Project Name: Musical Themes Across Decades

Team 026: Rhea Chitanand, Sumana Chilakamarri, Sarah Friedrichs, Mitali Bidkar, Kavya Golamaru

Introduction:

This project examines the thematic evolution of music over the decades. It is valuable to researchers, historians, sociologists, music enthusiasts, and songwriters by revealing how musical themes align with historical and cultural shifts. Researchers can explore trends in themes like love and social justice, while historians and sociologists can link these shifts to major events such as wars and social movements. Music enthusiasts gain insight into how music reflects different eras, and songwriters can leverage machine learning to analyze past trends, predict future directions, and even inspire composition based on evolving thematic patterns. Unlike existing approaches, which often focus on lyrical sentiment, genre evolution, or audio features, our approach uniquely combines thematic trend analysis, machine learning classification, and interactive visualization to better understand how music reflects cultural and societal shifts over time. We plan to gauge the success of the project by conducting user tests on the created visualizations to ensure they are insightful to relevant audiences.

Problem Definition:

This project aims to examine the thematic evolution of music over the decades by analyzing predefined theme scores associated with songs. By tracking how themes such as love, heartbreak, politics, violence, and social justice change over time, we aim to understand broader cultural and societal shifts reflected in music. We will visualize the prominence of different themes across decades to identify trends and transformations in musical content. Additionally, we will develop a classification model that predicts the decade of a song based on its thematic scores, using machine learning techniques. This approach enables us to explore the relationship between musical themes and historical periods, offering insights into how music evolves.

Literature Survey:

Previous ML models in music worked to classify music genres through various classification algorithms and evaluated their performance using F1 scores, accuracy, and RMSE (Shelke & Patil, 2024). One study provides a three-part framework involving domain definition, feature extraction, and emotion recognition for song analysis. Feature extraction emphasizes the importance of capturing low and high-level features and highlights the evolution from traditional ML models to advanced deep-learning models for emotion recognition. However, it also addresses the subjective nature of emotional perception (Han et al., 2022). Robust algorithms include SVM, which was found to outperform other machine-learning models (Kim et al., 2010) and pre-trained BERT (Revathy et al., 2023); however, they mainly rely on audio data without considering lyrics as a feature (Silla et al., 2008). Another common algorithm in music genre classification is the CNN. One study used a CNN that used a gradient boost method, which performed well; however, it was performed on data with 10-second audio clips from YouTube, which could lead to errors (Bahuleyan, 2018). This model is trained on the MEL spectrogram of the audio signal. One study used both audio and lyrical analysis of data to perform a KNN classification to classify songs into different emotional states, but a limitation to this was that many words had multiple meanings, and they didn’t take these into account (Jamdar et al., 2015). Studies incorporating K-Nearest Neighbors (kNN) have combined lyrical sentiments with audio features to classify a song based on its emotion. Some limitations, however, include the fact that lyrical analysis involves analogies, sarcasm, and other lyrics that may have multiple meanings that the ML model may not account for.

Another study used a similar approach, evaluating NN, SVM, and CNN models, but found that the dataset used may not be able to accurately classify new genres as it is limited to Latin music (Prabhakar & Lee).

Chaudhary et al. (2022) utilize large-scale machine learning for genre classification but disregard lyrics, which our project can improve by utilizing a dataset with features that measure thematic elements in the songs and time-based analysis. One study compares the performance of various machine learning and deep learning techniques and reveals that kNN provides the highest classification accuracy. A potential shortcoming to consider is that music genres evolve, making it difficult for any model to accurately classify new genres without being retrained (Ndou, Ajoodha, & Jadhav, 2021). However, another study compared MLP, CNN, KNN, and Random Forest, and showed that Random Forest outperformed the others. These findings suggest that ensemble methods are particularly effective for music genre classification tasks. However, the paper relied on a dataset with mislabeled tracks and genre ambiguity, which could make it less applicable (Mogonediwa, 2024). Another study supported the ensemble method approach by using SVR combined with k-plane piecewise regression to musical features, achieving higher accuracy compared to single-method approaches (Xia & Xu, 2022). However, this study was used for musical emotion classification, which may not align with our focus. In one study examining emotional musical trends, Kwon et al. (2021) categorized top songs from 1998-2018 into positive, negative, and neutral themes and found a rise in neutral themes over the years. Downfalls of this study, however, include the researchers' definition of each "theme", the focus on only top-charting songs, and the shorter time range of the dataset. Castillo & Flores (2021) introduced a web-based timeline visualization for real-time genre classification but relied on 10-second segments, limiting full-song context. Mauch et al. (2015) analyzed the evolution of popular music in the U.S. from 1960 to 2010, using statistical methods to track changes in harmony, timbre, and chord progressions. Their study identified three major stylistic revolutions in modern music but was limited to harmonic and timbral features, without incorporating thematic content. Our project builds upon all of these studies by incorporating both audio and thematic features from a large dataset spanning the years 1950-2019, allowing us to explore historical shifts in song themes more holistically.

Proposed Method:

Our approach would be better than the state of the art since it introduces temporal analysis to lyrical theme research studies, which is currently unexplored. We also introduce a novel decade classification concept, which uses random forest models to predict the decade of a song's release based on lyrical themes. In the past, most research has primarily focused on genre classification or sentiment analysis using audio features, but we wanted to provide data-driven insights into how musical themes reflect cultural and societal shifts of the time. Our approach is more meaningful and interpretable to those who come from both technical and non-technical backgrounds, and incorporates interactive visualizations to demonstrate the real-world usability of our work.

<u>Data Cleaning:</u>

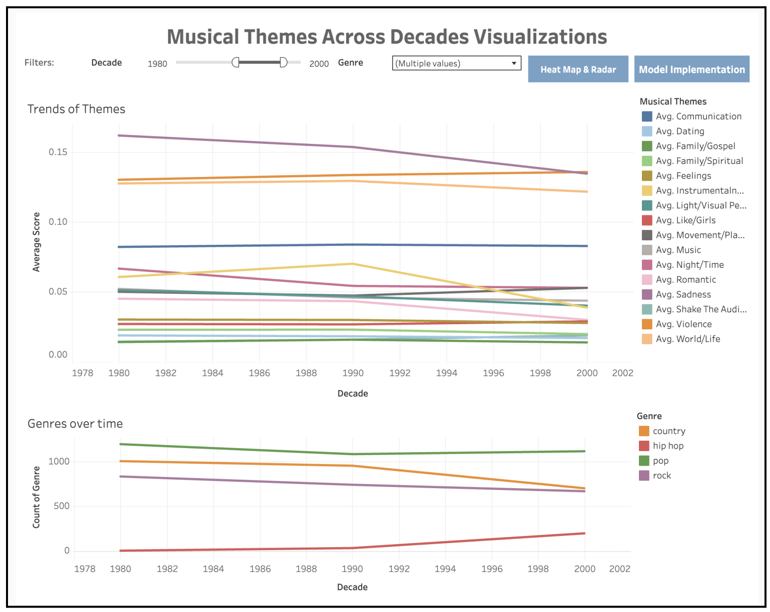
First, we cleaned the data to eliminate null values in the release data column and any of the lyrical theme score columns. We also created a new column to group the songs by release decade. To balance the dataset, we removed duplicates and downsampled the more prominent decade groups (2000s and later), and then split the data into train and test sets for modeling.

<u>Model Implementation:</u>

Random Forest Model: To analyze and predict the evolution of musical themes across decades, we used RFC since it is robust to noise and can model non-linear relationships. We used a 22-dimensional feature space with 16 lyrical theme scores and 6 musical features. Since the data was imbalanced, we downsampled recent decades to increase the fairness of the training. To improve model performance, we conducted hyperparameter tuning using GridSearchCV, testing combinations of parameters such as the number of trees, maximum depth, and minimum samples per split and leaf. The best configuration, selected based on the weighted F1-score, yielded approximately 39% accuracy—a reasonable baseline based on findings from related work.

XGBoost Model: We implemented an XGBoost classifier due to its ability to identify complex relationships within data and its robustness to overfitting. Additionally, we aimed to compare its performance against the Random Forest classifier to evaluate which approach better captures patterns within our dataset. We used the same 22-dimensional feature space, consisting of 16 lyrical theme scores and 6 musical features. To address class imbalance, we downsampled overrepresented recent decades to ensure fairer training. We then performed hyperparameter tuning using GridSearchCV, testing combinations of parameters such as learning rate, maximum tree depth, and the number of boosting rounds. The best configuration, selected based on weighted F1-score, achieved approximately 39.4% accuracy and allowed for direct comparison with the Random Forest classifier.

<u>*Model Visualizations - Dashboard in Tableau:*</u>



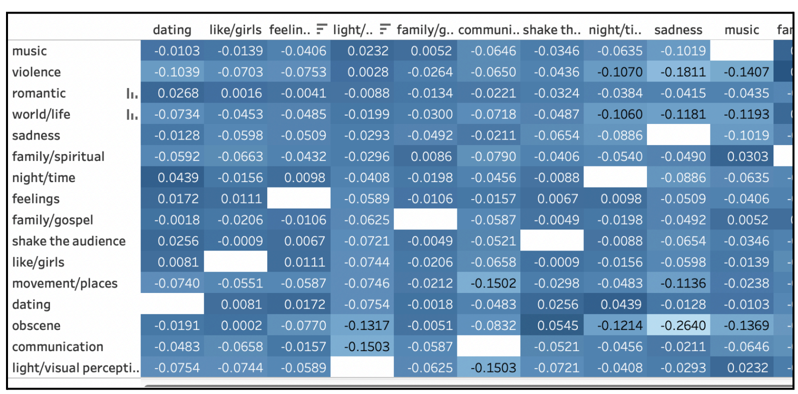
We created a dashboard to the right in Tableau to visualize trends over time through different views ([Tableau Link]). The following charts are included:

Line Visualizations:

The line chart to the right visualizes the evolution of lyrical themes over time across selected genres from 1950 to 2010. Each line represents the average prevalence of a particular theme per decade, with its associated average score being plotted. Users can filter by individual themes and genres, as well as adjust the time range using the decade slider, enabling dynamic exploration of specific trends. This analysis helps reveal long-term thematic trends, such as the steady rise of Violence and Obscene themes and the decline of World/Life and Family/Gospel themes. This visualization exemplifies how the patterns in thematic scores help in distinguishing songs across decades. We also created an additional line chart to show genres over time. From this chart, we can see that Hip hop became more prevalent after the 2000s. (See Appendix A for detailed view)

Heat Map:

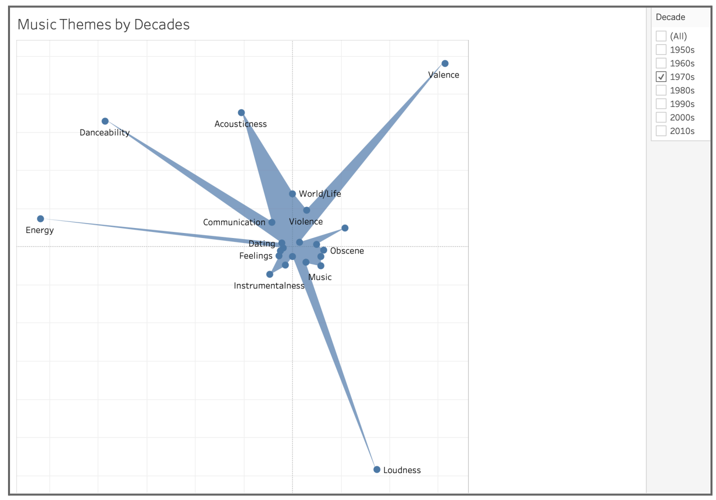
This component of the project is focused on visualizing the relationships between different themes across decades through correlation heat maps. We computed the correlation values between each pair of themes to understand how often themes appear together in songs through the use of the normalized thematic scores in our dataset.



	dating	like/girls	feelin..	light/..	family/g..	communi..	shake th..	night/ti..	sadness	music
music	-0.0103	-0.0139	-0.0406	0.0232	0.0052	-0.0646	-0.0346	-0.0635	-0.1019	
violence	-0.1039	-0.0703	-0.0753	0.0028	-0.0264	-0.0650	-0.0436	-0.1070	-0.1811	-0.1407
romantic	0.0268	0.0016	-0.0041	-0.0088	-0.0134	-0.0221	-0.0324	-0.0384	-0.0415	-0.0435
world/life	-0.0734	-0.0453	-0.0485	-0.0199	-0.0300	-0.0718	-0.0487	-0.1060	-0.1181	-0.1193
sadness	-0.0128	-0.0598	-0.0509	-0.0293	-0.0492	-0.0211	-0.0654	-0.0886		-0.1019
family/spiritual	-0.0592	-0.0663	-0.0432	-0.0296	0.0086	-0.0790	-0.0406	-0.0540	-0.0490	0.0303
night/time	0.0439	-0.0156	0.0098	-0.0408	-0.0198	-0.0456	-0.0088		-0.0886	-0.0635
feelings	0.0172	0.0111		-0.0589	-0.0106	-0.0157	0.0067	0.0098	-0.0509	-0.0406
family/gospel	-0.0018	-0.0206	-0.0106	-0.0625		-0.0587	-0.0049	-0.0198	-0.0492	0.0052
shake the audience	0.0256	-0.0009	0.0067	-0.0721	-0.0049	-0.0521		-0.0088	-0.0654	-0.0346
like/girls	0.0081		0.0111	-0.0744	-0.0206	-0.0658	-0.0009	-0.0156	-0.0598	-0.0139
movement/places	-0.0740	-0.0551	-0.0587	-0.0746	-0.0212	-0.1502	-0.0298	-0.0483	-0.1136	-0.0238
dating		0.0081	0.0172	-0.0754	-0.0018	-0.0483	0.0256	0.0439	-0.0128	-0.0103
obscene	-0.0191	0.0002	-0.0770	-0.1317	-0.0051	-0.0832	0.0545	-0.1214	-0.2640	-0.1369
communication	-0.0483	-0.0658	-0.0157	-0.1503	-0.0587		-0.0521	-0.0456	-0.0211	-0.0646
light/visual percepti..	-0.0754	-0.0744	-0.0589		-0.0625	-0.1503	-0.0721	-0.0408	-0.0293	0.0232

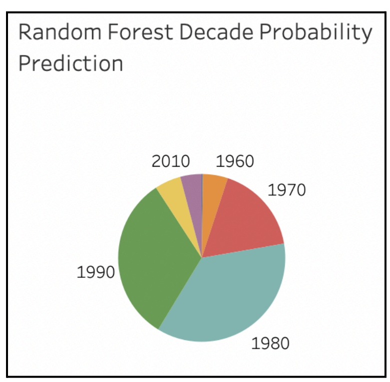
These correlation coefficients are valuable as they display underlying patterns in the thematic content of songs and how these patterns evolve over time as music changes. To calculate the correlation coefficients, we utilized both pandas and NumPy libraries in Python and exported these results into Tableau for visualization purposes. In Tableau, we designed an interactive correlation heat map where users can filter by genre and select a range of years, and will instantly see the heat map to explore how lyrical relationships change and co-exist. Filtering by genre will allow users to see how impactful certain themes are on genres, and changing the range of years shown will allow users to more easily analyze thematic evolution. The color scheme of the heatmap ranges from light to dark blue. Darker shades indicate stronger positive correlations, and lighter cells indicate weaker negative correlations.

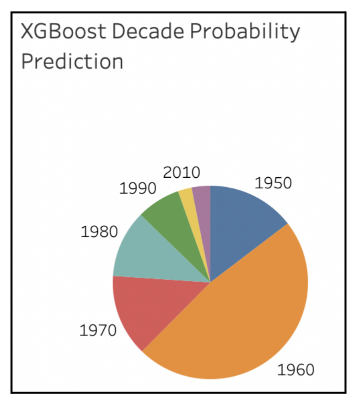
Radar Chart Visualization:



The radar chart in this project visually represents the distribution of thematic elements (like romantic, violence, sadness, etc.) within individual songs, allowing users to quickly grasp the emotional and narrative profile of a track. By comparing the intensity of each musical theme as spokes on the chart, it reveals how songs differ in their lyrical focus. Longer and wider spokes signify a larger presence and vice versa. We first pivoted the dataset so that the musicality features became rows, resulting in musical features and corresponding score columns. Several calculated fields were created in Tableau to measure the angle of each feature in radians and its distance from the center in terms of the average score. After the main polygon was created, a dual axis was used to show the exact locations of the themes for readability. When filtered by release date, the radar chart also helps highlight how musical themes have evolved over time, making it a powerful tool for both individual analysis and temporal comparison.

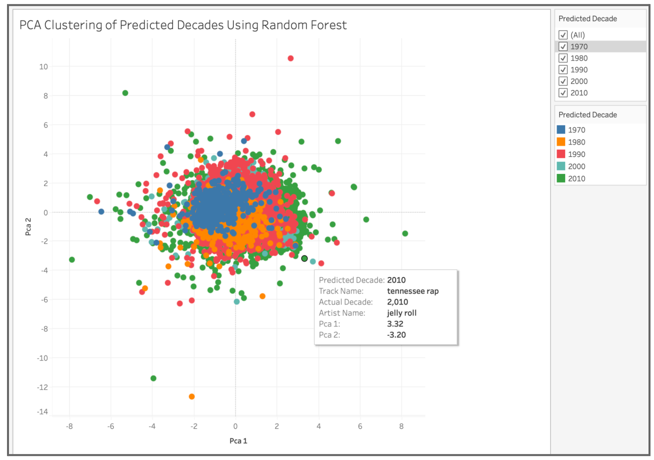
Pie chart:



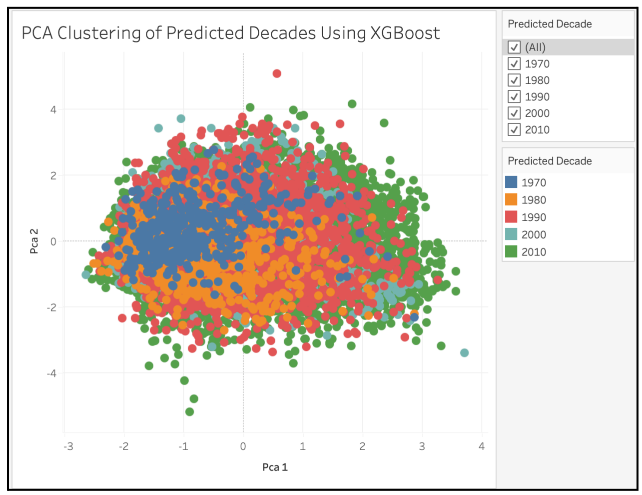


The pie charts are a visualization of our Random Forest and XGBoost models' predicted probabilities for the decade that a song belongs to. By breaking down the models' outputs into percentage shares across decades, users can see how confident the models are in their predictions for any given track. We first extracted the classification probabilities from both of our models in Python as the input datasets. The interactive filter allows users to input a song title and view its decade prediction distribution. As the models have been further fine-tuned, these predictions are reflective of musical patterns over time. Additionally, the charts serve as a tool for model validation: by comparing the predicted probabilities with the song's actual release decade, we can quickly assess the models' accuracy and identify areas for improvement.

Random Forest:
We applied PCA to our random forest model to reduce the feature space and visualize the two components in Tableau. Each scatterplot point is plotted by its PCA values and represents a song, with colors representing the predicted decade by RF. Hovering over each point reveals the track name, artist, predicted decade, actual decade, and PCA coordinates. There is an overlap in the clusters as expected due to the subjectivity of musical themes, but the more distant decades do show some separation. This visualization helps the user see the model's performance in terms of correct predictions and misclassifications, and the clustering demonstrates thematic similarities between decades.



XGBoost Classifier:
This PCA clustering visualization shows the distribution of songs based on their predicted decades using the XGBoost model. Each point represents a song projected into two principal components from the 22-dimensional feature space. The colors correspond to the model's predicted decades, revealing areas of overlap, especially between adjacent decades like the 1980s and 1990s, indicating shared thematic and musical characteristics.



Evaluation:
The dashboard was deployed to a set of users familiar with music and provided with a form to gather qualitative and quantitative feedback. The core questions our experiments answered included:

- How often do users interact with music?
- Are the visualizations intuitive and easy to interpret?
- Do users gain meaningful insights about musical trends and themes from the dashboard?
- Does the design help users understand the outputs of the ML model used for classification?
- Is the dashboard engaging enough for independent exploration?
- What improvements or pain points do users encounter while interacting with the visualizations?

We surveyed twelve users with varying levels of familiarity with music and machine learning, all of whom reported interacting with music daily. Most participants listened to music between 3–6 hours per day, indicating a strong interest. The ML familiarity scores ranged from 1 to 5 out of 5, which helped us assess whether the dashboard was understandable even to non-technical users. Most users rated the visualizations highly on clarity, with all giving scores of 4 or 5. This suggests that the thematic visuals were effective in communicating conceptually rich data. All respondents rated their ability to find insights about musical themes over time at 4 or 5, confirming the dashboard's analytical usefulness. Ratings on understanding the model's predictions were more varied, with scores between 3 and 5. This indicates room for improvement in explaining ML mechanics. Scores for visual engagement were consistently high (all 4s and 5s), and ten users said they would use the dashboard independently, with scores of 4 or 5; two users gave it a 3, pointing to hesitation possibly due to usability or technical understanding. There were also suggestions for clearer snapshot visualizations, additional clarity on average scores, more intuitive song selection, and a general interest in learning more about the classification algorithm. Overall, the user testing validated that the dashboard is effective for exploratory analysis but would benefit from minor design enhancements and additional transparency around model outputs.

To validate our visualizations' results against known musical trends, we cross-referenced the thematic patterns observed in our dashboard with established findings in history. For instance, the observed rise in Violence and Obscene themes aligns with the increased popularity of explicit lyrical content in hip hop and pop genres post-1990s. Similarly, the decline in World/Life and Family/Gospel themes reflects a cultural shift away from spiritual narratives toward more individualistic and provocative content. These historical parallels support the reliability of our thematic scoring and help confirm that our visualizations are capturing real-world lyrical evolutions.

Conclusions and Discussion:

Our project utilized machine learning classification models and interactive data visualizations to explore the thematic evolution of music over the decades. Throughout the course of the project, our team was successfully able to identify long-term shifts in musical themes, the introduction of specific genres, etc. Our Random Forest and XGBoost models were able to achieve strong predictive performance in classifying the decade of a song based on the thematic scores attached to each song. Ultimately, the XGBoost model performed slightly better, with a selection based on weighted F1-score. It achieved 39.4% accuracy and a weighted F1-score of 38%, compared to the Random Forest classifier, which achieved 39.0% accuracy and similar class-level performance. XGBoost demonstrated stronger recall for certain decades, particularly the 1960s and 1970s, making it a strong comparative model for analyzing thematic trends across time. In addition, based on our user testing results, our Tableau dashboard enabled users to interact with and understand the visualizations successfully.

Some limitations of this project include that our analysis was focused solely on lyrical themes and excluded information on tempo, valence, and other features related to the audio of each song. In addition, the thematic scores and the names of each theme are subjective and can be interpreted differently by users. For example, the "romantic" theme may be interpreted differently in different decades; this introduces bias in both the correlation analysis and classification. Future extensions include user testing to include professionals in the music industry or historians; future expansions for the project may involve utilizing more audio-based features, such as to further analyze thematic elements of songs over time, and enhancing the visualizations using the user study results.

All team members have contributed a similar amount of effort to the project.

References

Bahuleyan, H. (2018). Music genre classification using machine learning techniques. *arXiv preprint arXiv:1804.01149*.https://arxiv.org/abs/1804.01149

Castillo, J. R., & Flores, M. J. (2021). Web-Based Music Genre Classification for Timeline Song Visualization and Analysis. *IEEE Access*, *9*, 18801–18816. doi:10.1109/ACCESS.2021.3053864

Chaudhury, M., Karami, A., & Ghazanfar, M. A. (2022). Large-Scale Music Genre Analysis and Classification Using Machine Learning with Apache Spark. *Electronics*, *11*(16), 2567. https://doi.org/10.3390/electronics11162567

Costa, Y., Craig, A. P., Eerola, T., Goulart, A., Kumaraswamy, B., Nanni, L., Barachant, A., Breuer, V., Chen, J., Cheng, C., Cheng, Y. H., Hu, D. W., Duan, L., Feng, T., Girsang, A. S., Gwardys, G., Hamel, P., Holzapfel, A., Karunakaran, N., … Li, T. (2022, August 22). *Holistic approaches to music genre classification using efficient transfer and deep learning techniques*. Expert Systems with Applications. https://www.sciencedirect.com/science/article/pii/S0957417422016815

Han, D., Kong, Y., Han, J., & Wang, G. (2022). A Survey of Music Emotion Recognition. https://link.springer.com/article/10.1007/s11704-021-0569-4

Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *arXiv preprint arXiv:1506.05012*.https://arxiv.org/abs/1506.05012

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., & Turnbull, D. (2010). Music emotion recognition: A state of the art review. *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010).* https://www.researchgate.net/publication/228695729_Music_emotion_recognition_A_state_of_the_art_review

Kwon, L., Medina, D., Ghattas, F., & Reyes, L. (2021). Trends in Positive, Negative, and Neutral Themes of Popular Music From 1998 to 2018: Observational Study. *JMIR pediatrics and parenting*, *4*(2), e26475. https://doi.org/10.2196/26475

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of Popular Music: USA 1960–2010. *Royal Society Open Science*, *2*(5), 150081. https://doi.org/10.1098/rsos.150081

Mogonediwa, K. (2024). Music Genre Classification: Training an AI Model. https://arxiv.org/abs/2405.15096

Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music genre classification: A review of deep-learning and traditional machine-learning approaches. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 1–6. https://doi.org/10.1109/iemtronics52119.2021.9422487

Revathy, V. R., Pillai, A. S., & Daneshfar, F. (2023). Lyemobert: Classification of lyrics' emotion and recommendation using a pre-trained model. *Procedia Computer Science*, *218*, 1196–1208. https://doi.org/10.1016/j.procs.2023.01.098

Shelke, S., & Patil, M. (2024). Exploring machine learning techniques for music emotion classification: A comprehensive review. *IEEE Conference Publication.* IEEE Xplore. https://ieeexplore.ieee.org/document/10498480/

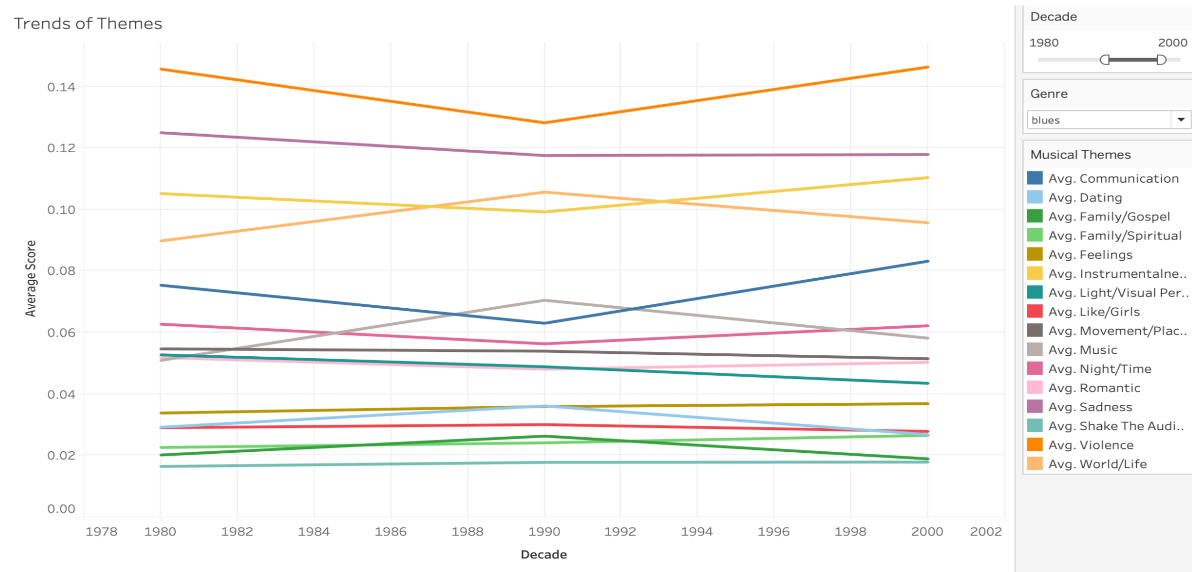
Silla, C. N., Koerich, A. L., & Kaestner, C. A. (2008). A machine learning approach to automatic music genre classification. *Journal of the Brazilian Computer Society*, *14*, 7-18.https://link.springer.com/article/10.1007/BF03192561#preview

Xia, Y., & Xu, F. (2022). Study on Music Emotion Recognition Based on the Machine Learning Model Clustering Algorithm. https://doi.org/10.1155/2022/9256586

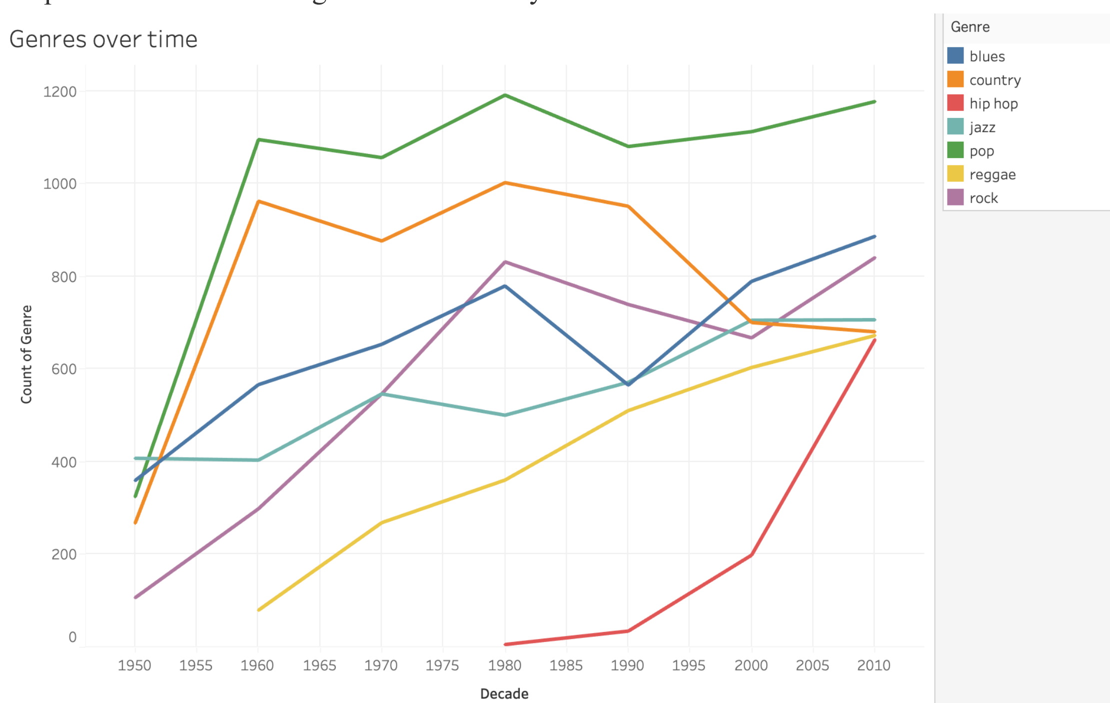
Appendix

A. Line Charts and Descriptions

a. This graph shows the trend of themes over time filtered by decade and genre.



b. Graph that shows trend of genres over time by count in our dataset.



B. Tableau Visualizations Extra Information

Input the Data Source Into Tableau: Take the different CSV inputs from the steps above and insert them to the data source tab on Tableau. Make any adjustments to variables such as converting release date to a date variable instead of a number. There should be 9 total datasets imported into Tableau.
Creating Each visual in Tableau: Below are descriptions of each sheet/chart we used to create our dashboard in Tableau. You can see the specific columns, rows, and filters we used to manipulate the output. Please see each description for recreation purposes
Theme Trends over time Line chart: The columns for this chart would be the Decade while the rows are the measure values ie the different theme scores averaged. There is a filter by genre and a slider for decades. The different colored lines each represent a different musical theme and you can show the graph key to the side of the chart.
Genres over time Line Chart: This graph represents the count of genres over time. The rows are the genre count and the columns are the decades. There is a filter for genre and also a slider to select which decades you want to view.

Heat Map:

Input both theme_correlations.csv and theme_correlations_by_genre.csv Data Sources into Tableau. These datasets represent correlations between themes across genres and are reshaped into a long format for visualization purposes.

Utilizing the theme_correlations.csv data source, drag 'Theme A' into Rows and 'Theme B' into columns. Drag 'Correlation' into both Color and Label. Set the Marks type to 'Square' using the dropdown menu. Our team chose a Blue-white color palette to easily identify the correlations between different themes, and set the center value to 0 so that more neutral correlations are displayed as white or with lighter shades while more intense correlations are displayed with darker shades of blue. There is also a filter for genre and a slider to select decades.

Radar Chart:

The original tcc_ceds_music dataset was duplicated to create the radar chart, since the data had to be pivoted so that the musicality features and their corresponding scores became rows instead of columns. The resulting dataset has Feature and Score columns. A dual axis was created to plot points of the exact locations of every musical feature used to form the spokes. There is a filter for the release decade. The data source is tcc_ceds_music 2.

Angle calculated field: This was created to ensure that the spoke for every musical feature appeared at the correct angle (in radians) across every decade.

- Calculation: RUNNING_SUM(2 * PI()/MIN({COUNTD([Feature])}))+ PI() / 2

Distance from the center calculated field: This was created to ensure that each musical feature's spoke was at the correct distance from the center based on the average score.

- Calculation: AVG([Score])

X calculated field: This was created so that every musical feature would be plotted at the correct X-coordinate based on the standard equation of a circle.

- Calculation: [Distance from the center] * COS([Angle])

Y calculated field: This was created so that every musical feature would be plotted at the correct Y-coordinate based on the standard equation of a circle.

- Calculation: [Distance from the center]* SIN([Angle])

Random Forest Classifier/XGBoost Decade Pie Charts:

The probabilities into which both models classified each song into decades were first extracted into separate CSV files by using the predict_proba function in Python on the best performing model (fine-tuned with the best hyperparameters). The dataset was then pivoted in Tableau to have a Decade column with a corresponding Probability column. There is a filter for the track name. The data source for the XGBoost pie chart is xgboost_decade_probabilities_scaled and the data source for the RFC pie chart is decade_probabilities

Dashboard Creation: Feel free to arrange the following sheets in any method you prefer on your dashboard. We choose to split up the visuals into three different dashboards and include a button to go between the three. The first dashboard shows musical trends over time by theme and genre. The second shows the correlation between themes and a radar chart by theme prevalence in each genre. The final dashboard shows the model visualizations and results for our Random forest and XGboost models.

Exporting Dashboard to Online Server: You can create a public version of the dashboard by using Tableau Public. You would click the server icon at the top, Then hover over tableau public, and then click save to tableau public.