# **Hit Song Science Problem**

**Team 132:** Carlos Gomez, Destiny Hartin, Kathleen Widjaja, Motiloluwa Ogunsina

## **Introduction**

For our project, we aim to tackle the Hit Song Science Problem by developing models that predict if a song is a success, based on a combination of musical features (tempo, loudness, etc.) along with external factors that can contribute to a song’s success such as social media following. Most research papers we saw looked at one aspect (often just audio features) of a song to predict its success, however, our project innovates on this approach by combining musical features with the number of followers the artist has on social media, if the song trends on TikTok, and whether the song has a music video. Our second improvement to similar studies is to evaluate the model's ability to predict multiple measures of song popularity depending on the goals of the user. Most other studies choose one measure of popularity to predict, or the entire study evaluates the relationship between multiple success measures. Another innovation our tool provides a recommendation for what could improve the popularity of the song. By revealing which factors most strongly predict a song's popularity, our model can guide artists and labels in targeting the right strategies to maximize success.

## **Problem Definition**

We want to figure out how to predict if a song will be popular on Spotify (Regression) and whether the song made it to the Billboard Hot 100 (classification), based on artist information and the song’s musical characteristics, so artists and labels can know what makes a hit. Given our dataset, where each song is represented by a feature vector (e.g., danceability, energy, loudness, etc.) and a target variable representing popularity (either a score from 0 to 100 or a label on whether the song is on the Billboard Hot 100s chart), our task is to develop two predictive models for each goal, that minimizes the error between predicted values and actual values . For popularity score, we aim to optimize where L is a loss function (e.g., mean squared error), and n is the number of songs. For predicting whether a song makes it to the Billboard 100, we test different models through cross validation and optimize for F1 score which aims to balance precision and recall. This is used to minimize both false positives and false negatives.

## **Literature Survey**

To define a song’s success, similar studies have used Billboard chart metrics (existence, duration, peak spot, performance over time). While using a combination of these success metrics is helpful to capture the full picture of a song’s initial and lasting success, it also introduces an additional level of complexity, requiring a separate model be built to determine weights for each success variable (Lee, J., & Lee, J.-S., 2018). Streams on services like Spotify have also been used, however, these studies limited their research to using the musical features of a small dataset to predict the number of streams and saw that the musical traits alone carried little predictive power (Nijkamp, R. 2018).

When deciding what factors to use in predicting success, we saw most studies on the Hit Song Science Problem analyze isolated factors like sound characteristics (Interiano, M et al., 2018) (Raj, Rishita et al. 2023)**,** uniqueness (Yu, Yulin et al. 2022), social media and playlist linking (Bischoff, K et al., 2009), or lyrics (Chiru, C. & Popescu, OG., 2017), using models such asRandom Forest, Bayesian Networks, and linear regression models to predict a song’s success. One study assessed if lyric simplicity affected song success by computing the text compressibility of Billboard Hot 100 songs over time (Varnum, M. E. et al., 2021). These approaches show varying accuracy, with Bischoff’s 2009 study achieving 80% precision using social-based playlist linking of songs to predict success. Limited research explores how *combining* these factors can improve the accuracy of the model’s predictions. Our approach is inspired by Yee & Raheem, who used Spotify audio features (like tempo and danceability) and social media features from YouTube to evaluate song success (achieving 22% accuracy with audio alone and up to 80% including social data) (Yee, Y. K. & Raheem, M. 2022).

Including factors like lyric complexity would not be feasible to implement in this project as it requires heavy analytics work to get the data. Another difficulty in this field is the processing of audio feature data seen in Bogdanov & Dmitry which discussed a standardized tool that can be used for signal processing of music (Bogdanov, Dmitry, 2013). While the usefulness of this would be helpful to obtain additional audio information, the framework is still being investigated. We will therefore need to find a dataset that already includes a song’s audio features.

Data from Last.fm or other music streaming platforms can be used to determine a song’s spread through its linkage in playlists (Su, A. T. et al.,2024), though this requires extensive additional analysis and data mining (Reisz, Niklas et al., 2022).

## **Proposed Methods**

#### **Data Collection & Cleaning**

Our original dataset was the Million Song Dataset; where you could access metadata on songs and their audio features, however, the site now only supports metadata for songs. We attempted to get audio features by connecting to APIs that were used in similar research studies like Spotify, Billboard, Last.fm, and MusicBrainz. However, these APIs have since changed to either provide less data, added a paywall, or cease to be available entirely. Since we could not find a way to get the data needed from the Million Song Dataset, we decided to use a Spotify based dataset from Kaggle (Pandya, Maharshi, 2023).

Our chosen dataset contains the song’s musical features as well as track popularity (a value between 0 and 100). Popularity is calculated by an algorithm and is based on the number of plays that a track has and how recent those plays are. This is already in the dataset and is not calculated within this project. To capture more information about the artist and the influence of social media on a song’s popularity, we connected the data to APIs (The Audio DB, Instagram Follower Count Scraper) and used a dataset on TikTok trending songs. From these, we combined the audio features with artist style & genre, artist’s gender, music video statistics from YouTube and Instagram follower count. A combination of social media sites like this was not seen in our research of similar analyses. We also used a dataset containing historical Billboard Hot 100 tracks to create a binary variable that marked if the song has ever been on the chart (Miller, Sean 2025).

Significant data cleaning was required to get additional information for a song as there are variations in song titles and how artists are listed. For example, the song title might include “(feat. Artist)” or “with Artist”, etc. We removed these strings from the song title. Similarly, multiple artists on a track can be delimited in different ways. We expanded the dataset into a song/artist level (vs song level) to normalize the dataset and avoid errors with any joins to other data. Our current dataset has over 158k rows, 55 columns, and is 73,172 KB. Data cleaning and extraction was done in Python and R.

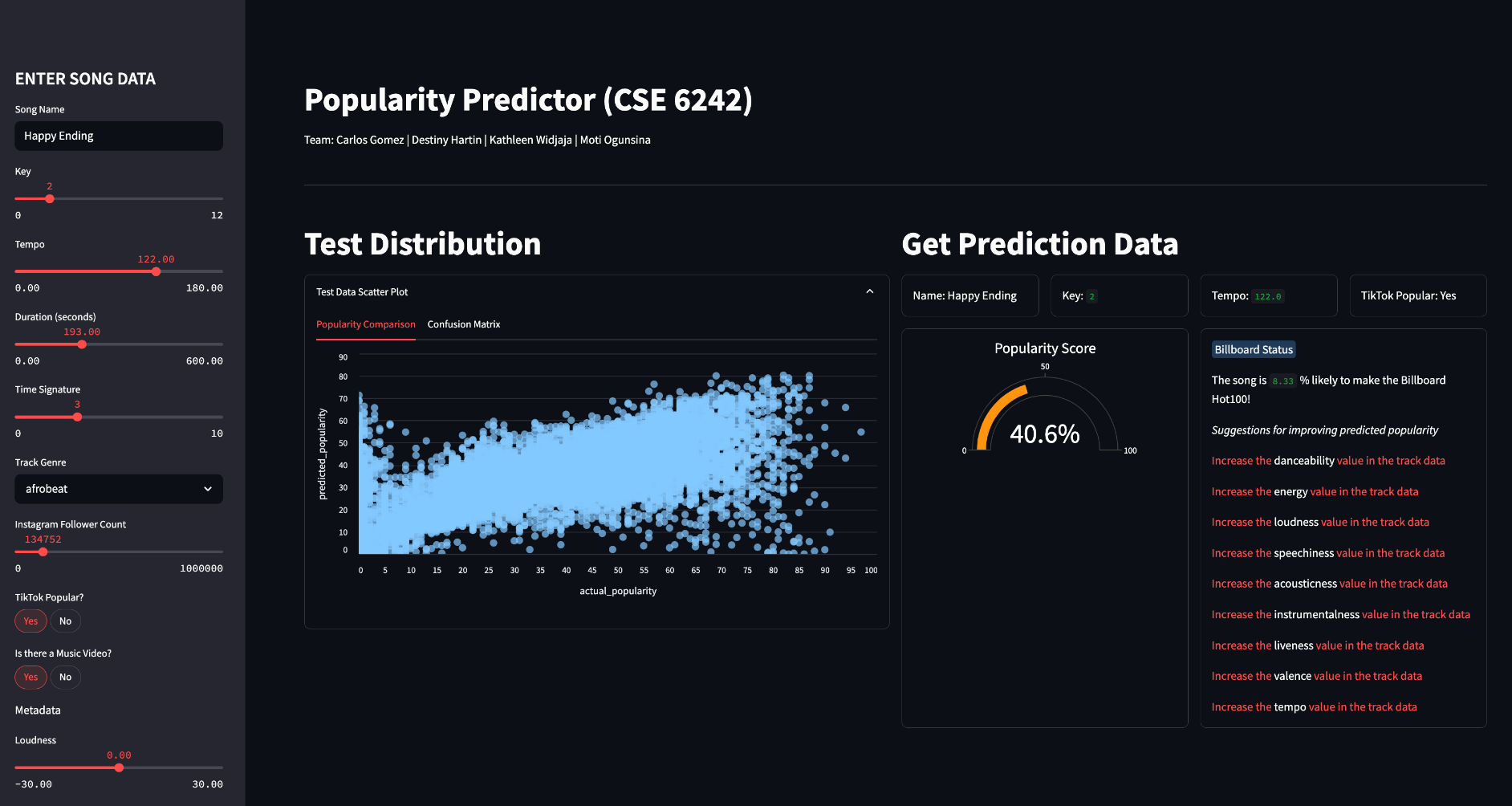
After joining all these datasets, there were significant number of rows missing for multiple columns. This was because not all of the songs and artists in our datasets existed in the APIs we used. Categorical features were label encoded, with an “Unknown” category. For numerical columns, only the number of Instagram followers was missing, the value was input with the median, and a missing mask was added to indicate to the model that the value was missing.

Another challenge with our dataset is that there is a significant class imbalance for whether a song made it to the billboard 100. To account for this, Synthetic minority oversampling was used in the training set for the classification algorithms (SMOTE). This improved the F1 score of all models. SMOTE works by creating synthetic points of the under sampled class through interpolation.

#### **Model Selection**

The goal of our project is to create a model that can accurately predict a song’s success. As mentioned above, there are two metrics that we used to define “success” of a song: a popularity scoring and song existence on the Billboard Hot 100 chart. To do this, we split our dataset into testing and training sets and trained the following algorithms to predict the song’s success: Random Forest, Gradient Boosting, Neural Networks (through MLPRegressor) and K- Nearest Neighbors (KNN). They were trained using a song’s audio features, artist information, TikTok trendiness, if a music video exists, and the artist’s Instagram followers to predict success. Each model was run twice, each time using a different predictor of success: popularity scoring (Regression) and probability of appearing on the Billboard Hot 100 chart (Classification). Each model was tuned through a grid search of hyperparameters, the best parameters were chosen using cross validation.

#### **Visualization**

The goal of our visualization is to provide an interactive tool that artists, producers, or record label executives can use to input information about a potential new song and see what the song’s success will look like. The visualization first shows information about the model’s accuracy including a comparison of actual values vs predicted values via an interactive scatterplot where users can hover over points to see information about the song.

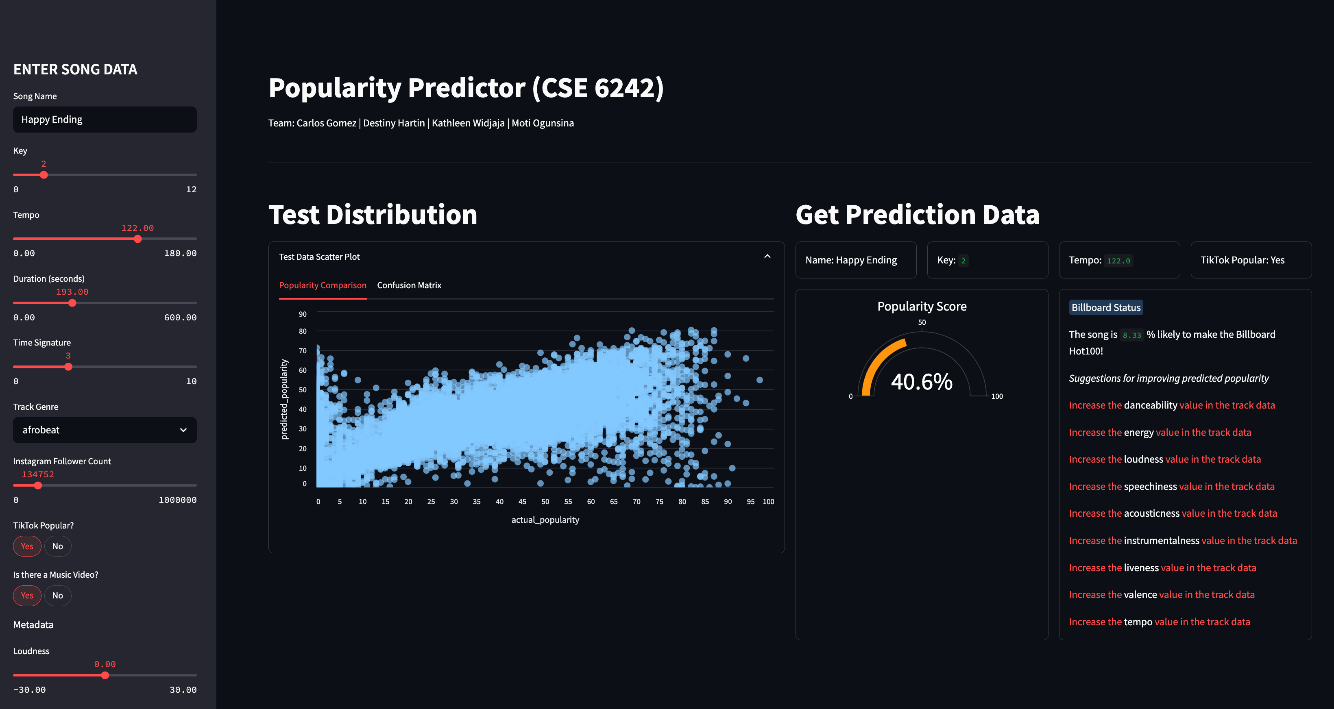
The user will be also able to input variables and run our model to predict the success of their song. An additional feature is recommendations on what could improve the success of the track. To do this, we look at the audio features as well as track genre and for each feature, we test different values (for audio features we test values from +.01 to +.1) to see how the changes improve the predicted popularity scoring. We then store the best predicted values and recommend updates that improve the predicted popularity score. The goal of our tool is to allow members of the music industry to “simulate” the success of a potential song before it is invested in. Similarly, we want to provide a tool that recommends what to invest in such as increasing the danceability of a song.

Figure 1 – Distribution of popularity scores vs the model’s prediction

Figure 2 - Full view of Song popularity prediction dashboard

1. **Evaluation**

Our evaluation questions are the test accuracy of our predictions, since end users need to ensure accuracy of the tool, as well as which variables are important in determining song popularity. For Spotify popularity, we used mean square error (MSE). For the Billboard classification model, we also used accuracy, precision, and recall evaluating how the model is classifying songs. Results below:

|  |  |  |
| --- | --- | --- |
|  | Popularity Score (0 – 100) | |
| Algorithm: | Test Mean Square Error |  |
| **Random Forest** | **205** | **0.51** |
| Gradient Boosting | 212 | 0.50 |
| Neural Network | 423 | 0 |
| KNN Regression | 285 | 0.32 |
| AdaBoost Regressor | 394 | 0.07 |

*Table 1 -Test metrics for popularity score*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Billboard 100 | | | |
| Algorithm: | Accuracy | Precision | Recall | F1 Score |
| **Random Forest** | **0.96** | **0.69** | **0.58** | **0.63** |
| Gradient Boosting | 0.88 | 0.26 | 0.67 | 0.37 |
| Neural Network | 0.05 | 0.05 | 1 | 0.09 |
| KNN | 0.91 | 0.33 | 0.59 | 0.42 |
| AdaBoost | 0.90 | 0.30 | 0.64 | 0.41 |

*Table 2 – Test metrics for Billboard classification*

## **Conclusion and Discussions**

In this project, we forecasted a song's success using both audio features and social media popularity. We evaluated and predicted two success measures: song popularity score and whether a song made it on the Billboard Hot 100. Based on measures of accuracy explained above, random forest emerged as the best performing models for both success metrics. The random forest regression to predict the popularity score resulted in a mean square error of 205 and a of 51% for regression. The random forest classification for the Billboard Hot 100 resulted in an accuracy of 96% and an F1 score of 62%. Both models for each success metric resulted in a moderate fit, but there is room for improvement in both models. While the accuracy of the Billboard classification model might look high, the class imbalance is the main reason for the high accuracy as there are many more songs in our dataset that do not make it onto the chart. The F1 score therefore gives us a better view of the performance, which shows that there is still some misclassification in the model. Similarly, for the popularity score, the low implies that there is still more variability in the data that our model and data does not fully capture. We would therefore not recommend that the current tool built be used for commercial use, however, we did see improvements to similar studies through the addition of the attributes in our model.

Through similar research studies, we saw that musical traits alone did not carry significant predictive power towards a song’s popularity (Nijkamp, R. 2018). We hypothesized that adding outside factors that influence a song’s popularity (such as the artist’s following, music videos for the song, or use of the song on social media apps) would add predictive power in a model. A similar study tested this theory by predicting a song’s performance on Spotify top charts and achieved 80% accuracy using musical attributes in tandem with YouTube data on the song (Yee, Y. K. & Raheem, M. 2022). While adding more diverse social media factors in our model (including YouTube, TikTok, and Instagram) did slightly improve the accuracy, the importance of TikTok and music video features was low.

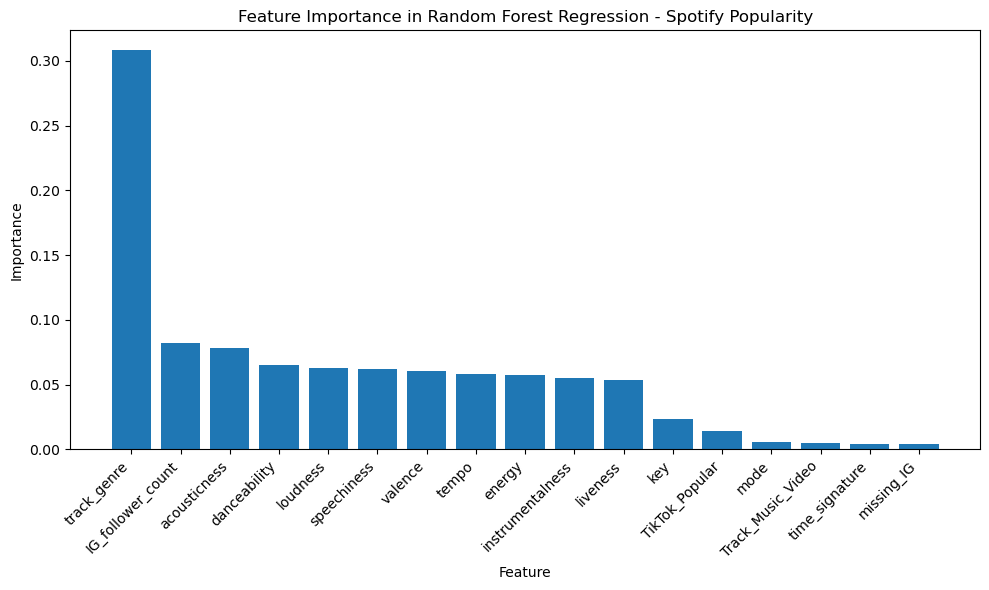


Figure 3 - Song Popularity Scoring Model feature importance

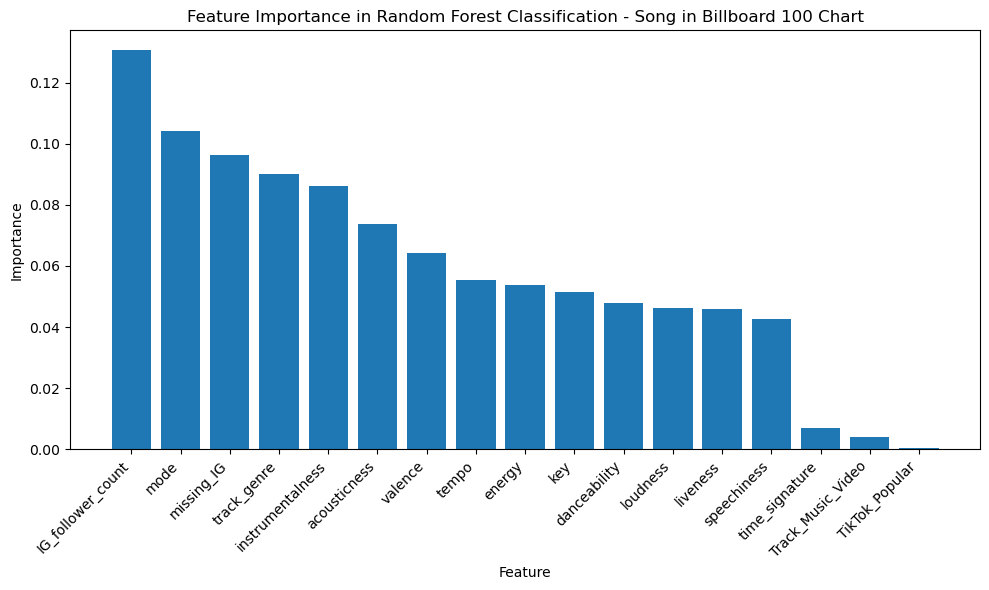


Figure 4 - Billboard 100 Model feature importance

The limitations we faced and potential improvements to the model include improved data collection and utilization of a larger and more balanced base-dataset of songs. APIs such as SongStats and ChartMetric include the number of plays on all streaming platforms for an artist and song, and follower counts for an artist on all major social media sites. They also have this data for a robust set of artists and songs. However, these require monthly subscriptions ranging from $250-700 for use of their service. Similarly, major streaming/social media platforms do not allow scraping or use of their APIs to get the information we needed for this analysis. The tool we chose to use did not include information for many of the artists and songs in our dataset, resulting in mostly null rows for the added social media fields which likely weakened the power of those features in our model.

In summary, our research found that adding additional features does add a small amount of predictive power to the model, however, we recommend improvements like making sure more rows in the data are not null for significant attributes that could add predictive power to the model, or testing the addition of even more potential contributors to song popularity such as lyric-based characteristics.

## **Distribution of work**

The work for this project was distributed across our team with all team members participating in the idea creation and research. Documentation was also split across the team with Destiny Hartin and Carlos Gomez working on all written reports, Motiloluwa Ogunsina creating the slides and video for the team, and Kathleen Widjaja creating the final poster for the team. Planning the workload for the team and creation of the GANNT chart was done by Kathleen. Destiny worked on data cleaning, researching additional data fields, and connecting to APIs. Kathleen also worked on data collection by adding multiple key social media features. Development, testing, and tuning of the models was done by Carlos. Motiloluwa built the interactive visualization tool and connected the visualization to the models.

1. **References**

## Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018). Musical trends and predictability of success in contemporary songs in and out of the top charts. Royal Society Open Science, 5(5). <https://doi.org/10.1098/rsos.171274>

Bischoff, K., Firan, C. S., Georgescu, M., Wolfgang Nejdl, & Raluca Paiu. (2009). Social Knowledge-Driven Music Hit Prediction. Lecture Notes in Computer Science vol 5678, 43–54. <https://doi.org/10.1007/978-3-642-03348-3_8>

Horsburgh, A. (2020, February 25). Metrics that matter: 23 KPIs for record marketing. Medium. <https://amberhorsburgh.medium.com/23-metrics-that-matter-229cfba81440>

Nijkamp, R. (2018). Prediction of product success: explaining song popularity by audio features from Spotify data. University of Twente, the Faculty of Behavioural, Management and Social Sciences. <https://essay.utwente.nl/75422/1/NIJKAMP_BA_IBA.pdf>

Chiru, C., Popescu, OG. (2017). Automatically Determining the Popularity of a Song. In: Polkowski, L., et al. Rough Sets. IJCRS 2017. Lecture Notes in Computer Science(), vol 10313. Springer, Cham. <https://doi.org/10.1007/978-3-319-60837-2_33>

Kishore N, Santhiya P, Askar, M., & Kirubaaharan S. (2024). *A Comprehensive Approach to Song Popularity Forecasting*. 1–5. <https://doi.org/10.1109/aimla59606.2024.10531405>

Reisz, N., Vito, & Thurner, S. (2022). *To what extent homophily and influencer networks explain song popularity*. ArXiv.org. <https://arxiv.org/abs/2211.15164>

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., & Serra, X. (n.d.). *ESSENTIA: AN AUDIO ANALYSIS LIBRARY FOR MUSIC INFORMATION RETRIEVAL*. Retrieved February 26, 2025, from <https://ismir2013.ismir.net/wp-content/uploads/2013/09/177_Paper.pdf>

PINARBAŞI, F. (2019). DEMYSTIFYING MUSICAL PREFERENCES AT TURKISH MUSIC MARKET THROUGH AUDIO FEATURES OF SPOTIFY CHARTS. *TURKISH JOURNAL of MARKETING*, *4*(3), 264–279. <https://doi.org/10.30685/tujom.v4i3.62>

Yee, Y. K., & Raheem, M. (2022). Predicting Music Popularity Using Spotify and YouTube Features. *Indian Journal of Science and Technology*, *15*(36), 1786–1799. <https://doi.org/10.17485/ijst/v15i36.2332>

‌ Lee, J., & Lee, J.-S. (2018). Music Popularity: Metrics, Characteristics, and Audio-Based Prediction. *IEEE Transactions on Multimedia*, *20*(11), 3173–3182. <https://doi.org/10.1109/tmm.2018.2820903>

Su, A. T., Aung, T. T., & Hlaing. (2024). *Trends and Patterns in Spotify’s Most Played Songs: A Comprehensive Analysis*. 1–6. <https://doi.org/10.1109/icait65209.2024.10754921>

Varnum, M. E. W., Krems, J. A., Morris, C., Wormley, A., & Grossmann, I. (2021). Why are song lyrics becoming simpler? a time series analysis of lyrical complexity in six decades of American popular music. *PLOS ONE*, *16*(1), e0244576. <https://doi.org/10.1371/journal.pone.0244576>

‌ Yu, Y., Cheung, P. Y., Ahn, Y.-Y., & Dhillon, P. (2022). *Unique in what sense? Heterogeneous relationships between multiple types of uniqueness and popularity in music*. ArXiv.org. <https://arxiv.org/abs/2207.12943>

‌ R. Raj, R. Wadhvani, A. Rasool and M. Gupta. *Quantile Regression Approach in Analysis of Factors Affecting Song Popularity | IEEE Conference Publication | IEEE Xplore*. (n.d.). Ieeexplore.ieee.org. <https://ieeexplore.ieee.org/document/10176824>

Apify. (2025, March 27). Instagram Followers Count Scraper · Apify. Apify. <https://apify.com/apify/instagram-followers-count-scraper>

TheAudioDB.com | Music Database and free Music API with High Quality Artwork and Metadata. (n.d.). <https://www.theaudiodb.com/>

Pandya, Maharshi. (2023), *Spotify Tracks Dataset* |Www.kaggle.com|

[www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset?resource=download](https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset?resource=download).

Miller, Sean (2025), Hot-100 | github.com|

[random-data/Music/hot-100 at main · HipsterVizNinja/random-data · GitHub](https://github.com/HipsterVizNinja/random-data/tree/main/Music/hot-100)