# *Features that Impact a Song's Popularity on TikTok*

***Group #18 – Chonel Chase, Shelby Carswell, Elias Yishak, Karishma Thakrar, Namrata Buxani***

## *Background*

TikTok is a social media platform created in 2016 that gained popularity during the pandemic in 2020. Music is a key part of the TikTok experience; users on the application are both resurfacing older music and developing a cultural sensation around newer trending songs. We have observed many now-famous musicians, such as Olivia Rodrigo and Loren Gray, start and accelerate their careers on the app and this has drastically changed the music production and advertising industries, remixing the world of sound. Brands are also capitalizing on leveraging sound in creative and strategic ways to drive digital marketing engagement (TikTok, 2019). Given the importance of audio, we're curious about why some songs are more successful than others. The data from this research can be used strategically by both artists and marketing agencies. Artists can utilize this information to cater songs towards key musical features to increase visibility on the platform, and marketing agencies can leverage songs likely to be popular in advertising to increase brand association and recall.

## *Problem Statement*

Our team plans to leverage regression analysis to determine how audio features impact rank and the trend of these variables via a time series analysis.

## *Primary Research Question*

Of the most top ranked songs, which audio features from Spotify's API contributed to a song being consecutively ranked on TrendPop (a music chart ranking the best performing songs on TikTok) and how does the importance of features change over time across the overall dataset, and potentially distinct subgroups? The tracked data variables observed include energy, mode, etc.

## *Literature Survey*

Studies by Royal Publishing have discussed various factors that contribute to the success of a song "moving" a user. In fact, previous research has supported the idea of factors like publishing company, artist, genre, tempo, danceability, liveness, and even the artists sales of the previous years as some of the top factors that increase probability of success and top ranks. This literature supports our hypothesis that happier or higher tempo songs will have a higher correlation with top rank and song popularity. In addition, songs that are more popular will have a higher brightness and danceability. Something of interest to keep in mind is that historical research states successful songs have a pattern of their own, which can make research outcomes challenging as the factors that influence a song's success evolves with time (Interiano, 2018).

A study in Towards Data Science sought to make quantitative sense of music and understand if a song's attributes could predict a track's popularity. The study utilized Spotify track data and measured popularity based on the total number of plays and how recent plays were of a track. While this study uses a different measure and source of popularity, it supports our method of pulling feature data as predictors. After initial data exploration, the study first attempted a linear regression model to predict popularity based on all factors. With a low R-Squared score produced by the linear regression model, the next models attempted were a decision tree and a random forest model. After seeing slight improvements with the random forest model, the study moved towards a classification approach with KNN to sort songs into bins of low, medium, and high popularity. Classification yielded a much higher accuracy score, leading to the conclusion that predicting buckets of popularity is more successful than attempting to predict discrete popularity scores. (Peker, 2021).

### *Initial Hypotheses*

During our data exploration process, we anticipate multicollinearities between variables like energy and liveness, as well as danceability and tempo. We assume that some features will need to be removed before we create our model. Our research also states that song lengths are getting shorter (Lutz, 2021). We hypothesize that the duration of a song will have a strong influence on popularity as well. We predict that songs that are more danceable and have higher tempos tend to be more popular, therefore being the highest significant variables. Billboard's Hot 100 from 1958-2013 found that a higher tempo and danceability often get a higher ranking on the Billboard Charts, supporting our hypothesis (Hoda, 2022).

### *Key Variables*

The Spotify API included the spotify track name and artist as well as the following features: acousticness, danceability, duration, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, and valence. Our TrendPop data gave us the following information for each top ranked song: date of top rank, likes, comments, shares, views, and rank. These two datasets were combined to create the one final dataset and the above features will be used as our key variables. Our dependent variable is the popularity rank pulled from TrendPop. We used key variables like Spotify ID, song name, and artist name to map the two datasets together to create a final descriptive dataset.

### *Data Consolidation & Cleaning*

This dataset came from two sources – Spotify API data and TrendPop website. The Spotify API data provided the features listed above in *Key Variables* that measure different audio characteristics of a song and are detailed fully in the API documentation. An example of one of these features is valence, which measures musical positiveness conveyed by a track (Spotify, 2023). This data was combined with TrendPop data that provided both the rank and date of rank that songs on TikTok as well as the engagement metrics listed in *Key Variables.* These two datasets were combined via a fuzzy algorithm which will be discussed later. The first step in the cleaning process was removing songs TrendPop had removed from the top 300 songs and labeled "Missing" within their website. From there, we removed the songs titles that have "Original Sound" as the song name as that indicates the sound was created on TikTok instead of it being pulled from a music platform. We utilized iterative code that looked through each record to remove parentheses from the sound title to increase matching of scraped song titles to Spotify API data. We then added audio features for the songs using the Spotify API and fuzzy match algorithm. Based on the song title and artist from TrendPop's data, this algorithm matched to a song in Spotify's database which had the most similar song and artist names. From our empirical research, we determined a threshold of 60% fuzzy score to pull in the Spotify API categorical variables; below this threshold yielded many incorrect matches. One consideration while using the fuzzy match algorithm was how we would handle duplicates that appeared for a small subset of songs. We considered songs with the same title and artist in TrendPop as duplicates which we then removed. For songs that had a very similar title and artist but with slight distinctions, we considered those distinct songs. An important call out is duplicates of the same songs were kept for the weekly significant feature analysis trend over time. For example, if a song was in the top 300 songs for various weeks, hence having "duplicate" records, these were all treated as individual records to capture the significance of features and how they change over time.

### *Exploratory Data Analysis*

The exploratory process included analyzing data types, cleaning the data, dropping null values, standardizing data, and determining any patterns in the data. During the exploratory data analysis, we looked at the standard deviation of some of the variables to determine if there was a large variance in the values of given Spotify features and completed a box plot analysis to determine outliers that may need to be investigated or removed. We also visualized the number of popular songs per artist and the number of times we saw popular song names. We analyzed various correlations between features and noted our observations, such as how likes and rank are correlated. We conducted unsupervised clustering with specific variables that we predicted would be significant such as danceability and tempo and grouped top ranking songs into four clusters. In addition, we analyzed the VIF score for each feature to determine if there was any multicollinearity.

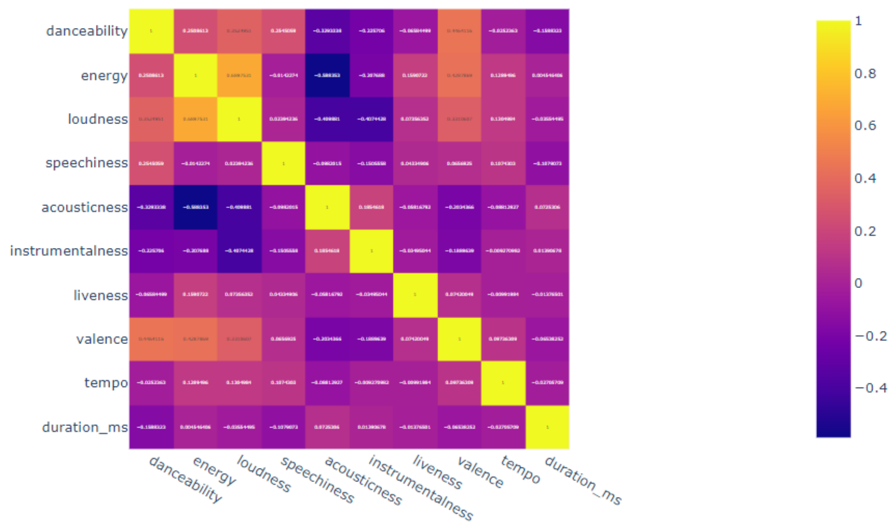



*Figure 1 - Correlation Matrix for all the features from the Spotify API*

**Principal Component Analysis (PCA)**

PCA was utilized to analyze patterns in multivariate datasets. We took the approach of loading our features that have numerical values and creating a graph that shows the projections of variance as seen below:

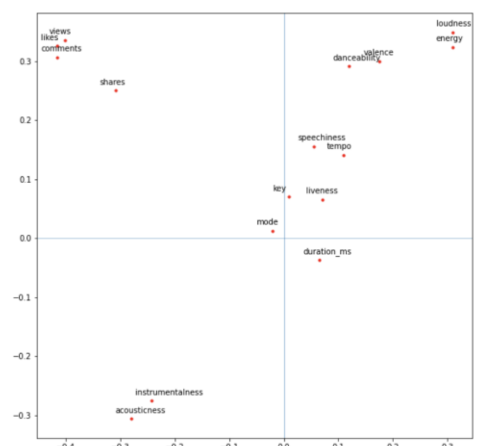



*Figure 2 – PCA Analysis*

Our PCA analysis is as promising as one can expect loudness and energy to be correlated or similar in value. Valence, which describes the musical positivity of a track, is correlated with danceability as well because we tend to dance to more happy, upbeat, positive tracks. While we are not planning to use likes, views, and comments in the modeling, it was also validating to acknowledge that these features were all similarly placed on the graph as well, with shares being close to these three features.

### VIF Analysis for Multicollinearities

VIF values were analyzed to assess any multicollinearity amongst the data. As shown in the table below, while energy and loudness have the highest VIF values, no variables indicate multicollinearity at a threshold of 5.

| Predictors | VIF |
|---|---|
| danceability | 1.6190 |
| energy | 2.9363 |
| loudness | 2.3814 |
| speechiness | 1.1357 |
| acousticness | 1.7002 |
| instrumentalness | 1.2659 |
| liveness | 1.0533 |
| valence | 1.5229 |
| tempo | 1.0621 |
| duration_ms | 1.0393 |

*Table 1 - Variable VIF values*

### K-Means Clustering

Based on given variables, we were curious about if K-Means would find patterns of clustering songs into genres or song categories. For preliminary analysis, we utilized danceability and tempo as the two features to build a K-Means clustering algorithm with. We utilized the elbow method to define the optimal number of clusters and where the centroids should be located. It was very interesting to see four distinct groupings of songs. We can see that the two clusters in the middle are close to each other, almost overlapping.

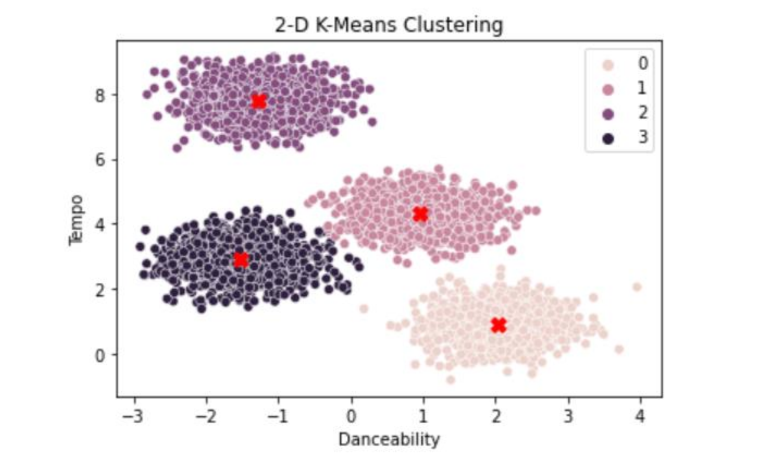



*Figure 3 – K-Means Exploratory Analysis*

## *Modeling Approach Overview*

### Logistic Model Approach & Analysis

Our original approach involved calculating the likelihood of a song being popular based on the track features from the Spotify API. Popularity would be a binary variable that was assigned a value of one, only if the track trended in the top 300 for at least two weeks consecutively. The data had a training and test split of 20%. The data was standardized via StandardScaler(). The 10 key variables that were tested in the logistic regression were the following: 'danceability', 'energy', 'loudness', 'speechiness', 'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo', 'duration_ms'. We added a historical popularity column indicating the number of previous weeks the song was listed on TrendPop in a row. If a song has two weeks of consecutive top rank, the song gets a 1, otherwise 0. We tested the precision of the model to accurately classify if a song would be classified with a 1 for popular or 0 otherwise. The model was only able to classify the 1's precisely 59% and 0's 60% of the time. The top significant variables were the following at 95% Confidence Interval: Danceability (0.209492), energy (-0.144431), loudness (0.138057), and liveness (0.124290). Both the scale and sign of the coefficient convey important information into the variables. For example, as the energy score decreases, there is a higher likelihood of it being considered a popular song. However, the weight of the coefficient is important too. The weight was not very large, which indicates that songs that are less noisy and intensely fast are more likely to be rated higher. In addition, we can look at the positive weights of danceability, energy, and loudness. These all have similar weights that indicate as the value of these features increase the rank increases by a similar factor. While the model produced significant factors, the R-Squared value was extremely low at 0.0153.

**Hyperparameter Tuning Approach & Analysis**

To increase the logistic model's performance, we altered our definition of popularity by conducting hyperparameter tuning of three variables. We redefined popularity by incrementing the number of weeks a song must trend concurrently to be considered popular and the minimum rank a popular song should have. Then we adjusted the threshold of the prediction value in 5% increments of the logistic model to increase accuracy. For example, a model with a threshold of 0.45 would classify less songs as popular than a model with a 0.60 prediction threshold. The full range of parameters can be seen in the hyperparameter matrix table, which created 3,528 models that we compared using ROC/AUC analyses at varying prediction threshold values, coefficient estimation of predictors, and predictor significance.

| | |
|---|---|
| Threshold | 0%, 5%, 10%... 95%, 100% |
| Minimum Rank | 25, 50, 75… 275, 300 |
| Consecutive Weeks at Rank | 2, 3, 4… 14, 15 |

*Table 2 - Hyperparameter matrix*

The Period and Minimum Rank hyperparameter matrix shows the matrix of minimum rank and consecutive weeks being varied. Each cell in the figure is comprised of 21 logistic models at different prediction threshold values. The value shown in each cell is the AUC for those 21 models. The best performing parameters have been highlighted blue and can be observed as a minimum rank of 50 and consecutive weeks at the given rank for 5 weeks.

| period | min_rank | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 25 | 50 | 75 | 100 | 125 | 150 | 175 | 200 | 225 | 250 | 275 | 300 |
| 2 | 59.96% | 59.99% | 59.85% | 59.96% | 59.91% | 59.16% | 59.15% | 58.36% | 58.67% | 58.76% | 59.04% | 58.70% |
| 3 | 59.95% | 60.11% | 59.95% | 60.08% | 60.56% | 59.63% | 59.47% | 58.68% | 58.66% | 58.59% | 58.88% | 58.51% |
| 4 | 60.83% | 60.42% | 59.15% | 58.55% | 59.40% | 59.54% | 58.69% | 58.66% | 58.25% | 58.53% | 58.31% | 58.21% |
| 5 | 51.91% | **62.75%** | 60.80% | 60.06% | 61.02% | 59.83% | 59.15% | 58.63% | 58.48% | 58.17% | 57.76% | 57.90% |
| 6 | 50.31% | 60.12% | 59.87% | 61.62% | 57.74% | 58.23% | 58.66% | 59.50% | 58.52% | 56.98% | 56.74% | 56.76% |
| 7 | 51.57% | 53.15% | 59.01% | 59.94% | 59.23% | 60.57% | 58.97% | 58.83% | 58.24% | 57.68% | 57.38% | 58.35% |
| 8 | 51.28% | 51.25% | 56.68% | 56.53% | 58.55% | 61.92% | 61.49% | 61.25% | 61.26% | 60.18% | 59.30% | 58.67% |
| 9 | 54.02% | 52.25% | 51.98% | 56.71% | 55.63% | 58.93% | 58.11% | 61.19% | 61.11% | 60.39% | 60.84% | 60.28% |
| 10 | 49.83% | 52.63% | 52.62% | 53.57% | 55.56% | 56.74% | 58.01% | 58.96% | 58.43% | 58.72% | 58.82% | 59.06% |
| 11 | 52.61% | 51.94% | 54.75% | 54.11% | 54.73% | 54.27% | 57.08% | 58.68% | 59.10% | 58.62% | 58.06% | 58.59% |
| 12 | 52.93% | 52.32% | 53.56% | 53.35% | 53.61% | 53.80% | 53.93% | 57.99% | 58.50% | 58.17% | 57.46% | 57.23% |
| 13 | 59.82% | 53.81% | 54.57% | 55.19% | 54.59% | 54.34% | 54.66% | 55.94% | 57.91% | 58.80% | 59.58% | 58.22% |
| 14 | 57.01% | 54.87% | 55.60% | 54.87% | 54.99% | 54.81% | 54.73% | 56.99% | 55.03% | 57.20% | 58.87% | 57.54% |
| 15 | 58.20% | 55.91% | 58.10% | 55.68% | 55.60% | 55.40% | 55.52% | 56.08% | 55.97% | 56.12% | 56.24% | 55.42% |

*Table 3 - Period and Minimum Rank hyperparameter matrix*

This results in models that produce an AUC value of 62.75%. This was the best performing model, but a good logistic regression model would have at least 75%. We further analyzed this model by varying the prediction threshold that yields the best performance when fixing minimum rank at 50 and consecutive weeks at rank to 5.

In the chart below, we can see details about the predictors of the 5% threshold model. By looking at the coefficients and p-values, we can see that the intercept, danceability and energy were found to be significant in explaining the variability found in the popularity indicator variable.

| Predictors | Coefficients | P Values | Significant (alpha < 0.1) |
|---|---|---|---|
| danceability | 3.1153730 | 0.0000004 | TRUE |
| energy | -1.7464058 | 0.0052008 | TRUE |
| key | 0.0230302 | 0.2554975 | FALSE |
| loudness | -0.0048759 | 0.8629382 | FALSE |
| mode | -0.1716895 | 0.2491413 | FALSE |
| speechiness | -0.3559780 | 0.5668697 | FALSE |
| acousticness | -0.2086337 | 0.5319779 | FALSE |
| instrumentalness | -0.0626016 | 0.8593000 | FALSE |
| liveness | 0.2682312 | 0.6184655 | FALSE |
| valence | -0.0334736 | 0.9243488 | FALSE |
| tempo | 0.0003861 | 0.8921658 | FALSE |
| time_signature | 0.4624705 | 0.1350348 | FALSE |
| const (intercept) | -5.9747028 | 0.0000662 | TRUE |

*Table 4 - Coefficients for model with highest AUC*

**Hyperparameter Analysis: Predictor P-Value Significance**

Although it was observed that the logistic regression models built in the previous section were not as performant as anticipated, an analysis of the predictors can help explain which predictors were more significant than others. However, due to the large number of models tested for this report, only a few will be focused on at specific fixed values to help illustrate the findings.

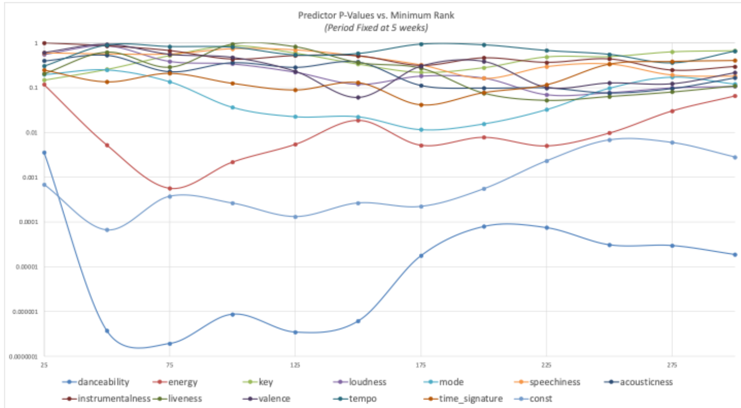



*Figure 4 - Predictor P-Values vs. Minimum Rank (Consecutive Weeks fixed at 5)*

The above figure shows the p values for each of the predictors and the intercept of our model for varying levels of minimum rank while holding the consecutive weeks fixed at 5 weeks. When looking for values of 0.1 or less, it can be observed that the predictors that remain significant are danceability, energy, and the intercept. Mode also seems to be significant as the minimum rank increases and then drops out of significance after a minimum rank of 225.

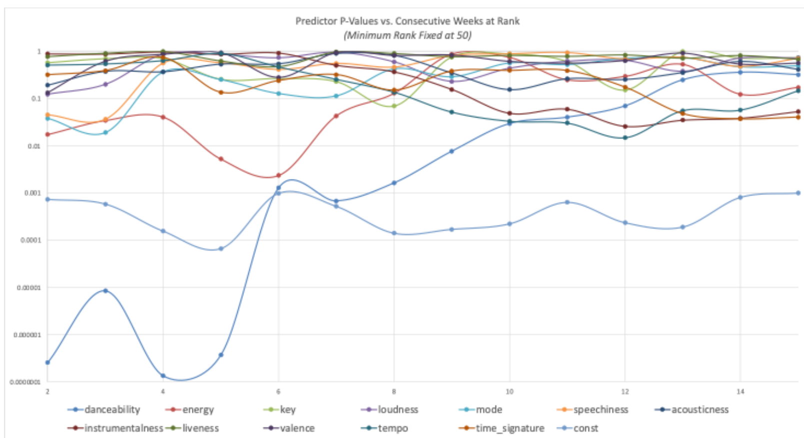



*Figure 5 - Predictor P-Values vs. Consecutive Weeks (Minimum Rank fixed at 50)*

The above figure is similar to Figure 4, except in this case, the minimum rank is held fixed at 50 while varying the consecutive weeks at rank. Similar to the previous figure, we see the same three predictors of danceability, energy and the intercept maintaining a level of significance for the majority of the different consecutive week levels. There are additional predictors that also drop into the significance level as well for larger values of consecutive weeks at rank.

**Random Forest Classifier**

Like the above logistic model, the data was scaled and then run with a Random Forest Classifier. As we know, the random forest creates an array of various decision trees that together have a better prediction accuracy than the individual trees on their own. The trees protect each other from the errors of the other models. The model was only able to classify the 1's precisely 56% and 0's 52% of the time. With an accuracy score of 0.5426 and classification percentages lower than the original logistic model, we did not move forward with the random forest model.

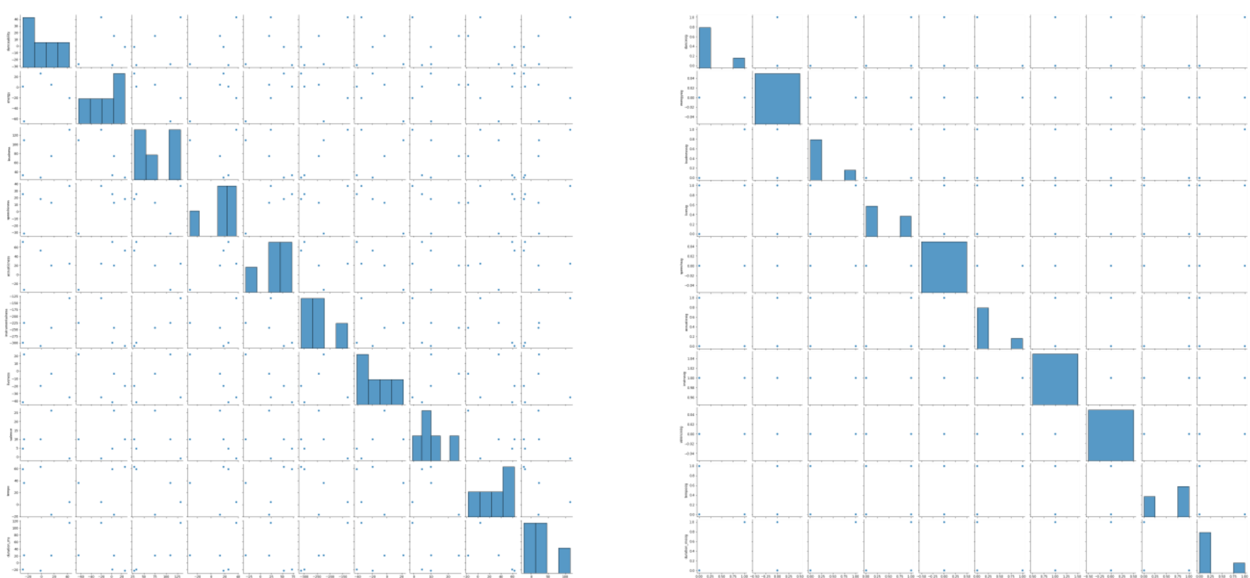
**Linear Model: Weekly Trend Over Time Approach & Analysis**

Our final model approach was to create an iterative time trend analysis that analyzes weekly records as an individual event. For each week, we created a sub-data frame of the records that applies to that week and creates a linear regression model that predicts the rank against other feature variables from the Spotify API. The data going into the linear model was trained via a split of 20% and scaled accordingly. The coefficients of the significant variables were documented for the date range that was run, and then trended over time to see if there was a trend in the weights of variables over time, as well as trend in which variables were most significant over time. The analysis was done via plotting the data in a linear graph and identifying trends to the naked eye first. From there, we aggregated the data from our weekly models to determine the frequency and seasonality of feature significance. We used a 90% Confidence Interval to classify a feature as significant in each period.

Conducting the linear regression over time allowed us to see how significant variables interact with one another and how they can even change as time and trends change. Our team started the analysis via utilizing all the Spotify API features in the linear regression model. However, in doing so, the R-Squared values were extremely low, and the prediction quality was lacking. With further research, we decided to remove the key and mode features as the numerical values associated with these are not correlated with a value to regress on, but were rather arbitrary. For example, the C Minor Key may be a value of 100, while the D minor key could be 50. Including this feature made the fitting process challenging.

After running linear regression models for each week of our data, we then consolidated the data across the linear models and created plots to show correlations and distributions across the predictors. As shown in the left plot below, many of the predictor distributions appear skewed in both directions and there is little correlation across predictors for the average coefficient distributions. The distributions on the right plot below are right skewed and again show little correlation across predictors for the count significance.

*Distribution of the Coefficient Values from Linear Regression and the Frequency Variable is Significant with a 90% Confidence Interval*



*Figures 6 & 7: distribution along diagonal and scatterplots against other predictors for average coefficient values and count of significance over all linear regression models*

We then analyzed each model's R-Squared value. The average R-Squared value across the models was 0.2198. We then set an R-Squared threshold of 0.7 and only 5 of the weekly models performed above that threshold with an average R-Squared value of 0.7827. The significant counts, average coefficient values, and average p-values are shown below for each predictor across the 5 performing models. Instrumentalness is shown to be the most significant predictor as it was significant in all 5 models and has a high p-value and the negative coefficient implies that as instrumentalness decreases, a song's rank will increase. However, as only 5 models performed at a reasonable R-Squared value, we cannot conclude that instrumentalness is truly significant across the dataset.

| **Predictors** | **Count of Significance** | **Average Coefficient Value** | **Confidence (1-P)** |
|---|---|---|---|
| danceability | 1 | 0.4283 | 0.5227 |
| energy | 0 | -10.3842 | 0.3741 |
| loudness | 1 | 75.3483 | 0.6441 |
| speechiness | 0 | 12.4646 | 0.6098 |
| acousticness | 1 | 27.0309 | 0.6365 |
| instrumentalness | 5 | -242.1174 | 0.9931 |
| liveness | 2 | -15.4504 | 0.6458 |
| valence | 0 | 10.0116 | 0.2687 |
| tempo | 3 | 29.2307 | 0.6872 |
| duration_ms | 1 | 23.3132 | 0.5305 |

*Table 5 - Predictor Significance, Avg Coefficient Value, and Average P-Value across the top 5 linear regression models*

### *Challenges with Data and Modeling*

Data challenges that we faced were regarding scaling likes and comments to the same scale as the other categorical features from Spotify. In addition, we were challenged with the duplicates, as discussed in the cleaning section of the paper. We also had challenges with incorporating the virality data from TrendPop such as likes, shares, and views. These were incorporated in a preliminary iteration of the daily linear regression. However, they decreased the accuracy of the model, leading the R-Squared to be 0.69. This may be suggestive of the sporadic pattern of likes, shares, and views that may not lead to a guaranteed rank increase. For example, songs that are very bad may also get many views and shares, as users are discussing it.

We also faced challenges with our initial approach of defining popularity as a certain rank for a consecutive number of weeks with a logistic model. Our logistic model did not correctly predict popular songs and categorized all songs as non-popular. We attempted to resolve this issue by tuning three parameters: the model threshold value, the minimum song rank to be considered popular, and the minimum number of weeks a song must be in the top rank threshold to be classified as popular. However, we saw little success here and, in some cases, saw that lowering the model threshold too much caused the model to incorrectly classify the non-popular songs as popular. After additional poor results with random forest, we decided to change our approach and look at variables over time with multiple linear regression models. While we saw some success with this approach, we still had very few of the linear models perform at a reasonable R-Squared value.

### *Novelties in Research*

A key takeaway our team had from this research was understanding how difficult and challenging it is to define virality and popularity in the terms of music. This is because the way music is being marketed and shared is constantly changing, even within the same platform, like TikTok. As an example, within the TikTok app, there are sub-threads for dancing trends, retail trends, political trends, etc. How a song goes viral can be dependent not only on what sub-threads are most popular at the time, but also based on hashtags and the current events that are going on in the user's environment or country. These features are hard to capture objectively and therefore the model is missing this pivotal data. We recognize that this is why this research has never been completed in a very accurate way prior that can fully assist musicians the way we envisioned. For this model and research to be applicable, the model would have to iterate constantly to update to the most significant features and introduce new features that are subjective if possible. Finally, in a future iteration of this research, hashtag metadata would be important to be pulled in as that is a major part of how songs get exposure based on trending tags (Admin, 2019).

### *Further Research*

Further research that would be interesting to explore would be to assess linear models across varying time horizons and analyze if different trends can be identified, and if there is any seasonality that impacts what types of songs will be popular. Another research avenue that could be promising is to label records into clusters and to assess if there are distinct subgroups of popular songs, as this method produced reasonable results in a similar study (Peker, 2021).

### *Results & Conclusion*

Based on the various approaches and outcomes outlined above, we can conclude that predicting the popularity or rank of a song on TikTok can be quite difficult. The table below shows each model approach along with the metric used to measure performance.

| Model Approach | Metric of Significance |
|---|---|
| Logistic Regression | R-Squared: 0.0153, AUC Score: 58.45% |
| Logistic Regression with Hyperparameter Analysis | AUC Score: 62.75% |
| Random Forest | Accuracy: 0.5426 |
| Linear Regression – Weekly Data | Avg R-Squared (across all models): 0.2198 |
| Linear Regression – Top 5 Models | Avg R-Squared: 0.7827 |

*Table 6 - Model approaches and metrics used to measure performance*

The most successful model approach used was the linear regression for weekly data. While the overall R-Squared average for the weekly linear models was low, the subset of the top 5 linear models performed at a relatively high R-Squared average. These models indicate that instrumentalness is the most important feature in determining the rank of a song, however broadening the scope to the changes in coefficients and significance across all the models further indicate that popular songs with a high rank may have vastly different features that are constantly changing. This conclusion supports our literature research that states how popularity and rank are hard to track and narrow down to specific feature qualities (Interiano, 2018). Music is consistently transforming over time and therefore, looking at data over a course of a few years, we can see that features will vary. While the models created in this analysis may not provide predictive answers to what will make a song popular, they do provide insight into how vast and distinct "popular' music can be and how different types of music can still become popular in social media trends.

### *References*


Admin, T. T. (2019, September 2). *Ways to get your music to go viral on TikTok*. Influencer Marketing Factory. Retrieved April 16, 2023, from https://theinfluencermarketingfactory.com/5-ways-to-get-your-music-to-go-viral-on-tiktok/#:~:text=TikTok%20Hashtags%20work%20almost%20as,be%20based%20on%20your%20music.

*Get track's audio features*. Web API Reference | Spotify for Developers. (n.d.). Retrieved April 1, 2023, from https://developer.spotify.com/documentation/web-api/reference/get-audio-features

Hoda Khalil, T. C. (2022, October 21). *Is it the lyrics? is it the energy? Big Data tries to predict what makes a song popular*. Scroll.in. Retrieved April 1, 2023, from https://scroll.in/article/1034737/is-it-the-lyrics-is-it-the-energy-big-data-tries-to-predict-what-makes-a-song-popular

Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018, May 16). *Musical trends and predictability of success in contemporary songs in ...* The Royal Society Publishing. Retrieved April 1, 2023, from https://royalsocietypublishing.org/doi/10.1098/rsos.171274

Lutz, B. Q. (2021, October 5). *If you've noticed songs have been getting shorter, you might be right*. Dance Music NW. Retrieved April 1, 2023, from https://www.dancemusicnw.com/science-behind-shorter-songs-2019/#:~:text=Over%20the%20last%20five%20years,3%20minutes%20and%2030%20seconds.

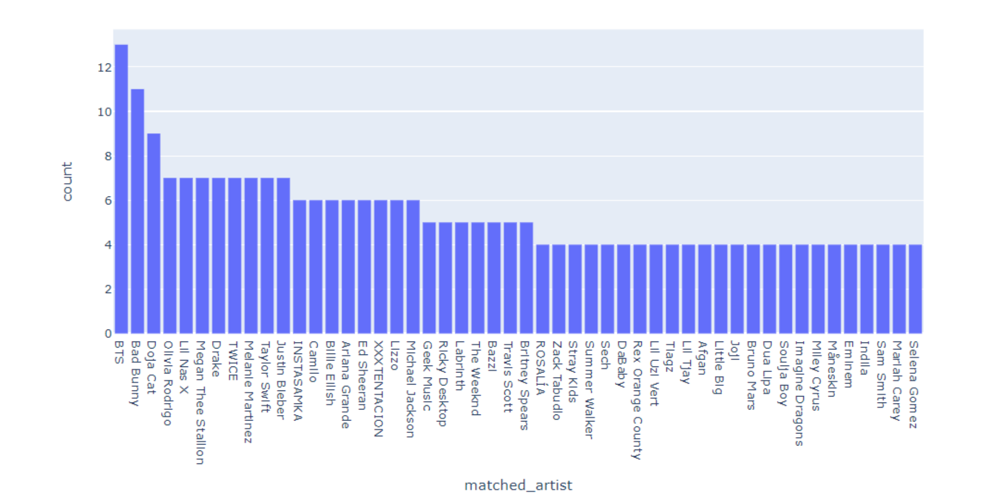
Peker, P. (2021, November 17). *Predicting popularity on Spotify - when data needs culture more than culture needs data*. Medium. Retrieved April 1, 2023, from https://towardsdatascience.com/predicting-popularity-on-spotify-when-data-needs-culture-more-than-culture-needs-data-2ed3661f75f1

TikTok. (2019, August 16). *New studies quantify TikTok's growing impact on culture and music*. Newsroom. Retrieved April 1, 2023, from https://newsroom.tiktok.com/en-us/new-studies-quantify-tiktoks-growing-impact-on-culture-and-music

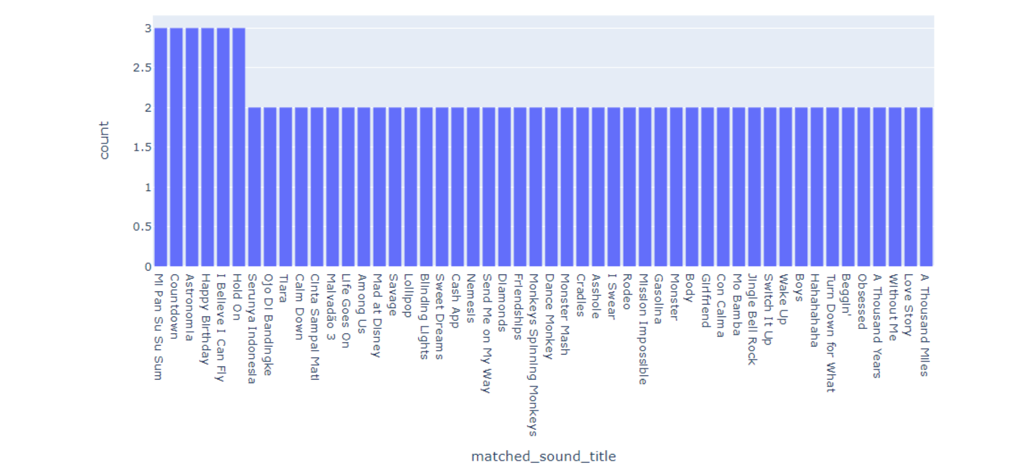
## *Appendix*

| Variable | Summary |
|---|---|
| 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. |
| 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. |
| 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. |
| 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. |
| acousticness | A confidence measure from 0.0 to 1.0 of whether the track is acoustic. |
| 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". |
| 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. |
| 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). |
| tempo | The overall estimated tempo of a track in beats per minute (BPM). |
| duration_ms | The duration of the track in milliseconds. |
| 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. |
| 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. |

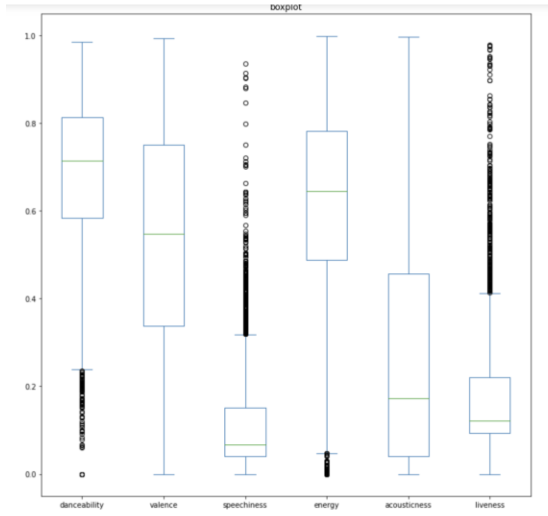
*Appendix Table 1 – Summary of metric definitions from Spotify API data*




*Appendix Figure 1 – Count of songs by artist name. We can see from this analysis who the most popular artists are in the data set.*



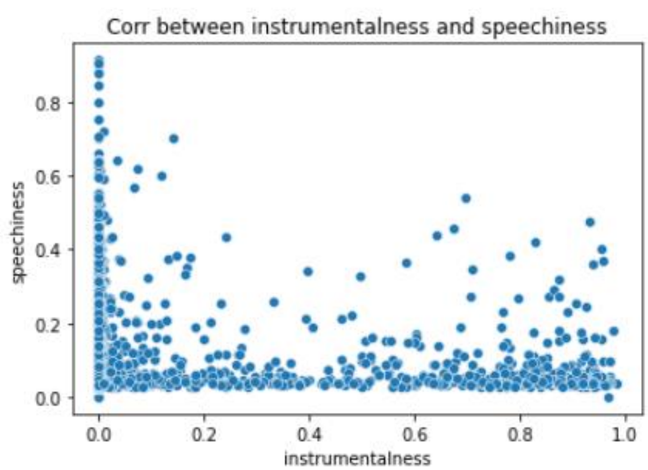

. *Appendix Figure 2 – Count of song titles. This analysis shows any song titles that have duplicate names, but might not be true duplicates due to having different artists.*



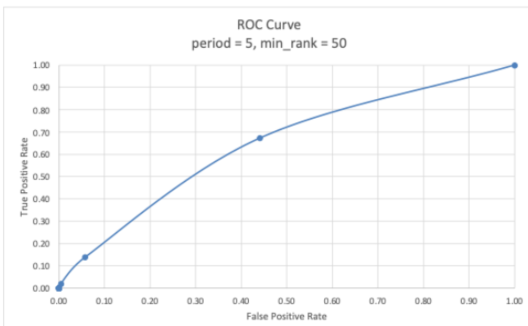

*Appendix Figure 3 - Box plot of outliers. This analysis shows potentially significant outliers for multiple features.*

```
danceability        0.172615
energy              0.212419
speechiness         0.118724
valence             0.254407
mode                0.493056
instrumentalness    0.267338
liveness            0.142718
acousticness        0.291616
dtype: float64
```

*Appendix Figure 4 – Standard deviation of key variables that are in similar scalar forms. We can see that the variation is the highest for the mode, and lowest for the speechiness.*




*Appendix Figure 5 – Simple correlation of exploratory analysis. While this is just one of the various permutations, the key takeaway is that most of the variables did not have strong correlations with each other.*




| threshold | true_positive_rate | false_positive_rate |
|---|---|---|
| 0% | 100.00% | 100.00% |
| 5% | 67.33% | 44.14% |
| 10% | 13.86% | 5.72% |
| 15% | 1.98% | 0.44% |
| 20% | 0.00% | 0.03% |
| 25% | 0.00% | 0.00% |
| 95% | 0.00% | 0.00% |
| 100% | 0.00% | 0.00% |

*Appendix Figure 6 - ROC Curve with Minimum Rank = 50 and Consecutive Periods at Rank = 5. Threshold values greater than 20% have been omitted from the table since the curve has collapsed to the graph's origin. It can be observed from the curve that the point that maximizes true positive rate while minimizing false positive rate is at the threshold value of 5%.*