

To perform this cluster analysis, I will need the location and the type of music venues in São Paulo. So, I'll be using retrieved data from Foursquare based [on their music venues categories](#).

I'll be also working with a list of all São Paulo's districts, as well as their geographical coordinates, since these data will be needed to make the correspondence with data gathered from Foursquare. The coordinates data are those mapped and made available with a public domain license [in Kaggle as a JSON file](#).