IMPROVING THE TRAVEL EXPERIENCE TO EUROPE

APPLIED DATA SCIENCE CAPSTONE

IMPROVING THE TRAVEL EXPERIENCE TO EUROPE

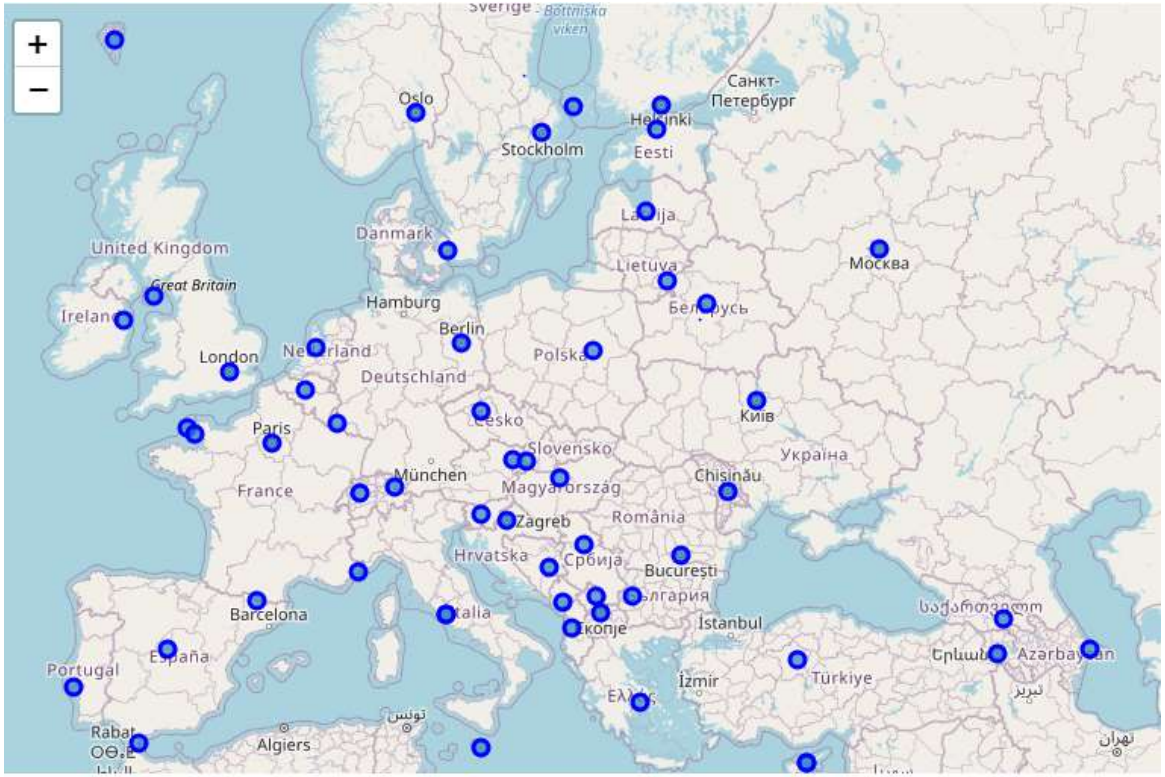
When people travel to Europe, they want to visit different places so they can have a wider experience of the European culture.

In the following project I will use k-means cluster classification with Foursquare location data to determine which European capital cities are different enough to be a great travel destination route.

Data Acquisition and Clean

I obtained the coordinates of every capital city in the world in the Kaggle website, the first thing to do is filtrate de data so we only get the European cities.

	CountryName	CapitalName	CapitalLatitude	CapitalLongitude	CountryCode	ContinentName
0	Somaliland	Hargeisa	9.550000	44.050000	NaN	Africa
1	South Georgia and South Sandwich Islands	King Edward Point	-54.283333	-36.500000	GS	Antarctica
2	French Southern and Antarctic Lands	Port-aux-Français	-49.350000	70.216667	TF	Antarctica
3	Palestine	Jerusalem	31.766667	35.233333	PS	Asia
4	Aland Islands	Mariehamn	60.116667	19.900000	AX	Europe



Data Acquisition and Clean

ONCE WITH THE EUROPEAN CITIES WE CAN PLOT THE LOCATIONS IN A MAP, THERE ARE 57 CITIES IN TOTAL

NOW USING THE FORSQUARE API, I GOT THE VENUES OF EVERY CITY IN A RADIUS OF 10 KM

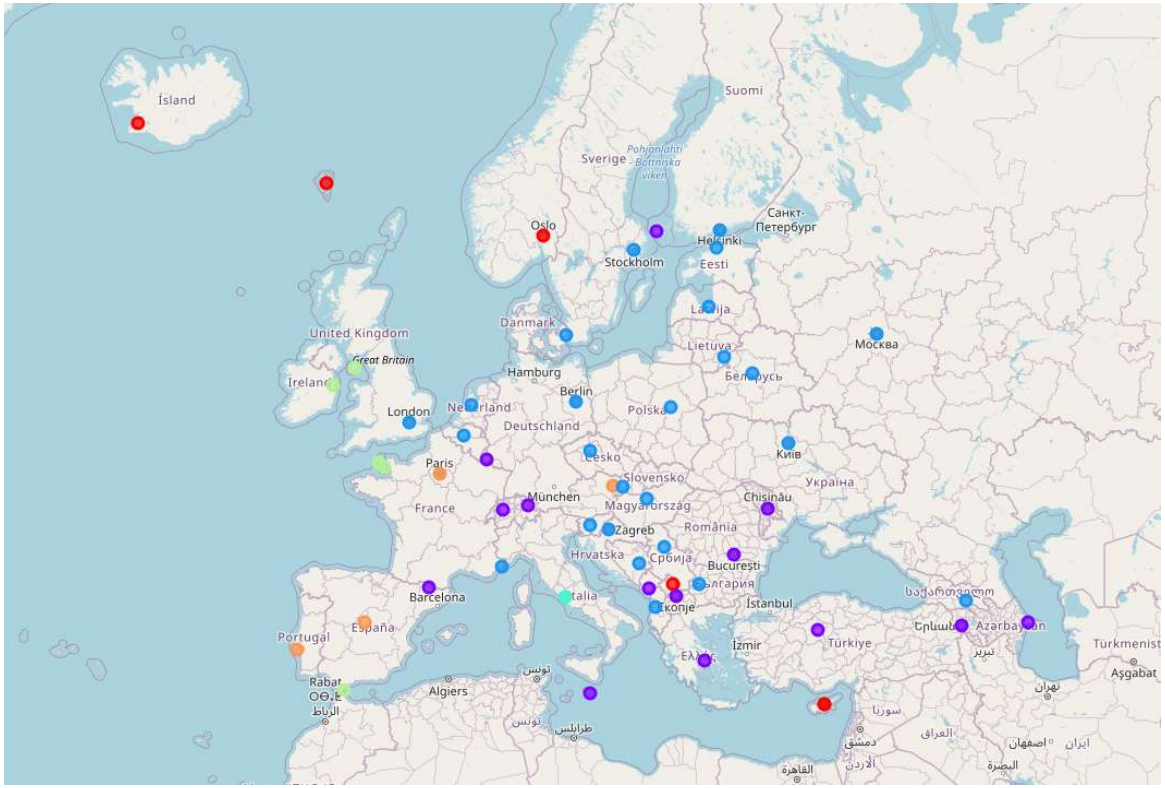
City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Mariehamn	60.116667	19.9	Ålands Sjöfartsmuseum	60.097170	19.926457	Museum
Mariehamn	60.116667	19.9	ÅSS Marina	60.098890	19.924264	Harbor / Marina
Mariehamn	60.116667	19.9	Indigo	60.100092	19.942963	Lounge
Mariehamn	60.116667	19.9	Sjökvarteret	60.104880	19.945174	History Museum
Mariehamn	60.116667	19.9	Träningsverket	60.127640	19.932688	Gym

	Venue Category	City Count
1	Café	54
3	Hotel	52
4	Park	51
2	Coffee Shop	48
0	Bakery	45
5	Restaurant	45

Now let’s see which are the most common venues in each city

AS WEE SEE, THERE ARE SEVERAL VENUE CATEGORIES TOO GENERAL THAT WON'T HELP THE MODEL TO MAKE A GOOD CLASSIFICATION.

I DECIDED TO DELETE THIS VENUE CATEGORIES TO HELP THE ALGORITHM IN MAKING A BETTER CLASSIFICATION



RESULTS

APPLYING THE K-MEANS ALGORITHM TO THE DATA WE OBTAINED 6 CLUSTERS, EACH CLUSTER REPRESENT A GROUP OF CITIES THAT WITH SIMILAR LOCATION CHARACTERISTICS

CONCLUSIONS

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 results 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.

THANK YOU