Coursera Capstone

IBM Applied Data Science Capstone

Opening a new campground around Ardèche, France

Author: Eleonora Balbi

April 2020

Background and description of the problem

The Ardèche Department in France is well known for its summer activities such as hiking, rock climbing and canoeing. Campgrounds are a flourishing business there and a lot of people go to campgrounds for a cheap and family friendly way to sleep and stay in the region for holidays. As a result, there are many campgrounds in the Ardèche region and more open every year. In any case, the location of the campgrounds is one of the most important decisions that will determine whether the campground will be a success or a failure.

The objective of this capstone project is to analyse and select the best locations in the Ardèche department in France to open a new campground. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the following business question: Where is the best place to open a new campground in the Ardèche region in France?

Data and Methodology

To solve the problem, we will need the following data:

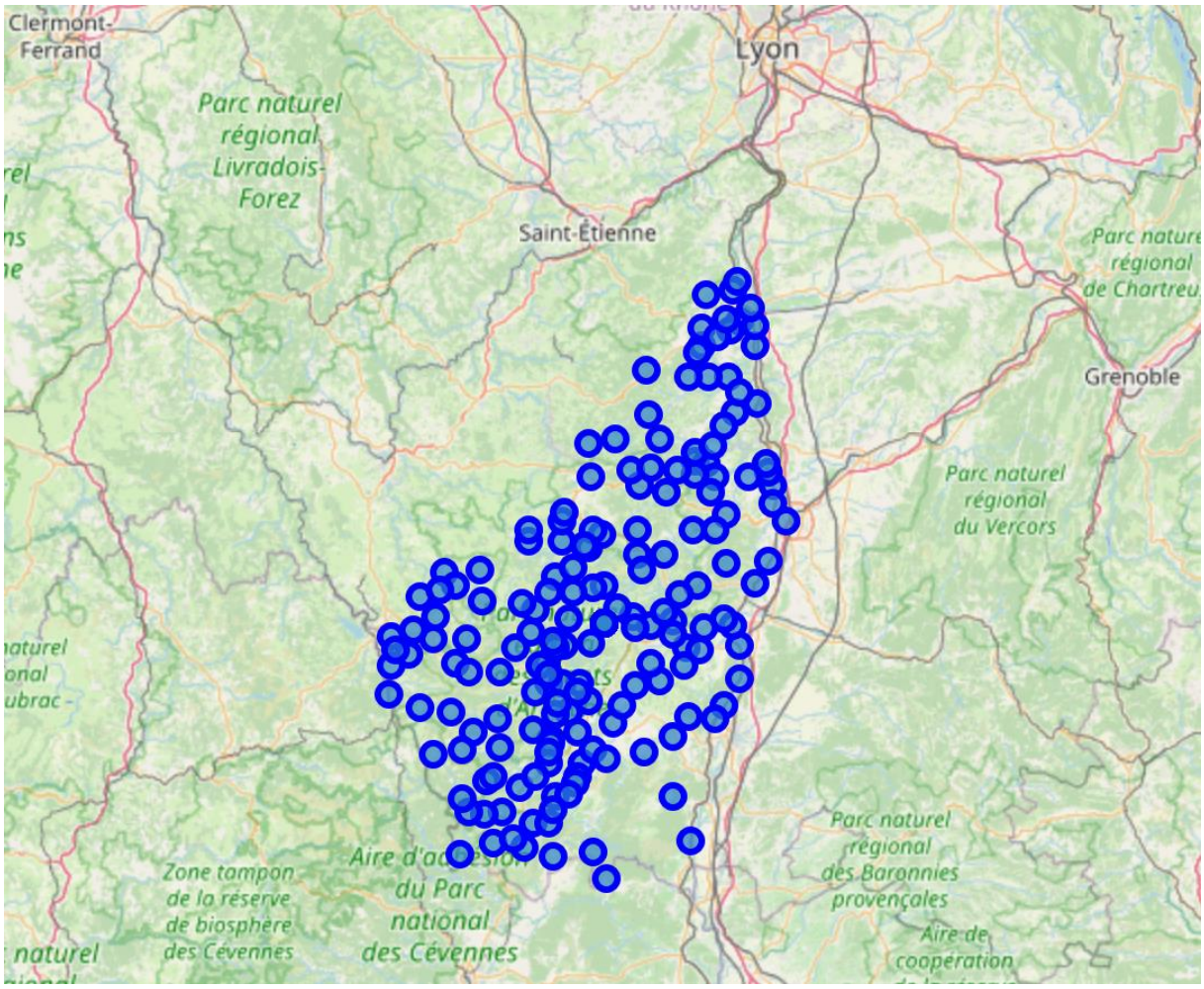
1) List of communities in the Ardèche department, scrapped from a Wikipedia page (https://en.wikipedia.org/wiki/Category:Communes_of_Ardèche). The list contains 199 communities of the Ardèche department. To perform web scraping techniques for data extraction from the Wikipedia page, the Python libraries Beautifoulsoup4 and pandas are used to create a data frame.

2) We will need latitude and longitude coordinates of these communities. This is required in order to plot the map and to get the venue data. The geographical coordinates of the communities are retrieved with the help of Python Geocoder package.

3) Venue data, particularly data related to camping grounds that are within an 8km radius of the community will be gathered from Foursquare API. We will use this data to perform clustering on the communities and select the best cluster to open a new campground.

Exploring the data

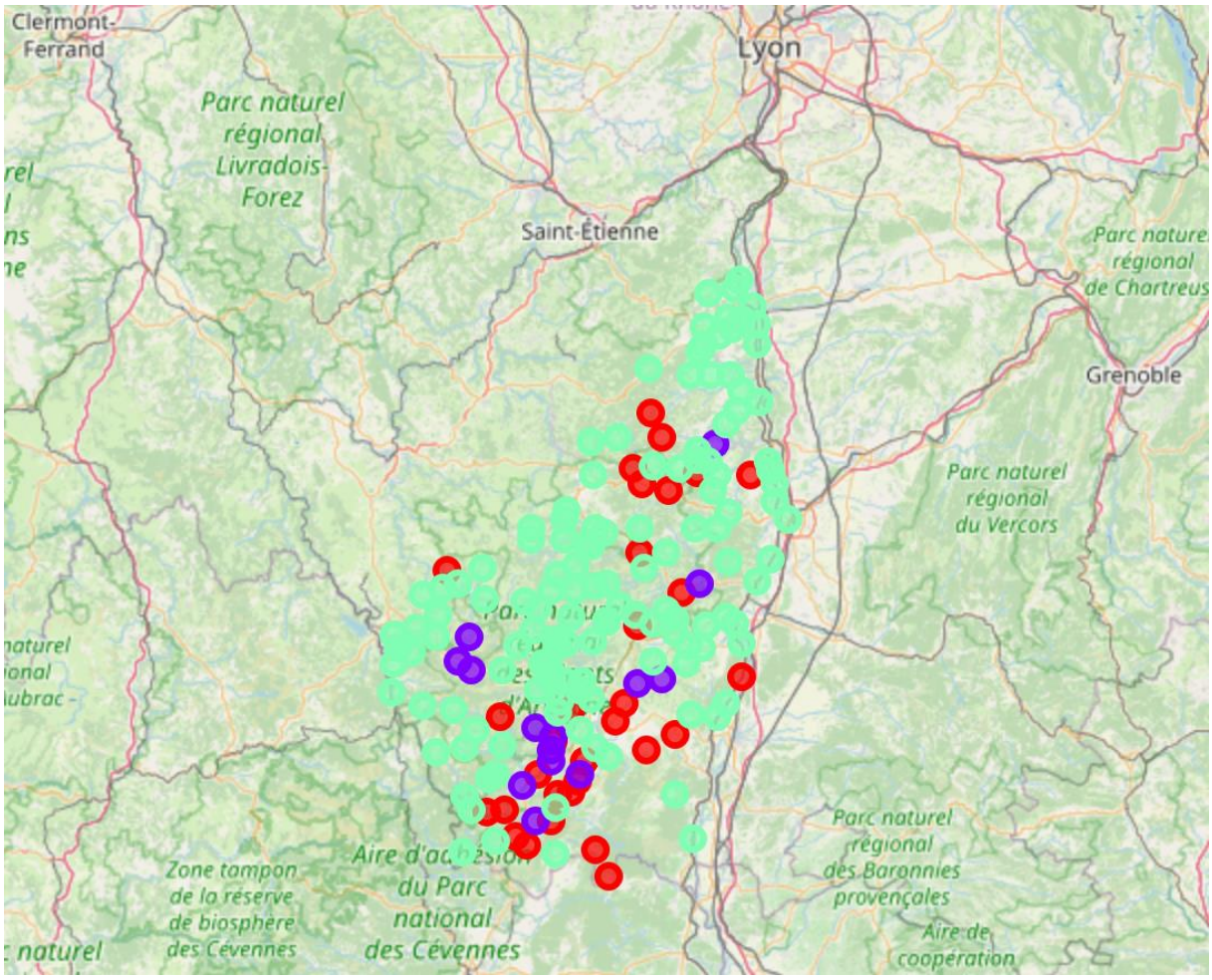
Thanks to the Folium library, we were able to plot a leaflet map of all the 199 communities of Ardèche department, where it could be possible to open a new campground. This map is shown below.



From the Foursquare data, we were then able to see that there are 141 unique categories for the whole Ardèche department. We could see that campgrounds were included as a category in the data. In the whole Ardèche region, there are already 32 campgrounds.

Clustering the Campground data

We tried to cluster the different communities based on the frequency of campgrounds and used k-means clustering. The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour and cluster 2 in mint colour.



Results

The results of the exploratory data analysis and clustering are summarized below:

- There are a total of 32 campgrounds in the Ardèche region, spread over 199 communities.
- We can categorize the communities into 3 clusters based on the frequency of the occurrence of campgrounds.
    - Cluster 0: Communities with moderate number of campgrounds
    - Cluster 1: Communities with high number of campgrounds
    - Cluster 2: Communities with low number of campgrounds
- Most campgrounds can be found in the southern area of Ardèche.

Discussion

Most of the campgrounds are concentrated in the southern area of Ardèche, with the highest number in cluster 1 and medium number in Cluster 0. Cluster 2 has a very low number to totally no campgrounds in the communities. This represents a great opportunity and high potential areas to open new campgrounds as there is very little to no competition from existing campgrounds. Meanwhile, campgrounds in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of campgrounds. From another perspective, this also shows that the oversupply of campgrounds mostly happened in the southern area of the department, whereas the

central and northern areas still have fewer campgrounds. Therefore, this project recommends people with an interest to open new campgrounds to open new campgrounds in communities in cluster 2 with little to no competition. In case of unique selling propositions to stand out from the competition, they can also open new campgrounds in communities in cluster 0. Lastly, property developers are advised to avoid communities in cluster 1 which already have high concentration of campgrounds and are most probably suffering from intense competition. In any case, it is still suggested to look for further "to do activities" in close reach of the property, such as rock climbing spots, rivers and restaurants in a reachable place, to satisfy more campground visitors. Unfortunately there is not a lot of data available on Foursquare for this region.

Limitations and Suggestions for future research

In this project, only one factor (i.e. frequency of occurrence of campgrounds) was considered. But there are other factors such as occurrence of leisure time activities and nature that could influence the location decision of a new campground. Further research could therefore use a methodology to include these data into their project to determine a preferred location for a new campground.

Conclusion

After identifying the business problem, specifying the required data, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities and lastly providing recommendations for the locations to open a new campground in the Ardèche region, we can conclude, that communities in cluster x are the most preferred locations to open a new campground. The findings of this project will help people who want to open a new campground in the Ardèche department in France.