

Are open-air festivals and
temperature related?

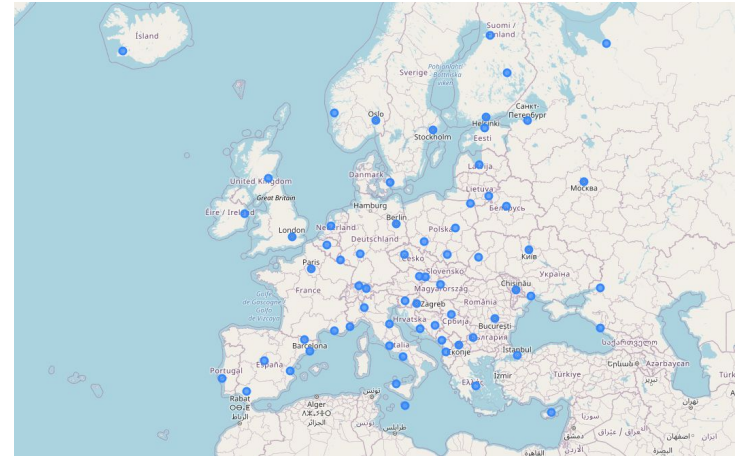
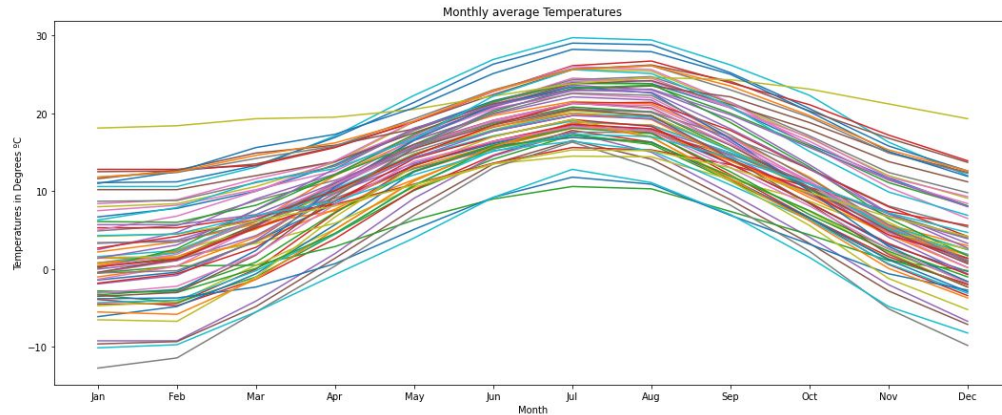
Introduction

An investor wants to create a new open-air festival in Europe and, among other aspects, they want to see if there is any correlation with temperature and the number of existing festivals.

We will conduct a clustering analysis in different European cities and use Foursquare data to extract the number of festivals per location.

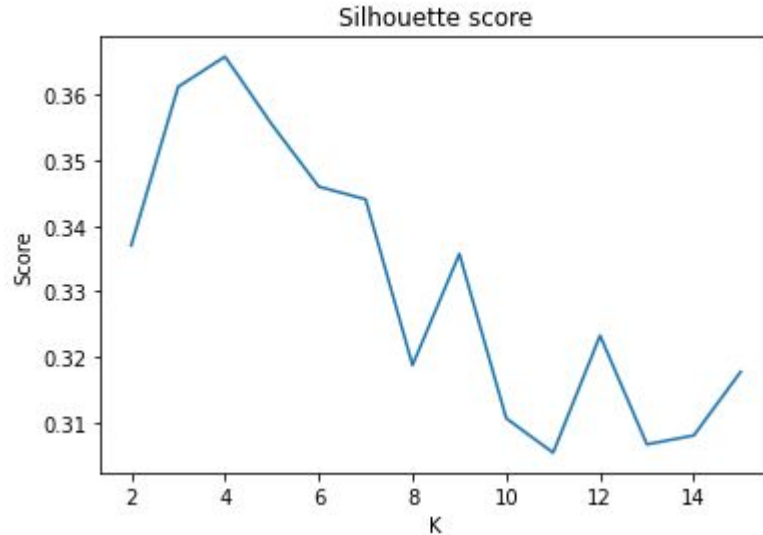
Data

We used mean monthly temperatures extracted from [Wikipedia](#) and Foursquare data of festivals.



Analysis

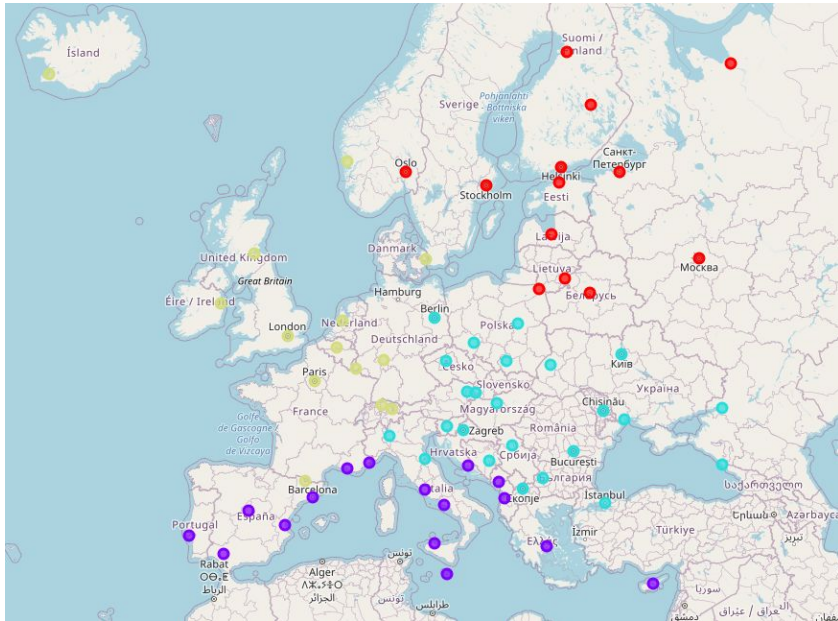
We used KNN clustering to aggregate the locations by temperature. Using the Silhouette score we evaluated the best K.



We chose 4 as it shows a maximum in the Silhouette Score

Analysis

The clustering resulted in a map that correlates quite well to the different climates in Europe.

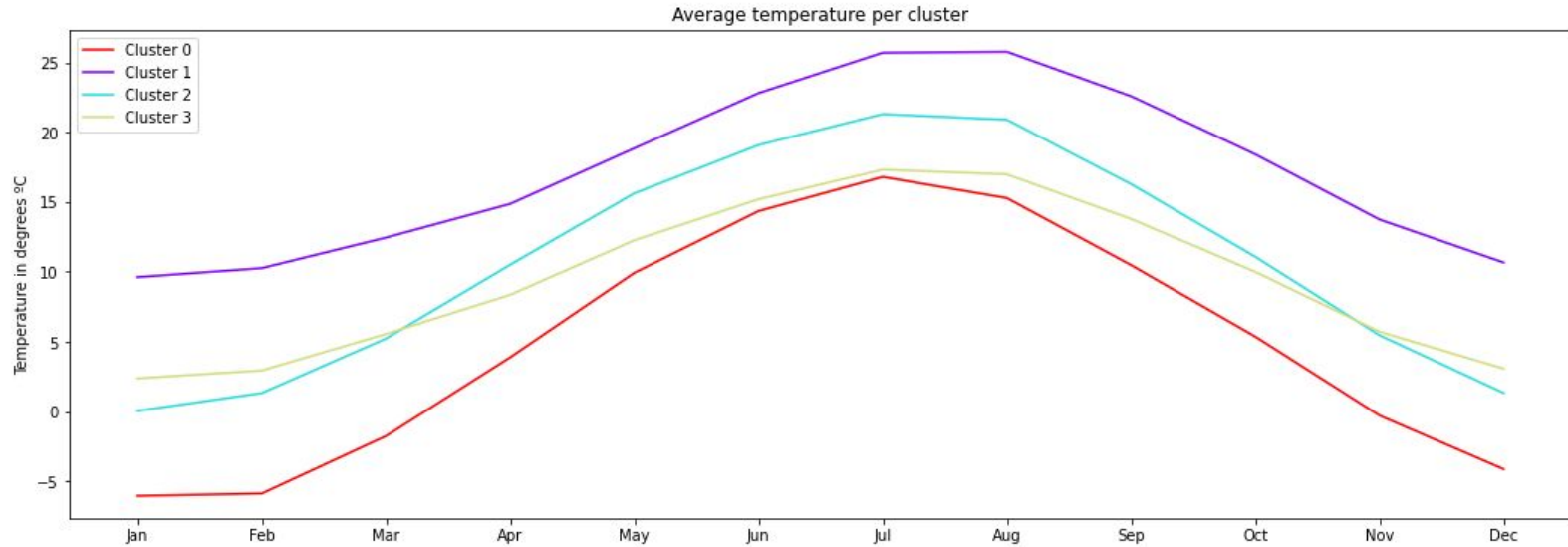


European climates



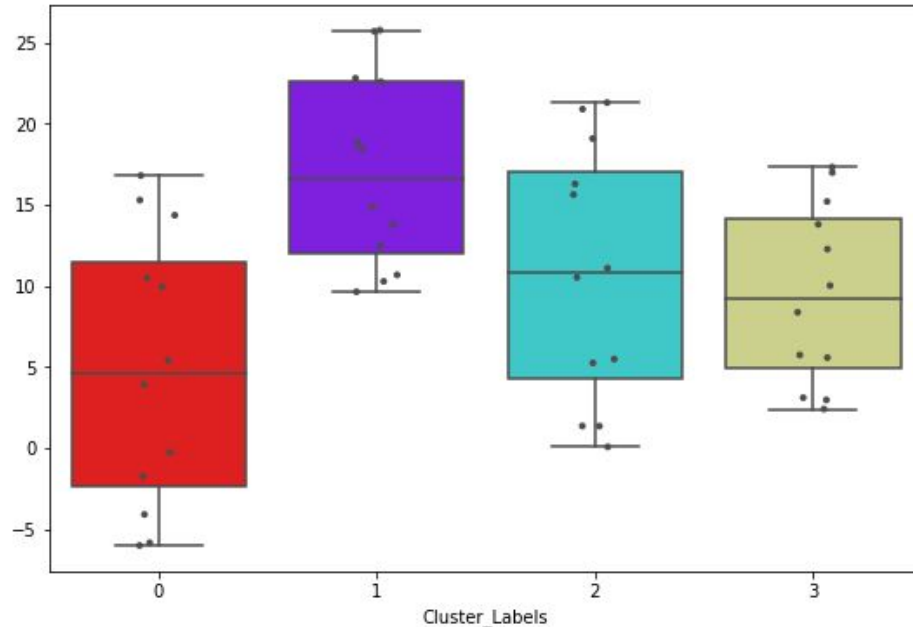
Analysis

Plotting the average temperatures of the clusters shows the main differences between clusters



Analysis

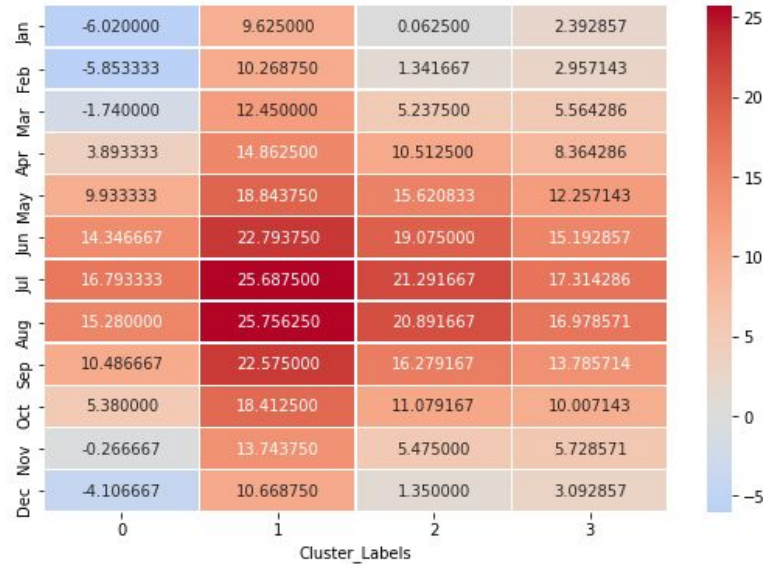
In order to have a more detailed view of the clusters we decided to use a Boxplot



We can clearly see the differences between cluster 0 and 1. Clusters 2 and 3 are similar but vary in the range of temperatures.

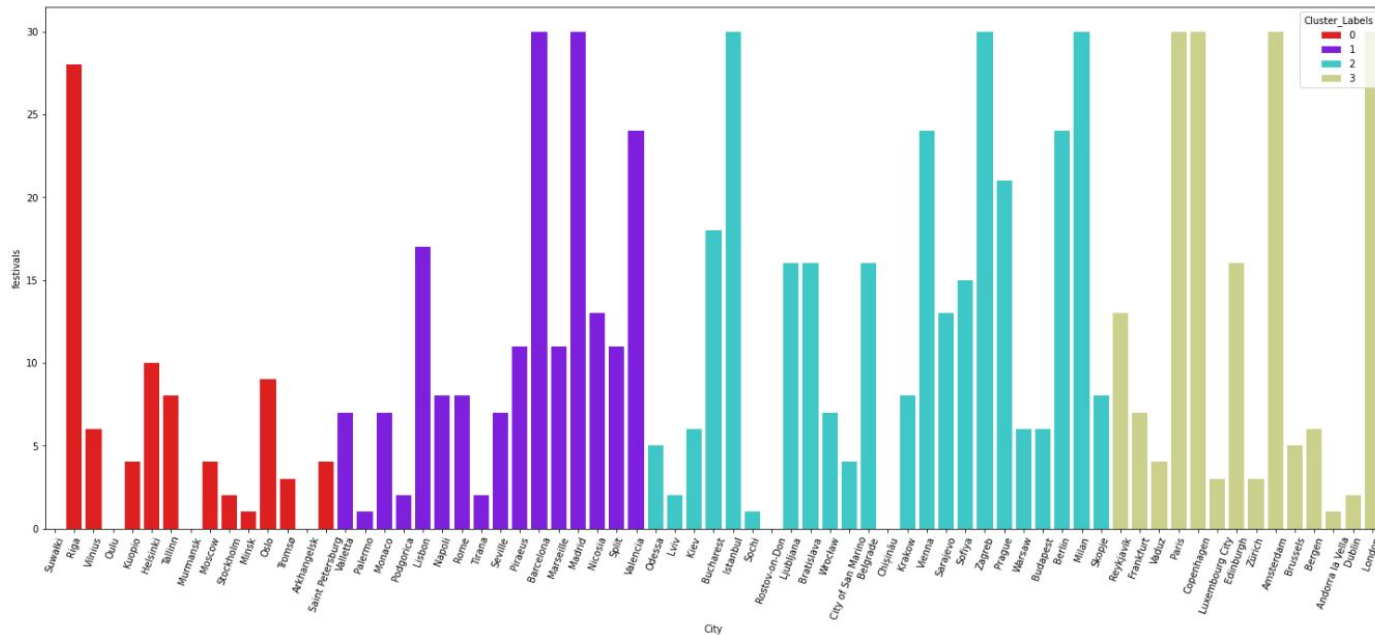
Analysis

Finally, the heatmap gives another clear visualization of the differences in temperature between the clusters



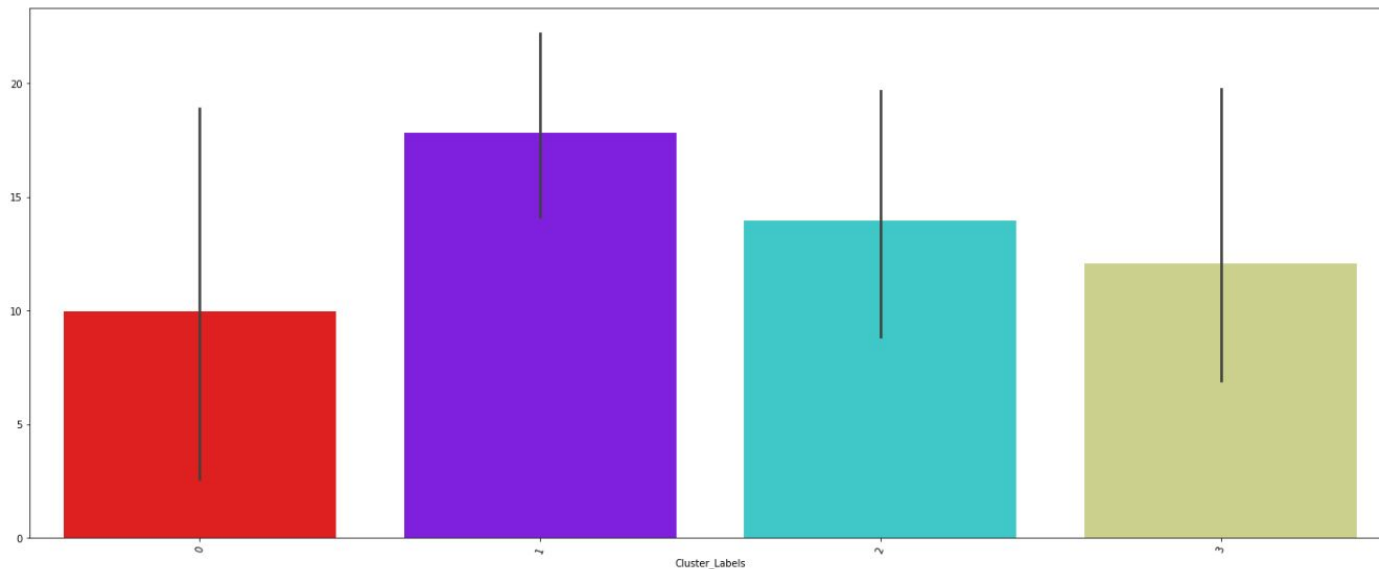
Results

We used Foursquare data to get the number of Festivals at each location. We aggregated them by clusters and used a bar plot to visualize it.



Results

Using this visualization does not help us much. Moreover, since we don't have the same amount of cities per cluster, it unbalances the results. Instead we computed the average of festivals per cluster.



Discussion

The analysis shows a positive correlation between temperature and the number of festivals. We were able to show only using data that, on average, regions with warmer weather held more open-air festivals than regions with colder temperatures.

Conclusions

The purpose of the analysis was to validate the hypothesis that regions with warmer temperatures held more open-air festivals. Using KNN clustering allowed us to cluster locations by similar temperature and check at the geographical distribution. A cross validation with a climate map of Europe supports the correctness of the clustering. Finally, segmenting the festivals by cluster lead into the final conclusion that areas with warmer temperatures held on average more festivals.