in a Jupyter environment.

A **clustering** method (most likely k-means) will be used to compare the cities. The algorithm shall be optimized by using different evaluation methods.

3 Methodology

3.1   Data Collection

"Where is the data coming from (identify all sources) and how will you get it?"

70 cities were selected for data analysis:
- The 21 biggest cities of Italy
- The 21 biggest cities of Spain
- The 21 biggest cities of France
- The 7 biggest cities of Portugal

For each city coordinates were retrieved by using the Nominatim library from geopy.geocoders. City, country, latitude and longitude were stored in a dataframe.

The cities of interest were displayed on a map which was centered around the centroid of all coordinates:

3.2   Data Understanding and Data Preparation

"Is the data that you collected representative of the problem to be solved? What additional work is required to manipulate and work with the data?"

Get venues for each city center.
Venues for each city were retrieved from Foursquare. A radius of 2000 around the city center was defined, the amount of venues were limited to 100 results.
A function was defined to automatically get the venues of all the cities. It includes creating a list and basically requests venues by using a Foursquare API.
The amount of venues was calculated for each city.

Analyze each city.
A dataframe was created which displays the top 10 venues for each city. To do so the following main methods were applied:
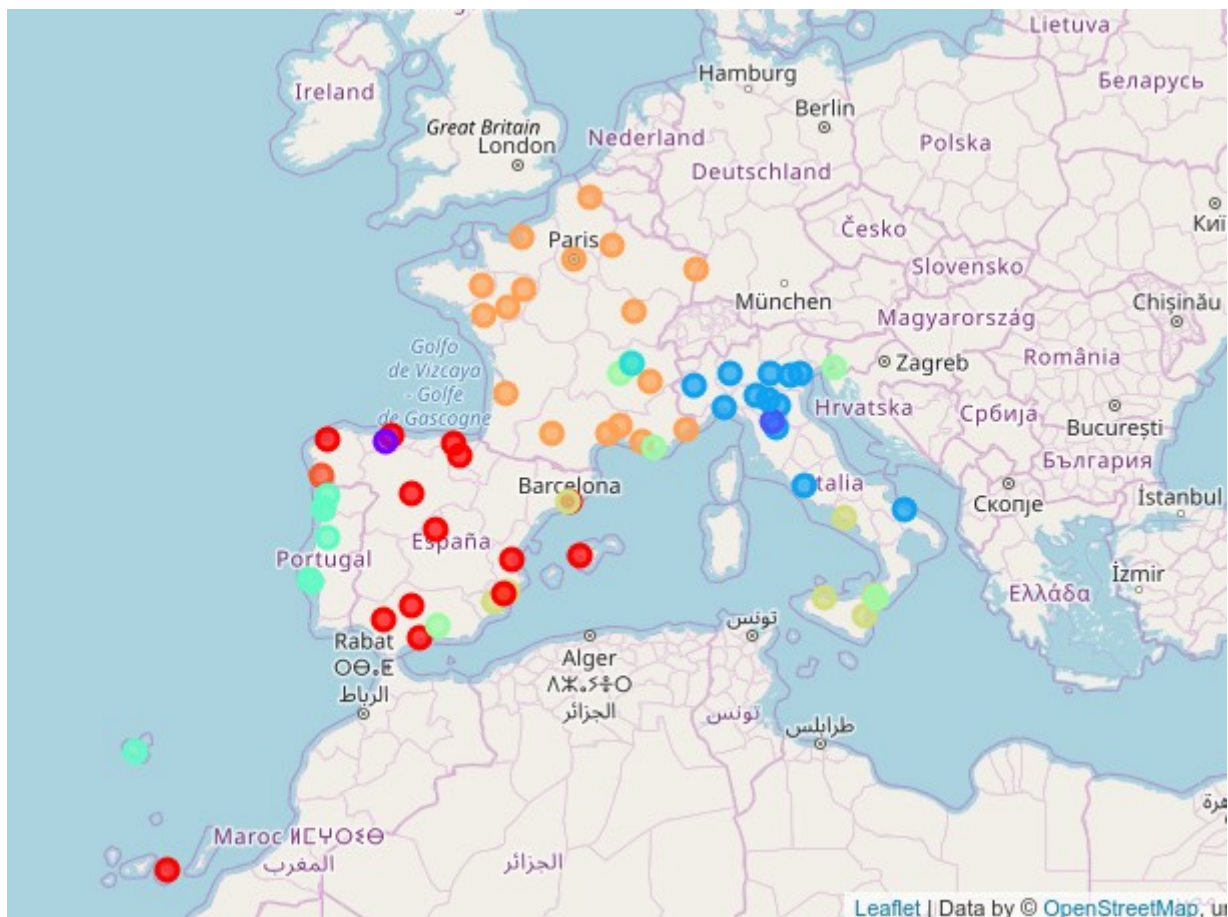- One hot encoding for the venue types
- Grouping the venues by city

3.3   Modeling

Using the dataframe **k-means clustering** was performed:
- Numbers of clusters = 10
- Random state = 0

The calculated clustering labels were added to the dataframe. The clustered cities were displayed on a map:

4 Results

4.1 Calculated Clusters

From the 10 clusters **4 clusters (1,3,5,9) were excluded** for city selection due to the following reasons:
- they only included 1 city each (Oviedo/Spain, Prato/Italy, Villeurbane/France, Vigo/Spain)
- to each city less than or equal to 5 venues were retrieved.

Cluster 0 contains most of the spanish cities. Most of their city centers have at least one Spanish or Tapas Restaurant, a Plaza and some bars in the city center.
Cluster 3 contains most of the italian cities. Most of their city centers have at least an Italian Restaurant, a Pizza Place, an Icecream Shop, a Plaza and a hotel.
Cluster 5 contains all portoguese cities. Most of their city centers have at least a Portoguese restaurant, a plaaza and a bar.
Cluster 8 contains most of the french cities. Most of their city centers have at least a French Restaurant, bars, a plaza and some cultural sites like museums or art gallery.

Cluster 6 contains similar french, italian and spanish cities, with a central Plaza, a Hotel and cafés.
Cluster 7 contains similar italian and spanish cities, with pizza places, some cultural sites and bars.

4.2 Selected Cities

For **Italy** the following cities were selected:
- *Rome* (first entry in cluster 3)
- *Trieste* (first italian entry in cluster 6)
- *Naples* (first italian entry in cluster 7)

For **France** the following cities were selected:
- *Paris* (first entry in cluster 8)
- *Nice* (cluster 8 too)
- *Lyon* (first french entry in cluster 6)

For **Spain** the following cities were selected:
- *Madrid* (first entry in cluster 0)
- *Granada* (first spanish entry in cluster 6)
- *Murcia* (first spanish entry in cluster 7)

For **Portugal** *Lisboa* (first entry in cluster 5) was selected.

5 Discussion

Clustering showed that city centers in the same country are "similar" based on Foursquare data, because k-means clustering produced 4 big cluster for France, Italy, Spain and Portugal each.

The main reason is that the city centers differ especially when looking at the restaurant types, which are very specific to each country, e.g. Italian Restaurants and Pizzas in italian cities, tapas places in Spanish cities.

The following map shows the selected cities:

6 Conclusion

The goals were:
- To **compare** about 70 **cities** or their city centers in Italy, France, Spain, Portugal
- To **select** 3 italian, 3 french, 3 spanish and 1 portugese **cities** as such so that they are as **diverse** as possible.

This was possible for Italy, Portugal and Spain. For France 2 cities from the same cluster had to be chosen. The decision was based upon "Feature Knowledge" – Nice was chosen alongside with Paris because it is a well-known and must see city.

The results of this analysis might be useful for tourists planning to visit European cities.