# Final Report

## Preprocessing Techniques:

I implemented the following preprocessing techniques on the dataset:

1. Removal of duplicates: I removed duplicate entries from the dataset using the `drop_duplicates()` function.
2. Handling missing values: I dropped rows with missing values using the `dropna()` function.
3. Conversion of categorical columns to numerical: I converted categorical columns ('Artist Names', 'Artist(s) Genres', 'Song Image') using `astype('category').cat.codes`.

## Data Analysis:

I performed correlation analysis and mutual information analysis to study the relationships between features. The correlation matrix and top features selected using mutual information were visualized to understand feature importance and relationships.

## Regression Techniques:

I used two regression techniques: Random Forest Regressor and Gradient Boosting Regressor to predict the popularity of songs.

## Model Comparison:

- Random Forest Regressor: Achieved Mean Squared Error, Mean Absolute Error, and R-squared Score of [insert results].
- Gradient Boosting Regressor: Achieved Mean Squared Error, Mean Absolute Error, and R-squared Score of [insert results].

## Features Used:

I selected the top 10 features based on mutual information for creating regression models.

## Training and Testing Sets:

The dataset was split into training and testing sets using a standard ratio of 80:20 or as required.

## Further Techniques for Improvement:

I used feature selection techniques like mutual information to improve model performance.

## Resultant Regression Line Plots:

Include screenshots of the regression line plots generated by the models if applicable.

## Conclusion:

In conclusion, the project aimed to predict song popularity using regression techniques. Through thorough preprocessing, feature selection, and model evaluation, it was observed that [provide insights from the analyses]. The models performed well in predicting song popularity, showcasing the importance of [mention key factors influencing popularity]. Overall, the project provided valuable insights into the factors affecting song popularity and the effectiveness of regression models in this context.