

Introduction, Bussiness problem

In this project i will compare all capitals from around the world with regards to their bussinesses. The capitals will be ordered according to their bussinesses. I want to find out if this ordering will be more reflective of the continent in which a city is located.

This can be helpfull for anyone looking for international bussiness opportunities. Someone who has succesfull bussines ventures in one capital might concentrate his next adventures in cities within the same cluster, so as to make good use of his cultural experiences.

Data

The data for the bussinesses comes from 4square, this is a service where users contribute to the database. A downside of this is that there will be data missing from countries where 4square is less popular or where people are overall less inclined to use digital services. However, we only need the comparative frequency of different venues and there is no reason to think that those frequencies will be affected by those problems.

The coordinates of the capitals comes from the website <http://techslides.com/list-of-countries-and-capitals>

Methodology

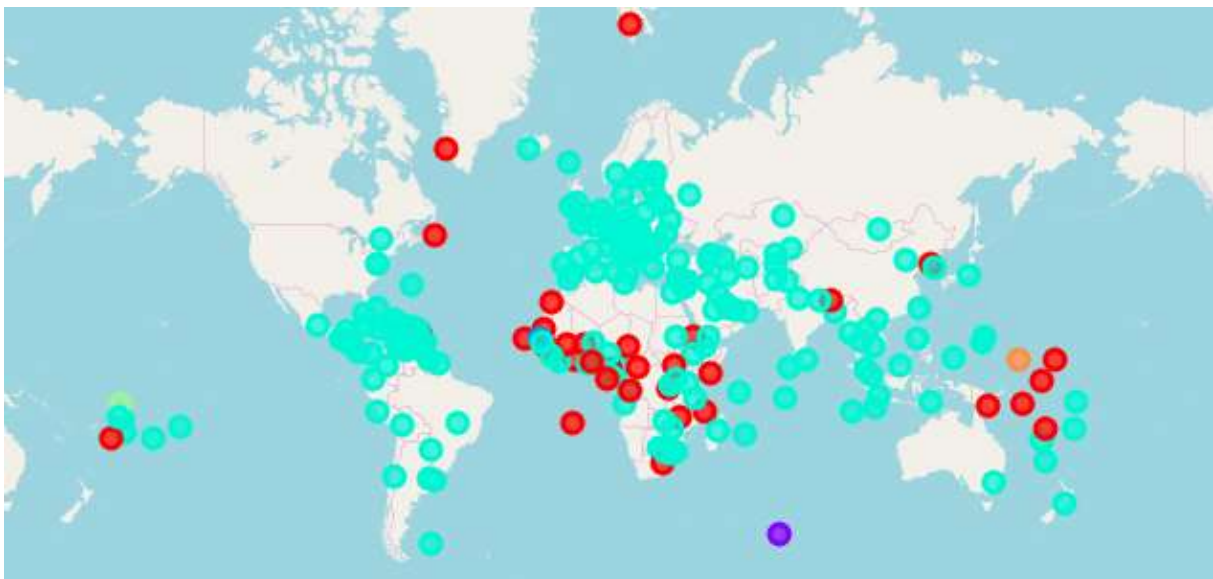
For each capital a list of venues is created. These are ordered according to their frequency, so as to obtain a list from '1st most common venue, 2nd most common venue, etc'.

A k-means algorithm is than used to group the capitals according to this list. This algorithm is chosen because it makes it easy to match the number of clusters with the number of continents. I choose to make 5 clusters.

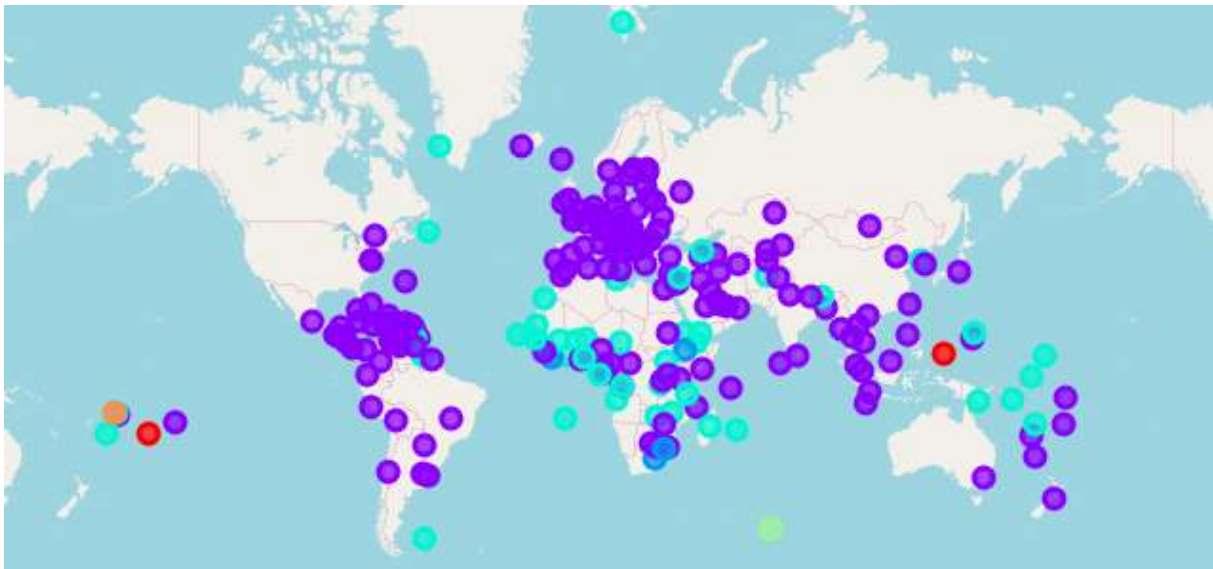
Results

It turns out that it matters how big of a circle you draw around each city. After all some cities will have big suburbs while others are surrounded by farmland and in some cases the ocean. This does however not affect the overall shape of the biggest clusters.

This is the result with a radius of 10km.



And this the result with a radius of 2 km



Discussion

Although the clustering does not neatly correspond with the different continents it is not completely random either. There are 2 main clusters: one spanning the Americas, Europa and south-east-asia, and another in Africa, the middle east and oceania. The first cluster seems to have bars and cafés as its main services, and the second cluster hotels.

Conclusion

It turns out that capitals that are geographically close are alike in terms of their 4square-registered venues. There are 2 big clusters and 3 small clusters which only very roughly correspond to any particular continent.