Kaleb Tsegaye

GEOG788P

**Initial project proposal: Visualizing Concert/Live Performance Data Using Songkick’s API**

Potential sources of data

The primary sources of data for this project are Songkick’s API and Spotify’s API. Songkick is a website for information about live performances. Songkick’s API allows a user to collect data on a catalog of over 6 million concerts. Spotify is a very popular music streaming service with an API that contains an abundance of metadata on its vast catalog of artists and music. In the case that these may not work out, I may also find other sources useful like Ticketmaster’s API or Apple Music’s API. Data cleaning should not be much of an issue for these sources, since they will be retrievable in a format that allows simple usage. The main concern with these APIs is rate-limiting. If my requests are deemed excessive, my access to these platforms may be limited or ceased temporarily. This would make creating larger maps with a lot of data unideal.

Problem

The goal of this project is to create a tool that will allow users to visualize live concert information. Users would be able to input information like genre, artist, or a tour name and a map will be plotted based on this input. Choropleth maps would be created for queries involving an artist or genre to show regions where they are the most frequent in. If given a tour name, a map of the path take can be created. Related artists’ tours could also be generated and compared to see patterns in their paths. Through this project, I plan to also examine the patterns that lesser-known artists show compared to more established performers. The target audiences for this project is small artists that may want visualized information to get an idea of suitable places to consider for marketing music and touring and music listeners who will be able to enjoy looking at information for their favorite artists and genres.

Python tools

For this project, the modules I expect to use include the requests module, Geopandas, matplotlib, and ESDA. Requests will be required for reading from the APIs, since this information will be retrievable through their respective websites. Of course, Geopandas, matplotlib, and ESDA will be needed for analyzing and plotting the data I will retrieve though the APIs. Custom code is expected to be quite extensive. I will need to be able to take many different types of input from the user and parse API requests to retrieve information. This code should be able to compile and organize pertinent data and create the plots.

Potential Challenges

As mentioned before, the issue of making large requests to the APIs could lead to limitations. Also, making larger plots could be too time-consuming to process and I may have to pivot and focus on visualizing smaller areas. Also, I am not totally sure how extensive data is outside of North America and Europe. If the data is minimal, plots of areas outside of these continents may be pointless.