Sean Ryan

DS210

Final Project

In this project, I observed a dataset featuring different songs on Spotify and performed a regression analysis to find the R^2 value for different characteristics of the songs to see if there was any correlation between tempo and the song's popularity on the platform. My project features a dataset containing more than 114,000 songs that can be found on the music platform Spotify. The dataset also gives ratings for different elements of songs, such as tempo or energy. For this project I focused on comparing the popularity rating for each Spotify track to 3 different other variables from the dataset to test for correlation. The popularity rating is a score given to each song between 0 and 100 where a larger number means a more popular song. The score is based on the total number of plays and number of recent plays it has on Spotify.

The algorithm I created finds the residual sum of squares and divides it by calculating the total sum of squares and subtracting that from 1. My hypothesis for this test is that songs with higher tempo are more likely to affect the popularity of the song. My goal is to see if tempo possesses significant correlation to popularity. After performing the analysis, I was unable to prove that tempo has an effect on its popularity on Spotify. Performing this experiment led to results that cannot determine if there is any significant correlation between tempo and popularity of a song on spotify. The results produced from this test may be a sign of popularity taking into account far more factors than tempo and therefore a wide range of tempos and rhythms in songs can be popular or unpopular.