

Applied Data Science – Data Science

INTRODUCTION

Business problem

Suppose you have a month's holiday, and you want to spend it on a trip around Europe, you will visit 5 capitals of European countries, but you don't want to visit 5 similar cities, you want a wider experience, so you ask yourself, if I visit London, does it make sense to visit Madrid? Or should I visit Rome?

Using the cluster classification algorithm (k-means) and foursquare location data, I want to improve the travel experience in Europe, I think many travel agencies should do this to make better vacation packages for their customers to have a full European experience and not the classic trip to Paris, Rome, London.

DATA

- To solve this problem, we will need the name and coordinates of every European capital city, example:

CountryName	CapitalName	CapitalLatitude	CapitalLongitude
Switzerland	Bern	46.916667	7.466667
Turkey	Ankara	39.933333	32.866667
Ukraine	Kyiv	50.433333	30.516667
United Kingdom	London	51.500000	-0.083333
Northern Cyprus	North Nicosia	35.183333	33.366667

Fig 1. Capital city data example

- Foursquare location data of every city.

Venue	Venue Latitude	Venue Longitude	Venue Category
Ålands Sjöfartsmuseum	60.097170	19.926457	Museum
ÅSS Marina	60.098890	19.924264	Harbor / Marina
Bagarstugan Cafe & Vin	60.099727	19.943189	Café
Indigo	60.100092	19.942963	Lounge
Sjökvarteret	60.104880	19.945174	History Museum

Fig 2. Foursquare location data example

METHODOLOGY

Using the data available for free in kaggle website we imported the csv file as a framework and explore the data, the first thing to do was to filter the data to work only with the European capitals.

CountryName	CapitalName	CapitalLatitude	CapitalLongitude	CountryCode	ContinentName
Aland Islands	Mariehamn	60.116667	19.900000	AX	Europe
Albania	Tirana	41.316667	19.816667	AL	Europe
Andorra	Andorra la Vella	42.500000	1.516667	AD	Europe
Armenia	Yerevan	40.166667	44.500000	AM	Europe
Austria	Vienna	48.200000	16.366667	AT	Europe

Fig 3. European capital cities dataframe

There were 57 cities in the dataframe, using this data with the foursquare API we get the location data of each capital city in Europe in a radius of 10 kilometers.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Mariehamn	60.116667	19.9	Ålands Sjöfartsmuseum	60.097170	19.926457	Museum
1	Mariehamn	60.116667	19.9	ÅSS Marina	60.098890	19.924264	Harbor / Marina
2	Mariehamn	60.116667	19.9	Bagarstugan Cafe & Vin	60.099727	19.943189	Café
3	Mariehamn	60.116667	19.9	Indigo	60.100092	19.942963	Lounge
4	Mariehamn	60.116667	19.9	Sjökvarteret	60.104880	19.945174	History Museum

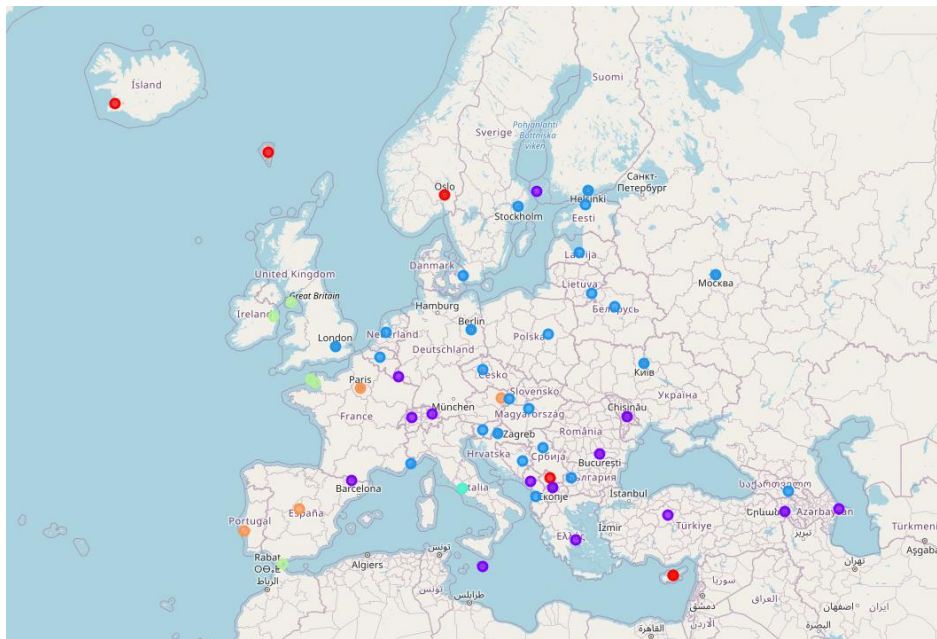
Fig 4. Foursquare location data dataframe

The main analysis I did with the data was to determine which were the most common venues of each city, I got the following result:

Venue Category	City Count
Café	54
Hotel	52
Park	51
Coffee Shop	48
Bakery	45
Restaurant	45

Fig 5. Most common venues

RESULTS



DISCUSSION

- **As a Tourist:** If you are thinking to go to vacation to Europe now you can select 3 or 4 different cities and ensure that they are different enough to gain a wider experience.
- **As a Travel Agency:** whit this information you can create new packages that improve the travel experience of your clients. One of the packages for example may include: Madrid, London, and Reykjavik (island).

Each capital of Europe has its unique characteristics, but some of them may not be as different as people think, using the location data of each capital of Europe we classified them in six different groups, with this people who want to spend their holidays travelling through Europe can identify which countries they should visit to have a different

experience in each stop. This information can also be used by travel agents who want to improve their travel packages.

It should be noted that this project was done only with the data of location of the cities but there are many other factors that can affect the decision such as weather, a special national date interesting for tourists, etc. So I recommend that this information be used as part of a larger research before making an important decision.

REFERENCES

Data Source: <https://www.kaggle.com/nikitagrec/world-capitals-gps>