MA 678

Mid-Term Project Proposal

Xinqiao (Vic) Zhang

Project Description:

People used to listen to music via vinyls and CD's when the Internet was less developed than it is now. With the rise of the Internet, many musical applications have grants the access to listening to music at any time to the audience. Spotify, as one of the largest music streaming companies in the United States, provides its streaming data to the public and for us to analyze the reaction between the streams and the characteristics of the song itself.

Dataset:

I found one dataset on Kaggle that would satisfy the my need of this study. There are 232725 tracks(sample) in the dataset which consists of several predictor variables and popularity as the outcome variable.

https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db/data

Response Variable:
    Popularity(scaling from 0 to 100).
Explanatory Variables:
    1. genre;
    2. artist_name;
    3. track_name;
    4. track_id:
    The Spotify ID for the track;
    5. acousticness:
    A confidence measure from 0.0 to 1.0 of whether the track is acoustic;
    6. danceability:
    Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity;

7. duration_ms:
The duration of the track in milliseconds;
8. energy:
Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity;
9. instrumentalness:
Predicts whether a track contains no vocals;
10. key:
The estimated overall key of the track;
11. liveness:
Detects the presence of an audience in the recording;
12. loudness:
The overall loudness of a track in decibels (dB);
13. mode:
Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived;
14. speechiness:
Speechiness detects the presence of spoken words in a track;
15. tempo:
The overall estimated tempo of a track in beats per minute (BPM);
16. time_signature:
An estimated overall time signature of a track;
17. valence:
A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.

Problem:

I would like to study the relation between the popularity and the characteristics of the track. Would a song from a certain artist be popular based on the popularity of that artist(this might take more knowledge on contemporary artists)? Would greater danceability make people stream the track more? What keys are the popular tracks generally on? It would be intriguing to know what it takes for a song to be popular on Spotify, and thus we could predict whether an upcoming song would be a hit or not. Furthermore, the concluding factors could be used for other artists in their later musical career.

Methods and Plans:

Understanding each of the variables could be essential to constructing a useful model. For instance, popularity was determined based on the streaming on Spotify but it is not in its original scale of streaming amount but in a scale of 0-100. Loudness are measured by logarithm of decibels and centered at mean. On the other hands, a single track could show up in different lines because it has multiple genres. All the entries besides genre were the same, so it would be essential to clean the data in a certain manner. I am not sure yet whether I should merge those lines together or I should leave it the way it is right now because the model is special itself. However, data cleaning should be done before constructing the models.

Once we are down to the model parts, I plan on fitting several logistic and generalized linear regression models and a multi-level model on the data to fulfill the analysis. I decide to use the genre variable as the multinomial factor for now because it seems important in the model and is hard to factor out. However, I might switch to another variable with the progress of this study. The desired result is that we could predict the popularity of a track based on the characteristics (even partially) given on a high confidence level.

Reference

https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db/data

https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/