Do Audio Features have A Causal Effect on Chart Placement?

Louis A. Gomez, Ayesha Parveen
Department of Computer Science
Stevens Institute of Technology, Hoboken, NJ

ABSTRACT

Objective: The way we consume music in recent years has changed from traditional mediums such as radio and physical sales to streaming via apps such as Spotify and Apple Music. This change is also reflected in the way the Billboard Hot 100 – the music chart that ranks the most popular songs in the country weekly – calculates its ranking by giving more weight to music streams. Performance on this chart is used as a proxy for artist success hence it is vital to understand what factors contribute to higher chart positions. In this work, we investigated whether audio features of tracks such as genre and tempo, have any causal effects on higher chart position (top 50) on debut.

Methods and Materials: We collected the past 5 years (2017-2021) of data (song title, artist name, and debut position) from Billboard charts for songs released in the summer and audio features of these songs from Spotify. To study the causal effects of the audio features on chart placement, we used a causal inference approach that computes the average causal significance for each of our features via pairwise testing while controlling for other potential audio features (causes).

Results: We found that features such as danceability, energy, r&b genre had positive causal effects while high-speechiness and duration had negative causal effects on higher chart placements. We further validated the relationships we found using a classification task where we trained a logistic regression model on data from 2017 to 2020 and tested on data from 2021. The model achieved an area under the receiver operating curve of 0.66.

Conclusion: In this work, we investigated the causal effects of audio features on chart placement. In performing causal inference, we found features that had positively and negatively causal effects on our outcomes. Our results show that more work is needed to further understand the complex relationships associated with what makes songs chart higher.

INTRODUCTION:

In the past decade, the way consumers use and interact with goods and services have become increasingly more digitized. This can be seen from the way we order food online using apps, sharing pictures on the internet, and consumption of news media via our phones instead of newspapers. An industry that has seen a major shift in the way goods are consumed is the music industry. Traditionally, songs were consumed either through plays on the radio or physical purchases. Due to the advent of streaming platforms such as Spotify, Apple Music, and YouTube you can listen to your favorite music anytime and anywhere. Although some things remain timeless like the power of a great song and the powerful connection between the artists and their fans, the way songs are now consumed by the public has drastically changed. According to the International Federation of Phonographic Industry (IFPI), streaming contributed 62.1% of global music revenue in 2020 up 19% from 2019 [1]. Fueled by these changes, music companies have also changed the way they measure the most popular songs in the country. The Billboard charts, who have been music's record keepers since the 1940s, keep track of important music metrics from the top-selling artists to the most played songs [2]. Previously, these calculations were done based on just radio plays and songs sales but with the changes brought by the internet and the usage of online streaming platforms, Billboard updated their calculation metrics to include the streaming of songs and digital song sales as well. Currently, Billboard has two defined types of streaming plays for the 'Billboard Hot 100' songs charts: on-demand based on streaming of songs on platforms like Amazon Music, Apple Music, Spotify, and YouTube and programmed plays based on Pandora and Slacker Radio with more weight given to the former category. Thus, streaming is the most dominant factor for a song to chart in the Billboard Hot 100 chart [2]. In this project, we are interested in investigating whether features of the songs contribute positively to a higher placement on the billboard charts. To quantify this, we estimate the causal effect of audio features (such as loudness and danceability) and artist metadata (such as genres) on whether songs chart in the top 50 or bottom 50 of the charts.

METHODS

Data Collection

We begin by describing the datasets used in our study. Due to the large number of songs that appear on the charts each year, we limit our scope using certain inclusion criteria. First, we restrict the search to the past 5 years (2021 - 2017) as this covers when music consumption favored streaming. Next, we selected only songs that were released in the summer to help mitigate the effects of seasonality. See Figure 1 for a distribution of debut position through the periods we study. Additionally, there are differences in the kinds of songs released at different times of the year which may affect the results of our analysis. To collect data, our first step was to query the publicly available Billboard API [6] for the Hot 100 weekly charts. This returned a table that contains information like chart position, song title, artist name, number of weeks on the charts, and last position of every song that has appeared on the charts. From this table, we select only songs that have a value of 0 for the last position as this indicates that it was the first time it appeared on the charts (i.e its debut position). After this step, we are left with song titles, chart position on debut), artist name, and the date. Next, for each song, we retrieved the audio features using the publicly available API from Spotify [4], the largest streaming platform both in the US and around the world [3]. To begin, we use artist name and track title as a query to the Spotify API and retrieve the unique track identifier and list of genres associated with the artist. Next, using the unique track identifiers, we retrieved the audio features which are a set of musical quantifiers used to characterize each song. The features include danceability, energy, acousticness, instrumentalness, liveness, mode, loudness, speechiness, tempo, duration, and valence. See Table 1 for the description of all audio features and their value ranges collected from Spotify.

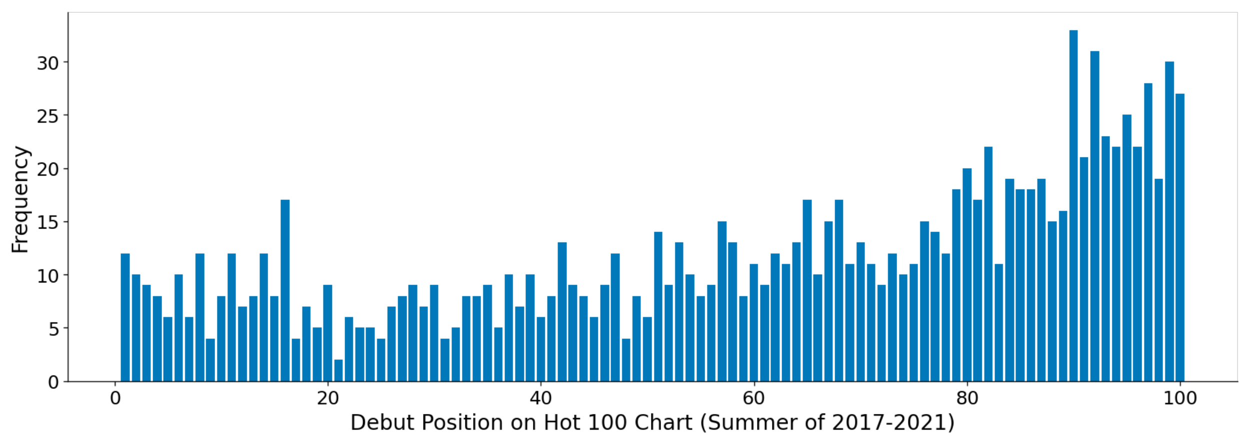

Figure 1. Distribution of debut positions for songs in the summer for the year 2017 to 2021

Data Processing

After data collection, we have 1197 songs where each song is associated with a vector that contains the audio features of the song. To process this data, we discretized numerical features into recommended ranges from Spotify and generated new features derived from the music genre feature. For numerical variables, we discretized them as follows. For danceability, energy, acousticness, valence, instrumentalness, liveness, duration, and tempo we create binary variables by using thresholds and assigning values to 1 or 0 based on the set thresholds. The threshold was 0.8 for liveness, 3.50 minutes for the duration, 120 beats per minute for tempo, and the rest were set to 0.5. Speechiness is transformed into three binary variables namely: low-speechiness (value < 0.33), mid-speechiness ($0.33 \leq$ value < 0.66) and high-speechiness (value ≥ 0.66). The last feature, mode, is unchanged as it is already a binary variable. To process the genre feature, we first mapped the list of genres associated with an artist into a high-level category to match one of the following genres: pop, hip-hop, RnB, country, Latin, dance, and alternative. We did this based on the highest frequency of the "base" genre.

Table 1: Audio features retrieved from Spotify

Audio Features	Description
acousticness	A confidence measure to track whether the is acoustic or not.
danceability	Suitability of the track for dancing based on tempo, rhythm stability, beat strength, and overall regularity and activity.
energy	Represents a perceptual measure of intensity and activity.
instrumentalness	Predicts whether a track contains no vocals
mode	Indicates modality (major or minor) of a track
liveliness	Detects the presence of an audience in the recording
valence	The musical positiveness conveyed by a track
tempo	The overall estimated tempo of a track in beats per minute
duration	The duration of a track in milliseconds
speechiness	Detects the presence of spoken words in a track

For example, the genres associated with Kendrick Lamar are conscious hip hop, hip hop, rap, and west coast rap. Based on our categorization, songs by this artist would be mapped to the hip-hop category. Secondly, we created seven new variables from the genre feature based on the high-level genre classes identified above by one-hot encoding each of the categories in the genre (e.g., pop-genre, Latin-genre). If a song is in that category, we assign a value of 1 for that genre category and 0 otherwise. This means that tracks by Kendrick Lamar will have values of 1 in the hip-hop genre but 0 in all other genre categories. Since we are interested in what leads to higher chart placement, we made the outcome (chart position) into a binary variable: (i) songs that debuted in the top 50 (0-50) and (ii) songs that debuted in the bottom 50 (51 - 100) of the charts. Songs in the top 50 are assigned a value of 1 and 0 for songs in the bottom 50. To reduce our feature space, we remove features that have a low variance because changes in variable values are vital for discovering causal relationships. Hence, we dropped instrumentalness and liveness.

Causal Inference

In this project, our main goal is to investigate the causal effects of audio features on chart placement (top or bottom 50) on the billboard hot 100 charts. While our data is inherently temporal, because we focus on only the debut position, our problem is simplified since we know our time window a priori (≅ 7 days) and hence do not need to test with different window sizes. Additionally, since audio features are static (i.e., they do not change) we only need to collect them at a single time point. To test if audio features have any causal effect on debut position, we use the causal inference approach developed by S. Kleinberg [7]. This method relies on the principle that causes happen before the effect and raise the probability of the effect happening. Formally, we performed:

$$\mathcal{E}_{avg}(c,e) = \frac{\sum_{x \in X \setminus c} \mathcal{E}_x(c,e)}{|X \setminus c|}$$

Where:

$$\mathcal{E}_x(c,e) = P(e|c \wedge x) - P(e|\neg c \wedge x)$$

This outputs a set of average causal significances $\mathcal{E}_{avg}$, for each audio feature. Due to the small set of potential causes, we didn't perform the usual step of multiple hypothesis testing to control for false positives. Using this approach, we performed pairwise testing to get the average difference our cause c makes to the effect e given a set of other potential causes $x \in X \setminus c$. The set of potential causes X, is our feature space which contains 17 variables where we iteratively exclude a cause c and compute the average causal significance. This was performed on the audio data from 2017 to 2020.

RESULTS

Causal Inference Results

Figure 2 shows the average causal significance on a scale for each of the 17 audio features we test for relationships. We found positive causal relationships in the following: alt-genre, r&b-genre, latin-genre, country-genre, low-speechiness, energy, and danceability. Conversely, negative causal relationships were found in the following: rap-genre, high-speechiness, mid-speechiness, duration_ms, and mode. Positive relationships mean that the presence of that variable increases the probability of the effect occurring and for negative relationships, this is inverted [7]. Features such as high-speechiness and mid-speechiness having negative causal effects makes intuitive sense as the billboard 100 charts focus on songs where (high or mid) speechiness alludes to media that is more of an audiobook or spoken word format. Additionally, the duration_ms feature also negatively contributes to the effect as in recent years, songs have become shorter in length to favor streaming [8]. Essentially, the shorter a song is, the more it can be streamed in a fixed period. Features including tempo, valance, and dance-genre, pop-genre and acousticness have approximately causal significance values of 0. According to [7] we cannot make any conclusive claims about whether these have any causal relationship or not because the positive influences each of these variables bring canceled out by the negative influences.

Evaluation

While there isn't established literature on the potential effects of audio features on chart placement, we evaluate our results via a prediction task using the potential causes with non-zero average causal significances using the area under the receiver operating curve (AUROC) as a scoring method. We used a simple logistic regression model with the chart placement (1 or 0) as the label and the following variables as features: high-speechiness, mid-speechiness, low-speechiness, r&b-genre, rap-genre, country-genre, energy, alt-genre, danceability, latin-genre, duration, and mode. We used data from 2017-2020 to train the model (data used to estimate causal effects) and data from the summer of 2021 as test data. Results for this classification showed an AUROC of 0.66 for predicting chart placement on the billboard Hot 100 charts.

DISCUSSION

In this project, we examined the causal effects of the audio features of a song on higher chart placement on the Billboard Hot 100 list. We collected Billboard data from the last 5 years (2017-2021) and then picked the songs that debuted for the first time on the charts. After making a list of those songs, we collect their audio features from Spotify. Using this data, we try to find the causal effects of the song's features on its placement on the charts. We use the causal method approach that relies on the principle that causes happen before the effect and they inherently raise the probability of the effect happening. Based on the outcomes we found that features like alt-genre, r&b-genre, latin-genre, country-genre, low-speechiness, energy and danceability had positive effects whereas features like rap-genre, high-speechiness, mid-speechiness, duration_ms, and mode seemed to have negative effects. To evaluate our results, we performed a binary classification task on the 2021 Billboard data which resulted in an AUROC of 0.67 which indicate that the features do have an effect, but they are not as predictive for us to be able to firmly confirm the causal relationships.

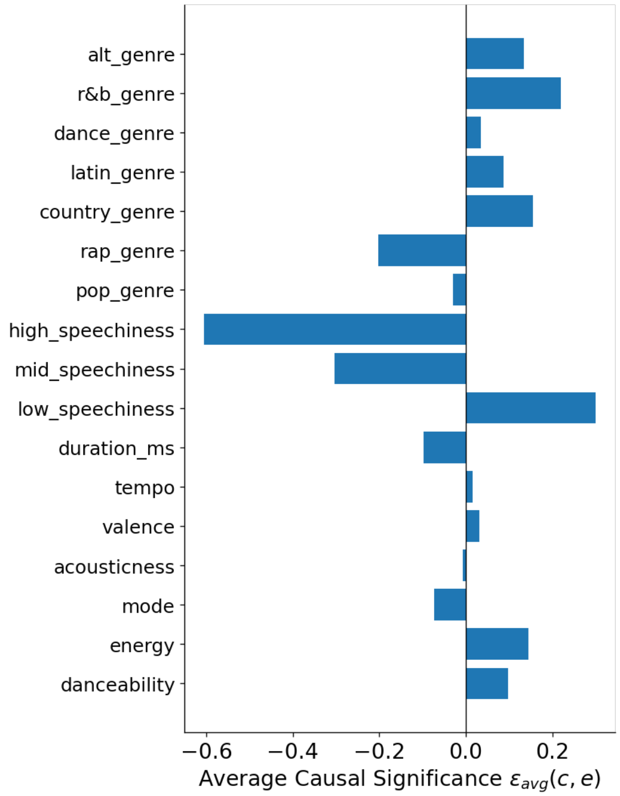

Figure 2. Bar graph showing the average causal significance for all potential causes in our feature space

Limitations and Challenges

A limitation we have is the kinds of features we examined. One of the main challenges we had was characterizing the popularity of an artist at specific time points. This was vital as one of our assumptions was that popularity is positively associated with performance on the Billboard charts - especially at debut. To quantify this, we tried to gather information on follower count as a proxy for popularity, but Spotify does not provide historical data on follower count. As an alternative, we tried to access the same data in different social media platforms such as follower count on Instagram and Twitter or number of subscribers on YouTube. Unfortunately, neither of these platforms provide historical information on followers/subscribers so we were unable to quantify that feature. Secondly, we had issues with data collection due to different naming standards in the Billboard and Spotify databases. This resulted in mismatches for the names of songs and artists on the tracks. Lastly, for some pairwise testing, due to a limited number of samples, we had positivity issues where the probability of a set of variables was 0 in our dataset. Hence, we were unable to estimate the relationships between those variables.

Future Work

In future work, we could include more variables that increase the model complexity which would enhance the performance. We could investigate subsequent chart positions (like bubbling hot 50 i.e., songs that did not make the hot 100 lists but had the potential to do so) over a duration of time to see how the positions could have changed and what could be the possible causes for that. Secondly, we could include other metrics that have a huge influence on the current music markets like song performances on Tik-Tok. Furthermore, an examination of how seasonality may change the causes for songs to chart would be interesting. For example, will we see the same relationships during the winter seasons (where people are more inclined to stream/listen to holiday songs or mellowed down songs) compared to the summertime where people are more inclined to listen to up-tempo songs? Since the relationship between chart positions and song characteristics extend beyond a song's audio features, more complex factors such as record-label data, radio plays, digital song sales could help uncover more of this complex relationship. This potentially gets us closer to answering the question: ‘What makes a song chart in the Hot 100?

REFERENCES:

1. “Global Music Report 2021”, *IFPI,* https://gmr2021.ifpi.org/report, March 23rd, 2021

2. Billboard Staff, “Billboard Finalizes Changes to How Streams Are Weighted for Billboard Hot 100 & Billboard 200”, *Billboard,* https://www.billboard.com/articles/news/8427967/billboard-changes-streaming-weighting-hot-100-billboard-200, May 1st, 2018

3. Mansoor Iqbal, “Spotify Revenue and Usage Statistics (2021)”, https://www.businessofapps.com/data/spotify-statistics/, September 23rd, 2021

4. ‘Spotify Developer API’, *Spotify,* https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features

5. ‘Python API for downloading Billboard Charts, https://pypi.python.org/pypi/billboard.py

6. Allen Guo, ‘billboard-charts’, https://github.com/guoguo12/billboard-charts

7. S. Kleinberg. *‘Causality, Probability and Time’*. Cambridge University Press, New York 2012.

8. Dan Kopf, *‘The economics of streaming is making songs shorter’,* https://qz.com/1519823/is-spotify-making-songs-shorter/ , January 17, 2019