Joshua Taylor, 261180779

1. I am choosing to use the “Spotify Recommendation” dataset from Kaggle for my project. I want to use this dataset in particular because its features are connected directly to a songs’ raw attributes rather than their creator or genre. Some examples are energy, time signature, valence, key, etc.
2.
    a. The dataset I chose seems feasible for a regression algorithm because it contains a multitude of numerical features. I won’t need to do any processing other than scaling the features.
    b. I want to use this dataset to make song recommendations using a K nearest neighbors regression algorithm. My end goal for this project is for the model to expand past the features listed in the dataset (including more from spotify’s API such as artist, genre, popularity, etc) so that a user can give the model 5 songs and would be returned with a new song that is most similar to the 5 they gave.
    c. The evaluation metric I will use will be mean squared error since this is a regression algorithm.
3. The final goal for an application is for a user to enter a website, give 5 songs recognized by spotify, choose a target genre, and then have a new song recommended to them in the target genre that has similar musicality to the songs they provided. The output will simply be displayed as text, perhaps an image of the cover that is used for the song, and finally a link to the song on spotify.