Project proposal

1. Choice of dataset

The following data set is to **test** our algorithm:
Kaggle Dataset

- The main data set will be an input playlist from which we identify the songs the client likes the most. An ideal playlist can be 2021's most listened songs. This is read using the spotify web API and choosing songs from the “Now songs” playlist for each (62) genres that spotify has.

2. Methodology

a. Data Preprocessing:

The data will include several details relating to the music.
We will use spotify’s web API to extract the data from users spotify account and will put the information into python dictionaries from json format. This will allow us to analyze the data using python.
We will use python and matplotlib/seaborn to visualize the data, namely all the features that the songs have. After we identify the quantity of the songs with the **closest** features, we will mark them to choose a 50 song playlist that has the closest relation to the favorite songs inputted.

b. Machine learning model:

The goal of our model is to produce a list of song recommendations based on an input playlist. Thus, our model will use features of those songs to find and recommend similar songs that would match a user’s taste in music.

We will use the k-means algorithm to mark the input data based on their features. We chose this machine learning model because it will allow us to detect patterns in a user’s playlist by clustering. The recommendations will be based on the closest neighbors of the cluster of the input playlist. The number of the clusters in the initial part of the algorithm will be a hyperparameter. Hyperparamter search must be done in an accurate way to have an accurate model.

c. Evaluation Metric:

- To evaluate the algorithm we will ask the user to listen to the output playlist and let the model know if they liked it or not. They can rate out of 10 and over 70% result will be an ideal rate.

- Moreover, we can run a quick survey gathering data about the user experience of the app while it is in its prototype form to evaluate the satisfaction with the generated playlist. We can do this using Google Forms.

3. Application:

What does the user input? How does the user provide inputs? (Is there a webcam? A way for users to submit images? text?)

The user will type a playlist of 62 or more of their favorite songs. There is no specific number of inputs. However since it's a machine learning model, the more, the better.

What does the user receive as output, how will the output be displayed?
The output is a list of 50 recommended songs which will be generated by the model based on the input playlist. The output will be displayed as a playlist format.

(should the user be able to play the song directly on the web app so they can hear these recommendations? Or just list names + info of the recommendations? We can figure that out when making the web app because this is more about the aesthetics and final touches, less about ML).

We are planning to use the spotify Web API to read the data and ask the user to input their data, thus we will have a web application that handles all these difficulties with a good UI.