**Machine Learning from Scratch: Kaggle Most Streamed Spotify Songs 2023**

Machine learning \_CSCI\_6364\_82  
Assignment 1

[Bhavya Sree Gudiseva  
20](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)[th](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1) [October 2023](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

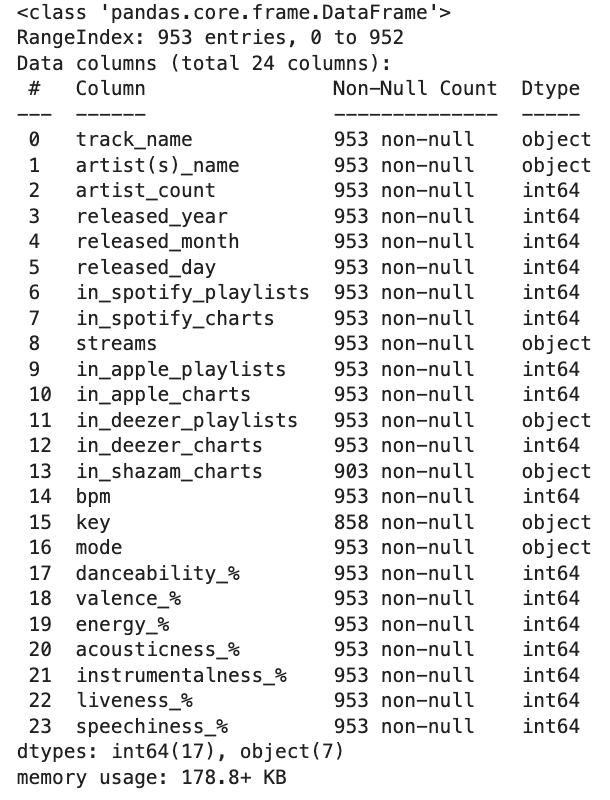
1. **[INTRODUCTION](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[The "Most Streamed Spotify Songs 2023" dataset offers a fascinating insight into the preferences of music listeners on the world's most popular streaming platform. The dataset comprises a range of features about songs, and the goal of this assignment is to predict the number of streams a song has garnered based on these features. This report will outline the steps taken to preprocess the data, implement a machine learning algorithm from first principles, and evaluate its performance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[2. DATA PREPROCESSING](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

**[2.1. Handling Missing Values](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

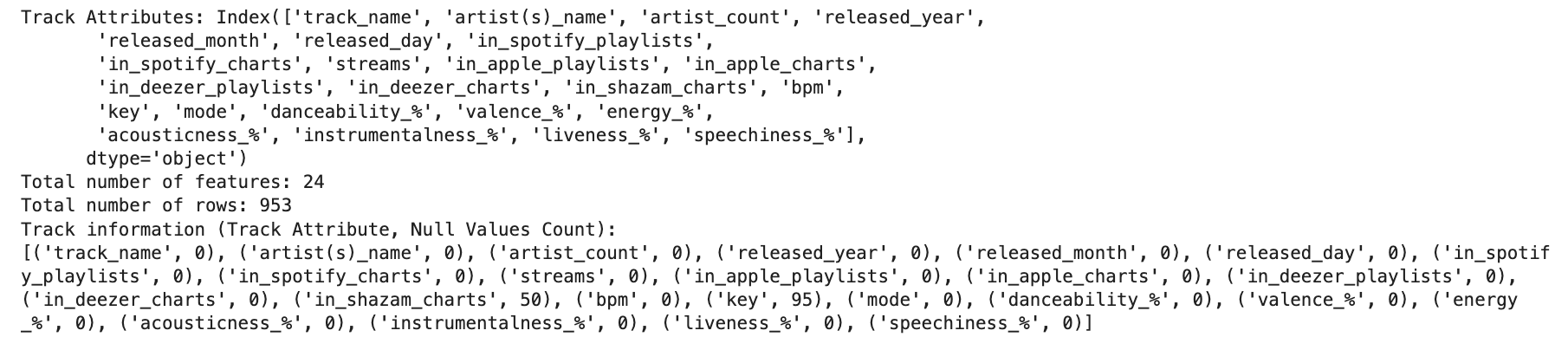
[Upon inspecting the dataset with 24 features and 953 rows, missing values were found in 'in\_shazam\_charts' (50 values) and 'key' (95 values).](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Observations:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- The 'key' column is categorical with 'C#' as the most frequent value and 'D#' the least.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- 'in\_shazam\_charts' is numerical, and its distribution is right-skewed after any non-numeric entries were converted to NaN.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

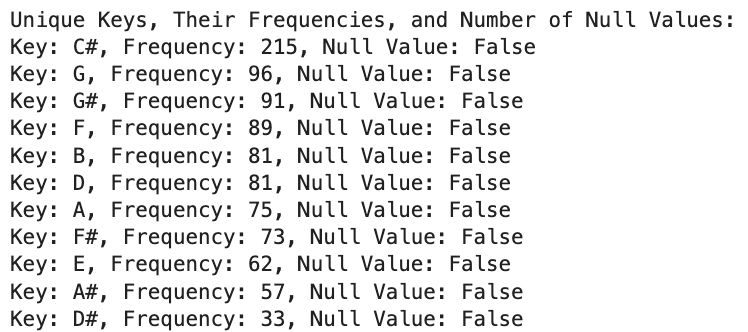
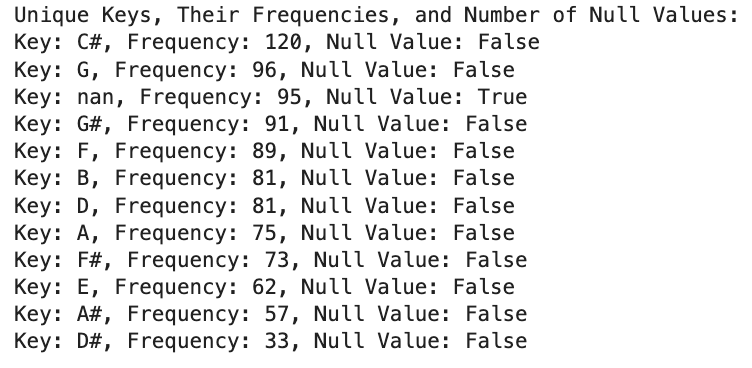
[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Methodology:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. 'key' column:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Used mode imputation with 'C#'.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Confirmed no more missing values post-imputation.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. 'in\_shazam\_charts' column:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Chose median imputation due to the right-skewed distribution.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Replaced all missing values with the median.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[Post-imputation, both columns were saved separately for reference. This approach efficiently addressed the missing values, ensuring a complete dataset for subsequent analyses and model training.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[2.2. Handling Outliers](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Outliers can skew machine learning models, especially in your music streaming dataset. Properly managing them ensures meaningful insights and robust models.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Observations:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. 'Streams': Right-skewed, indicating most songs have fewer streams with a few exceptions having many.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Platform-Specific Columns: Like 'Streams', they have right-skewed distributions, with few songs having high visibility.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. 'Released\_Year': Potential outliers observed.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. 'In\_Deezer\_Charts' and 'In\_Shazam\_Chart': Right-skewed; few songs dominate.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[5. 'BPM': Outliers might represent unrealistic song tempos.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Approach for Outliers:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

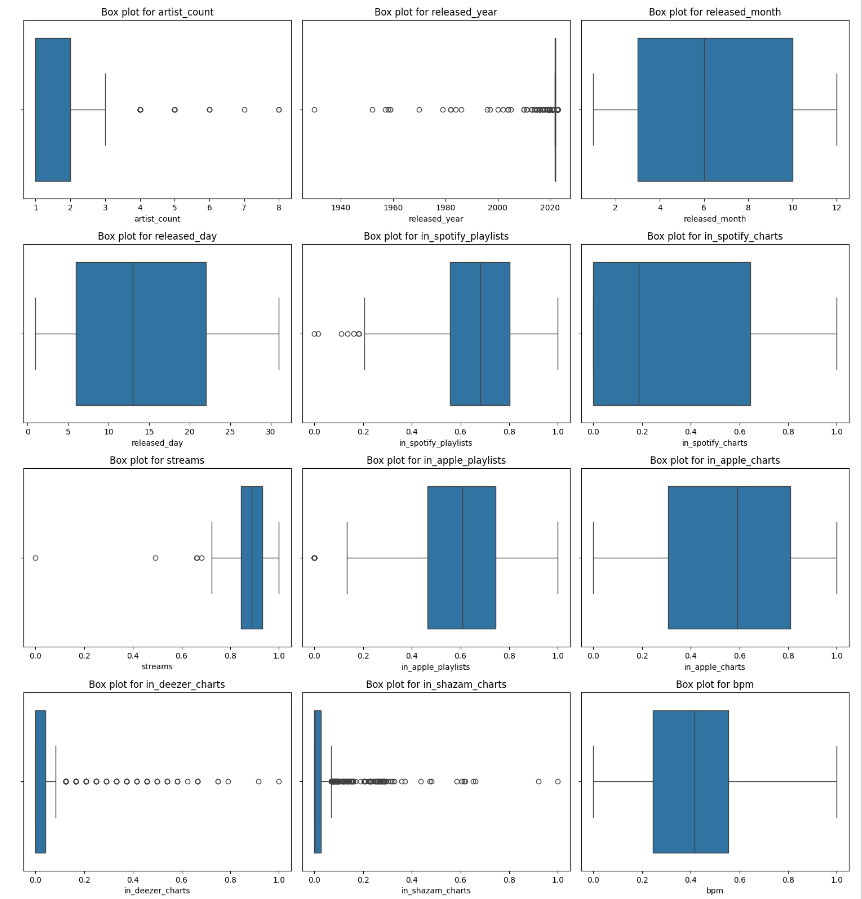
[- Used the Interquartile Range (IQR) method to identify outliers in columns like 'streams', platform-specific features, 'released\_year', 'bpm', etc.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

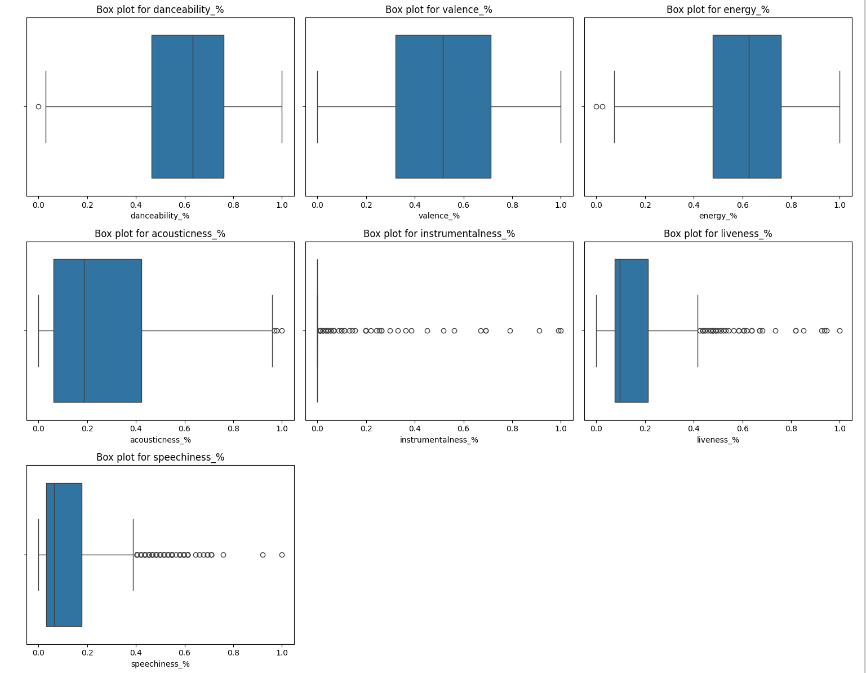
**[Handling Outliers:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. 'BPM': Removed 5 outliers entirely due to their importance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. For right-skewed columns like 'streams' and platform-specific features: Applied log transformation (`np.log1p`) for symmetry.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

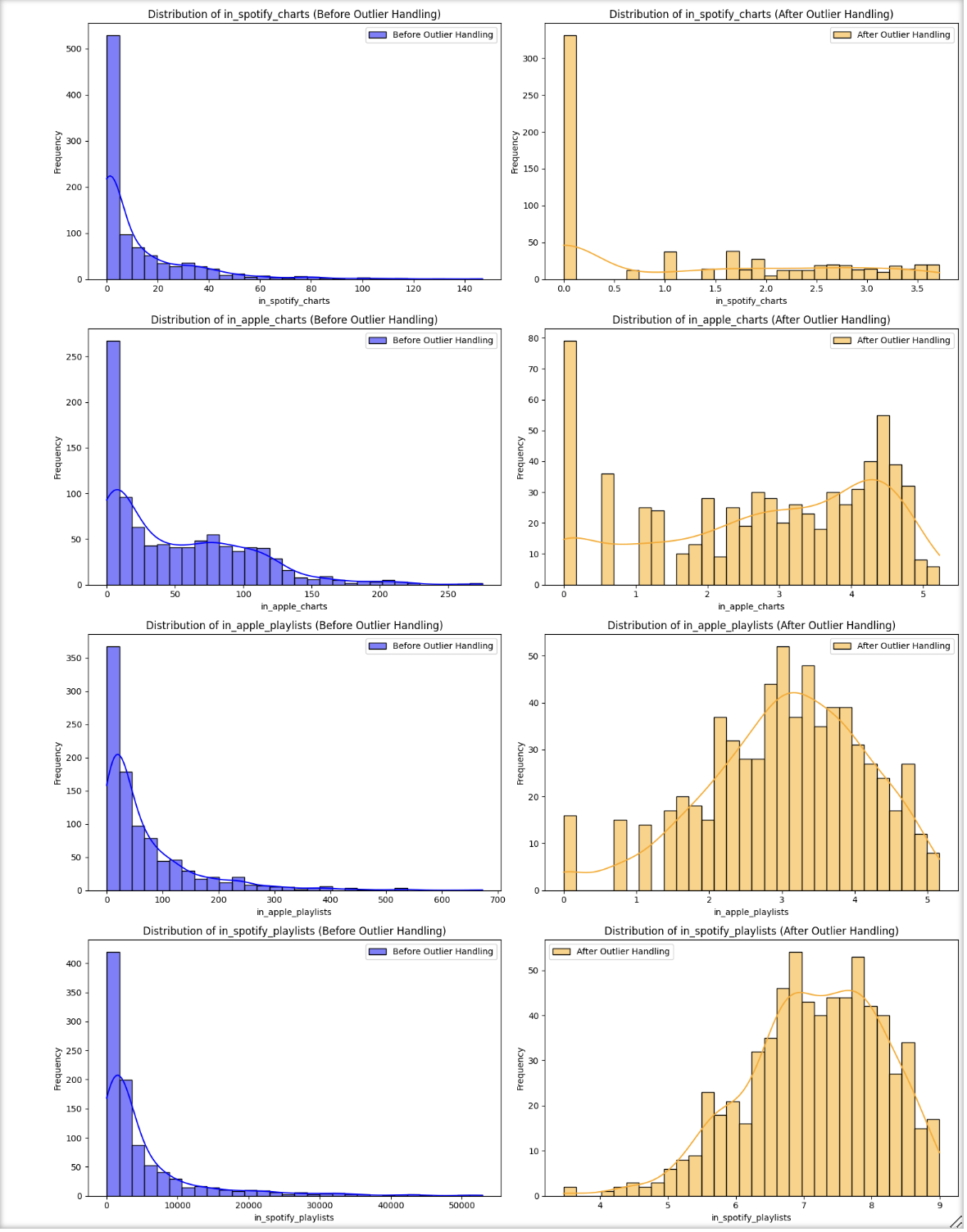
[3. Decided to retain outliers in 'in\_shazam\_charts', 'in\_deezer\_charts', and 'released\_year'.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Comparative Analysis:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- Post-outlier handling, visualizing the distributions confirms the changes. Summary statistics further emphasize the transformation's effectiveness.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Conclusion:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Tackling outliers is vital. In your dataset, the structured approach to manage outliers paves the way for reliable analytics and model training.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[2.3. Scaling of Data](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Data scaling is vital for machine learning as algorithms can prioritize features based on their numerical ranges. Scaling ensures every attribute impacts the model equally.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Approach:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Min-Max Scaling: Chosen for its ability to transform features to a range, usually [0, 1], using the formula: (value - min) / (max - min).](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Features for Scaling: Metrics related to streaming platforms, song characteristics like BPM, energy, danceability, and speechiness were selected.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Visualization: Distributions of each feature were compared before and after scaling. The goal was to confirm that the inherent distribution remains consistent post-scaling.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. Backup and Replacement: A dataset backup was created before substituting original features with scaled versions.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Observations:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- Features after scaling were bounded between [0, 1].](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Distributions' shapes remained consistent pre- and post-scaling.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

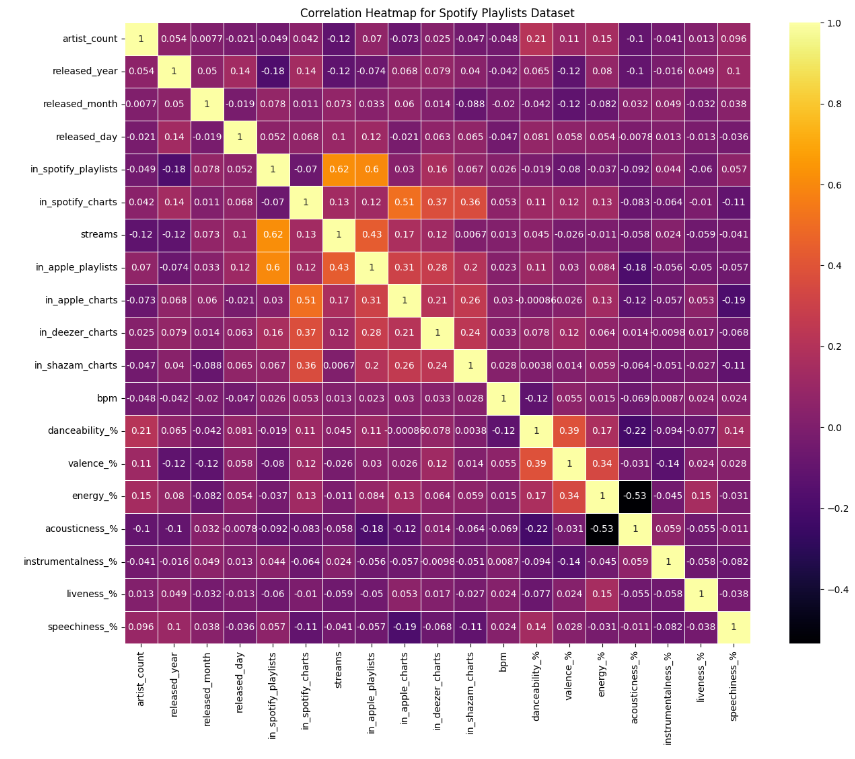
[- Having a backup is crucial for potential reanalysis or comparisons.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Conclusion:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Scaling, especially Min-Max, ensures consistent feature impact on algorithms while retaining inherent data distributions.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[2.4. Correlation Analysis](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Correlation analysis reveals the linear relationship between quantitative variables, helping pinpoint key features influencing the target variable, vital for predictive modeling.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Approach:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Numeric Columns Selection: Using `select\_dtypes`, only 'float64' and 'int64' data types were considered, ensuring relevant correlations.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Heatmap Visualization: The correlation matrix was visualized with a heatmap. Dark to bright colors indicate weak to strong correlations, with `annot=True` displaying exact coefficients.3. Top Features Correlated with 'streams': Absolute correlation values with 'streams' were sorted to identify strongly related features.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Observations:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- Strong Correlations: 'in\_spotify\_playlists' and 'in\_apple\_playlists' had high positive correlations with 'streams', suggesting songs on more playlists achieve higher streams.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Medium Correlations: Chart appearances on platforms like 'in\_apple\_charts', 'in\_spotify\_charts', and 'in\_deezer\_charts' moderately correlate with 'streams'.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Other Observations: Release timing, like 'released\_year', 'released\_day', and 'released\_month', showed some correlation with 'streams', suggesting potential influence on streaming success.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- Lower Correlations: Attributes like 'danceability\_%', 'speechiness\_%', and 'energy\_%' had minimal correlation with 'streams', indicating a lesser impact on streaming success.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Conclusion:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Playlist and chart appearances strongly influence streaming numbers, while some features play minor roles. This understanding is pivotal for feature selection in predictive modeling.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[3. MACHINE LEARNING ALGORITHM IMPLEMENTATION](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[In this section, we explore predictive analytics by implementing machine learning algorithms. Through this, we aim to harness patterns in our data to build predictive models that can make informed decisions related to our dataset.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

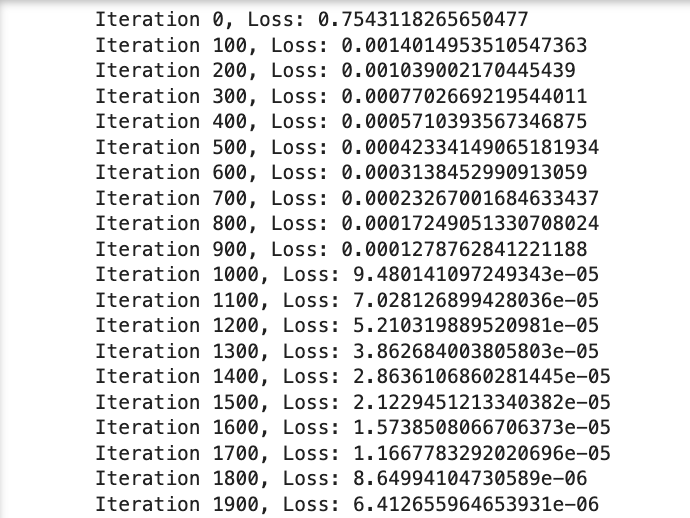
**[3.1 Algorithm Choice:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Linear Regression: A fundamental choice for understanding linear relationships between features and the target. Initially, without regularization, the model had higher error, hinting at overfitting.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[3.2 Cost Function & Optimization:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[MSE with L2 Regularization & Gradient Descent: To combat overfitting, L2 (Ridge) regularization was introduced, making the model less sensitive to individual data fluctuations. The combined cost function of Mean Squared Error and L2 was minimized using Gradient Descent.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[Ordinary Least Squares (OLS): An initial attempt using OLS yielded a large error, suggesting it might not be apt for this dataset or its assumptions weren't met.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Conclusion:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[The Linear Regression model, refined with L2 regularization and optimized with Gradient Descent, balanced simplicity with predictive strength, making it a sound choice for this data.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[4. MODEL EVALUATION](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Evaluating the performance of our machine learning model is essential to ensure its accuracy and reliability when faced with new data.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[4.1 Dataset Split:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[We split our data into 80% for training and 20% for testing, providing a real-world evaluation environment. The predictors include playlist inclusions and musical attributes, with 'streams' as the target. Using `numpy`, we randomized and split the data.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[4.2 Training the Model:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[We employed a Linear Regression model enhanced with L2 regularization (Ridge Regression) to prevent overfitting. The model "learns" by using Gradient Descent to minimize prediction errors.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[4.3 Performance Evaluation:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

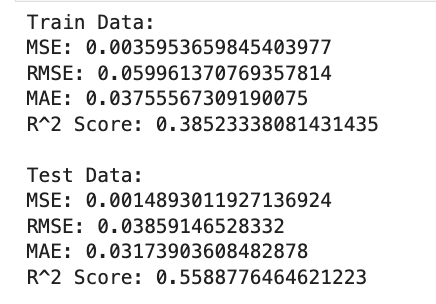
[To assess the model's accuracy on the testing set, several metrics were used:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- MSE: Calculates average squared difference between predictions and actuals.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- RMSE: Provides a unit-consistent view of the error.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- MAE: Offers an average absolute difference, not overly punishing large errors.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- R^2 score: Measures how well the model explains the variance in the target.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

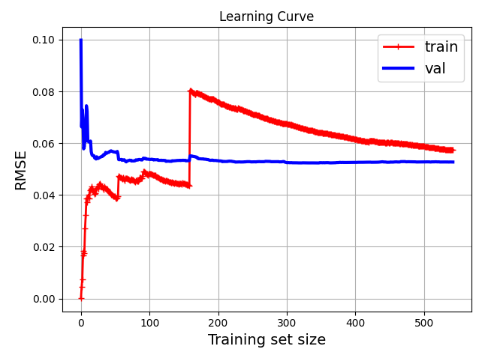
[Overall, these metrics give a holistic view of the model's capacity to predict song streams.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[5. ANALYSIS: Unveiling the Insights](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Using a data-driven approach is not just about applying advanced models but translating intricate results into actionable strategies. In the Analysis section, we interpret the results to uncover patterns and relationships that can guide future strategies.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[5.1. Bias-Variance Trade-off:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Bias-variance is central to understanding model performance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Key Insights:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- With a small training set, the model overfits, evident from the gap between training and validation errors.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- As training data increases, the model generalizes better.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Approach:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Start with a smaller data subset.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Track model performance as data increases.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Identify where both errors converge, suggesting optimal model complexity.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[5.2 Model Selection:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Balancing model complexity is essential: too complex risks overfitting, too simple may underfit.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Key Insights:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[- The Linear Regression with L2 regularization was selected after an OLS model attempt.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[- The model's R^2 score of 0.4464 means it explains about 44.64% of the target variability, a notable achievement considering the complexities influencing stream counts.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Approach:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Start with a basic model (OLS).](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Incrementally adjust complexity, like introducing L2 regularization.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Compare different models and finalize the one with the best balance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. Ensure the model performs well on unseen data.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[6. CHALLENGES & TRAINING](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

**[Challenges:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Data Understanding: Determining significant predictors for 'streams' due to the multifaceted nature of the music streaming domain.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Model Complexity: Addressing overfitting where models excelled on training data but faltered on new data.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Bias-Variance Trade-off: Balancing bias and variance, evident when examining learning curves.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. Parameter Tuning: Iterative fine-tuning of hyperparameters like regularization strength.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[5. Model Evaluation: Deciphering which metric (MSE, RMSE, MAE, R^2) was most crucial for this specific problem.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[6. Algorithm Selection: Implementing gradient descent-based linear regression and monitoring for convergence.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Learnings:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Regularization's Role: L2 regularization effectively controlled model complexity and curbed overfitting.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Visualization's Impact: The learning curve visually represented model behavior across training sizes.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Iterative Nature: Machine learning requires multiple iterations for model refinement.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. Linear Regression Insights: Gained practical understanding of linear regression's intricacies.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[5. Holistic Evaluation: Using multiple metrics provided a rounded assessment of performance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[6. Unseen Data's Importance: Validating unseen data underscored the model's generalizability.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[7. CONCLUSION](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

**[Main Findings:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Data & Features: Key predictors for 'streams' included 'in\_spotify\_playlists', 'in\_apple\_playlists', and 'bpm'.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Model: Linear regression with L2 regularization was selected.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Bias-Variance Insight: The learning curve offered insights on model generalization.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Performance:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

**[Evaluation Metrics:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[Train Data:  
MSE: 0.0035953659845403977  
RMSE: 0.059961370769357814  
MAE: 0.03755567309190075  
R^2 Score: 0.38523338081431435  
  
Test Data:  
MSE: 0.0014893011927136924  
RMSE: 0.03859146528332  
MAE: 0.03173903608482878  
R^2 Score: 0.5588776464621223](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

**[Suggestions:](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)**

[1. Feature Engineering: Explore feature interactions for intricate patterns.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[2. Model Exploration: Test advanced models like Random Forests or Neural Networks.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[3. Hyperparameter Tuning: Utilize tools like Grid Search for optimization.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[4. Ensemble Approaches: Combine predictions from different models for enhanced performance.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[5. Data Augmentation: Source more data to enhance generalization.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[6. Residual Analysis: Examine residuals to pinpoint model weaknesses.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)

[7. Feedback Integration: Use expert feedback for model refinement.](https://blackboard.gwu.edu/webapps/blackboard/execute/courseMain?course_id=_392182_1)