

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](#).

Foursquare divides the venues into categories and subcategories, and music has its special place in this taxonomy.

The API lists many music-related categories, such as record stores, music shops, music schools, karaoke bars and so on.

But since this analysis is interested in getting to know São Paulo's live music scene, the scope will be limited to music venues where musicians perform live.

Therefore, music venues in São Paulo will be retrieved in the following nine categories: Amphitheater, Concert Hall, Music Venue, Jazz Club, Piano Bar, Rock Club, Music Festival, Nightlife Spot and Nightclub. Their location and neighborhood will also be retrieved.

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](#), that contains the latitude and longitude of all 96 districts in São Paulo.