Detecting Stylistic Variation in Pop Production*

Michaela Drouillard

November 10 2023

We explore whether we can predict if a song was produced by Jack Antonoff using the audio features provided by Spotify's API. Our logistic regression model has an accuracy score of 76%, a precision score of 39%, and a recall of 76%. Our random forest model has an accuracy of 61%, and a No Information Rate of 43%. We especially found that danceability, a metric trained on human-tagged data, was a key predictor. Our results have implications for computational explorations of culture, how the combination of human subjectivity and machine listening can come to create an approximate for style.

1 Introduction

We are interested in understanding how to create an approximate for production style. To do this we used Spotify API's audio features dataset to explore different audio feature metrics and identify which were the most import predictors in identifying pop producers' style. We focus on Jack Antonoff, as he is a prolific and widely used producer, who also has his own solo acts.

By way of background, Antonoff began his career as a guitarist in fun., and is the frontman in Bleachers. He has produced and co-produced music for Lorde, Taylor Swift, Florence and the Machine, The Chicks, Clairo, The 1975, Grimes, Zayn, Pink, Lana Del Rey, Olivia Rodrigo, the Minions: The Rise of Gru Soundtrack, and more. His style is considered identifiable, and his collaborators are influential – as Andrew Marantz writes in the *New Yorker*: "When his band releases an album, the world responds politely. When he produces one by Lorde or Lana Del Rey or Taylor Swift, the world wobbles on its axis" (Marantz).

We are interested in whether Spotify audio features can be used to train a model that distinguishes whether a song was produced by Antonoff. We obtained tracks from six artists who've collaborated with Antonoff, both before their collaboration and during. We also obtained

*Code and data are available at: https://github.com/michaeladrouillard/spotify_pop.

tracks from similarly prolific pop producers: Joel Little, Ariel Rechstaid, Max Martin, Greg Kurstin, Paul Epworth and Rick Nowels. Our data set contains 3738 tracks in total.

We built two models to explore the data. The first model, a logistic regression, draws on the dataset of artists who have collaborated with Antonoff, with a binary variable to tag whether or not the song was an Antonoff production. With this model, we explore which variables are stronger predictors of Antonoff's style, developping a better understanding of whether the Spotify data can trace Antonoff's involvement in an artist's sound over time. We built a second model, a Random Forest (RF), to complement the insights drawn from the logistic regression. The RF trains on the full data set, including other major pop producers, to classify tracks according to who produced them, and return information on which predictors were most useful in doing so.

We find that `danceability` is the strongest predictor of whether or not a song has been produced by Antonoff. Furthermore, when `danceability`'s correlation scores with other features in the dataset are considered, the variance in correlation scores tends to be strongest classifier of whether a song was produced by Antonoff. In other words, danceability in concert with other features in the dataset provides the best predictions.

Our results suggest that we can capture something about Antonoff's music with these variables, and it might, in the end, just be that subjective "ick" that some people feel, and others enjoy, and that we're all just trying to describe in using different methods and tools. With the understanding that what we understand as "data-driven" in recommendation systems is often deeply intertwined with human subjectivity -- engineers' understandings of taste and taste-making shape the recommendation systems that they build ("Seaver) -- we contribute to a more informed understanding of how style and tastes can been algorithmically rendered using Spotify's audio features.

2 Data

We collected data from the Spotify API, which we accessed using the `spotifyr` package in R (Thompson et al.). We acquired data on Lorde, Taylor Swift, St. Vincent, Lana Del Rey, The Chicks, Florence and the Machine, and Bleachers discographies. HAIM, Marina and the Diamonds, Maggie Rogers, Sharon Von Etten, and Mitski were included to represent other contemporary pop artists with overlapping fan bases who have never collaborated with Antonoff. Of the artists who did collaborate with Antonoff, we chose artist who have produced multiple albums, and at least one album with Antonoff.

We acquired the discographies of comparable contemporary pop producers' by scraping their respective discography Wikipedia pages. The producers include Epworth, Rechstaid, Max Martin, Nowels, Joel Little, and Greg Kustin. These producers were chosen based both on their popularity, and because some have collaborated with the same artists as Antonoff.

Table 1: Counts of Tracks Per Producer in Original Dataset

Producer	Number of Tracks
antonoff	326
epworth	321
kurstin	559
little	90
martin	778
nowels	42
other	1558
rechtshaid	64

The data was manually validated, and live performances, karaoke editions, international versions or translations, and remix albums were removed. Where deluxe albums were available, original albums were deleted to avoid duplicate rows.

We created a binary variable which contains 1 if Jack Antonoff produced or co-produced the song, and 0 if it was produced by somebody else. The information underpinning this discography was drawn from the "Jack Antonoff Production Discography" Wikipedia page (Wikipedia).

In total, we collected 3738 tracks from 828 different artists. If tracks were produced by producers other than the main producers included in our study, they were marked as "other".

2.0.1 Audio Features Dataset

Spotify provides variables for each song, according to variables developed by Spotify, which makes for simple comparisons across artists. The final audio features in our data set are: `artist_name`, `track_name`, `energy`, `danceability`, `key`, `loudness`, `mode`, `speechiness`, `acousticness`, `instrumentalness`, `liveness`, `valence`, `tempo`, `jack`.

From the `get_artist_audio_features` documentation (*Get Track's Audio Features*):

- **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.
- **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. 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.

- **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.

- **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.

- **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. If no key was detected, the value is -1.

- **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 typically 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.

- **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.

- **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).

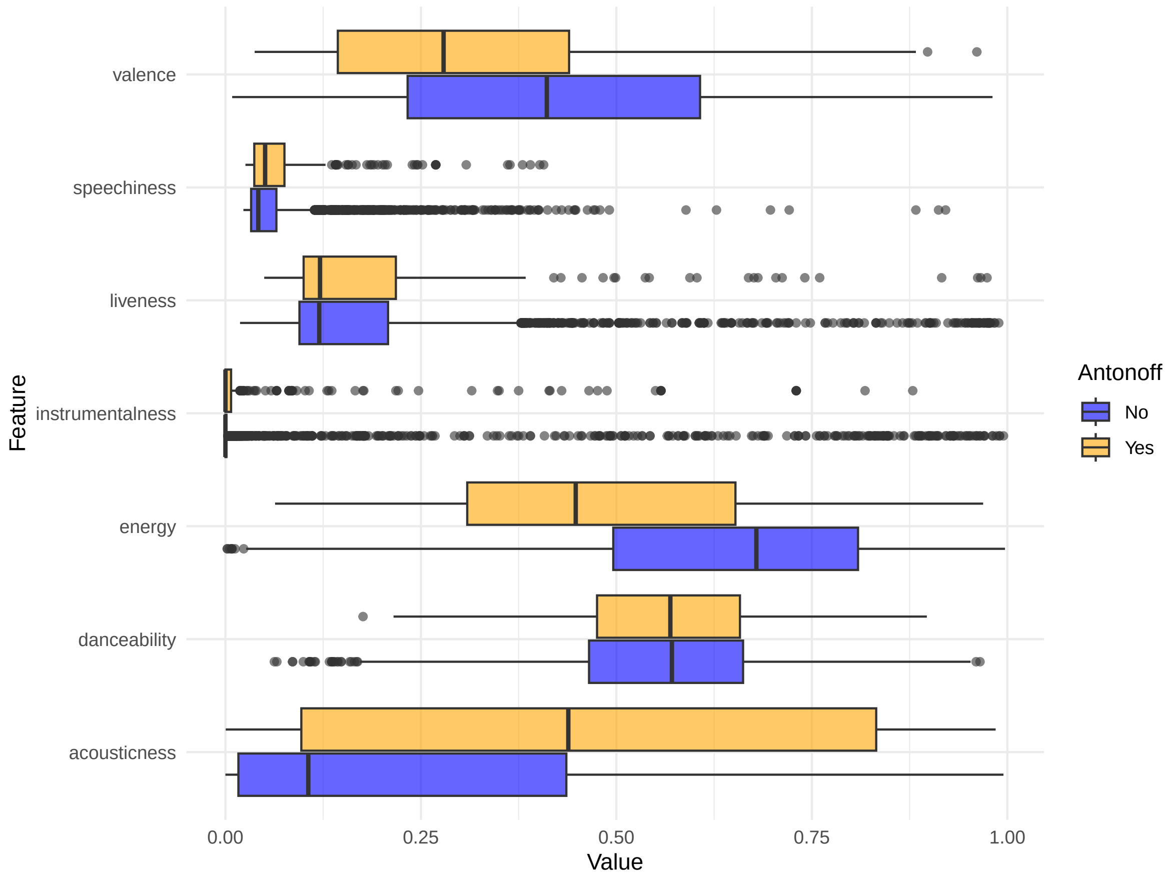


Figure 1: Comparing the Mean Values of Each Variable

By analyzing the cumulative distribution functions of various audio features, particularly when categorizing songs based on whether they were produced by Jack Antonoff (Figure 2), distinct patterns emerge. Notably, while attributes such as liveness, instrumentalness, and speechiness exhibit relative homogeneity, there are slight disparities in features like danceability, energy, loudness, acousticness, and valence.

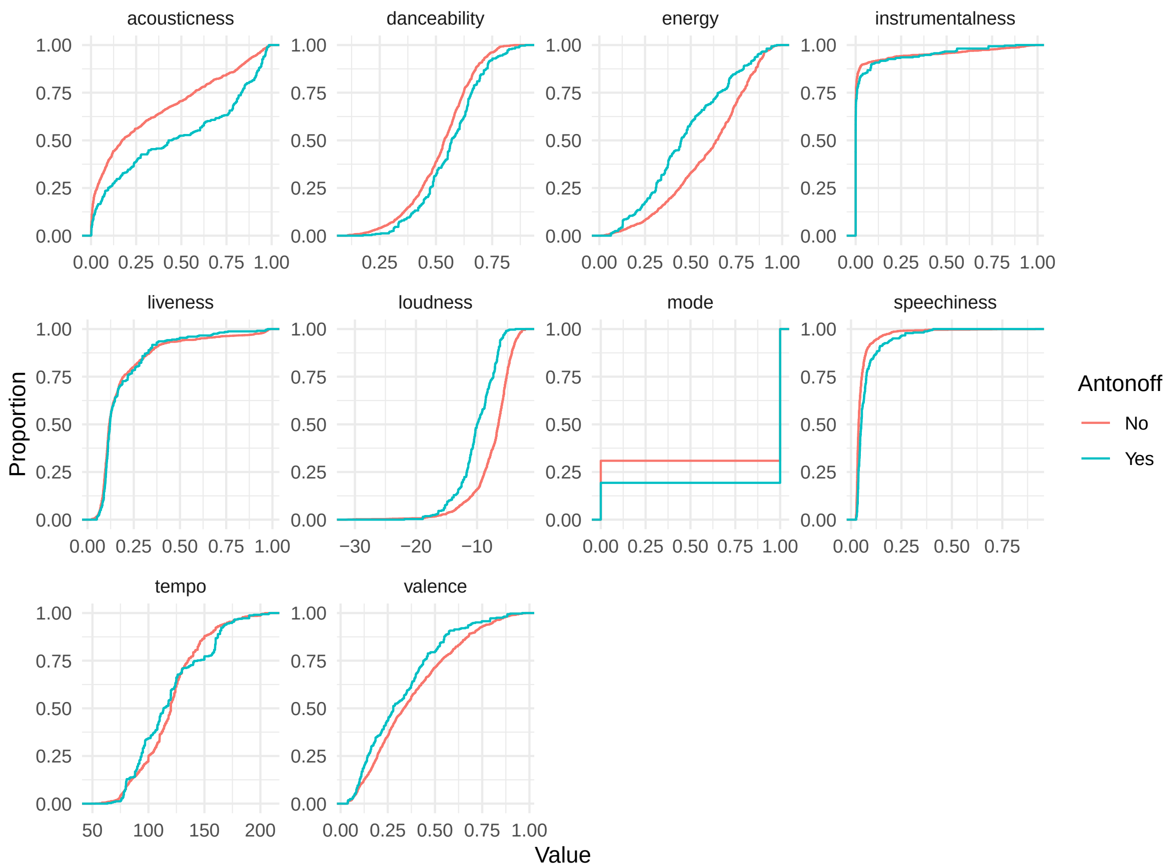


Figure 2: Cumulative Distribution Function of Each Variable

Examining the temporal evolution of artists' sounds, especially in correlation with their collaboration with Antonoff, provides insights into the broader trends in pop music. Each point in the following plots represent one track.

In Figure 3, we observe a downward trend in energy scores among artists collaborating with Antonoff. However, this trend appears consistent with the overall trajectories of these artists' works, suggesting broader shared trends in this area of pop, as exemplified by Lana Del Rey and Taylor Swift. Sharon Von Etten, Maggie Rogers, and Mitski, display upward energy trajectories in their discographies.

No significant trends are obersable in danceability over time across the artists' collaborations

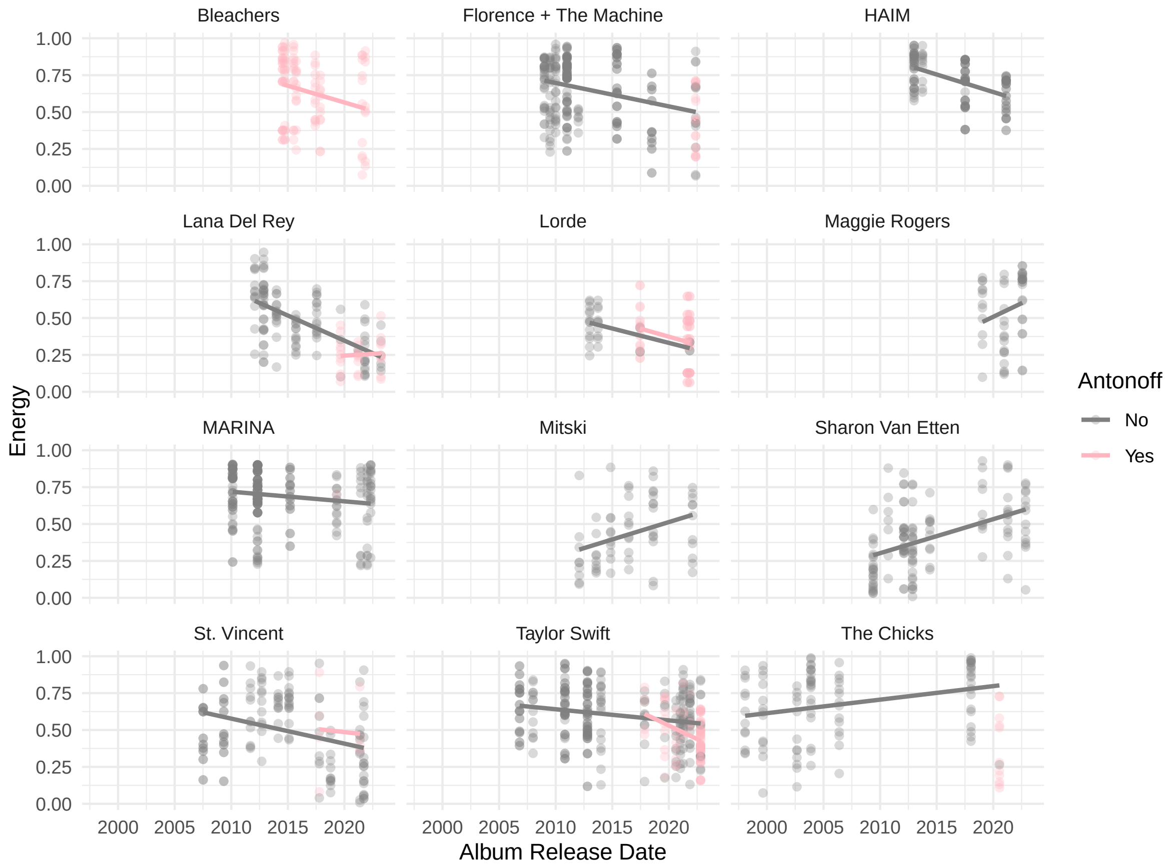


Figure 3: Energy Variable

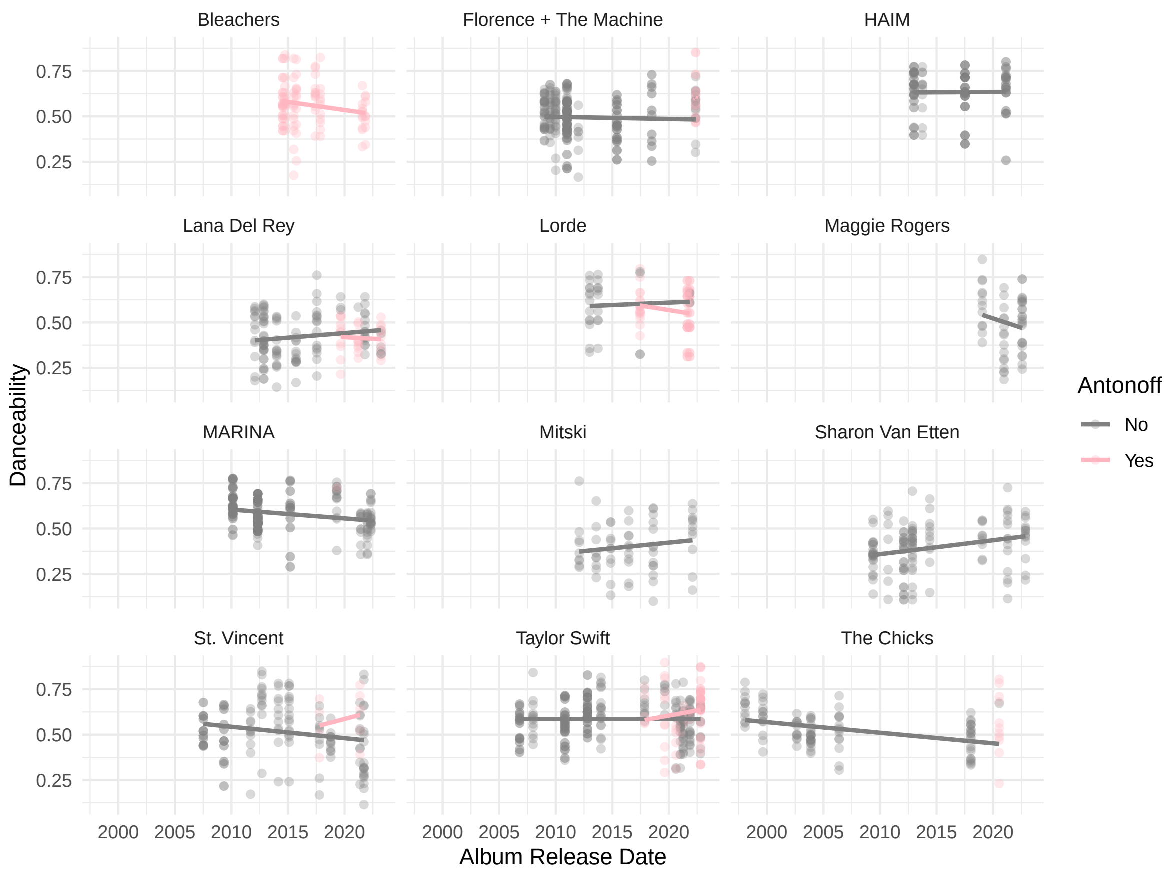


Figure 4: Danceability Variable

with Antonoff, as demonstrated in (Figure 4), although there is observable an variance in danceability scores across songs in every album, across all artists. This suggests that, over time, danceability scores across tracks in every album are consistently varies over time. The Chicks earlier work, between 2002 and 2005, presumable the country album “Home” have the smallest amount of variance in danceability. This could be due to genre difference.

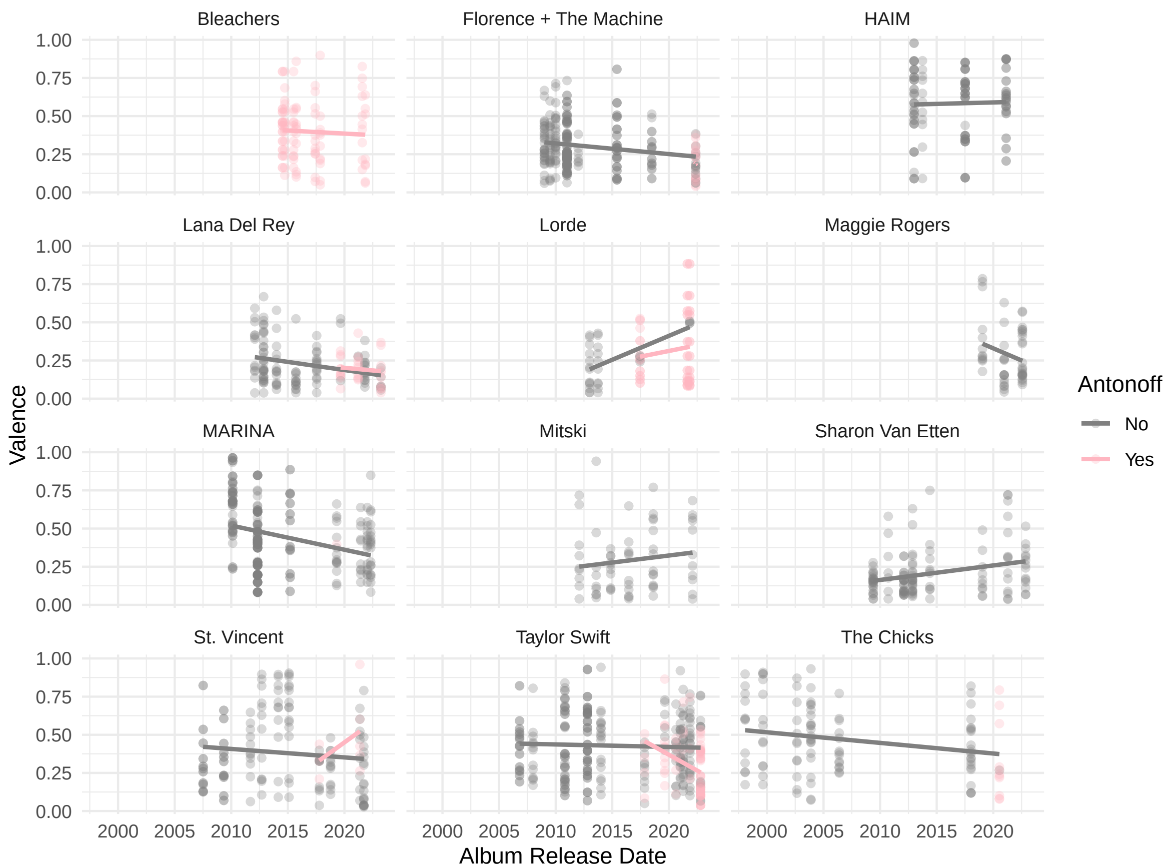


Figure 5: Valence Variable

Figure 5 reveals a notable shift in valence scores in Lorde’s discography, where her tracks display a slight upward trend and a greater range in valence scores over time. Otherwise, similarly to danceability, the valence scores display significant variance across tracks in every album.

Figure 6 underscores a trend toward increased acousticness in the works of Taylor Swift, Lorde, and HAIM, contrasted with declining trends in artists like Mitski, Sharon Von Etten, and Maggie Rogers.

Figure 7 exhibits no significant correlations, and scores are consistently low for most artists, excluding St. Vincent, whose tracks display both a consistently high degree of variance and a

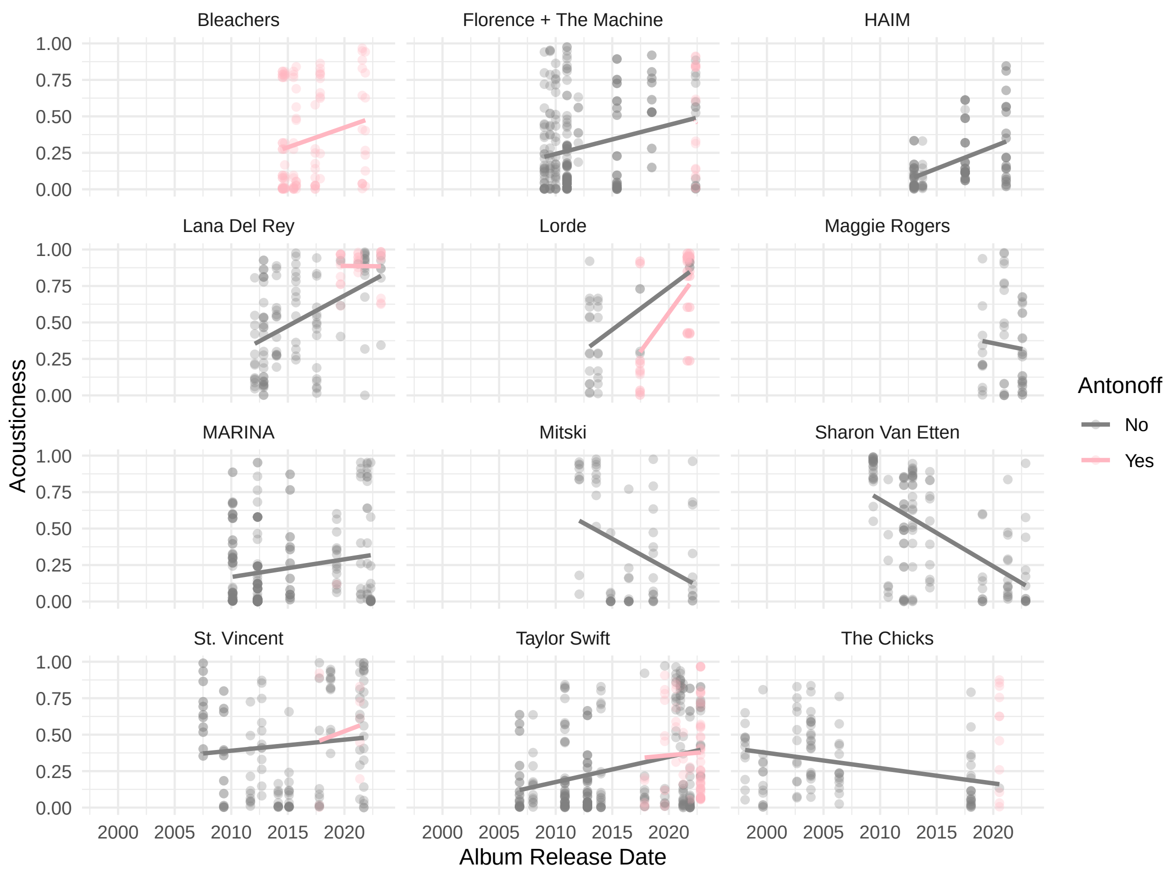


Figure 6: Acousticness Variable

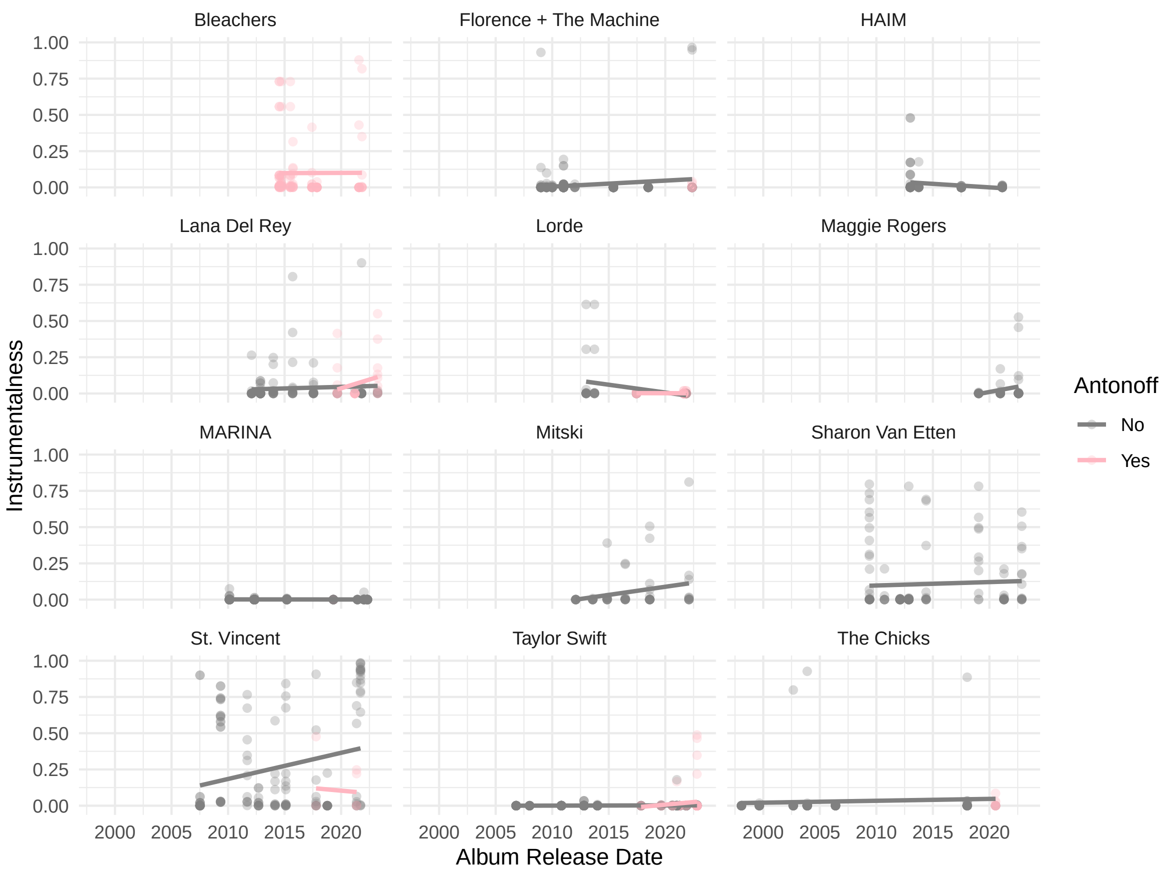


Figure 7: Instrumentalness Variable

noticeable upward trend over time.

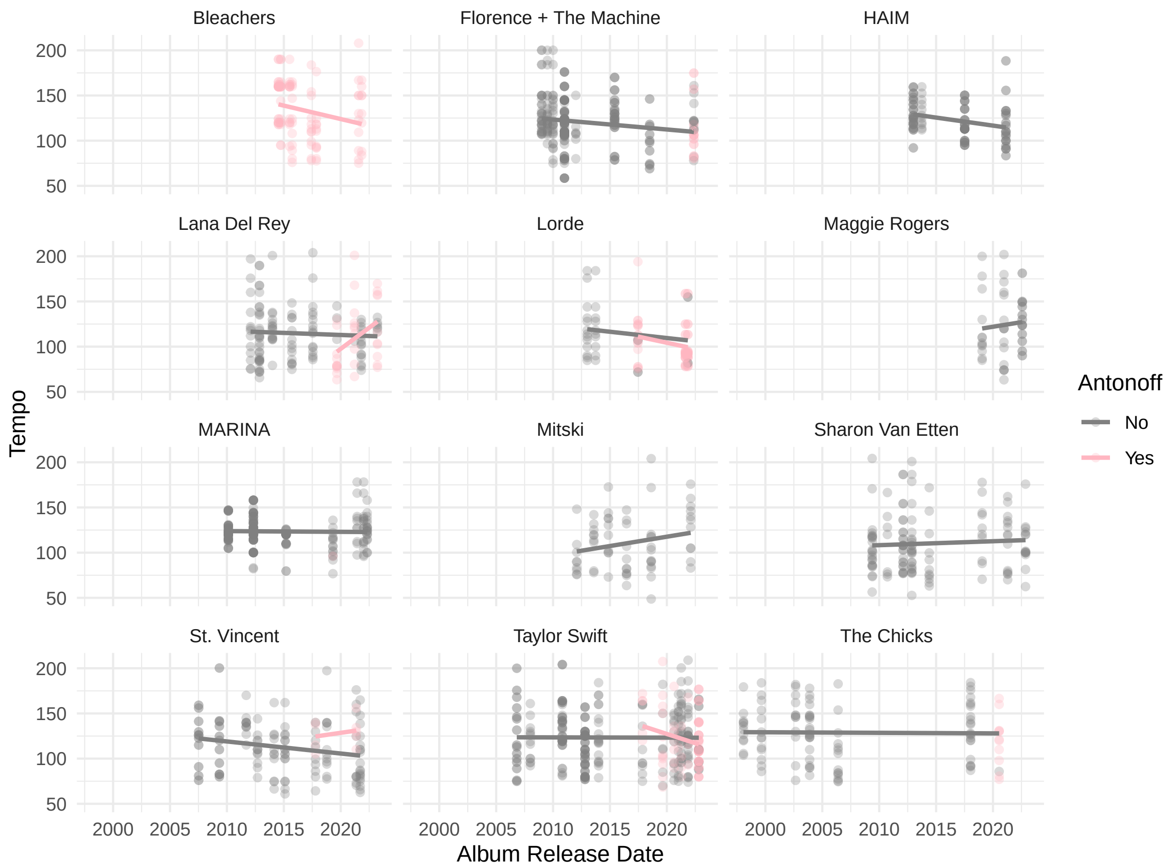


Figure 8: Tempo Variable

Finally, tempo trends, as shown in Figure Figure 8, are generally stable across artists, with notable exceptions in collaborations involving Lana Del Rey and Taylor Swift. In Lana Del Rey's tracks, her collaborations with Antonoff see an upward trend in tempo. In Taylor Swift's tracks, her collaborations with Antonoff exhibit a downward trend.

We've observed subtle trends in variables associated with songs produced by Jack Antonoff, as evidenced by our cumulative distribution functions and minor trends in the evolution of artists' sound characteristics, when collaborating with Antonoff (particularly energy).

To be clear, we do not suggest that Antonoff's involvement is the cause for the change in artist's sound. Arguments of "Antonoffication" risk removing agency and artistic credit from (female) artists, and attributing both their critical success and failures to Antonoff (Wilson). In an industry so dominated by male producers, and using features so limited in their ability to describe anything about the process of creation, it would not be appropriate the use these trends to suggest that Antonoff's involvement is the cause behind changes in sound. The purpose of these visualizations is to observe the distribution of variables across artists, over time, with a focus on Antonoff.

This exploratory data analysis sets the stage for a) employing logistic regression to quantify the likelihood of Antonoff's involvement in a song based on these audio features, b) classifying tracks produced by a wider range of producers using a Random Forest, c) identifying the most influential variables in this classification to better understand how style can be inferred from the provided data set features.

3 Model

3.1 Logistic Regression

The aim of the logistic regression is to determine whether we can predict Antonoff's influence in artists' sound. We select a total of 1465 tracks from the dataset from 12 artists: Lorde, Taylor Swift, St. Vincent, Lana Del Rey, The Chicks, Florence and the Machine, HAIM, Marina and the Diamonds, Maggie Rogers, Sharon Von Etten, Mitski, and Bleachers (Antonoff's own band).

The binary response variable, called `is_antonoff`, has a value of 1 if the song was produced by Antonoff, and a 0 otherwise. There are 10 predictor variables, all of which are continuous: **`danceability`**, **`energy`**, **`loudness`**, **`mode`**, **`speechiness`**, **`acousticness`**, **`instrumentalness`**, **`liveness`**, **`valence`**, and **`tempo`**.

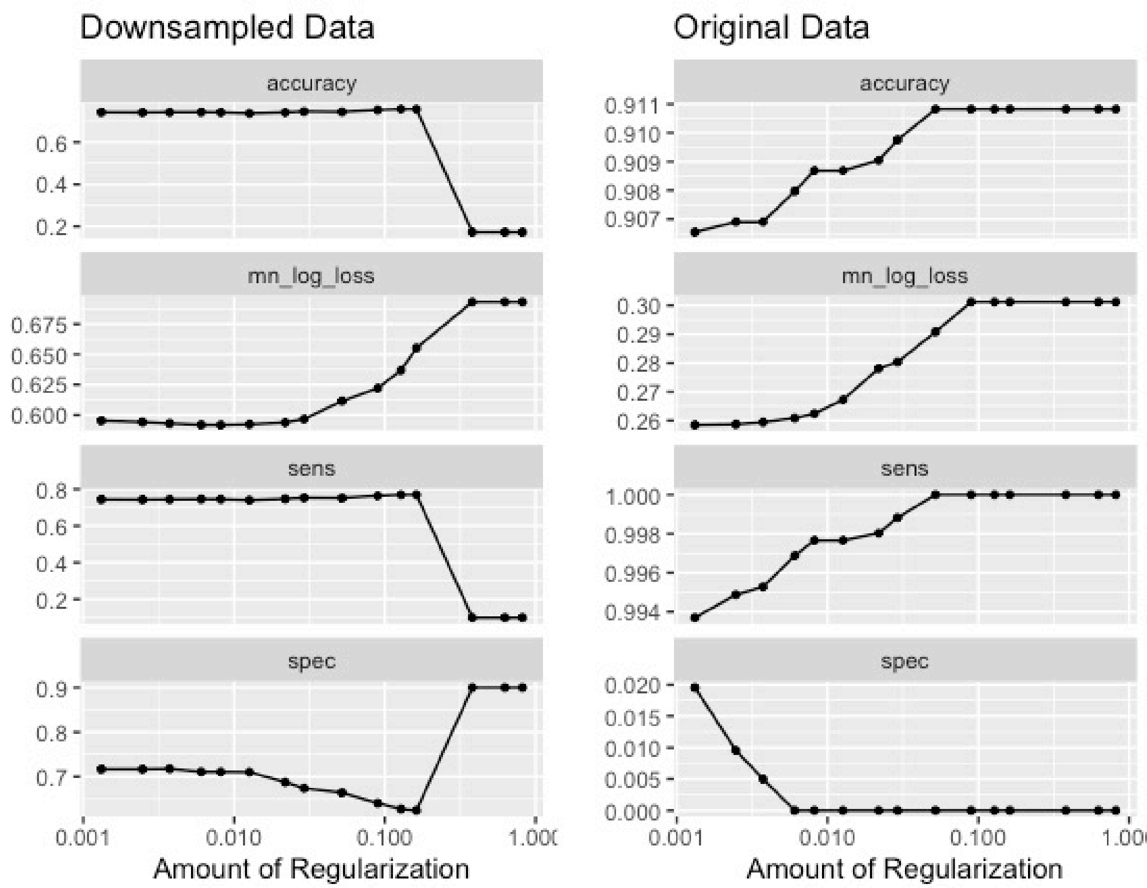
The formula for this logistic regression can be written as:

$$Pr(y_i \ = 1) = logit - 1(\beta_0 \ + \beta_1 \ \times danceability + \beta_2 \ \times energy + ... + \beta_{10} \ \times tempo)$$

where $\Pr(y_i = 1)$ is the probability of a song being produced by Antonoff, β_0 is the intercept, and $\beta_1, \ldots, \beta_{10}$ are the coefficients corresponding to the 10 predictor variables.

After some minor pre-processing using the `tidyverse` R package (Wickham et al.). We followed Silge to build a model that deals with the imbalanced classes.

We used the `tidymodels` to fit the model (Kuhn and Wickham). To deal with the imbalanced classes, we downsampled the majority class using the **`step_downsample()`** function from the **`themis`** package. Then we used 5-fold cross-validation with stratification by the **`is_antonoff`** variable to estimate the performance of the model.



3.2 Random Forest

Random forests (RF) build an ensemble of decision trees during the training phase, and then use majority voting for classification tasks. Each tree in the ensemble is trained on a random sample of data taken with replacement, and only a random subset of features are considered when splitting at each node. This method ensures that each tree in the ensemble is unique, avoiding overfitting and capturing non-linear relationships and more complex patterns in the data.

The training process for the RF model can be summarized as follows:

1. For each iteration ($b = 1$) to (B):

    a. Draw a bootstrap sample (Z^*) of size (N) from the training data.

    b. Grow a random-forest tree (T_b) to the bootstrapped data by recursively performing the following steps for each terminal node of the tree until the minimum node size (n_{min}) is reached:

i. Select (m) predictors at random from the (p) predictors.
ii. Determine the best predictor and split-point among the (m).
iii. Split the node into two daughter nodes.

2. The ensemble of trees $(\{T_b\}_{b=1}^{B})$ constitutes the random forest model.

For a new instance (x), the RF model prediction $(\hat{C}^{RF}(x))$ is determined by the majority vote from the ensemble of trees:

$$\hat{C}^{\mathrm{RF}}(x) = \text{majority vote}\left\{\hat{C}_b(x)\right\}_{b=1}^{B}$$

Where $(\hat{C}_b(x))$ is the class prediction of the (b)-th tree.

The categorical feature `producer`, which contains the names of the producer for each track, was one-hot encoded to create binary columns for each producer (i.e. `is_antonoff`, `is_elworth`, etc.). The producer feature was then converted into a factor variable, and the dataset was split into training and tests sets.

The RF was trained using the (**ranger?**) package. The 'importance' parameter was set to 'impurity', which sets the model to interpret variable importance using Gini impurity.

Gini impurity is a measure of misclassification. At every split in a decision tree, the algorithm tests how well each feature splits the data. The features that split the data the most accurately, and then with the least impurity, would have the lowest impurity score. In Random Forests, a variable's importance can be calculated by looking at how much the tree nodes that use that variable reduce impurity on average.

Gini impurity was chosen on the grounds that this study is interested in not only classifying producers, but, more importantly, in understanding how the underlying data informs these decisions. By calculating the Gini impurity not only for the entire model, but specifically for each producer, we can better understand how the variables in our dataset describe the differences between producers. Without song structure, instruments, sound, or other more ephemeral components of style, these differences can be taken as a proxy for style.

4 Results

4.1 Logistic Regression

The model's accuracy score is 0.757, and the ROC AUC score is 0.819. The precision score is 0.3851, meaning that approximately 38.51% of songs predicted to be produced by Antonoff were actually produced by him. The recall of 0.7561 indicates that about 75.61% of songs produced by Antonoff were actually correctly identified in the model.

Table 3: Importance Scores of Each Variable in Determining Antonoff's Tracks

Variable	Importance	Sign
danceability	5.176	POS
valence	2.976	NEG
instrumentalness	2.603	NEG
energy	1.537	POS
acousticness	1.032	POS
mode	1.017	POS
speechiness	0.811	POS
liveness	0.415	POS
loudness	0.347	NEG
tempo	0.006	POS

Table 2: Confusion Matrix for Logistic Regression Results

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	62	99
Predicted Negative (0)	20	308

Table 3 shows variables' respective importance in determining whether a track was produced by Antonoff.

`Danceability`, with a positive importance score of 4.142, had the highest level of importance in determining whether a song was produced or co-produced by Jack Antonoff. `Valence` was the second most important, with a negative score of 2.559. The third, `instrumentalness` correlates negatively with Antonoff tracks with a negative score of 2.309.

In the 20 songs in Jack Antonoff's discography with the highest `danceability` scores, Bleachers tracks account for 35% of this list, and Taylor Swift tracks account 40% of this list (Table 4). St Vincent, The Chicks, Lorde, and Florence and the Machine are also featured.

4.2 Random Forest

Table 5: Overall Performance Statistics

Metric	Value
Accuracy	0.6111
95% Confidence Interval	(0.579, 0.6425)
No Information Rate	0.4274

Metric	Value
P-Value (Acc > NIR)	<2.2e-16
Kappa	0.4223

The accuracy of the model (0.6111) being significantly higher than the No Information Rate (0.4274) means that it is doing more than just blind guessing based on class prevalence. With a p-value below 0.05, we can confirm the model's statistical significance over a naive approach.

The model's accuracy and Kappa statistics suggest a moderate level of performance. However, there is enough demonstrated statistical significance to suggest that there are underlying patterns in the data that differentiate producers.

Correlation heatmaps can help us better understand the differences between each producer's "style". The variables with the most variance were danceability, loudness, and energy.

Variable	Min Value	Max Value	Range
energy	-0.779	0.821	1.600
loudness	-0.728	0.562	1.290
danceability	-0.484	0.553	1.037
speechiness	-0.445	0.300	0.745
acousticness	-0.449	0.271	0.720
mode	-0.295	0.289	0.584
instrumentalness	-0.235	0.330	0.565
valence	-0.212	0.202	0.414
liveness	-0.245	0.085	0.330

Similar to the logistic regression, danceability was also, across all producers, the most important predictor in the classification process.

Table 4: Top 20 Most Danceable Antonoff-Produced Songs

Artist Name	Danceability	Track
Taylor Swift	0.897	I Think He Knows
Taylor Swift	0.875	Vigilante Shit
Florence + The Machine	0.852	Heaven Is Here
Bleachers	0.838	I Wanna Get Better - Demo Version
Taylor Swift	0.824	Cornelia Street
Bleachers	0.824	Hate That You Know Me - MTV Unplugged
Bleachers	0.818	Wake Me
Bleachers	0.814	Shadow
Taylor Swift	0.811	Paper Rings
The Chicks	0.805	Julianna Calm Down
Lorde	0.796	Sober
The Chicks	0.784	Texas Man
Bleachers	0.776	Hate That You Know Me
St. Vincent	0.774	Daddy's Home
Bleachers	0.769	All My Heroes
Taylor Swift	0.766	Look What You Made Me Do
Taylor Swift	0.751	Question...?
Taylor Swift	0.743	Lavender Haze
Taylor Swift	0.739	False God
Bleachers	0.732	Goodbye

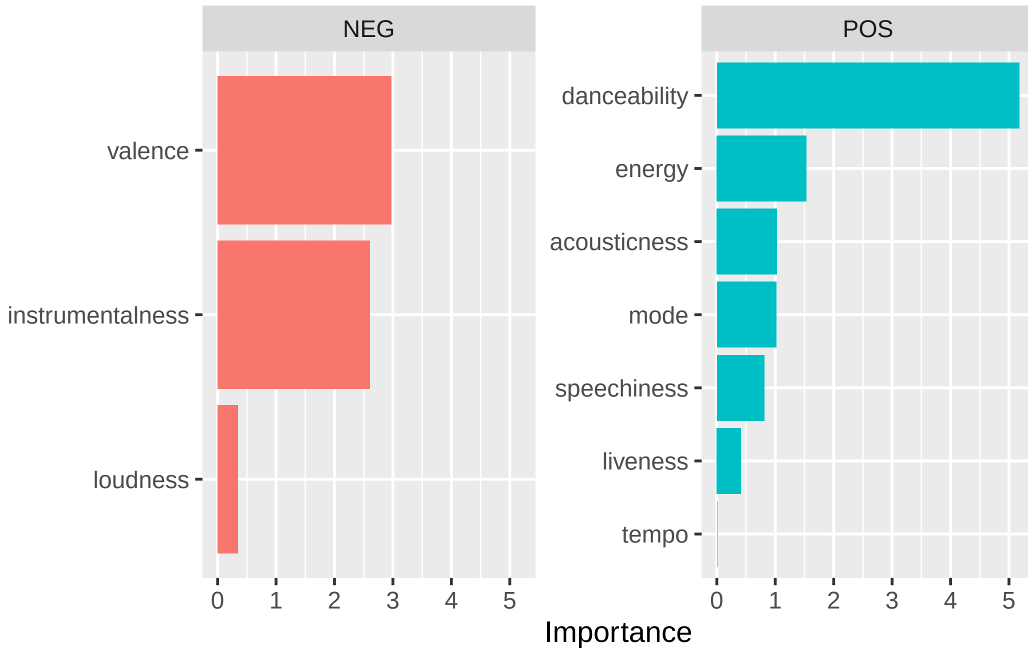
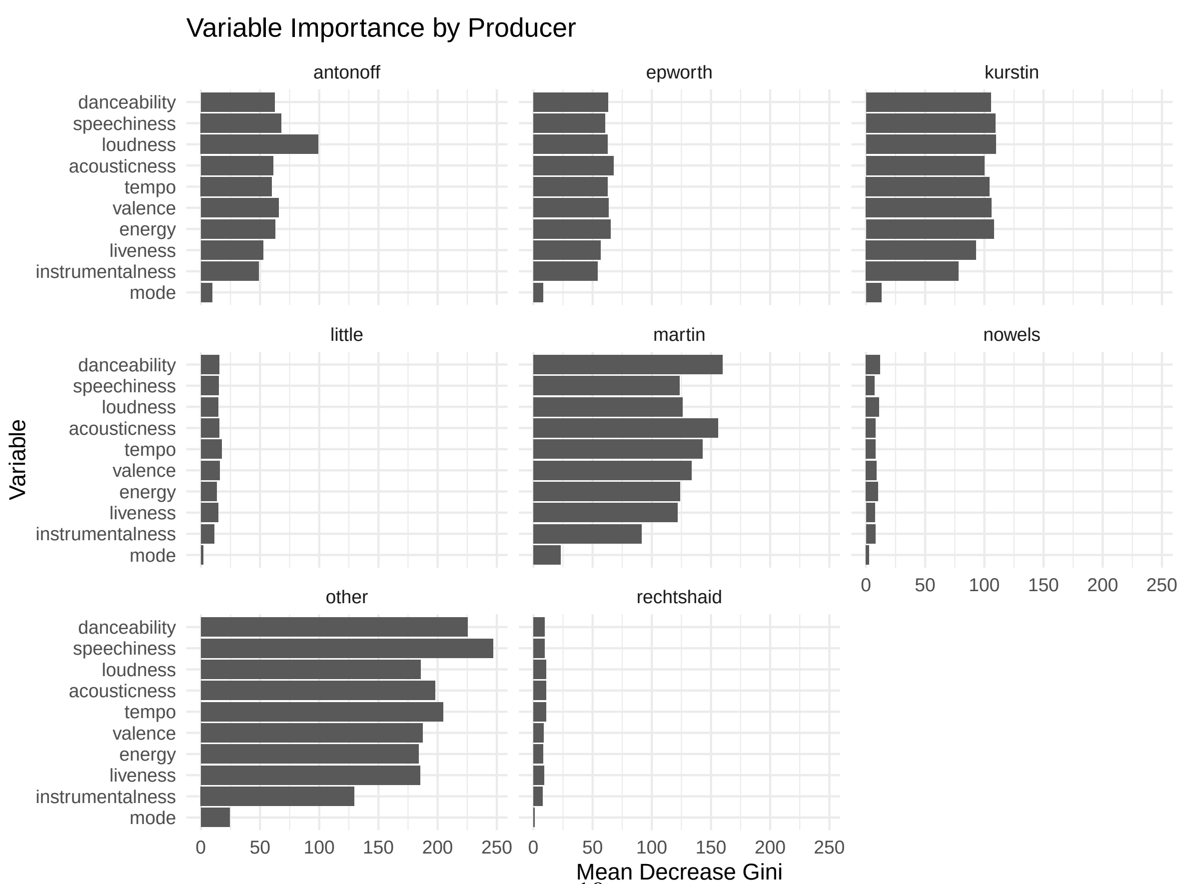


Figure 9: Importance Scores of Each Variable in Determining Antonoff's tracks



Relative to other producers, Antonoff's danceability has the least positive correlation with valence (0.311), the most negative correlation with acousticness (-0.0733), and significant positive correlations with speechiness (0.336), loudness (0.366), and energy (0.274).

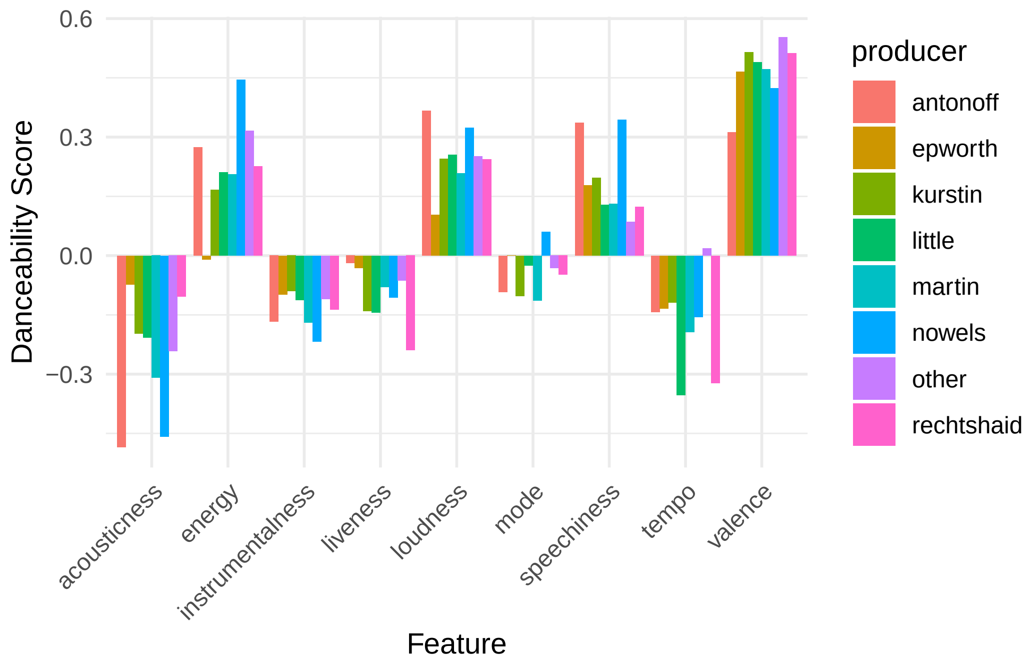


Figure 10: Danceability Scores Across Producers

5 Discussion

Our logistic regression model was not able to strongly predict whether or not a song was produced by Jack Antonoff using the Spotify API's audio features data. These data are limited, providing static metrics that stand in for entire tracks, which may not be detailed enough for intra-genre classification. However, when classifying the tracks, `danceability` was the strongest predictor.

Our Random Forest results, which were moderately strong at classifying tracks, indicate that Antonoff's music, characterized by its `danceability` score, exhibits a significant positive correlation with factors such as `speechiness`, `loudness`, and `energy`, while showing a negative correlation with `acousticness`, and a less pronounced positive correlation with `valence`, compared to other music producers. This profile—marked by speechiness, loudness, energy, and a blend of positivity with a touch of nostalgia—aligns with both detailed auditory analysis and critical assessments of Antonoff's sound.

Antonoff's tracks have been described as anthemic (Rosen), with a distinct vocal treatment that makes it sound, as Caleb Gamman, a Youtuber who went viral for his critique of Jack Antonoffs sound, observes, "the way it would sound to the person who's singing them" (Gamman).

> In the treble mix of the song, like in the very upper end, sort of above where the vocals are – he often really crushes that down. It's lower in volume than anything else, which is sort of a weird effect. When you hear someone speaking, you hear a lot of that, like, noise in their voice. When you yourself are speaking, you hear less of it, right? You hear more of your own voice bouncing around in your head. And so it sort of creates this effect of – **like he mixes his vocals the way it would sound to the person who's singing , them. Which is sort of strange.** He often has no vocals on that upper end, which is very unusual, and then often there are random little bits of noise happening up there which is the sort of thing her likes to do.

Perhaps, without further documentation on the `speechiness` feature's construction, clearing the noise and isolating the middle range of the vocals may make for stronger overall `speechiness` score in Antonoff's tracks. Antonoff also tends to collaborate with singer-songwriter artists, whose music is more lyric-oriented than dance hit pop.

Even Antonoff-produced songs with the highest `danceability` scores aren't exactly dance hits, compared to production styles like that of Max Martin.

For instance, the highest scoring song in danceability*, I Think He Knows* by Taylor Swift, seems danceable, yet the next highest-scoring songs – Taylor Swift's *Vigilante Shit* or Florence and the Machine's *Heaven is Here* – don't resonate as dance tracks. However, they aren't understated, ambient tracks either – the exhibit some sort of compelling quality, like they're coercing your body to do something. Like Gamman's description of Antonoff's vocal design sounding like how they would sound to the person who's singing them, these tracks seem like they're trying to get inside of your skin. But they don't have the rhythm of a dance track.

I Wanna Get Better by Bleachers typifies the Antonoff style as identified by our model. This song is more anthemic, resonating with a sort of neo-Springsteen aesthetic that Antonoff is often characterized (or admitting) to striving for, than danceable. It also does sound like it's trying to demand your attention, and the speechiness, loudness, and energy which are distinctly correlated with danceability in Antonoff's RF results, are all apparent. The redemptive, broken-but-getting-better tone can also be heard in the lyrics, but it still doesn't have the kind of rhythm for dancing – which maps onto how Antonoff's "valence" score isn't as a high as other pop artists in Figure 10.

Cornelia Street by Taylor Swift further exemplifies this near-danceable category. While melodically catchy, the lyrics themselves are wistful and longing. The vocal treatment, as described by Gamman, is there – Swift's vocals sound magnified, carrying the bridge of the song.

What does the danceability variable then signify in Antonoff's music? Our findings suggest that his most "danceable" songs are better understood through the lens of correlated features, even though danceability emerges as the primary predictor. This crowd-sourced metric, in conjunction with machine-listening-derived features, offers a unique insight into an artist's style, capturing an intuitive understanding of a song's tone and energy.

We spoke with Glen McDonald, a Principal Engineer at Spotify who worked at The Echo Nest, a music intelligence start-up that was acquired by Spotify in 2014. While Macdonald did not build the features, he has worked with them, and provided information on their provenance. Macdonald revealed that `valence` and `danceability` were created by giving tracks to college interns and asking them to tag whether a song was positive or gloomy, or danceable or undanceable. Variables like `energy` or `instrumentalness`, on the other hand, were determined primarily through machine listening techniques, with human subjectivity being applied to fine-tune the features[1] (McDonald).:

> "You could imagine writing a formula for energy that combines loudness and tempo and degree of harmonic variation or something. So that feature was [machine learning], but that one's more like a human helping a machine figure out a formula. Whereas valence is teaching the machine to try to reproduce a purely human thing. The same with danceability. I mean, danceability is whether a human can dance to it. The machine's not gonna dance, so the machine can't have any opinion on that. The computer could have an opinion on energy. And the computer can definitely have an opinion on loudness. So there's a spectrum from, loudness as purely analytical, and then energy is a little like loudness, with a little more subjectivity. And then danceability and valence are purely subjective."

Glen confirmed that this process didn't account for lyrics, which made valence a particularly difficult variable to build. It could, in theory, pick up on aspects of vocal performance, but it didn't process anything about languages or words or the meanings of songs. Take an upbeat, happy-sounding Elliot Smith song with devastating lyrics: the machine might register it as happy, where human listeners understand that it's sad. Furthermore, two humans might even disagree on the song's valence: "Plus, we have the confounding factor of like, a song that seems happy today could be sad tomorrow because the singer was killed in a plane crash. The song didn't change, but our world changed and our reactions changed" (McDonald).

Macdonald suggests combining energy and valence to create quadrant, which usually works "fairly well" to describe music (see Figure 11) (McDonald) . High energy and high valence could be generally happy, cheerful, upbeat. Low energy, low valence could be sad or downbeat. High energy, low valence is sort of angry. Low energy, high valence is serene or calming.

Our analysis reveals that the amalgamation of crowdsourced metrics (like danceability) and machine-listening variables offers a nuanced approximation of musical style. This approach not only mirrors the multifaceted nature of auditory perception but also enriches our model's

[1] A notable example of this being that bluegrass songs were given very high `speechiness` ratings. Because there were no banjos in the training data, the instruments were registered as human speech. The engineers had to go back and add more songs with banjos to the training data.

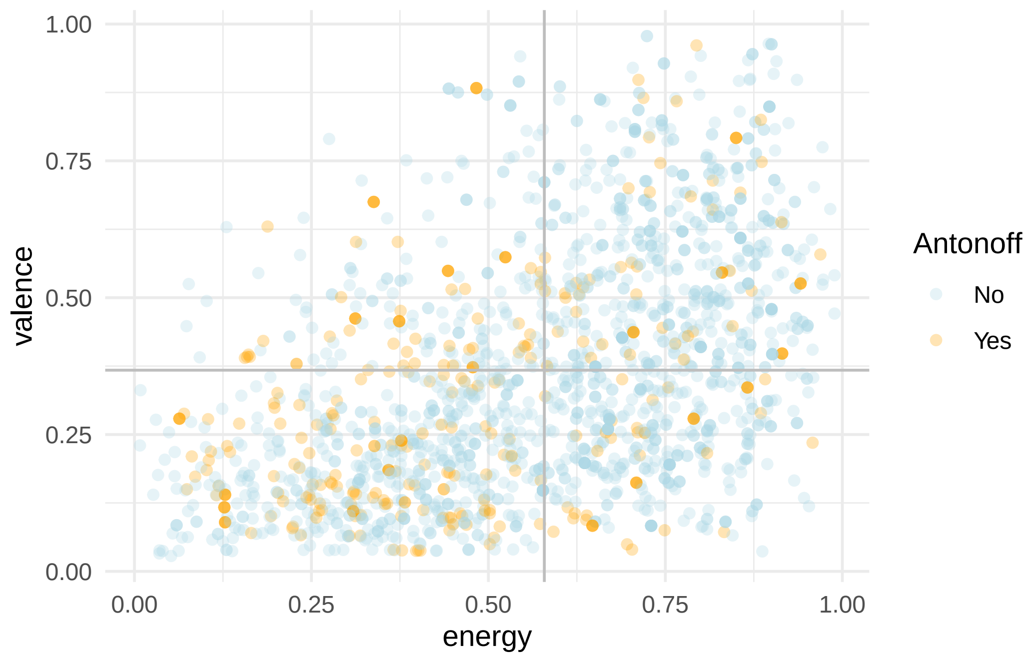


Figure 11: Valence and Energy Quadrant

capacity to capture the essence of an artist's style. Interestingly, the subjective nature of crowdsourced metrics, which might initially seem rudimentary, plays a crucial role in this process. They shape a more intuitive and embodied understanding of music, enhancing the overall analysis.

For instance, the work of Martikainen et al. illustrates this point effectively (Martikainen et al.). They combine open-source audio tools with qualitative, crowd-sourced listener data to understand audio-based stylistic variation in podcasts. This synergy between human subjectivity and algorithmic precision underlines the importance of blending these methods for a comprehensive analysis.

However, it's important to note that while these features offer a quantitative lens to examine music, they cannot fully encapsulate the nuanced and subjective aspects of musical style and creativity. The role of human subjectivity is pivotal in interpreting and experiencing music, as evidenced by the crowdsourced nature of certain features like danceability and valence. This aspect becomes even more significant when considering the broader scope of music recommendation systems.

As Nick Seaver's anthropological study of music recommendation companies shows, engineers "develop ways of thinking about musical preference and software, attempting to reconcile them with each other" ("Seaver). This extends to a more nuanced understanding of recommendation systems as a whole, highlighting the intricate balance between computational measures and human interpretation in the evolving landscape of music analysis and appreciation.

Future studies should look into reproducing crowdsourced danceability or valence variables to better understand the underlying user preferences that motivate these metrics. In doing so, we can build more robust ways to computationally model more ephemeral and subjective aspects of music such as style.

References

Gamman, Caleb. *Interview Conducted by Michaela Drouillard on 23 March 2023*. 2023.

Get Track's Audio Features. https://developer.spotify.com/documentation/web-api/reference/get-audio-features.

Kuhn, Max, and Hadley Wickham. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. 2020, https://www.tidymodels.org.

Marantz, Andrew. "Jack Antonoff's Pop-Music Collaboration Machine." *The New Yorker*, May 2022, https://www.newyorker.com/magazine/2022/05/23/jack-antonoff-pop-music-collaboration-lorde-taylor-swift.

Martikainen, Katariina, et al. *Exploring Audio-Based Stylistic Variation in Podcasts*. 2022, pp. 2343–47, https://doi.org/10.21437/Interspeech.2022-10871.

McDonald, Glen. *Interview Conducted by Michaela Drouillard on 11 April 2023*. 2023.

Rosen, Jody. "Jack Antonoff Is Only Making Music with Friends." *The New York Times Magazine*, 2020, https://www.nytimes.com/interactive/2020/03/11/magazine/jack-antonoff-profile.html.

"Seaver, Nick". *"Computing Taste:algorithms and the Makers of Music Recommendation"*. University of Chicago Press, 2022.

Silge, Julia. *To Downsample Imbalanced Data or Not, with #TidyTuesday Bird Feeders*. Julia Silge, 2023, https://juliasilge.com/blog/project-feederwatch/.

Thompson, Charlie, et al. *Spotifyr: R Wrapper for the 'Spotify' Web API*. 2022, https://CRAN.R-project.org/package=spotifyr.

Wickham, Hadley, et al. "Welcome to the tidyverse." *Journal of Open Source Software*, vol. 4, no. 43, 2019, p. 1686, https://doi.org/10.21105/joss.01686.

Wikipedia. *Jack Antonoff Production Discography — Wikipedia, the Free Encyclopedia*. 2023, https://en.wikipedia.org/wiki/Jack_Antonoff_production_discography.

Wilson, Carl. "The Real Reason the Internet Hates Jack Antonoff." *Slate*, 2023, https://slate.com/culture/2023/11/taylor-swift-producer-jack-antonoff-1989-taylors-version.html.