

Data Science Capstone Report

Esperanza Garijo

February 2021

1 Introduction

We have been tasked with finding the best spot to do a music festival in Germany. It must be a place that will be near as many people as possible, but it must be outside of a city. We also want it to be accessible, as it can't be some spot in the middle of the forest, so we're looking for a spot that has a camping site to be in the vicinity. We believe that more people will come to our festival if there is a strong music scene in the nearby cities, and that the music scene in any given city can be measured by the number of music venues. Thus, we aim to find the camping site in Germany which is close to as many music venues as possible.

This project would be of interest to anyone interested in starting a festival in Germany. The code can also be easily tweaked for it to work in any other country, and with some minor changes it could also be used to find spots where other types of businesses are concentrated, which would be useful for anyone looking to open a business in an area with high density of potential customers.

2 Data

We will be using Wikipedia data to find the cities of Germany. We will then use the Foursquare API to find all of the music venues in each city. Finally, we will use Foursquare to find the camping site that is nearest to the perfect spot we found.

3 Methodology

First, the data was scraped from Wikipedia and cleaned. The process to find the music venues was tested first on the city of Berlin to check that it worked. Since all of the venues returned had something to do with music, it was concluded that the methodology of looking for the highest density of music venues was in fact representative of the music scene in the city. A function was created to fetch the venues from a list of cities, which would work for any type of venue. It was then applied to specifically music venues in all the cities contained in

the dataframe. K-Means Clustering was subsequently used to create 5 clusters of venues which were geographically close to each other. The number 5 was arrived at through visual inspection of the data. The centre of the cluster was chosen as the perfect spot. Using the Foursquare API, the camping site nearest the chosen spot was found, and this was determined to be the perfect spot for the festival.

4 Results

The data showed that the largest cluster was in North Rhine-Westphalia, a very densely populated area in Germany. The chosen camping site was the Campingplatz Unterbacher See/Nord, in the vicinity from Düsseldorf.

5 Discussion

The spot makes sense given that North Rhine-Westphalia is a very densely populated area, with a big portion of the population of Germany living there. It is also near the city of Cologne, which has a vibrant music scene. As this area is very urbanised, it is very accessible, and the camping is in a pleasant spot near a lake.

6 Conclusion

In conclusion, the project produced the perfect camping site where one could organise a festival in Germany, and this camping site is the Campingplatz Unterbacher See/Nord.