TOP UNIVERSITIES' BATTLE OF NEIGHBORHOODS

By Arina Voorhaar

INTRODUCTION

In this project I decided to study to which extent the neighborhood can affect the quality of a University. For this, I have downloaded the list of top 1000 universities in the World according to the QS ranking, have collected their geodata, and using Foursquare I have explored the neighborhoods of each University and clustered the Universities according to the facilities that their neighborhoods can offer. Then I analyzed whether there is any correlation between the clustering I obtained and the rank standings.

The target audience can include potential students that are choosing a university to apply to. Of course, when we choose a university, we pay a lot of attention to its reputation and quality. But what if the options are similar in terms of ranking? What criteria would we look at? Neighborhood plays a big role here. Is it a university in the middle of nowhere or in a large city? Are there may places to go and many ways to entertain yourself or not? This project would give answers to some of those questions and potentially help the student to make a good decision.

DATA USED

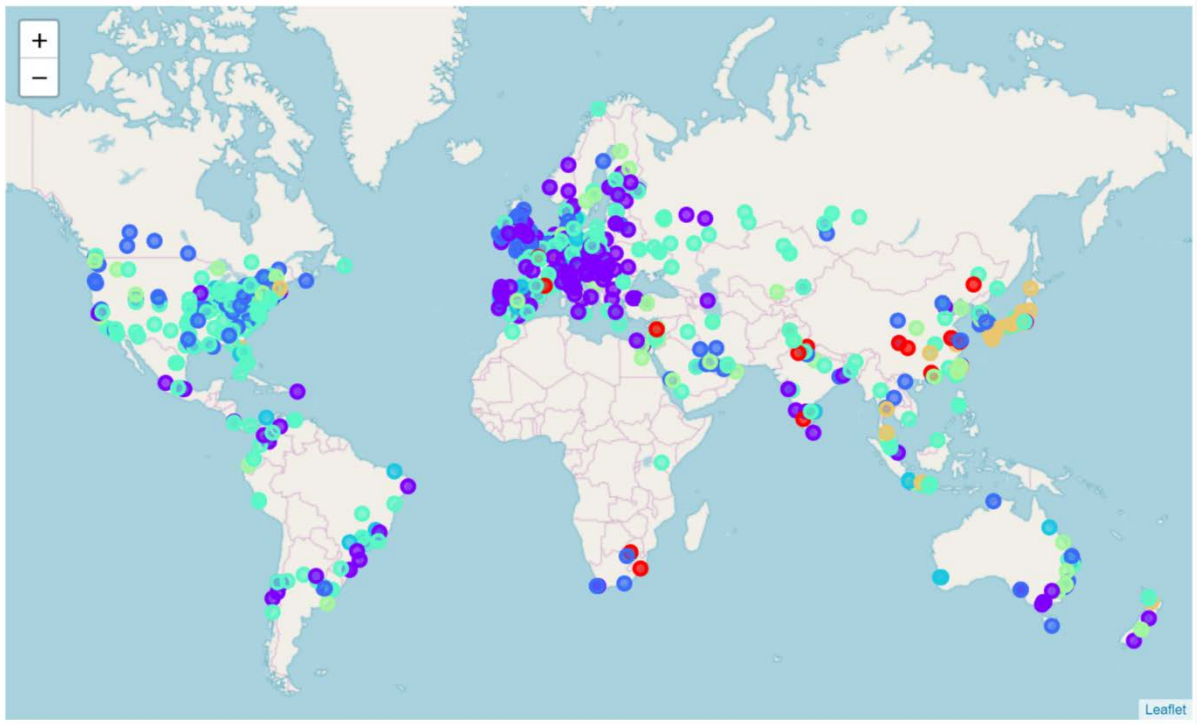
I downloaded the list of top 1000 Universities of the World, according to the QS ranking of 2018. From that list I took the following information: name of University, its standing and the country it is located in. Then, using Google API geocoder, I obtained the geographical data (latitude, longitude) of each of the Universities. And finally, I used Foursquare to explore the neighborhoods of the Universities to find out what facilities they offer.

METHODOLOGY

At first, I tried to build all sorts of basic models, uncluding Linear Regression, Decision Tree, etc that would learn to guess, based on the neighborhood data, how high this or that university would be in the ranking. The result was negative, as one could expect. Indeed, the quality of the neighborhood cannot determine the quality of the University. Still, to complete this project, I decided just to try to cluster the Universities, so that I can see what do their neighborhoods have in common. This can also be useful for a potential student, since they can choose between options with similar positions in the ranking but still might have some preferences when it comes to the neighborhoods where they will have to spend most of their time for several months/years. So, I used the 'elbow algorithm' to determine the reasonable number of clusters and then used the KMeans algorithm.

RESULTS

The clustering we obtained can be visualized as follows:



What one can immediately see on the map is that even though the clustering algorithm did not employ any data different from the information about most common places in the neighborhood, we see that some clusters are represented by universities of some certain region: e.g., the purple cluster is represented by European Universities, the orange one – by Japanese universities, the blue cluster has a lot of British and American Universities, etc. At the same time, it is nice to see the Universities located in a region different from the most common one. For example, if a student prefers universities that belong to the purple cluster, but does not want to move to Europe, they still might find a similar kind of University in a more preferable region.

DISCUSSION

Unfortunately, we realized that there is no significant correlation between the neighborhood's facilities and the quality of the University. Still, it would be interesting to try to come up with other criteria and data that can potentially be in correlation with the University rankings.