Chart Toppers: Billboard Top 100 Songs

Athena Ru, Khushmeet Chandi, Zaid Muqsit

2023-12-02

Introduction

The Billboard Hot 100 is the music industry standard record chart in the United States for songs, published weekly by Billboard magazine. Chart rankings are based on sales (physical and digital), radio play, and online streaming in the United States. We are interested in what makes a song popular/successful (as determined by Billboard rankings) in the US musical landscape. We investigate three research questions: (1) what makes a song stay on the chart; (2) what factors influence how high a song peaks on the chart; and (3) what characteristics of songs have led to success in each decade from 1950 to 2020? By examining longevity, magnitude, and time, we can investigate the chart behavior and audio features of popular songs.

We build three models. Model 1 is a random forest that predicts the number of months a song will stay on the Billboard Top 100. Model 2 is a cumulative logit model that predicts what quantile of the chart a song will peak in. Model 3 is a multinomial model that predicts what decade a Billboard Top 100 song was released. These models are also useful for inference, as we can see what characteristics of songs are related to success.

Our findings are useful for music producers and artists who are determined to make music that will be a hit with the masses. Since the data contains only Billboard Top 100 songs, these models are most useful for a song that just debuted. For example, if a song just debuted in position 74, how likely is it to peak in the second quantile (positions 25-50) based on audio features and Spotify popularity? Will it be a one-hit-wonder or will it stay on the chart, demonstrating extended success? And then more broadly for people who are studying or for those who are just interested: how has music changed over the decades? What features proved to be most popular across decades, or only in certain decades?

Data

The data comes was compiled by Sean Miller who uploaded it on Data.World. Sean obtained one dataset on every weekly Hot 100 singles chart from Billboard.com and another dataset on the audio features of these songs from the Spotify Web API. We obtained his data from the TidyTuesday repository on GitHub. We dropped all NA values from the two datasets and merged them on `song_id`. We then dropped all songs where genre was not given (an empty bracket). The resulting dataframe consists of 12,055 observations. The data ranges from 08/09/1958 to 05/29/2021.

For each song, the following new variables were created:

- `debut_position` (numerical): chart position when the song first entered the Billboard 100
- `max_peak_position` (numerical): highest position the song peaked in
- `peak_quantile` (ordered categorical): quantile for `max_peak_position`
  - 1 for #1-25; 2 for #26-50, 3 for #51-75; 4 for #76-100

- `months_on_chart` (numerical): floor(`weeks_on_chart` / 4.33) assuming an average of 4.33 weeks in a month
- `decade` (ordered categorical): 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020
- `genre_list` (list): each sub genre in `spotify_genre` is replaced with the main genre
  - The main genres are the top genres (five most frequent) in the dataset: rock, pop, soul, country, and rap
  - ("dance pop", "pop", "uk pop") becomes ("pop", "pop", "pop")
- `main_genre`: the most frequent top genre in `genre_list`; the more frequent top genre for tiebreakers; "other" if none of the top genres appear in `genre_list`
  - "country" is assigned if `genre_list` is list("country", "country", "country", "rock")

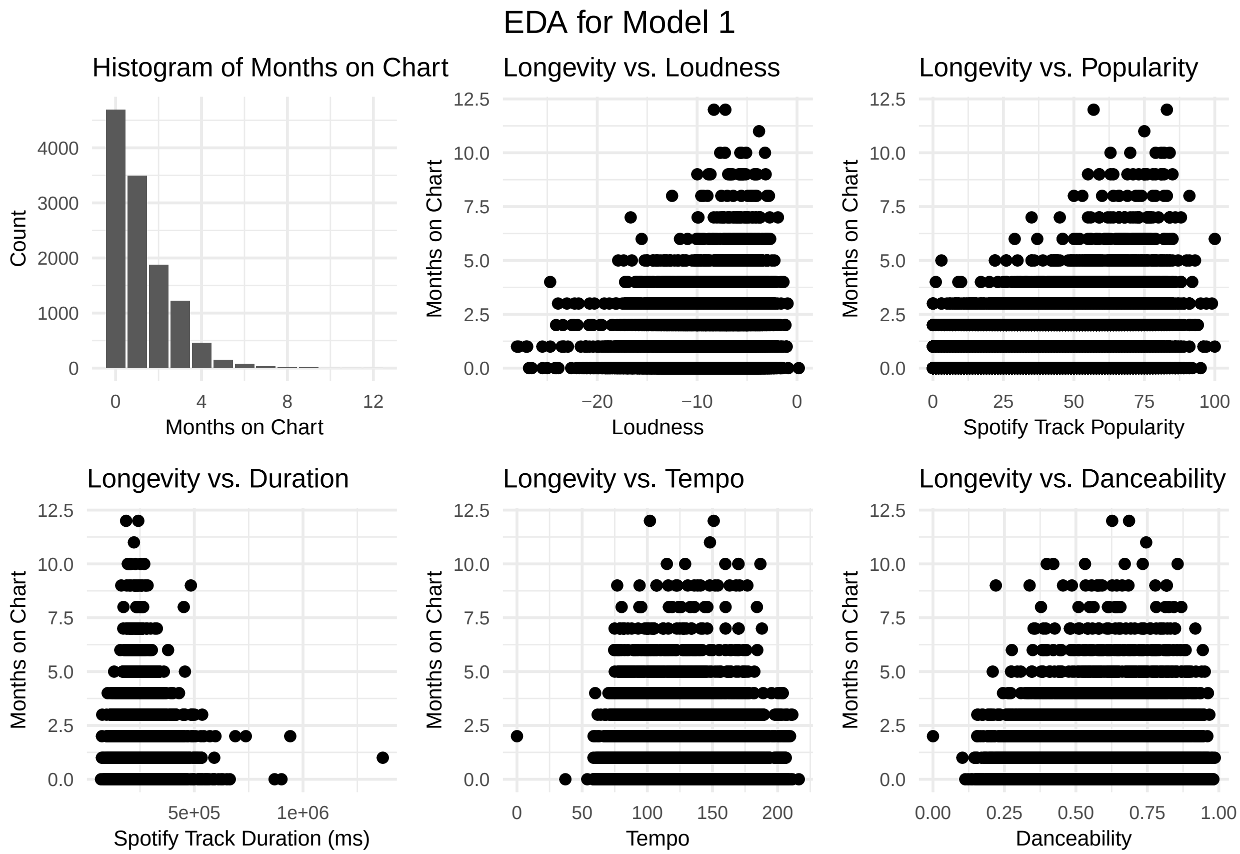
There are 33 columns after feature engineering.

For Model 1, the response variable is `months_on_chart` and the predictors are `danceability` (numerical), `spotify_track_popularity` (numerical), `tempo` (numerical), `loudness` (numerical), and `spotify_track_duration_ms` (numerical).

For Model 2, the response variable is `peak_quantile` and the predictors are `spotify_track_popularity` (numerical), `main_genre` (categorical), `spotify_track_explicit` (Boolean), `debut_position` (numerical), and `danceability` (numerical).

For Model 3, the response variable is `decade` and the predictors are `danceability` (numerical), `energy` (numerical), `loudness` (numerical), `speechiness` (numerical), `acousticness` (numerical), `instrumentalness` (numerical), `valence` (numerical), `spotify_track_popularity` (numerical), and `main_genre` (categorical). Please see the Appendix for a complete data dictionary.

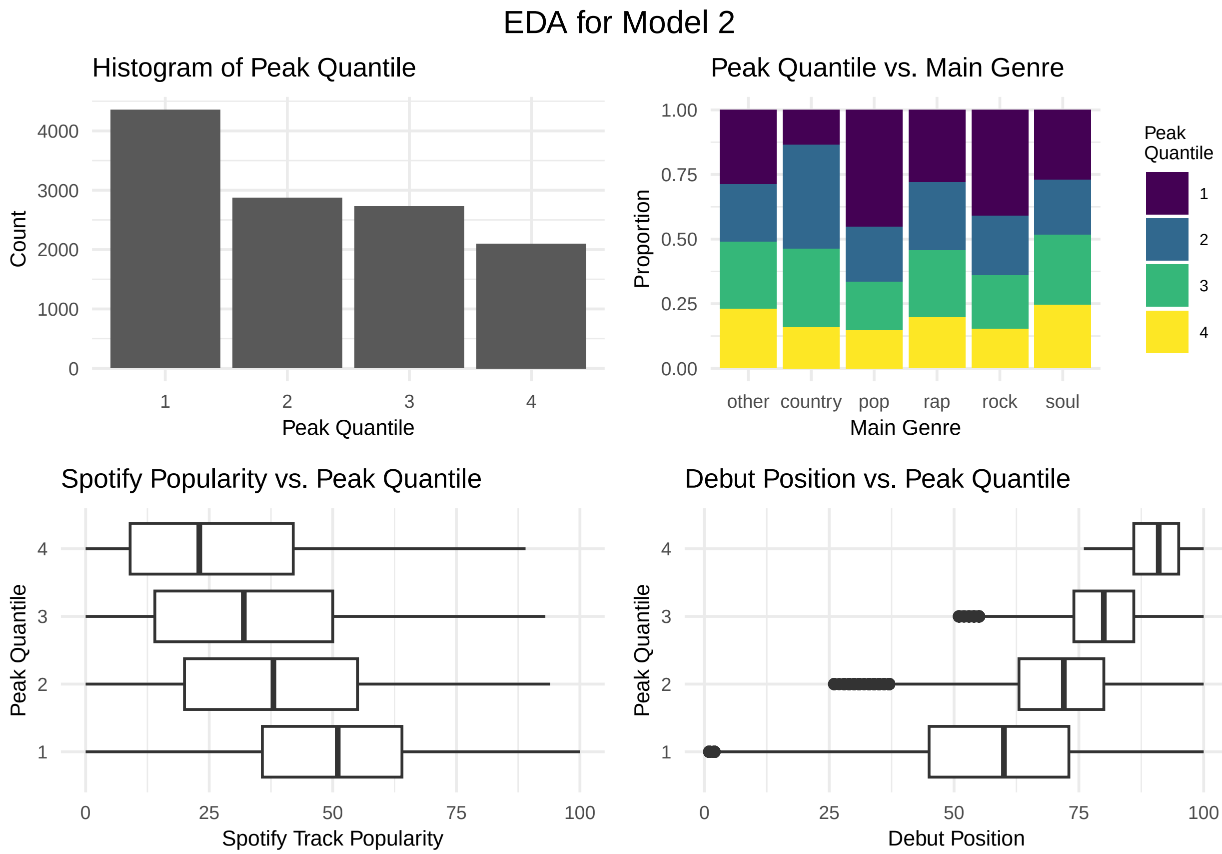
EDA



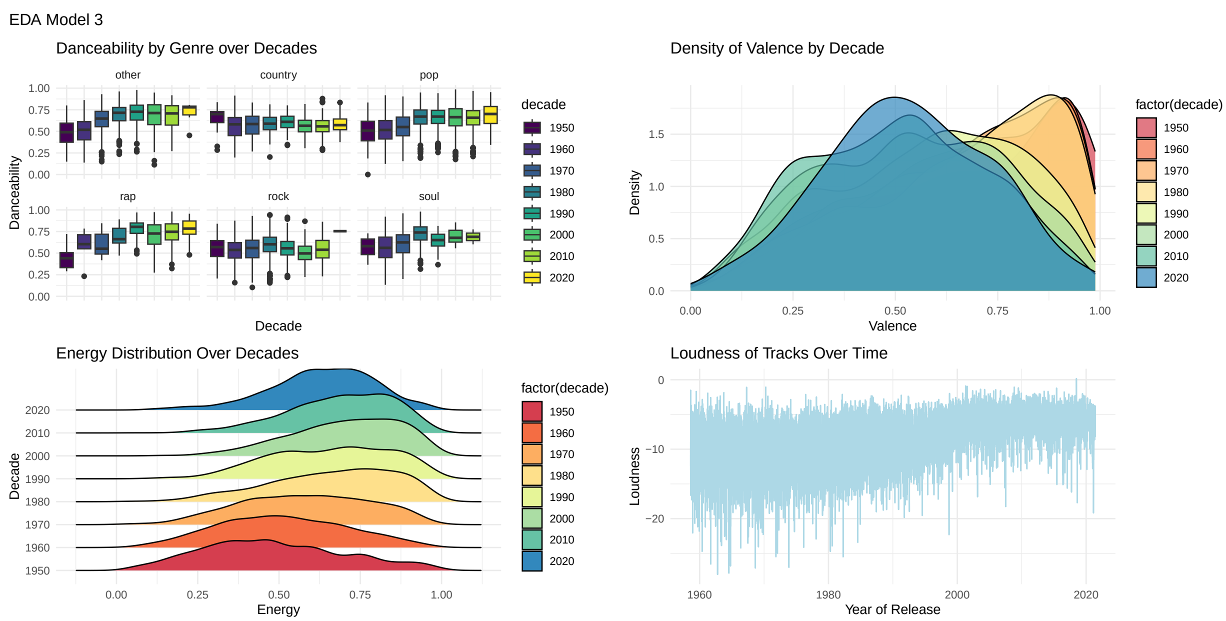
For ease of interpretation, we calculated the (full) months that a song is on the chart using $months\_on\_chart = \lfloor weeks\_on\_chart/4.33 \rfloor$ as a measure for longevity. For example, if a song stays on the chart for 5 weeks, it is categorized as 1 month on the Billboard.

The histogram of the `months_on_chart` shows a right skew with a majority of the songs staying on the Billboard for a month or less. The histogram also shows that it is rare for a song to stay on the Billboard for longer than a month but some songs have stayed from 4-12 months on the Billboard.

There appears to be a positive relationship between `loudness` and `months_on_chart`, and also between `spotify_track_popularity` and `months_on_chart`. According to the scatterplots, it also seems that songs remaining on the chart for longer also tend to be shorter, but this seems to be a weaker relationship. The relationship between `tempo` and `months_on_chart` also appears to be weak, but there is a slight pattern in that tempos in the range 75 to 175 tend to have higher `months_on_chart`. A similar relationship between `danceability` and `months_on_chart` is seen in the respective scatterplot where songs that remain on the Billboard the longest tend to have a `danceability` between 0.25 and 0.8. In general, most of these relationships appear to be non-linear.



The histogram of peak quantile shows that there are more songs that peaked in the first quantile while the other three quantiles are relatively evenly split. Pop and Rock have the greatest proportion of first quantile (top 25) songs while Country has the least. There seems to be a positive relationship between Spotify track popularity and peak quantile. Similarly, the higher a song debuts on the Billboard 100, the more likely it is to peak in an upper quantile. Interaction effects were also examined, but none of them appeared significant.



We created four plots for the decade model's EDA. The first compares danceability across decades by genre.

This depicts how we chose one of our interactions for model three, as each genre has different fluctuations across decades for this score. The other three EDA plots demonstrate various shifts in main effects. We see that valence has decreased over time, signalling that popular songs are stabilizing at a good medium for valence nowadays. We see a slightly opposite picture for energy, where the general trend is an increase in energy. Popular songs over the decades show more and more energy. And lastly, in terms of loudness, we see an approach to a healthy medium. Songs went from having high deviations to being similar in loudness. The maximum did not change, just the ranges.

Methodology

Model 1

Model 1 predicts the longevity, how long a song will stay on the Billboard, in terms of months based on different song features. Originally, a Generalized Additive Model (GAM) was chosen because the response variable is continuous and the distribution of the response variable is not a normal distribution but rather right-skewed. A GAM would also allow flexible nonlinearities in several variables, but retains the additive structure of linear models. However, there was still some variability in the response that could not be explained with this GAM.

After exploring a random forest model, this ensemble learning technique provided a more flexible model that seems to better capture the trends/pattern of the data. Variable selection was informed by an initial LASSO and then observing node purity contributions per predictor, along with the R^2 and accuracy values. The dataset was split into 80% training data and 20% testing data. We also tuned the number of trees in the forest by plotting MCE versus number of trees.

One key trade-off to note is that the more flexible random forest model does come at a slight loss of interpretability. The model is particularly useful for prediction, specifically predicting how long a song will stay on the Billboard given certain features.

Model 2

For Model 2, we use a cumulative logit model to predict what quantile of the Billboard Top 100 a song will peak in. We use this model because the response (`peak_quantile`) is ordinal categorical. The motivation behind creating such an ordinal categorical variable is that there is not much difference between peaking at position 34 versus position 35. Thus, splitting the chart into quantiles gives more meaningful analysis. The independent variables are all either continuous or categorical. `main_genre="other"` was chosen as the baseline level because it allows us to interpret how the top 5 genres fare on the Billboard 100 compared to all other genres. To check for multicollinearity, VIF was calculated for the continuous predictors, and none of them were above 10. This model predicts what quantile of the Billboard Top 100 a song will peak in. A key assumption is the proportional odds property which states that the effect of a parameter is identical for all 3 cumulative logits. The dataset was split into 80% training data and 20% testing data. 5 fold cross-validation was conducted on the training data and the accuracy was used as the performance metric. Interactions effects did not yield any substantial increases in accuracy, so the model will solely consist of main effects for easier interpretability.

Model 3

After making models demonstrating the qualities of songs that have longevity on the billboard and also that peak on the billboard, we move toward another perspective. For our third model, we used various predictors and interactions to help formulate a model that would predict whether a song was from a given decade. This model demonstrates another perspective on popular characteristics, as it analyzes the best songs throughout a decade, and determines the characteristics that make them successful within that time period. It will essentially highlight the characteristics of a song that essentially “make” a decade, and show how characteristics of songs either changed or stayed the same over the course of time. For this, we decided to use a multinomial model, as we have various categories (the decades) that may have very different relationships with the predictors, and there is no "order" (a given trait wouldn't consistently increase or decrease through decades, they have their own unique characteristics). For variable selection, there were several steps taken (not shown due to limited space, in appendix). Initially, we set up a multinomial CV-LASSO with lots of main effects and predictors, and analyzed what LASSO removed. From there, we actually made a random forest model with the left over predictors, and saw the importance that that model assigned, and iteratively removed (and added back in as necessary) predictors that were viewed as not as important as others. Once we had the main effects, we brainstormed interactions. Lots of the numeric to numeric interactions would probably be more difficult to interpret, so we decided to prioritize categorical on numeric interactions, and genre was a variable that proved to be very significant in that regard. As can be imagined, various aspects of a song can change drastically based on the genre. We ultimately landed with our final set of predictors, which we then put in through a regular mutinomial regression, and obtained our model. One key assumption for this model is the assumption of independence of irrelevant alternatives; that is, the probability of being in decade A or B shouldn't depend on whether decade C is included or not as a potential option.

Results

Model 1

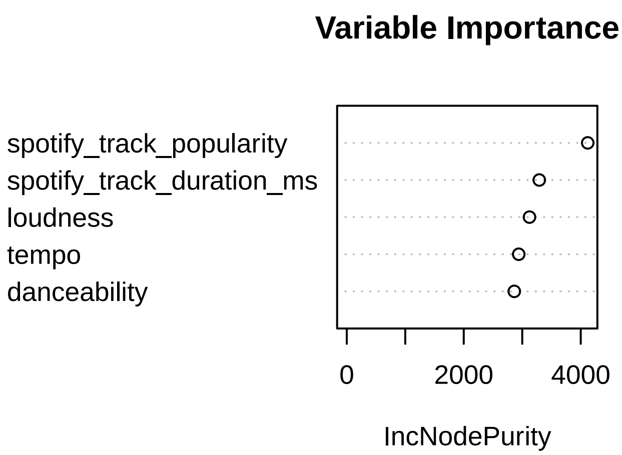


Table 1: Model 1 - Node Purity Increase

	IncNodePurity
danceability	2864.106
spotify_track_popularity	4119.686
tempo	2940.162
loudness	3124.234
spotify_track_duration_ms	3291.100

As seen in the variable importance plot, in order of decreasing importance, we have the variables `spotify_track_popularity`, `spotify_track_duration_ms`, `loudness`, `tempo`, and `danceability`. In this

Table 2: Model 1 - Results

Metric	Value
Training R-squared	0.785
Test R-squared	0.131
Test Mean Squared Error (MSE)	1.607
Accuracy on Test Data	60.606 %

case, since this is a regression random forest model, the node purity is measured by RSS and `IncNodePurity` refers to the total decrease in node impurity from splitting on that variable (averaged across all the trees).

By inspecting the trees further, we can see some general trends. It appears that there is a positive relationship among `spotify_track_popularity` and longevity. There are also certain ranges for `spotify_track_duration_ms` and `danceability` that are associated with higher time on the Billboard. The relationship between longevity and `loudness` and longevity and `tempo` is more complicated and cannot easily be inferred upon using the random forest model.

In general, this model is better suited for prediction rather than inference. We can predict how long a song with certain values for these five predictors will stay on the Billboard Top 100 (assuming that it is going to place in the Billboard Top 100).

Model 2

Table 3: Model 2 Output

	Estimate	2.5 %	97.5 %	OR
spotify_track_popularity	-0.035	-0.039	-0.031	0.966
main_genrecountry	0.419	0.086	0.752	1.520
main_genrepop	0.028	-0.231	0.286	1.028
main_genrerap	0.532	0.036	1.029	1.703
main_genrerock	0.225	-0.041	0.492	1.252
main_genresoul	0.113	-0.198	0.425	1.120
spotify_track_explicitTRUE	1.137	0.733	1.541	3.116
debut_position	0.103	0.096	0.110	1.108
danceability	0.039	-0.516	0.595	1.040

$logit[P(Y \leq j)] = 5.399 + 7.040 + 8.843 - 0.035 * spotify\_track\_popularity + 0.419 * main\_genrecountry + 0.028 * main\_genrepop + 0.532 * main\_genrerap + 0.225 * main\_genrerock + 0.113 * main\_genresoul + 1.137 * spotify\_track\_explicitTRUE + 0.103 * debut\_position + 0.039 * danceability$

The accuracy on the testing set is 55.168%. A parameter is statistically significant if the 95% confidence interval does not cross 0. Thus, `spotify_track_popularity`, `main_genrecountry`, `main_genrerap`, `spotify_track_explicitTRUE`, and `debut_position` are significant. The coefficient estimates for `main_genrecountry` and `main_genrerap` are positive which means that songs with a main genre of country and rap are 1.5 and 1.7 times more likely to to be in an upper quantile (1 or 2), respectively, than songs with a main genre of other. For every one unit increase in `spotify_track_popularity`, the odds of peaking in an upper quantile (1 or 2) decreases by 3.4%, holding all else constant. This is counterintuitive and pretty strange. Perhaps this could partly be due to TikTok or viral songs that explode in popularity on social media and digital platforms, but do not have as much physical sales and presence which is a component in calculating performance on the Billboard 100. Similarly, for every one unit increase in `debut_position`, the odds of peaking in an upper quantile increases by 10.8%, holding all else constant. This makes sense. For explicit songs, the odds of peaking an an upper quantile is 3.116 times that of not explicit songs, holding all else constant. Thus, `spotify_track_explicit` is the most influential predictor in this model.

Model 3

Table 4: True vs. Predicted Decades (Truth along Side)

	1950	1960	1970	1980	1990	2000	2010	2020
1950	0	0	1	0	0	1	0	0
1960	55	422	111	11	6	2	2	2
1970	1	59	225	104	35	9	1	0
1980	0	11	101	216	99	12	4	0
1990	0	9	19	51	103	41	14	0
2000	0	4	7	16	53	181	46	0
2010	0	3	1	6	19	71	236	25
2020	0	2	1	1	0	0	7	4

1960 Model:

$log(P(1960)/P(1950)) = 3.2953 + 0.4523 * danceability + -0.6911 * energy + -0.0192 * loudness + -0.7603 * speechiness + -2.5577 * acousticness + 0.2771 * instrumentalness + -0.6315 * valence + 0.005 * spotify\_track\_popularity + -1.4101 * main\_genrecountry + -1.3263 * main\_genreother + 5.2399 * main\_genrerap + 0.6908 * main\_genrerock + 1.5588 * main\_genresoul + -0.9859 * main\_genrecountry : spechiness + 1.7541 * main\_genreother : spechiness + 4.0097 * main\_genrerap : spechiness + -1.9556 * main\_genrerock : spechiness + -5.133 * main\_genresoul : spechiness + -0.0543 * main\_genrecountry : loudness + -0.0593 * main\_genreother : loudness + 0.9408 * main\_genrerap : loudness + 0.0061 * main\_genrerock : loudness + 0.0011 * main\_genresoul : loudness + -0.0286 * main\_genrecountry : acousticness + 0.4906 * main\_genreother : acousticness + 1.6463 * main\_genrerap : acousticness + -1.3043 * main\_genrerock : acousticness + -0.2602 * main\_genresoul : acousticness + 0.2497 * main\_genrecountry : danceability + 0.0522 * main\_genreother : danceability + 3.7141 * main\_genrerap : danceability + -0.9984 * main\_genrerock : danceability + -0.2535 * main\_genresoul : danceability$

We came out with a total of seven models (we had eight decades and the first was a reference category). For reference, we have the model for the 1960s here, and because each is very long, the rest are in the appendix. We will refer to the 1960s model for some interpretations (as well as those models in the appendix for comparisons. Two coefficient interpretations are as follows. For the decade of 1960, the coefficient for energy is -0.6911, which suggests that for this particular decade, as the energy of a track increases, the log-odds of the track being from the 1960s (as opposed to the reference category, the 1950s) decrease. This indicates that songs with higher energy are less likely to be from the 1960s, relative to the 1950s, when holding all other variables constant. For the decade of 1970, the coefficient for the interaction term between the country genre and danceability is 1.705. This suggests a multiplicative interaction effect on the log-odds of a song being from the 1970s (versus the 1950s). Specifically, for country songs, as danceability increases, the log-odds of the song being from the 1970s increase more steeply than what would be predicted by the main effect of danceability alone. This indicates that for country songs, higher danceability is particularly characteristic of the 1970s as opposed to the 1950s. Additionally, we can pick out characteristics that seem to be important at certain times or identify when trends start. For example, danceability seemed to be an important and defining characteristic starting in the 1980s, as its coefficient significantly rose then and stayed at that. Additionally, on the flip side, in 1970, it seems like speechiness reached an all time low, where an increase in it, specifically in the rap genre, heavily decreases the log odds of being in 1970 with reference to 1950. Our models predict a decade with a 57% accuracy. While this may seem low, there are actually eight decades, which means the average guess would be right 12.5% of the time, and we top that by just about 5x. Additionally, if you analyze the confusion matrix, it can be seen that a majority of the incorrect predictions were from immediately neighboring decades (IE predicting 1970 or 1950 for a 1960s song). This brings up an important point: the idea that decade boundaries are not hard boundaries in which song elements change - these things change over time and can be split across decades. For example, a song from 1971 might be very similar in terms of its elements to one from 1969, even though they are in different decades. Taking this into account, a lot of the error becomes explainable. Additionally, we provide specificity, sensitivity, and various other helpful analytics for our model in the appendix for further accuracy analysis.

Conclusion

The project explored the attributes that contribute to the success of songs in the U.S. Billboard Hot 100 across decades. Three models were developed to predict a song's Billboard longevity, peak quantile, and the decade of release, using features such as genre, danceability, and Spotify popularity. The models used, including random forest and multinomial, provide insights for music producers and artists aiming for chart success, as well as those trying to analyze past music trends. The analysis highlighted the evolving trends in music preferences, with implications for the music industry's marketing and production strategies.

A common predictor across all three models is `spotify_track_popularity`, showing that digital performance is a strong indicator of Billboard chart performance. Model 1 suggests that Spotify artists with higher popularity scores tend to have songs that last longer on the Billboard. On the other hand, Model 2 indicates higher popularity scores are slightly less likely to peak in an upper quantile. This discrepancy may suggest that artists who are more underrated, not as widely popular, tend to quickly peak for a shorter period of time. While artists who are more well-known and popular tend to stay on the Billboard longer even if their songs do not peak as high on the Billboard. This discrepancy may have been exacerbated by promotions of songs through TikTok and other social media that tends to use clips of songs. Short clips might prioritize catchy or memorable segments of a song that make the song peak higher yet the overall artist may not be as popular so the song's popularity changes or drops quickly (in a short period of time) once the song's clip is no longer viral.

`spotify_track_explicit` and `genre` were both significant predictor variables for peak quantile, however, they were not significant for longevity. This suggests that the genre and whether a song is explicit is more helpful in determining the peak quantile of a song rather than how long it will stay on the Billboard.

Model 1 and Model 3 offer differing insights into what makes a song popular. While a lot of different predictors, including interaction effects, were significant in predicting decade as included in Model 3, there were little predictors that appeared to be significant in predicting longevity (as seen in LASSO as well) in Model 1. This might suggest that other features not included in this data might be useful when predicting longevity, an area for potential future work. In general, this disparity underscores the nuanced nature of factors that influence song popularity, emphasizing the importance of considering various dimensions when analyzing and forecasting musical trends.

Producers, musical artists, and music critics can use these models to compare and contrast what success in music means to them. If they are aiming to please a certain generational audience or if they just want to place at the highest possible position on the Billboard, they can use these models to inform themselves. For example, an application of Model 2 may be a musical artist using the prediction of how long a song will stay on the Billboard to inform their marketing and promotional strategies.

The models, while performative, also have the major limitation of the data being from the Billboard 100. That means that this data is highly skewed towards those in the US, not internationally. Additionally, because it only grabs the top 100 songs, it does not take into account broader music trends outside the songs that may be the most listened to (which may be similar, and those not listened to as much very different), so it would be hard to generalize to all music. Future research would attempt to analyze more board music trends, in hopes to provide more holistic and applicable results.

Appendix

Data Dictionary

Table 5: Billboard Information

variable	description
url	Billboard Chart URL
week_id	Week ID
week_position	Week position 1: 100
song	Song name
performer	Performer name
song_id	Song ID, combo of song/singer
instance	Instance (this is used to separate breaks on the chart for a given song. Example, an instance of 6 tells you that this is the sixth time this song has appeared on the chart)
previous_week_position	Previous week position
peak_position	Peak position as of that week
weeks_on_chart	Weeks on chart as of that week

Table 6: Audio Features

variable	description
song_id	Song ID
performer	Performer name
song	Song
spotify_genre	Genre
spotify_track_id	Track ID
spotify_track_preview_url	Spotify URL
spotify_track_duration_ms	Duration in ms
spotify_track_explicit	Is explicit
spotify_track_album	Album name
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. A value of 0.0 is least danceable and 1.0 is most danceable.
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
key	The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation.
loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track.
mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived.
speechiness	Speechiness detects the presence of spoken words in a track. Values above 0.66 describe tracks that are probably made entirely of spoken words.
acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
instrumentalness	Predicts whether a track contains no vocals. Values above 0.5 are intended to represent instrumental tracks.
liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.
tempo	The overall estimated tempo of a track in beats per minute (BPM).
time_signature	Time signature
spotify_track_popularity	Popularity

The Rest of Model 3's Multinomial Models:

1970 Model:

$log(P(1970)/P(1950)) = -5.5288 + 0.51 * danceability + 1.1969 * energy + -0.2078 * loudness +$

$$-4.9755 * speechiness + -4.3317 * acousticness + -0.1476 * instrumentalness + -1.0163 * valence + 0.0107 * spotify_{t}rack_{p}opularity + -0.2628 * main_{g}enrecountry + -2.7827 * main_{g}enreother + -9.0596 * main_{g}enrerap + 0.6047 * main_{g}enrerock + 2.336 * main_{g}enresoul + -3.1058 * main_{g}enrecountry : spechiness + 9.2831 * main_{g}enreother : spechiness + -7.3211 * main_{g}enrerap : spechiness + 5.1368 * main_{g}enrerock : spechiness + 2.2475 * main_{g}enresoul : spechiness + 0.1427 * main_{g}enrecountry : loudness + 0.0486 * main_{g}enreother : loudness + -0.6554 * main_{g}enrerap : loudness + 0.1023 * main_{g}enrerock : loudness + 0.1487 * main_{g}enresoul : loudness + 2.9443 * main_{g}enrecountry : acousticness + -0.3741 * main_{g}enreother : acousticness + -4.4072 * main_{g}enrerap : acousticness + -0.0851 * main_{g}enrerock : acousticness + 0.0142 * main_{g}enresoul : acousticness + 1.705 * main_{g}enrecountry : danceability + 4.0021 * main_{g}enreother : danceability + -9.3302 * main_{g}enrerap : danceability + 1.3878 * main_{g}enrerock : danceability + 0.8997 * main_{g}enresoul : danceability$$

1980 Model:

$$log(P(1980)/P(1950)) = -10.3033 + 7.7816 * danceability + 4.0161 * energy + -0.2065 * loudness + -14.3539 * speechiness + -5.4317 * acousticness + -2.1458 * instrumentalness + -4.4328 * valence + 0.0225 * spotify_{t}rack_{p}opularity + -2.4413 * main_{g}enrecountry + -2.0745 * main_{g}enreother + 3.844 * main_{g}enrerap + 1.5594 * main_{g}enrerock + -0.1727 * main_{g}enresoul + 1.262 * main_{g}enrecountry : spechiness + 13.281 * main_{g}enreother : spechiness + -2.8916 * main_{g}enrerap : spechiness + 7.8854 * main_{g}enrerock : spechiness + 0.2124 * main_{g}enresoul : spechiness + -0.0783 * main_{g}enrecountry : loudness + -0.0257 * main_{g}enreother : loudness + 0.547 * main_{g}enrerap : loudness + 0.1051 * main_{g}enrerock : loudness + 0.1629 * main_{g}enresoul : loudness + 2.2781 * main_{g}enrecountry : acousticness + -1.0437 * main_{g}enreother : acousticness + -2.9391 * main_{g}enrerap : acousticness + -0.1998 * main_{g}enrerock : acousticness + 0.428 * main_{g}enresoul : acousticness + 0.2551 * main_{g}enrecountry : danceability + 1.607 * main_{g}enreother : danceability + 0.9944 * main_{g}enrerap : danceability + -0.4825 * main_{g}enrerock : danceability + 2.0955 * main_{g}enresoul : danceability$$

1990 Model:

$$log(P(1990)/P(1950)) = -7.7779 + 8.3081 * danceability + 2.1606 * energy + -0.063 * loudness + -1.1221 * speechiness + -6.7169 * acousticness + -1.7483 * instrumentalness + -6.5493 * valence + 0.0361 * spotify_{t}rack_{p}opularity + 3.9303 * main_{g}enrecountry + -1.6879 * main_{g}enreother + 2.8991 * main_{g}enrerap + 0.9987 * main_{g}enrerock + 2.755 * main_{g}enresoul + -8.6779 * main_{g}enrecountry : spechiness + 7.7092 * main_{g}enreother : spechiness + 7.1969 * main_{g}enrerap : spechiness + -2.7829 * main_{g}enrerock : spechiness + -1.0664 * main_{g}enresoul : spechiness + 0.3312 * main_{g}enrecountry : loudness + 0.0304 * main_{g}enreother : loudness + 0.6787 * main_{g}enrerap : loudness + 0.1599 * main_{g}enrerock : loudness + 0.2187 * main_{g}enresoul : loudness + 0.3539 * main_{g}enrecountry : acousticness + -1.0344 * main_{g}enreother : acousticness + 3.6 * main_{g}enrerap : acousticness + -0.5036 * main_{g}enrerock : acousticness + 1.3613 * main_{g}enresoul : acousticness + 0.3364 * main_{g}enrecountry : danceability + 2.3041 * main_{g}enreother : danceability + 3.8208 * main_{g}enrerap : danceability + -0.4434 * main_{g}enrerock : danceability + -3.7446 * main_{g}enresoul : danceability$$

2000 Model:

$$log(P(2000)/P(1950)) = -0.8813 + 7.2424 * danceability + -1.746 * energy + 0.4754 * loudness + 7.6145 * speechiness + -7.0696 * acousticness + -2.0077 * instrumentalness + -7.3522 * valence + 0.0656 * spotify_{t}rack_{p}opularity + 5.9418 * main_{g}enrecountry + -0.6901 * main_{g}enreother + 6.1528 * main_{g}enrerap + 2.0198 * main_{g}enrerock + -4.1794 * main_{g}enresoul + -5.5287 * main_{g}enrecountry : spechiness + -0.1747 * main_{g}enreother : spechiness + -1.6863 * main_{g}enrerap : spechiness + -10.5875 * main_{g}enrerock : spechiness + 0.4726 * main_{g}enresoul : spechiness + 0.2434 * main_{g}enrecountry : loudness + 0.1332 * main_{g}enreother : loudness + 0.6762 * main_{g}enrerap : loudness + 0.3168 * main_{g}enrerock : loudness + -0.0219 * main_{g}enresoul : loudness + -1.1022 * main_{g}enrecountry : acousticness + -0.4061 * main_{g}enreother : acousticness + 2.5502 * main_{g}enrerap : acousticness + -2.5652 * main_{g}enrerock : acousticness + 2.5992 * main_{g}enresoul : acousticness + -1.6598 * main_{g}enrecountry : danceability + 1.5687 * main_{g}enreother : danceability + 1.0071 * main_{g}enrerap : danceability + -1.0273 * main_{g}enrerock : danceability + 2.4974 * main_{g}enresoul : danceability$$

2010 Model:

$$log(P(2010)/P(1950)) = -1.5182 + 7.2937 * danceability + -2.1352 * energy + 0.5654 * loudness + 8.0869 * speechiness + -7.2529 * acousticness + -1.8501 * instrumentalness + -8.7048 * valence + 0.1371 * spotify_track_popularity + 6.4445 * main_genrecountry + -1.4319 * main_genreother + 3.9797 * main_genrerap + -3.3771 * main_genrerock + -3.6829 * main_genresoul + 6.1198 * main_genrecountry : spechiness + -2.2688 * main_genreother : spechiness + -0.9553 * main_genrerap : spechiness + -7.3191 * main_genrerock : spechiness + 0.2048 * main_genresoul : spechiness + 0.3367 * main_genrecountry : loudness + -0.0808 * main_genreother : loudness + 0.5108 * main_genrerap : loudness + -0.081 * main_genrerock : loudness + -0.3566 * main_genresoul : loudness + -2.0688 * main_genrecountry : acousticness + -0.5294 * main_genreother : acousticness + 5.3205 * main_genrerap : acousticness + -1.833 * main_genrerock : acousticness + -0.4592 * main_genresoul : acousticness + -1.6202 * main_genrecountry : danceability + 1.1619 * main_genreother : danceability + 2.5183 * main_genrerap : danceability + 1.7311 * main_genrerock : danceability + -1.8554 * main_genresoul : danceability$$

2020 Model:

$$log(P(2020)/P(1950)) = -3.1763 + 7.3038 * danceability + -2.3451 * energy + 0.4795 * loudness + 8.4173 * speechiness + -5.8675 * acousticness + -1.377 * instrumentalness + -8.1562 * valence + 0.1768 * spotify_track_popularity + 6.6622 * main_genrecountry + -1.6113 * main_genreother + 3.2819 * main_genrerap + -6.0408 * main_genrerock + -2.449 * main_genresoul + 1.9767 * main_genrecountry : spechiness + -2.6152 * main_genreother : spechiness + 1.9665 * main_genrerap : spechiness + -1.3107 * main_genrerock : spechiness + -0.971 * main_genresoul : spechiness + 0.1761 * main_genrecountry : loudness + -0.3097 * main_genreother : loudness + 0.7338 * main_genrerap : loudness + -0.4172 * main_genrerock : loudness + -0.5245 * main_genresoul : loudness + -2.6212 * main_genrecountry : acousticness + -1.7093 * main_genreother : acousticness + 5.7238 * main_genrerap : acousticness + -3.3213 * main_genrerock : acousticness + -1.7681 * main_genresoul : acousticness + -1.7668 * main_genrecountry : danceability + -0.8422 * main_genreother : danceability + 5.1181 * main_genrerap : danceability + 1.7415 * main_genrerock : danceability + -3.1294 * main_genresoul : danceability$$

Table 7: Model 3: Head of RF Importance Score

	Variable	MeanDecreaseAccuracy
spotify_track_duration_ms	spotify_track_duration_ms	85.74760
spotify_track_popularity	spotify_track_popularity	70.17760
loudness	loudness	53.14224
acousticness	acousticness	47.03562
valence	valence	46.60747
danceability	danceability	44.28058

Table 8: Accuracy Statistics for Model 3

	x
Accuracy	0.576
Kappa	0.490
AccuracyLower	0.555
AccuracyUpper	0.595
AccuracyNull	0.212
AccuracyPValue	0.000
McnemarPValue	NaN

Table 9: Additional Statistics for Model 3

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Precision	Recall	F1	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
Class: 1950	0.000	0.999	0.000	0.977	0.000	0.000	NaN	0.023	0.000	0.001	0.500
Class: 1960	0.827	0.901	0.691	0.951	0.691	0.827	0.753	0.212	0.175	0.254	0.864
Class: 1970	0.483	0.892	0.518	0.878	0.518	0.483	0.500	0.193	0.093	0.180	0.688
Class: 1980	0.533	0.887	0.488	0.904	0.488	0.533	0.509	0.168	0.090	0.184	0.710
Class: 1990	0.327	0.936	0.435	0.902	0.435	0.327	0.373	0.131	0.043	0.098	0.632
Class: 2000	0.571	0.940	0.590	0.935	0.590	0.571	0.580	0.132	0.075	0.127	0.755
Class: 2010	0.761	0.940	0.654	0.964	0.654	0.761	0.703	0.129	0.098	0.150	0.851
Class: 2020	0.129	0.995	0.267	0.989	0.267	0.129	0.174	0.013	0.002	0.006	0.562