Project Proposal

A Prediction of Song Popularity

Danish James

Domain

For many in the music business, trying to predict how a song will perform can be almost impossible to determine ahead of time. Grand collaborations between highly regarded artists can drop off the face of the charts within days, while a song that no one had even heard of could reign supreme on the charts for months at a time. This project intends to address the uncertainty and attempt to quantify what factors allow for a song to become popular and perform well. For this, we shall utilize information from https://spotifycharts.com as well as Spotify themselves, alongside a host of other music data sites.

An **MVP** would be capable of classifying what intrinsic factors about a song can determine its popularity based on a years worth of chart data. Some of the more complex mechanisms, such as lyrical analysis and scraping other sites for more detailed information about the artists and their features, would need to be dropped. In addition, there might be more reliance upon the API for work that could have otherwise been scraped.

Data

VARIABLE	TYPE	DESCRIPTION	USED IN MODEL?
DAILY STREAMS	Int	Amount of plays a song receives every day	Target Variable
SONG NAME	String	Song Identifier	N
ARTIST NAME	String	Artist Identifier	N
GENRE	String	Describes the type of song	Y
SONG LENGTH	Int	Total playtime of the song, converted to seconds	Y
AGE	Int	Number of days the song had been released for	Y
SINGLE	Boolean	Determines if the song was released and promoted as a single, vs being an unpromoted b-side.	Y
LANGUAGE	String	Language of Origin for the song	Y
ARTIST FOLLOWERS	Int	A measure of the artists popularity based on their social media follower counts	Y

ARTIST POPULARITY	Int	A separate measure of artist popularity based on their monthly listener count on Spotify, indicating how many people actively listen to this artist	Y
MODALITY	Boolean	Determines whether the song is in Major or Minor key	Y
LYRICAL COMPLEXITY	Float	A value utilized to insinuate how intricate and involved the lyrics of the song are, based off of a separately created algorithm	Y
ENERGY	Float	A value that tries to measure how intense the song is	Y
TEMPO	Int	Measures how fast the song is in Beats Per Minute (BPM)	Y
VALENCE	Float	Measure of the positivity of the song	Y
INSTRUMENTALNESS	Float	A value for how much of the song involves vocals, and how much is more purely instrumental sounds	Y
LOUDNESS	Float	Measures how loud the song is in decibels (dB)	Y
SPEECHINESS	Float	Measures how much of the song is spoken word as opposed to singing	Y
TOTAL			15 Variables Utilized