Started from the Bottom Now We Here:
Predicting the Top Releases in the Billboard Hot 100 Charts

Lander Basterra, Brady Fowler, Monica Rajendiran, Gabe Rushin

Executive Summary

In this paper we determine if logistic regression models can predict whether or not a track will rise to the top 10 of the Billboard Hot 100. The Billboard Hot 100 chart reflects the most popular current tracks across radio stations, streaming services and both digital and physical purchases. Billboard Charts have been used since the 1940s as a consistent means of ranking and comparing the popularity of specific tracks across a range of mediums.

Historically, tracks enter the chart at a low rating and move up in the rankings as they gain traction nationally. We postulate that chart conditions like the average age of a top-ten song or how many prior hits a artist has achieved may be important in predicting a new track's success. This is based on the concept that a new track may gain more traction if the current chart consists of "stale" songs since listeners may actively seek new material. We collected the chart-related data by programmatically scraping the Billboard Hot 100 website.

Furthermore, we hypothesized that specific audio traits may play a part in a song's success on the charts. Tunes that are more energetic and danceable may achieve a higher number of spins due to play counts from nightlife establishments and the propensity of energetic tracks to achieve virality. We retrieved a thorough array of audio descriptors by scraping the Spotify API, retrieving features like tempo, duration, energy, acousticness and more.

We used a variety of model selection techniques (hybrid stepwise and manual p-value inspection) to choose and validate the best models to predict chart success, measured as a binary outcome of reaching the top 10 or not. We then compared each model's predictive abilities by using cross fold validation and ROC curves based on a hold-out training set consisting of every track released from 2014 to the present. Our best model achieved an overall accuracy of 73.79% and AUC of 68.8% on our held out testing set.

While our models satisfactorily outperformed a random-coin flip model, we believe that additional features could further improve our ability to predict a track's rise. By incorporating other measures of social media traction we could incorporate currently unmeasured fanbase potential. Our models currently project only on the first-week chart state of a track's release, but if we expanded our focus to incorporate the performance/trajectory over the first few weeks we could likely achieve even better results.

Background

In today's streaming and playlist-centric environment, chart popularity serves as a signal to both fans and tastemakers as to what songs are attracting the most national buzz. In the **Cambridge Companion to Pop and Rock,** musicologist Simon Frith explains:

> *"Until quite recently (and certainly when compared to the film and television industries), the music industry did little formal market research, except in the form of the charts. And even in the charts, a measure of what is selling and what, given airplay figures, is likely to sell, are primarily used as a means of stock control…"*[1]

We use statistical methods to leverage historical Billboard weekly data to infer which factors influence a song's popularity (as determined by whether it reaches the top 10 of the charts). This objective has real-world applications; companies like Shazam and Pandora subsidiary Next Big Sound have combined social network trends and Billboard performance to predict artist popularity.[2,3] We believe our approach is novel in its narrowed focus to only produce snapshot "top-ten" or "not-top-ten" predictions for a track in its initial week on the charts.

Data

A large number of factors may influence a track's chart success. The Billboard Hot 100 website[4] provides a weekly list of the top 100 ranked tracks, including the chart date, title, artist name, peak position, last week position, rank, change in rank, and an option to listen to the track through the music streaming service Spotify. Our data collection began by using the 'billboard' Python module to create a program that extracted each Billboard chart from the present week back through January 1, 1995 and stored the information as comma separated values.

Each weekly chart contains features describing each song, including song title, artist name, the track's peak position over all time, last week's chart position, and number of weeks since release. The Billboard Hot 100 Chart enables listeners to interactively play tracks on their website through a Spotify plugin which relies on a unique Spotify track ID for each track. We used this identifier to ping the Spotify API[5] in order to retrieve additional features describing a song's "audio fingerprint". The features retrieved from Spotify included: acousticness, danceability, duration in milliseconds, energy, instrumentalness, key, liveness, loudness, mode, speechiness,

---

[1] Frith, Simon, Will Straw, and John Street. *The Cambridge Companion To Pop and Rock*. New York: Cambridge UP, 2001. Print. Page 34.
[2] "Next Big Sound." *Analytics and Insights for the Music Industry*. Next Big Sound, n.d. Web. 11 Oct. 2016. <https://www.nextbigsound.com/charts/predictions>.
[3] Knopper, Steve. "Can Shazam Predict the Next Big Hit?" *Rolling Stone*. N.p., 20 Feb. 2014. Web. 11 Oct. 2016. <http://www.rollingstone.com/music/news/can-shazam-predict-the-next-big-hit-20140220>.
[4] http://www.billboard.com/charts/hot-100.
[5] https://developer.spotify.com/web-api/

tempo, time signature, and valence. Spotify uses these measures internally in their music recommendation systems since collectively, the measures quantify a song's defining characteristics. For our purposes, these audio-fingerprint features provide a way to predict what audio-characteristics make a track more likely to achieve Billboard 100 top ten popularity. These musical uniqueness traits extracted through the Spotify python module were later merged with our previously retrieved data on the song's chart information.

Missingness

Initial inspection revealed that 8.86% of the tracks had missing features. Further analysis showed that the only feature with missingness in this data set is the Spotify ID (and consequently the audio features pulled from the Spotify API). We determined this pattern to be missing not at random (MNAR). The MNAR data is due to the fact that not all artists that are featured in the Billboard charts are similarly available on Spotify.

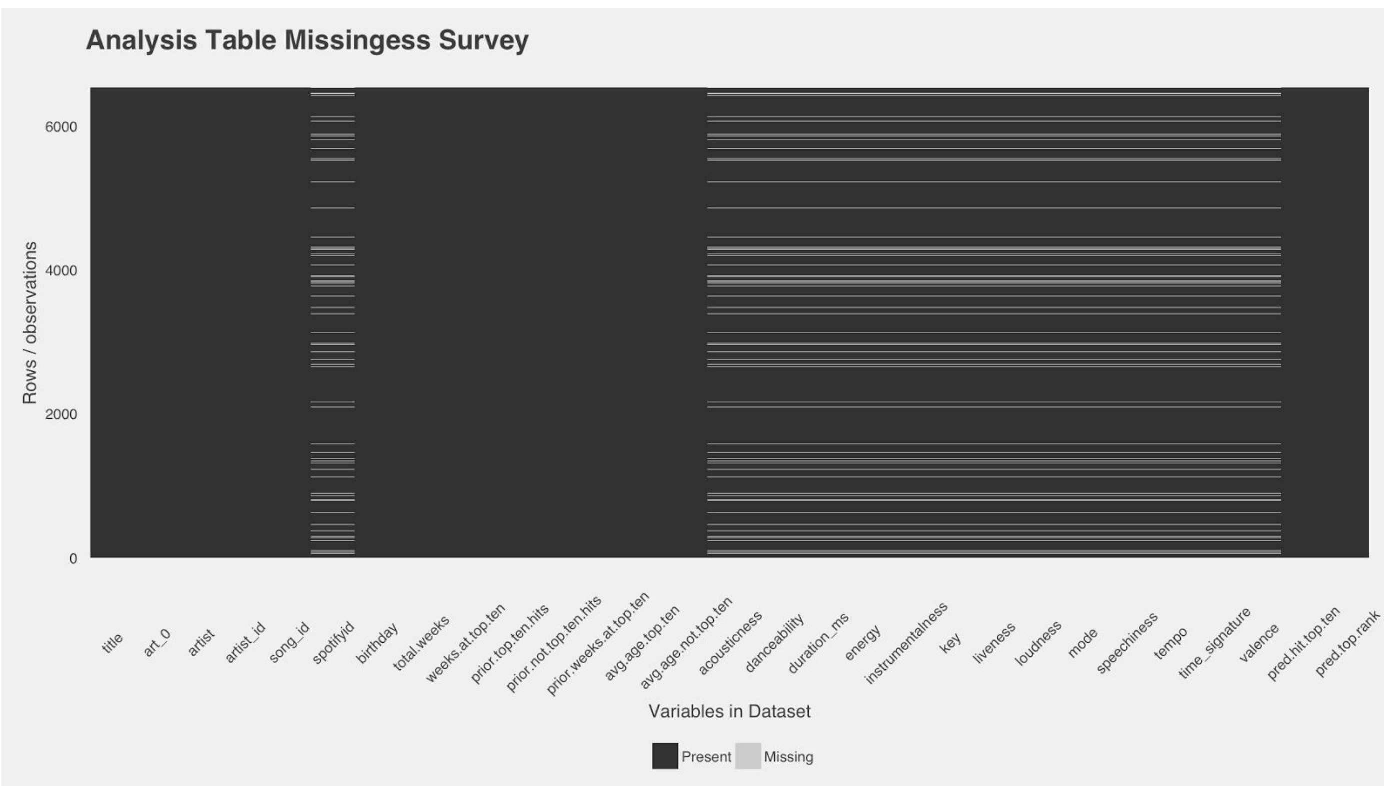


Figure 1. Missingness across all variables

Several notable popular artists not on Spotify (including Beyonce, Prince, and Taylor Swift) have chosen not to make their tracks available on the streaming service due to contract disputes over the payment model. We came to the conclusion that since artist songs stylistically and sonically differ greatly from one another, it does not make sense to try to impute audio traits for songs that were missing a Spotify artist ID. For example, using the Katy Perry discography to impute Taylor Swift tracks does not make conceptual sense. Furthermore, regression imputation is unusable because entire blocks of features are missing from an observation instead of just one (see chart above).

In order to calculate artist-level features (like number of previous top ten hits) we needed a reliable artist-level ID. Whenever a track had a Spotify ID we were able to rely upon the cleaned list of artist names from their API which we pulled while extracting the audio features. However, when a Spotify ID was absent, we had to create a parting function that split on key-words like 'featuring', 'with', 'duet with' etc in order to isolate the primary and secondary artists. Manual inspection of this parsing process revealed satisfactory cleaning and artist name separation.

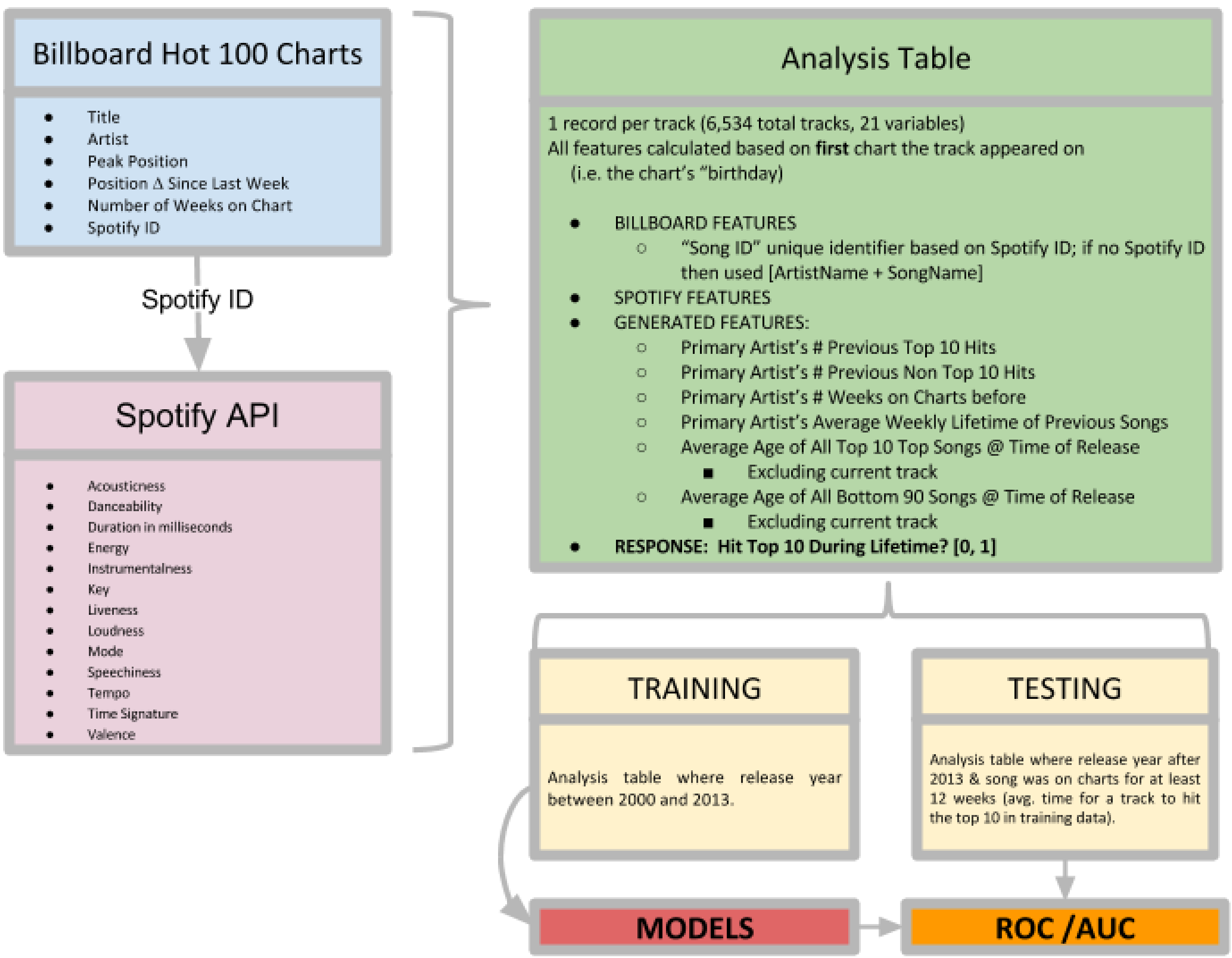


Figure 2. Predicting Top-Ten Success of a Song Workflow

After cleaning and creating the data set containing all chart and audio trait information for each song between 1995 and 2016 weekly charts, we pulled the data set into R for feature engineering subsequent modeling and analysis. We created a distinct list of songs for our base data set that would be used for analysis. We then engineered additional predictors based on the primary artist such as: the number of previous top 10 hits, the number of previous non top 10 hits, the number of weeks the artist had songs on the charts before, and the average weekly lifetime of previous songs on the chart. The full list of predictors both extracted from the chart information and Spotify, as well as those feature engineered are in Table A2 in the Appendix. The complete data flow is diagrammed above.

Analysis

Our main hypothesis is:

H_o: Features describing a Billboard Hot 100 song can **NOT** predict a song's top-ten success.

H_a: Features describing a Billboard Hot 100 song can predict a song's top-ten success.

Since we sought to predict the likelihood of a song's popularity given chart-conditions and audio features, we believed that logistic regression would be the most appropriate modeling technique. Through the steps described in the previous section, our data was adequately cleaned, missingness was removed and necessary features were engineered. To understand how the chart/audio features differ between top 10 and bottom 90 tracks over time we built a series of visualizations attached in the appendix. We then generated an array of logistic regression models trained on different subsets of our analysis data (which contains both chart and audio information for each song as-of its release day). For example, we looked at the sole effects of audio trait information (e.g. liveness, danceability) on predictability of whether a song would, be in the top 10 of the Billboard 100 charts.

We also looked at the possibility of adding interaction terms to our full data model. We only tested the ones that proved to be statistically significant (p-value < 0.05). After an extensive analysis of these significant terms, we proved that they did not improve the results of our existing model and decided not to include any since new models were performing worse or roughly the same. Results for the performance of interaction terms can be found in Table A1 in the Appendix. Finally, since the information sources and rules that dictate the chart have been modified over the years to better describe changes in music consumption[6], we chose to limit our

[6] "… a new "recurrent rule" goes into effect on the Hot 100: descending songs are now removed from the chart if ranking below No. 25 after 52 weeks, as well as if ranking below No. 50 after 20 weeks. Previously only the latter measure was applied."
http://www.billboard.com/articles/columns/chart-beat/6770427/adele-tops-hot-100-fourth-week-justin-bieber-alessia-cara

analysis to approximately the last decade: 2000 to 2016.[7] In order to test the predictive accuracy of the final chosen model, we held out any track released after 2014 that had been on the charts for at least 12 weeks.[8]

The logistic regression models that we created for each subset of features, along with their resulting evaluation metrics are listed in the table below. For our logistic regression models we subsetted the data to the corresponding information (see column "Subset") and then used hybrid stepwise model selection, removing insignificant variables to pare down the model to only significant variables. We also checked for multicollinearity and tested the inclusion of any potential interaction terms - all final models presented below did not have issues with multicollinearity so no interaction terms were included. The cross validated accuracy presented is computed from our training data while the prediction accuracy score is computed from our held-out test set (data after 2014).

Subset	**Model**	**CV Accuracy**	**Prediction Accuracy**	**AUC**
Full Data (Music Traits + Chart Info)	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + prior.top.ten.hits + mode + avg.age.not.top.ten + avg.age.top.ten + acousticness + energy + loudness + valence + duration_ms + time_signature	85.60%	73.79%	68.8%
Music Traits	**pred.hit.top.ten** ~ danceability + mode + duration_ms + acousticness + energy + loudness + valence	82.805%	67.96%	60.7%
Chart Info	**pred.hit.top.ten** ~ prior.not.top.ten.hits + prior.top.ten.hits	84.08%	75.73%	70.3%

Table 1. Logistic Regression Models and Evaluation Metrics

Conclusions

We do not reject our null hypothesis and conclude that features describing a Billboard Hot 100 song can in fact predict a song's top-ten success with better accuracy than a random coin flip. The residual deviance analysis of our best logistic regression models indicate that the inclusion of key audio and chart-condition features predict better than a model with nothing but an intercept. The receiver operating characteristics curves and associated AUC measures suggest that our models may not have the advanced ability to pinpoint hot tracks but could still be useful

[7] We included records dating back to 1995 during the feature engineering and missingness analysis stages but only used 2000-2016 records in our models. This lagged process was intentional since it allowed us to generate accurate calculations for the "number of previous hits" and other similar features for tracks on the 2000 charts.
[8] The 12 week lifespan rule was instituted to remove any new 2016 tracks that still might be on their upward trajectory. The average track in the training data that hit the top 10 took 12 weeks to reach that peak.

in ranking tracks by their predicted-hit likelihood. We discuss our final chosen models and their adherence to assumptions below:

Logistic Regression

The full data set used to predict a song's inclusion in the top 10 ranks of the Billboard 100 charts contained both audio traits (e.g. danceability) and chart information (e.g. prior top 10 hits by the respective artist) so we created three different logistic regression models to determine the effects of the audio trait predictors vs. the chart information predictors vs. both audio trait and chart information predictors. Based on each model's cross-validated accuracy, prediction accuracy, and AUC score, the full data model and the chart information model performed the best, with the chart information model performing marginally better.

Since the chart information model only performed marginally better than the full data model and contained only two predictors after stepwise hybrid selection, we decided to go with the full data model for determining the top-ten success predictability. The most significant predictors in our best model were danceability, prior.top.ten.hits (the number of songs by the artist that had been a top 10 hit), prior.not.top.ten.hits, mode, and energy. Acousticness, loudness, duration_ms (duration of the song in minutes and seconds), and valence were less significant while avg.age.not.top.ten (the average age of the artist's songs not in the top ten) and time_signature were not significant. Significance was determined by the coefficient value of each predictor and its corresponding p-value.

We can extract several inferences from our best model, based on the resulting coefficients produced. For example, our model indicates that, holding all other predictors at a fixed value, we can expect that for every additional unit of danceability (which is on a score of 0.0 to 1.0 with 1.0 being the most danceable) there will be 563.93% increase in the odds of a song being a top-ten success (odds ratio = 6.639). We can also expect that for every additional unit of energy, there will be an 80.27% decrease in the odds of a song being a top-ten success (odds ratio = 0.197).

```
Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)           -2.442e+00  8.973e-01  -2.721  0.00650 **
danceability           1.893e+00  3.428e-01   5.523 3.33e-08 ***
prior.not.top.ten.hits -6.034e-02  8.611e-03  -7.007 2.43e-12 ***
mode                  -3.535e-01  8.533e-02  -4.143 3.43e-05 ***
prior.top.ten.hits     1.626e-01  1.438e-02  11.306  < 2e-16 ***
acousticness          -7.915e-01  2.702e-01  -2.930  0.00339 **
energy                -1.623e+00  4.021e-01  -4.037 5.41e-05 ***
loudness               7.516e-02  2.764e-02   2.719  0.00656 **
duration_ms            2.113e-06  1.002e-06   2.109  0.03493 *
valence                4.788e-01  2.357e-01   2.031  0.04222 *
avg.age.not.top.ten   -3.543e-02  2.317e-02  -1.529  0.12624
time_signature         2.637e-01  1.806e-01   1.460  0.14428
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4223.9  on 4866  degrees of freedom
Residual deviance: 3875.4  on 4855  degrees of freedom
AIC: 3899.4

Number of Fisher Scoring iterations: 7
```

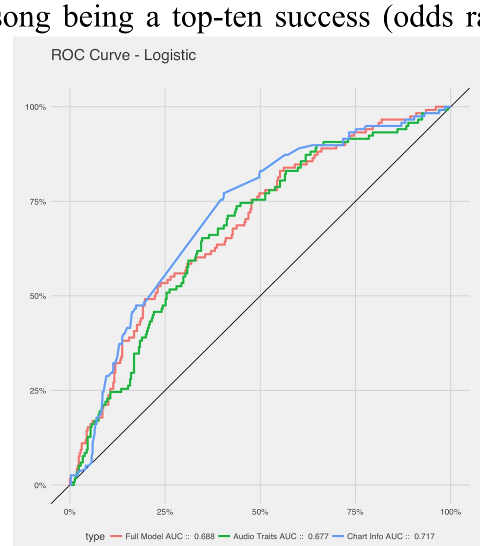


Figure 4. Coefficients for Best Model and Roc Curves for All Logistic Models

Random Forest

In addition to our logistic models we trained a random forest model to examine whether non-linear classifier might have increased success in determining whether a song would be a hit. While the random forest model did exhibit a slightly higher AUC and smoother ROC curve, it did not significantly outperform the hybrid stepwise selected logistic models. We believe that a series of weak classifiers trained on different feature subsets and ensembled with a random forest stacking classifier would have the best performance on our test set.

Random forest classifiers are also commonly used to model feature importance since the splitting algorithm determines which features best separate the observations into categories of "top 10 hit" or not. The variable importance ranking (determined through the mean decrease in Gini coefficient due to that variable) generally confirm the findings of our logistic models' significant variables. Danceability, speechiness and the current average age of the top ten at the time of release are all rated as the most influential variables and also are shown to be significant and impactful in our logistic regressions.

Boosting A Track's Hotness

Since our model's goal is to determine a track's likelihood to hit the top ten given current chart conditions and the audio-traits of that specific track we now show how it could be used by record labels or musicians to tweak the release-date or change it's sonic composition to best compete on the Hot 100. We chose the track "Sing" by Ed Sheeran which first hit the charts on 2014-04-26 and failed to reach the top 10, peaking at number 13.

"Sing" by Ed Sheeran" // Released: 2014-04-26, Peak: 13

prior.top.ten.hits	0	instrumentalness	1.22e-06	prior.not.top.ten.hits	2
key	8	prior.weeks.at.top.ten	0	liveness	0.0601
avg.age.top.ten	25.38462	loudness	-4.451	avg.age.not.top.ten	10.47222
mode	0	acousticness	0.304	speechiness	0.0472
danceability	0.818	tempo	119.988	duration_ms	235382
time_signature	4	energy	0.67	valence	0.939

We simulated changes to several of the track's features to determine how Mr. Sheeran could have increased the likelihood his track would reach the top 10 all else held equal. Our best model (chosen with hybrid stepwise selection) determined that "Sing" had a 26% chance of hitting the top 10. By decreasing the acousticness to 0, increasing the danceability to 1, decreasing the energy to 0.3 and increasing the loudness to -3.45 decibels our model would have put Mr. Sheeran's likelihood of reaching the top 10 at 55%. Obviously there are many other features at play here and changing features like energy have no guaranteed effect on popularity but this simulation still serves as an example of how a Billboard Hot 100 chart could be used in industry.

Appendix

	Dependent variable:		
	pred.hit.top.ten		
	(1)	(2)	(3)
danceability	1.889^{***}	2.204^{***}	
	(0.341)	(0.337)	
prior.not.top.ten.hits	-0.060^{***}		-0.067^{***}
	(0.009)		(0.009)
mode	-0.350^{***}	-0.385^{***}	
	(0.085)	(0.085)	
prior.top.ten.hits	0.164^{***}		0.178^{***}
	(0.014)		(0.014)
acousticness	-0.803^{***}	-0.973^{***}	
	(0.270)	(0.267)	
energy	-1.588^{***}	-1.712^{***}	
	(0.401)	(0.395)	
loudness	0.077^{***}	0.097^{***}	
	(0.028)	(0.027)	
duration_ms	0.00000^{**}	0.00000^{***}	
	(0.00000)	(0.00000)	
valence	0.478^{**}	0.414^{*}	
	(0.235)	(0.232)	
time_signature		0.314^{*}	
		(0.183)	
key		-0.017	
		(0.011)	
Constant	-1.750^{***}	-3.056^{***}	-1.675^{***}
	(0.538)	(0.876)	(0.052)
Observations	4,867	4,867	4,867
Log Likelihood	-1,940.014	-2,025.423	-2,001.170
Akaike Inf. Crit.	3,900.027	4,070.845	4,008.340
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure A1. Significant Predictors in Logistic Regression Models with (1) selected as the best model

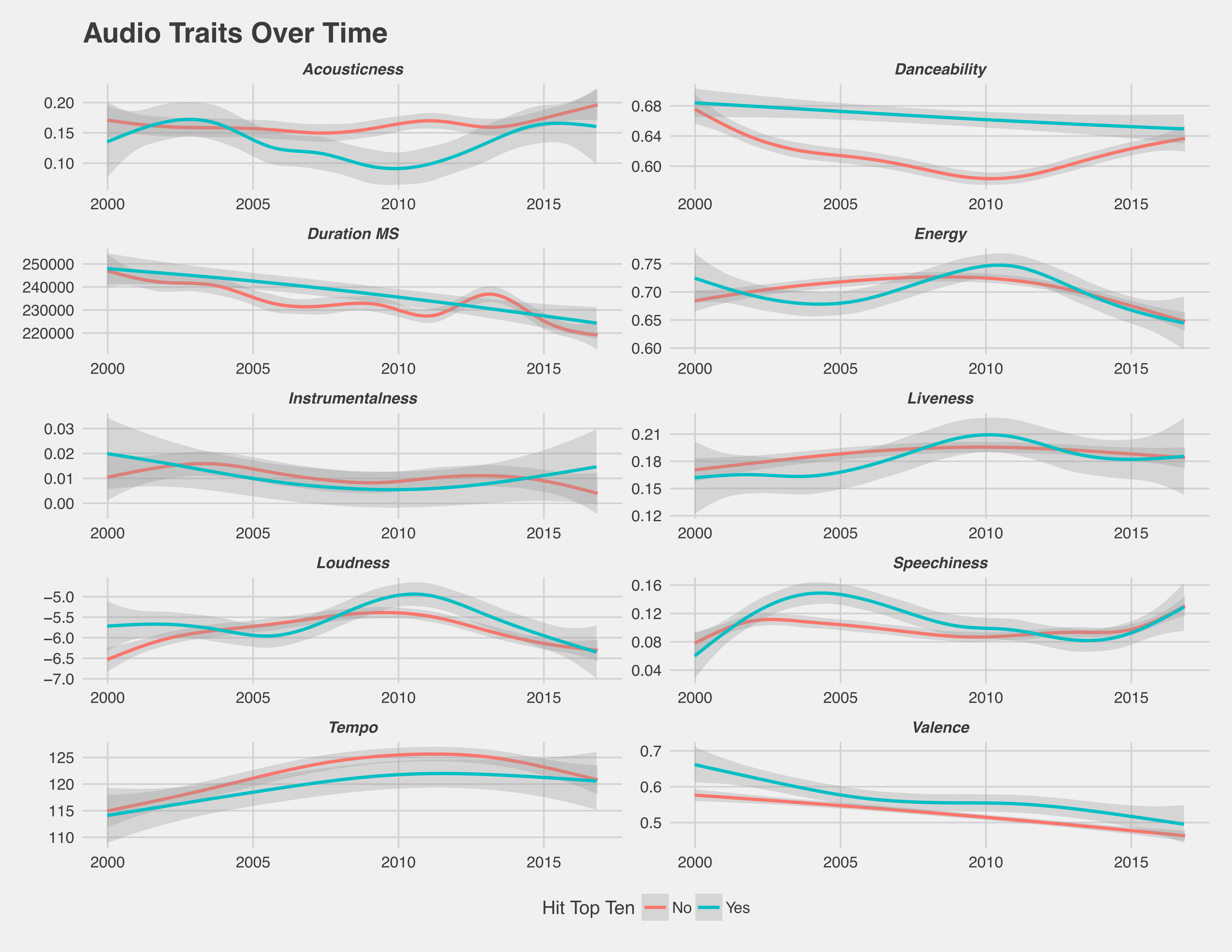
Interaction term	Model	CV Accuracy	Prediction Accuracy	AUC
	pred.hit.top.ten ~ danceability + prior.not.top.ten.hits + prior.top.ten.hits + birthday + mode + avg.age.not.top.ten + avg.age.top.ten + acousticness + energy + loudness + valence + duration_ms + time_signature	85.60%	73.79%	68.8%
energy:duration_ms	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + energy:duration_ms	82.17%	73.02%	68%
prior.not.top.ten.hits:valence	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + prior.not.top.ten.hits:valence	79%	72%	69%
prior.not.top.ten.hits:loudness	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + prior.not.top.ten.hits:loudness	81.5%	72%	69%
prior.not.top.ten.hits:acousticness	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + prior.not.top.ten.hits:acousticness	84.7%	72.6%	69%
	pred.hit.top.ten ~ danceability + total.hits + mode + prob.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature	82.8%	71%	68%
total.hits:energy	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + total.hits:energy	93%	72.6%	69%
total.hits:loudness	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + **total.hits:loudness**	89.2%	72.4%	68%
total.hits: prob.top.ten.hits	**pred.hit.top.ten** ~ danceability + prior.not.top.ten.hits + mode + prior.top.ten.hits + acousticness + energy + loudness + duration_ms + valence + avg.age.not.top.ten + time_signature + **total.hits:prob.top.ten.hits**	87.3%	73.4%	69%
	pred.hit.top.ten ~ energy + valence + danceability + prior.top.ten.hits + **danceability:prior.not.top.ten.hits** + **prior.not.top.ten.hits:loudness**	79.6%	69.6%	67%

Table A1. Interaction Table for Influential Terms

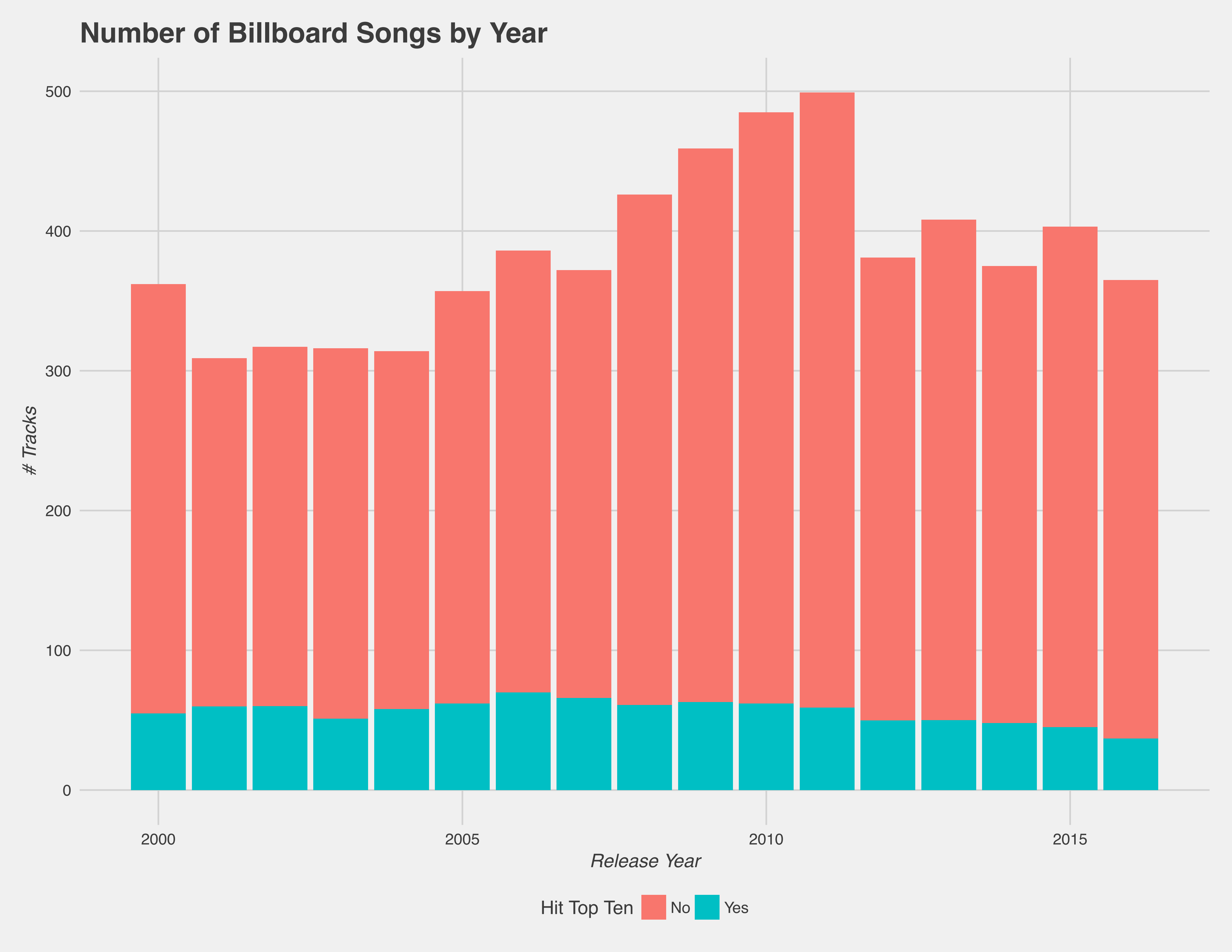
Features used for Predicting a Song's Position in Billboard Top 100 Charts	
Predictor	**Description**
This Week Position	The position of a song at the end of its second week
Last Week Position	The position of a song at the end of its first week
Title	Title of the Song
Artist (ID)	Primary contributor to song
Release Date	Day, Month, Year of song's release date
Song Duration	The duration of the track in milliseconds.
Song Tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Song Time Signature	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
Song Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
Artist's Prior Hits	The number of songs the artist has previously had in the top 10
Average Age of Top Ten	The average number of weeks this artist has had a song in the top 10
Accousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
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 Level	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. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
Key	The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on.
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live

Loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db
Mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0
Speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
Instrumentalness	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Engineered Features	
Prior Top Ten Hits	The number of songs by the artist that have been a top ten hit on the Billboard 100 charts.
Prior Not Top Ten Hits	The number of songs by the artist that have not been a top ten hit on the Billboard 100 charts.
Average Age of Top Ten Hits	The average age of songs by the artist that have been a top ten hit on the Billboard 100 charts.
Average Age of Non Top Ten Hits	The average age of songs by the artist that have not been a top ten hit on the Billboard 100 charts.

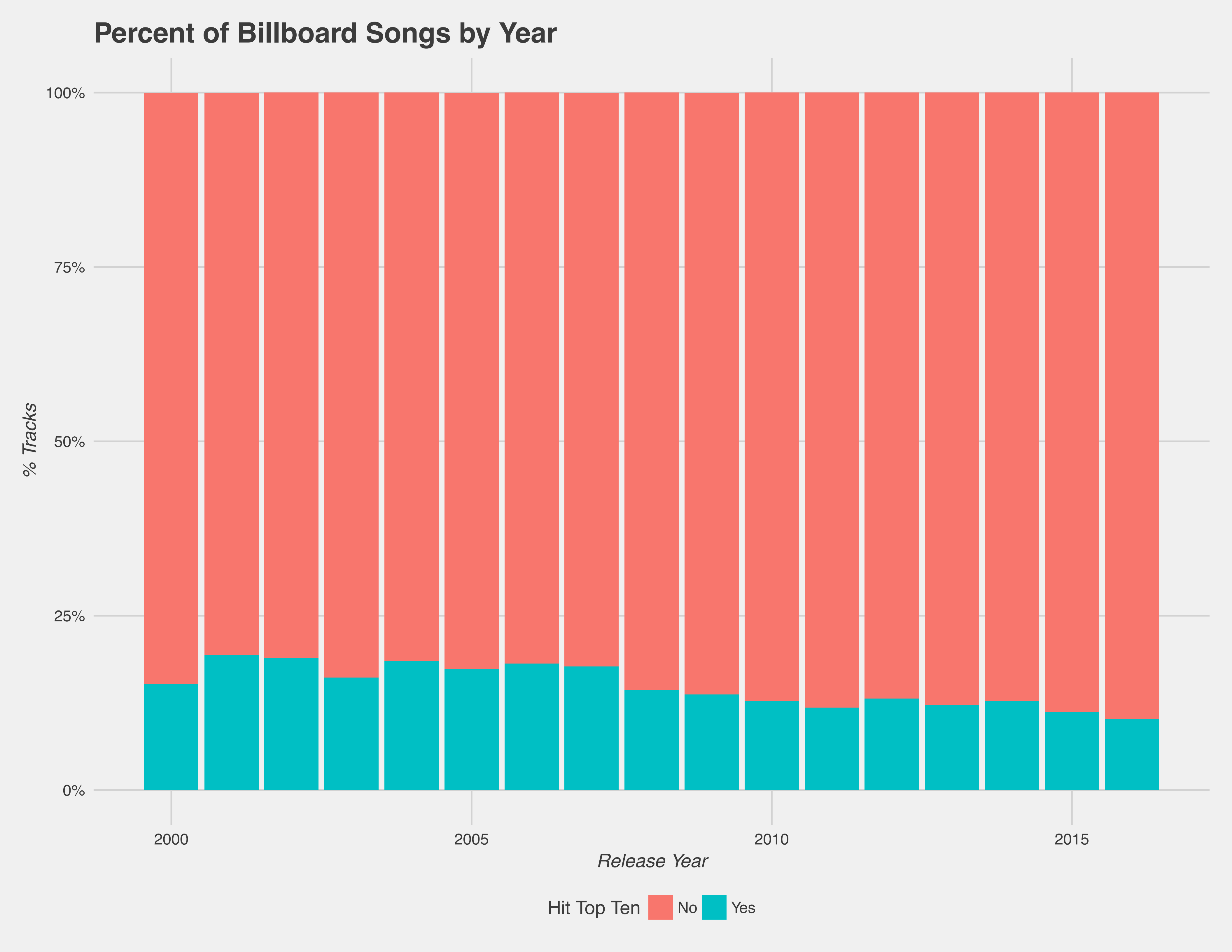
Table A2. Features used for predicting a song’s top 10 Billboard chart success



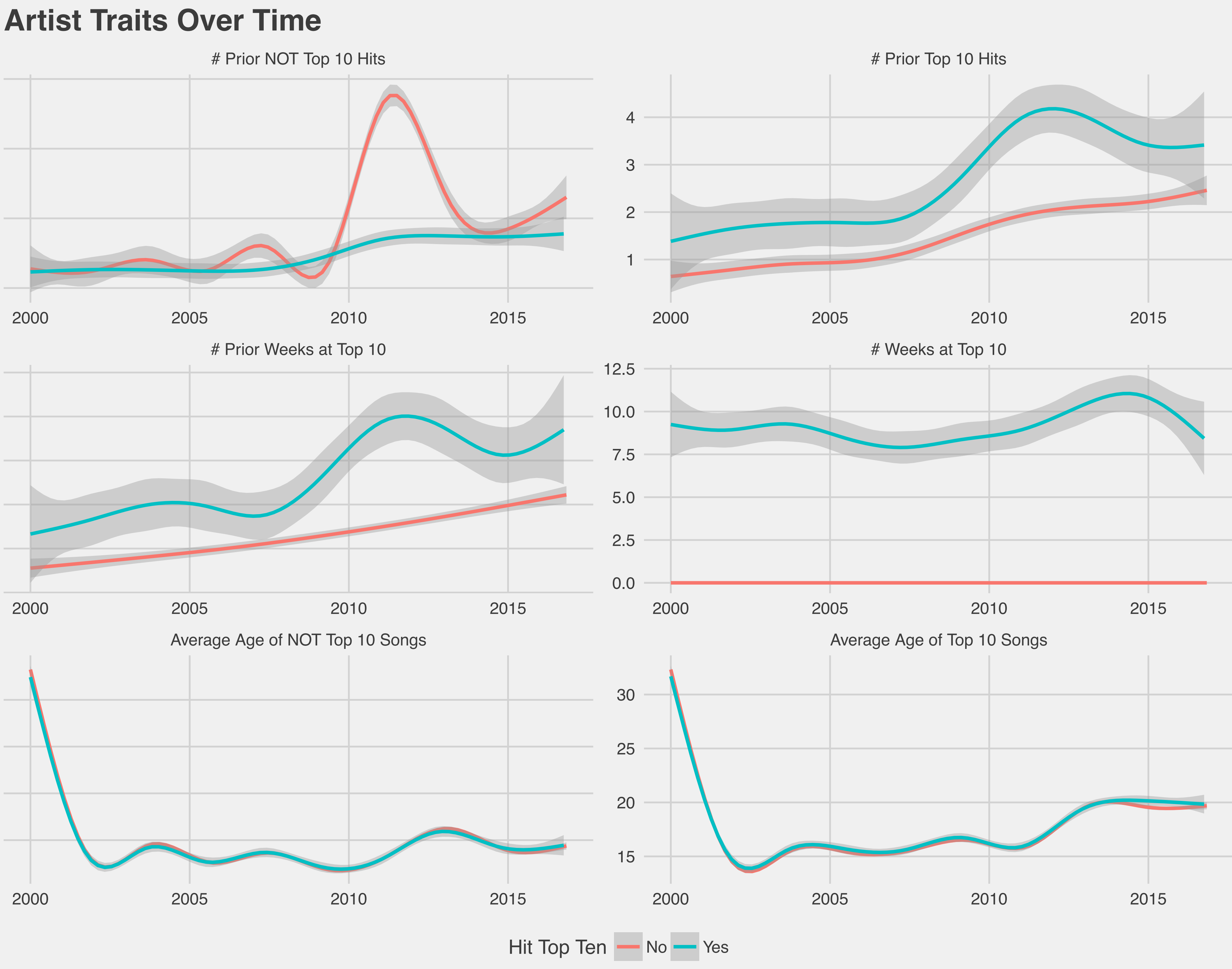
Audio Traits Over Time
Acousticness
Danceability
Duration MS
Energy
Instrumentalness
Liveness
Loudness
Speechiness
Tempo
Valence
2000
2005
2010
2015
Hit Top Ten
No
Yes



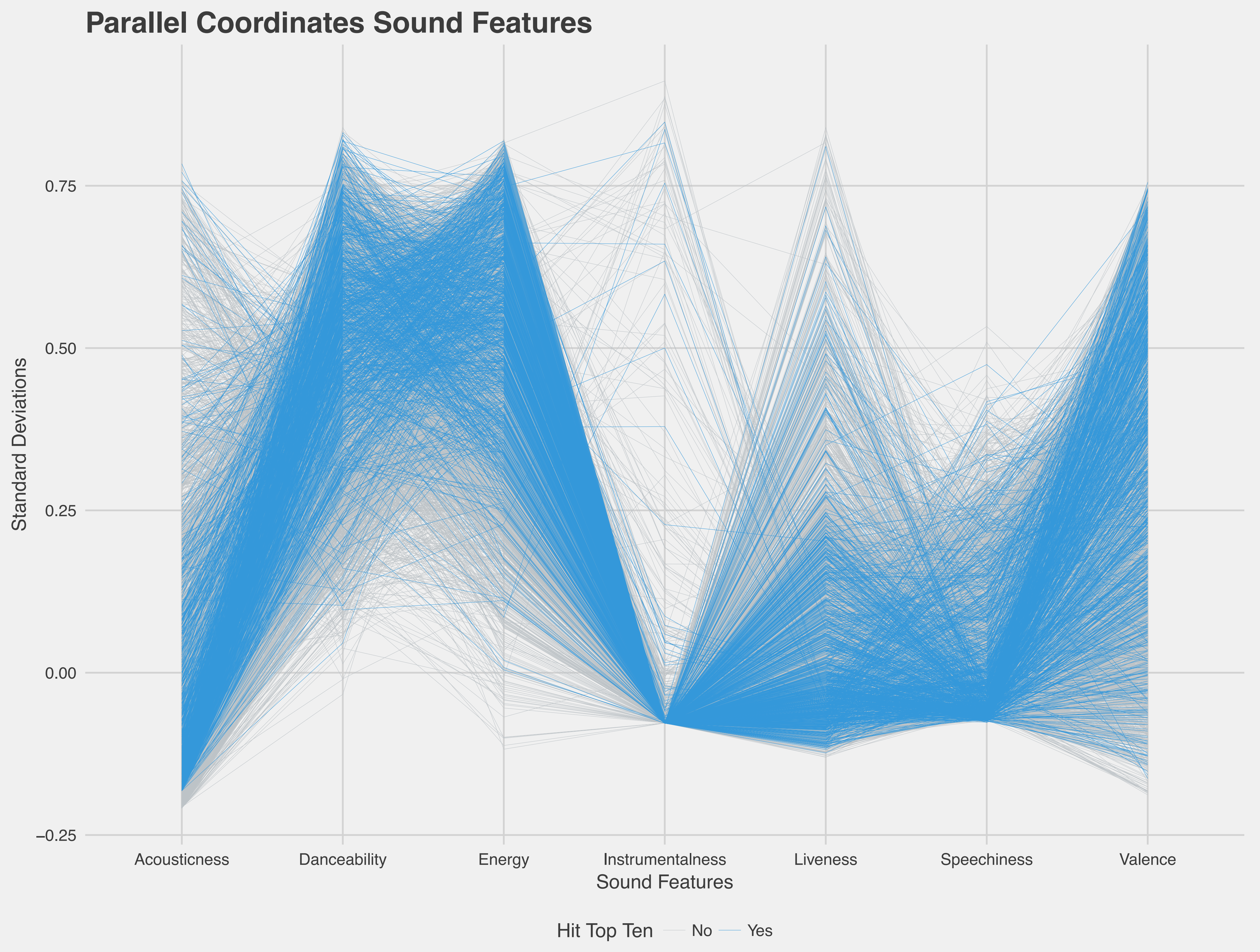
Number of Billboard Songs by Year
Tracks
500
400
300
200
100
0
2000
2005
2010
2015
Release Year
Hit Top Ten
No
Yes



Percent of Billboard Songs by Year
100%
75%
50%
25%
0%
% Tracks
2000
2005
2010
2015
Release Year
Hit Top Ten
No
Yes

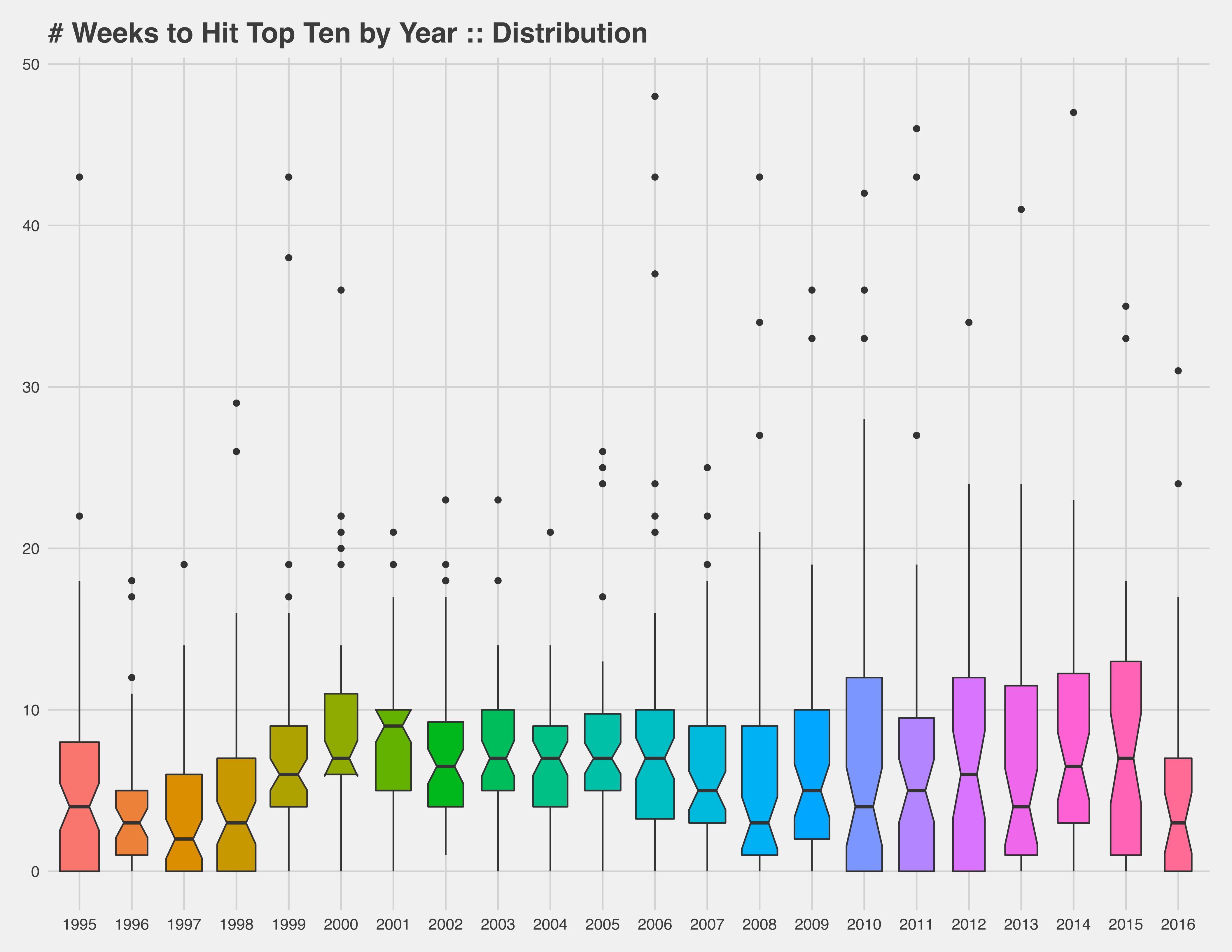


Artist Traits Over Time
Prior NOT Top 10 Hits
Prior Top 10 Hits
Prior Weeks at Top 10
Weeks at Top 10
Average Age of NOT Top 10 Songs
Average Age of Top 10 Songs
Hit Top Ten
No
Yes



Parallel Coordinates Sound Features
Standard Deviations
0.75
0.50
0.25
0.00
−0.25
Acousticness
Danceability
Energy
Instrumentalness
Liveness
Speechiness
Valence
Sound Features
Hit Top Ten
No
Yes


Sound Feature Distribution
Acousticness
Avg Age Not Top Ten
Avg Age Top Ten
Danceability
Duration_ms
Energy
Instrumentalness
Liveness
Loudness
Prior Not Top Ten Hits
Prior Top Ten Hits
Prior Weeks At Top Ten
Speechiness
Tempo
Valence
Hit Top Ten 0 1



Weeks to Hit Top Ten by Year :: Distribution
0
10
20
30
40
50
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016