Frost 2019 Project Proposal
An Investigation of Industry Influence on Music Art Content

By Daniel DeFoe

Advised By Kelly Bodwin

Goals of the Project: To use data and statistical analysis to help answer the question "What is the effect over time of music industry influence on musical artists, in terms of musical content?" This will be done by looking at popular music as a whole, as well as conducting case studies of particular musical artists. It is the goal to investigate whether there is a reason to believe that record label industry interests are influencing the actual musical content which artists create.

How the Goal will be Accomplished: To accomplish the goal of the project, popular artists with a reasonable tenure in the industry will have their 'character' as an artist tracked over time. For an artist to be considered to be a part of the study, that artist must have been active for at least 7 years (or have released his/her/their first album 7 years ago), have released at least 3 albums, and must have made an appearance on the Billboard Hot 100 in the last 20 years. Because artists are the units of the study, all information will be considered aggregated to the artist. The mentioned 'character' of an artist is a qualitative consideration of an artist's uniqueness, consistency of content, and complexity of content. A few general 'characters' will be defined as archetypes as a later portion of the project, but the important element in considering artist character in the project is to see if and how an artist's character changes over time, then to see if this change could potentially be attributed to the forces of the industry rather than a sole untampered choice of the artist. This could be suggested by more songwriters being involved in writing an artist's content, artists of a particular label making choices to cater to what is popular, or if an artist drastically changes their content and then is clearly rewarded by the industry. Potential explanation for changes in character and popularity trends will be discussed along the way during analysis, and data will ultimately help form a narrative in attempts to answer to answer the research question.

Potential Smaller Research Questions
- Is more music made by more people in more recent years?
- Have artists converged to the same genre or genres as one another?
- Do artists under the same label or parent label share any traits in how their content has changed?
- Have there been more featured rappers or other musical artists on more tracks in recent years?
- What is the commercial and industry performance of explicit songs in different genres?
- Can it be predicted that a songwriter who is not a song's artist worked on a song based on the song's linguistic and musical features?

Data Sources
-Billboard Hot 100 (Charlie)
-Genius (Charlie) for lyrics

-Spotify top plays
-Grammy awards official website
-RIAA official gold and platinum listings
-Spotify Analytics
-Ultimateguitartabs.com for chords

Planned Structure of the Data

The current plan is to structure the data in the following tables:

Table Name	Column Names
BillboardRankings **NOTE: this was created using Charlie's functions	Artists: (String)the song artists Name: (String)song name Weekly rank: (Numeric)the rank of the song on the given week Peak position: (Numeric)the highest rank the song has had in any week Weeks on chart: (Numeric)number of weeks the song has been on the Billboard Hot 100 Week: (Date)the concluding date that the week is of Date: (Date)date that the song was released Genre: (String)list of potential genres that the song can be classified as Writing Credits: (String)people considered songwriters for the track Lyrics:(String) the lyrics of the song Songwriter: (Boolean) if any of the 20 listed
SpotifyStreams	Position: (Numeric)the rank of that song in the top 200 streams of that week Track Name:(String) name of the song that is on the chart Artist: (String)the artist the song is by Streams: (Numeric)number of streams the given song got on Spotify in the given week Week: (Date) the date of the week
RIAACertifications	Title: (String)name of the song Artist: (String)name of the artist Status: (String)what the highest certification the track has received is Label: (String)the record label the song was released on BigThree: (String)which big three record label the label is under ('None' if not under one the big three)
GrammyWinners	Award: (String)grammy that was won GrammyYear: (Numeric)year for which the award was won Genre:(String) genre the award was given under SongTitle: (String)title of the song

	Artist: (String)name of the artist
SongSections	Title: (String)song title Arist: (String)name of the artist Section: (String)name of the section if known Progression: (String)chord progression in numbers in the key of C major EndDifferent: (String) 'None' if the section is completely consistent, otherwise, a string of chord progression in numbers in the key of C major of the end of a section if it differs from the progression ExtendedChords: (Numeric) number of chords with an extension (C7 or C9) NonDiatonic: (Numeric) number of non-diatonic chords (chords outside of the key, here the key of C) used in the section NumChords: (Numeric) number of unique chords used in the section
SongAttributes	Title: (String) song title Artist: (String) name of the artist Album: (String) Name of the album NumSections: number of sections in the song Song Attributes from spotify ----- ----- These would be spotify's attributes of a track
ArtistInfo	Artist: (String) name of artist Genres:String) genres that the artist plays NumAlbums: (Numeric) number of albums YearFirstAlbum: (Numeric) year of the artist's first album

Some further processing of features will occur likely to conduct the analysis effectively. This is how the data will be structured likely to start. These will be the starting tables.

Things to Consider
These are some points which will be considered to rationalize the data collection and analysis choices that are planned to be made.

- Industry and commercial popularity of an artist is essentially an explanatory, the response is the consistency of content in form and complexity
- There may not be chord data on some of the hip-hop/rap songs, so as far as musical analysis, spotify's rhythm data may be more effective for these, where more pop/rock/singer-songwriter songs will likely have chord data
-

More may be considered later, but these are just to start.

Artists Archetype Representatives: Along with the conduction of case studies in exploring how a particular artist's content can change or remain similar over time, the project will observe the trends of six predetermined artist archetypes. These labels are subjective but they are made in attempts to group similar types of artists and compare how their complexity changes or has changed over time. The archetypes will be considered as follows, being represented by the listed artists:

-"The Sellout": Maroon 5, Taylor Swift
-"The Rebel" : Foo Fighters, J Cole
-"The Pop Product": Justin Bieber, Britney Spears
-"The Reinvention":Justin Timberlake, Twenty One Pilots
-"The Classic": Elton John, Beatles
-"The Unknown":Khalid, Billie Eilish

The last smaller research question which will be addressed if possible is trying to predict which of the first five archetypes that "the unknowns" will end up being.