

Trenton vonHartitzsch

Pascal Wallisch

Principles of Data Science / TTR 12:30pm

14 May 2024

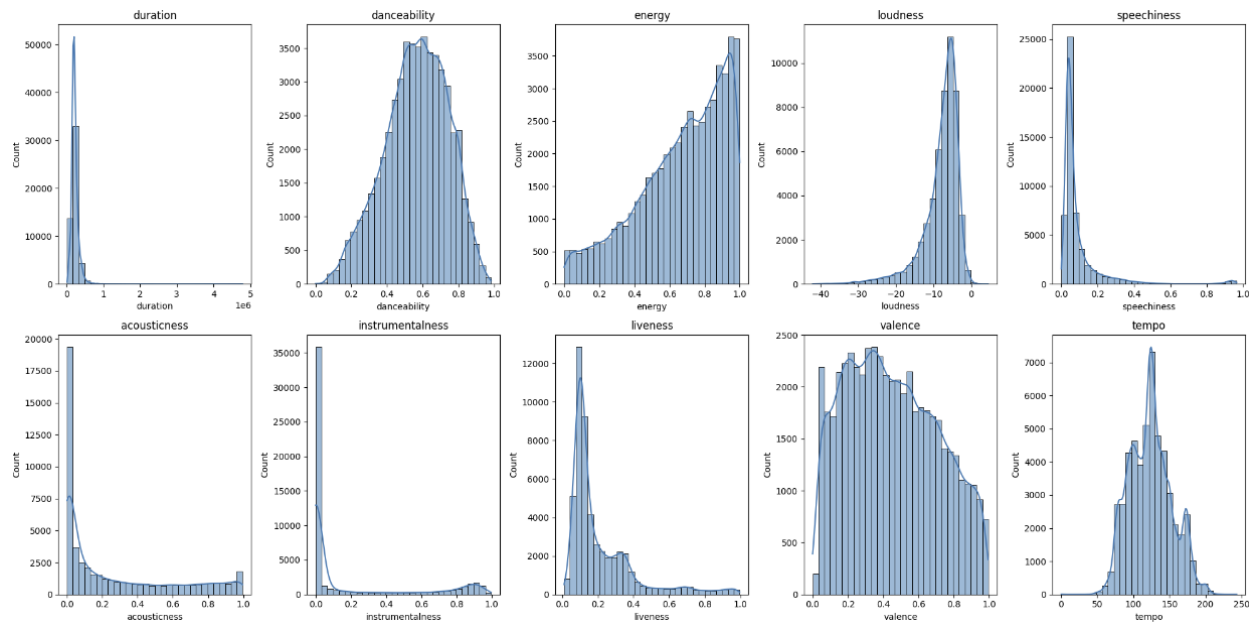
Captone Project (Spotify)

In this project, I aimed to analyze 52,000 songs from Spotify in order to gain a better understanding of what factors make music popular, as well as the audio features that make up specific genres of music. The dataset, consisting of data on 52,000 songs, was first loaded into a Pandas DataFrame. Initial inspection was conducted to understand the structure and completeness of the data. Following the loading and preparation of the data, Missing values were checked and rows containing any missing values in the key features were removed to ensure the robustness of the statistical tests performed throughout the project. In this instance, there were no NaN values so all data was able to be used in the project. To ensure reproducibility of the results, I seeded my N number (18098787) in a random number generator as a unique identifier. This not only matters for the specific train/test split or bootstrapping but also to protect my work from potential plagiarism that can occur.

Question 1:

Consider the 10 song features duration, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence and tempo. Is any of these features reasonably distributed normally? If so, which one?

In this question, I attempted to identify which of the ten song features—duration, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, and tempo—follow a normal distribution. The use of the identification of normally distributed features is important for subsequent statistical analyses that assume normality, such as parametric tests, which may be used later within the project. I employed a plot of the distributions to help with understanding the data and to see the distributions clearly. The results were as follows:



Based on visual inspection, none of the features shows a clear normal distribution. Some, like tempo, might approximate normality more than others, but each has distinct skewness. For formal analysis, statistical tests like Shapiro-Wilk were used to assess normality. The test's null hypothesis is that a sample comes from a normally distributed population. The Shapiro-Wilk test is suitable for this purpose, but even small deviations from normality can result in the rejection of the null hypothesis.

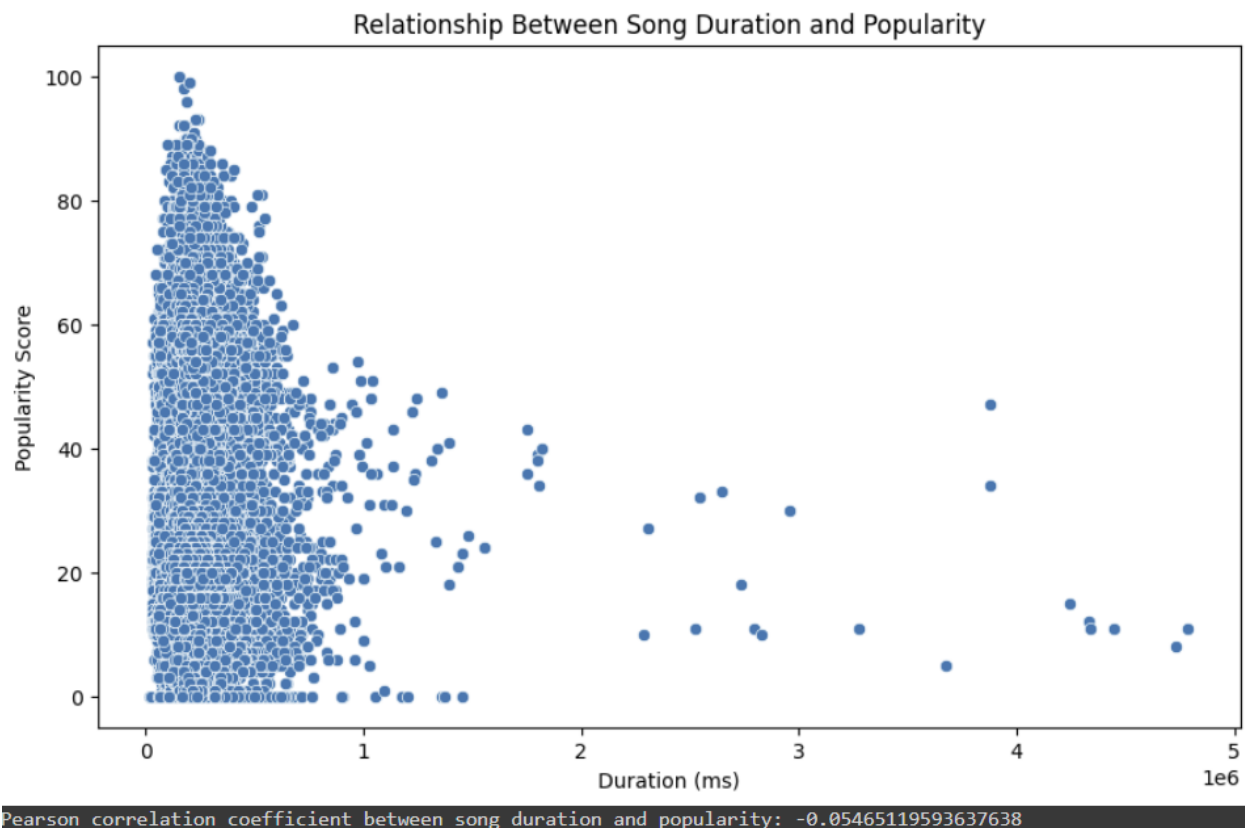
```
Shapiro-Wilk test p-value for duration: 0.0
Shapiro-Wilk test p-value for danceability: 1.4476362789222889e-18
Shapiro-Wilk test p-value for energy: 2.735334602362043e-42
Shapiro-Wilk test p-value for loudness: 0.0
Shapiro-Wilk test p-value for speechiness: 0.0
Shapiro-Wilk test p-value for acousticness: 0.0
Shapiro-Wilk test p-value for instrumentalness: 0.0
Shapiro-Wilk test p-value for liveness: 0.0
Shapiro-Wilk test p-value for valence: 2.82614702813396e-32
Shapiro-Wilk test p-value for tempo: 2.6710901095894868e-17
```

The results from the Shapiro-Wilk tests for normality show extremely low p-values for all the features. Duration, Loudness, Speechiness, Acousticness, Instrumentalness, and Liveness have p-values that are 0.0, strongly suggesting that these distributions are not normal. Danceability, Energy, Valence, and Tempo although not zero, have p-values that are very close to zero (well below any conventional alpha level such as 0.05), indicating these features are also not normally distributed. Therefore, the results indicate the rejection of the null hypothesis and none of the song features in the dataset are normally distributed.

Question 2:

Is there a relationship between song length and popularity of a song? If so, if the relationship positive or negative?

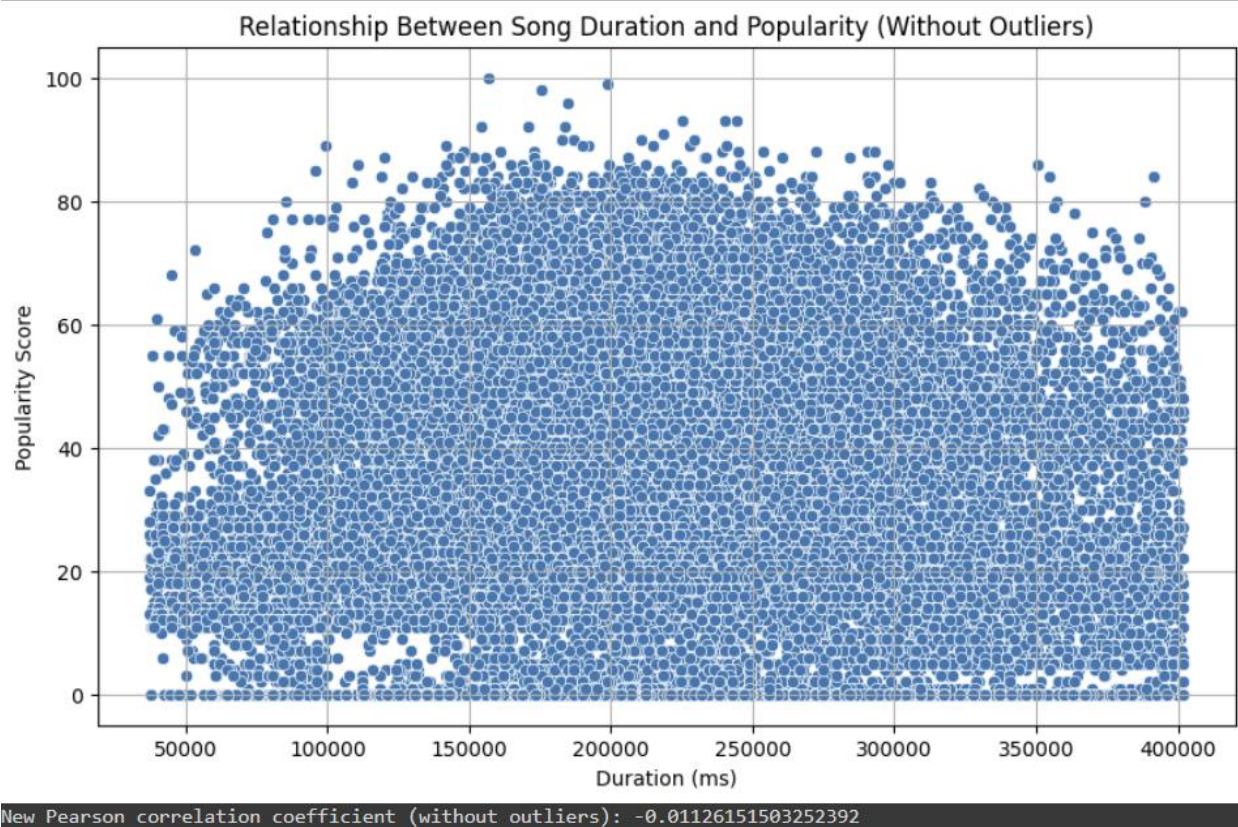
In this question, I attempted to identify if there is a relationship between the duration of a song and its popularity. Understanding this relationship is important for song production and marketing as it might inform decisions if longer or shorter songs tend to be more popular. To explore this relationship, I employed a scatterplot to visually assess the correlation between these two variables. A scatterplot was used as it is one of the most straightforward methods to visualize potential correlations or trends between two quantitative variables. To aid alongside the visual analysis, a statistical test was performed to quantify the strength and direction of the relationship between song duration and popularity. Pearson's correlation coefficient was calculated as it measures the linear correlation between two variables, providing both the strength and direction of the relationship.



The plot shows a wide dispersion of points without a clear linear trend. There is considerable variability in popularity scores across the range of song durations. There does not appear to be a strong visual correlation between song length and popularity. Pearson Correlation Coefficient: The computed correlation coefficient is approximately -0.055. This value is close to zero and

negative, indicating a very weak negative linear relationship between song duration and popularity. This implies that song length is not a strong predictor of its popularity on Spotify. The length of a song does not significantly influence how popular it will be, at least not in a linear manner.

However, there is a possible issue with including outliers to see the relationship. In order to try and remove the potential error that these outliers pose on the correlation, I first identified outliers as data points that lie beyond 1.5 times the interquartile range from the quartiles. Once removing the outliers, I performed another analysis via scatterplot and Pearson correlation coefficient to see if the relationship between duration and popularity changes significantly.

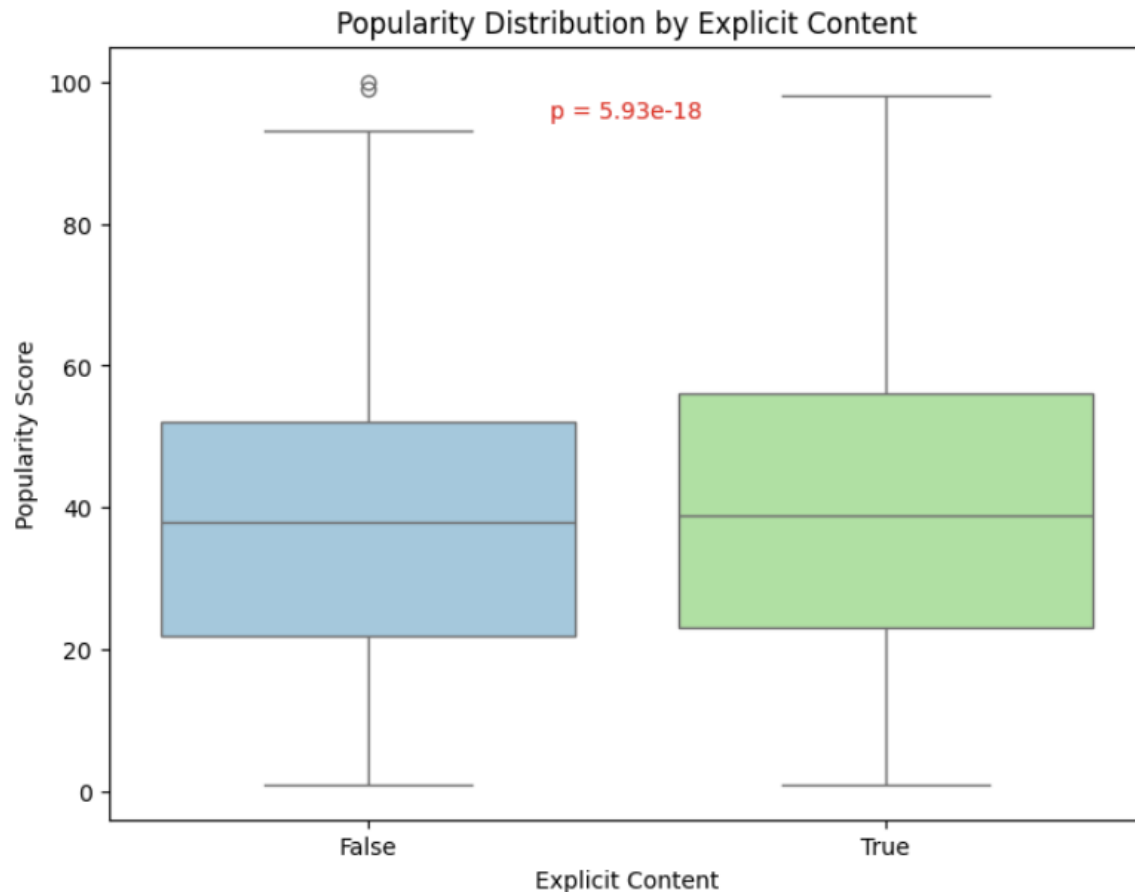


The scatterplot still shows a wide dispersion of popularity scores across the range of song durations, but without extreme values in duration. There is no obvious linear trend even with outliers removed. The new correlation coefficient is approximately -0.011, which is even closer to zero than before. This change indicates an even weaker negative linear relationship between song duration and popularity after removing outliers. Removing outliers has slightly altered the correlation coefficient, further weakening the already negligible relationship between song duration and popularity. This suggests that the length of a song, within a more typical range of durations, has very little to no impact on how popular it will be on Spotify, but shorter songs tend to lean to be slightly more popular.

Question 3:

Are explicitly rated songs more popular than songs that are not explicit?

In this analysis, I sought to determine whether explicitly rated songs are more popular than songs that are not explicit. The popularity of explicit content can influence decisions related to production, marketing, and distribution, particularly in targeting demographics that either prefer or avoid explicit material. This involved comparing the popularity of two groups: songs with explicit content and songs without explicit content. To do this, I first looked into the data and noticed a large number of songs had 0 as their popularity (6374 to be exact). A popularity of 0 does not make much sense, as even songs with the least popularity would scratch at least 1. Because of this irregularity, I removed all songs that had a popularity of 0, as it was most likely a substitute to a NaN value. After the cleaning of the data, I created boxplots to visually compare the distribution of popularity scores between explicit and non-explicit songs and then furthered that with the use of the Mann-Whitney U test. This non-parametric test is very helpful, as the distributions are not normally distributed. The null hypothesis for this analysis is that explicit and non-explicit materials have the same popularity.



Mann-Whitney U statistic: 108874531.5
P-value: 5.925200292635755e-18

The very small p-value of $5.93e-18$ (far below the conventional alpha level of 0.05) indicates a statistically significant difference in popularity between explicit and non-explicit songs, with explicit songs tending to be more popular, therefore we can reject the null hypothesis.

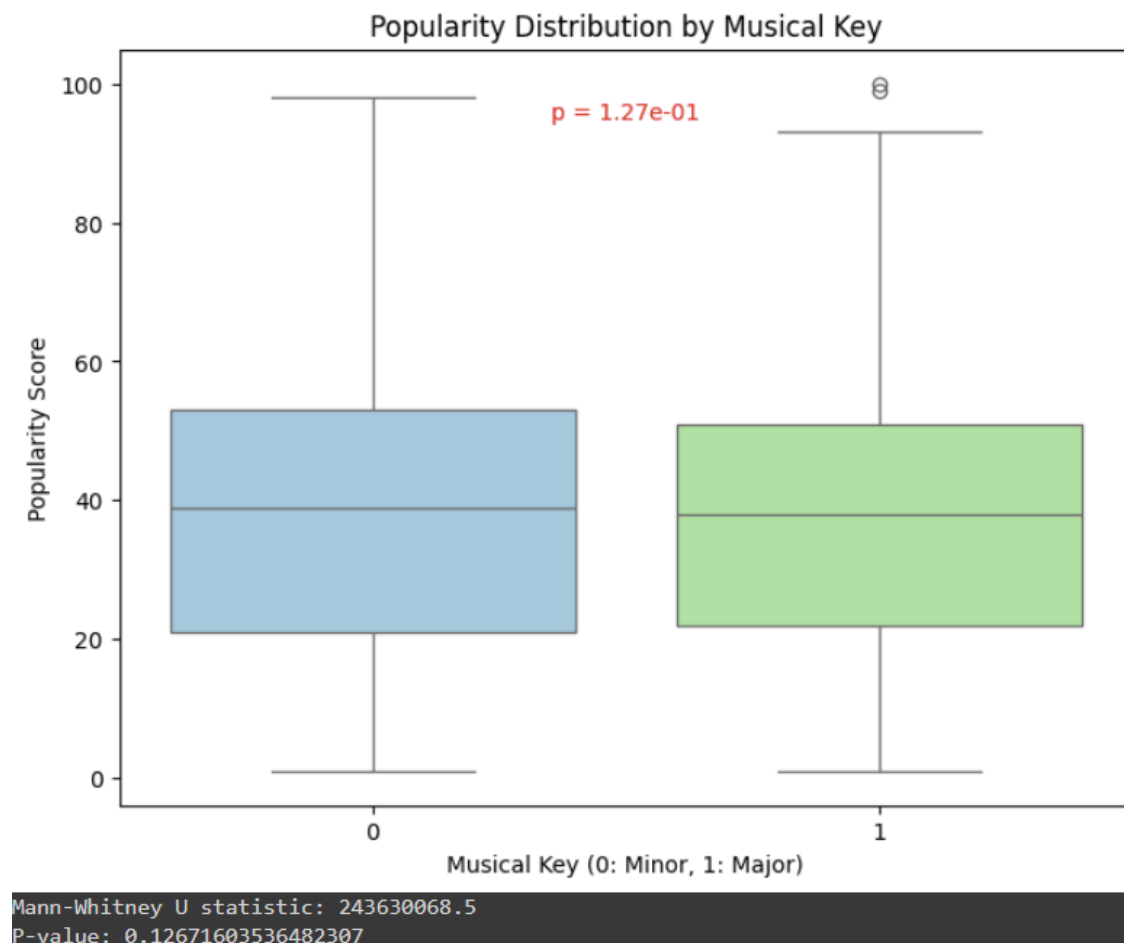
Question 4:

Are songs in major key more popular than songs in minor key?

In this analysis, I explored whether songs in a major key are more popular than songs in a minor key. This involved comparing the popularity scores of two groups: songs in major keys versus songs in minor keys. I first started by creating a boxplot to visually compare the distribution of popularity scores between songs in major vs minor keys. I then coupled it with a Mann-Whitney U test to see if there was a statistically significant difference in popularity based on key mode. It is important to note for this problem, the hypotheses are as follows:

Null - There is no difference in popularity between songs in major key and songs in minor key.

Alternative - There is a statistically significant difference in popularity based on key mode.

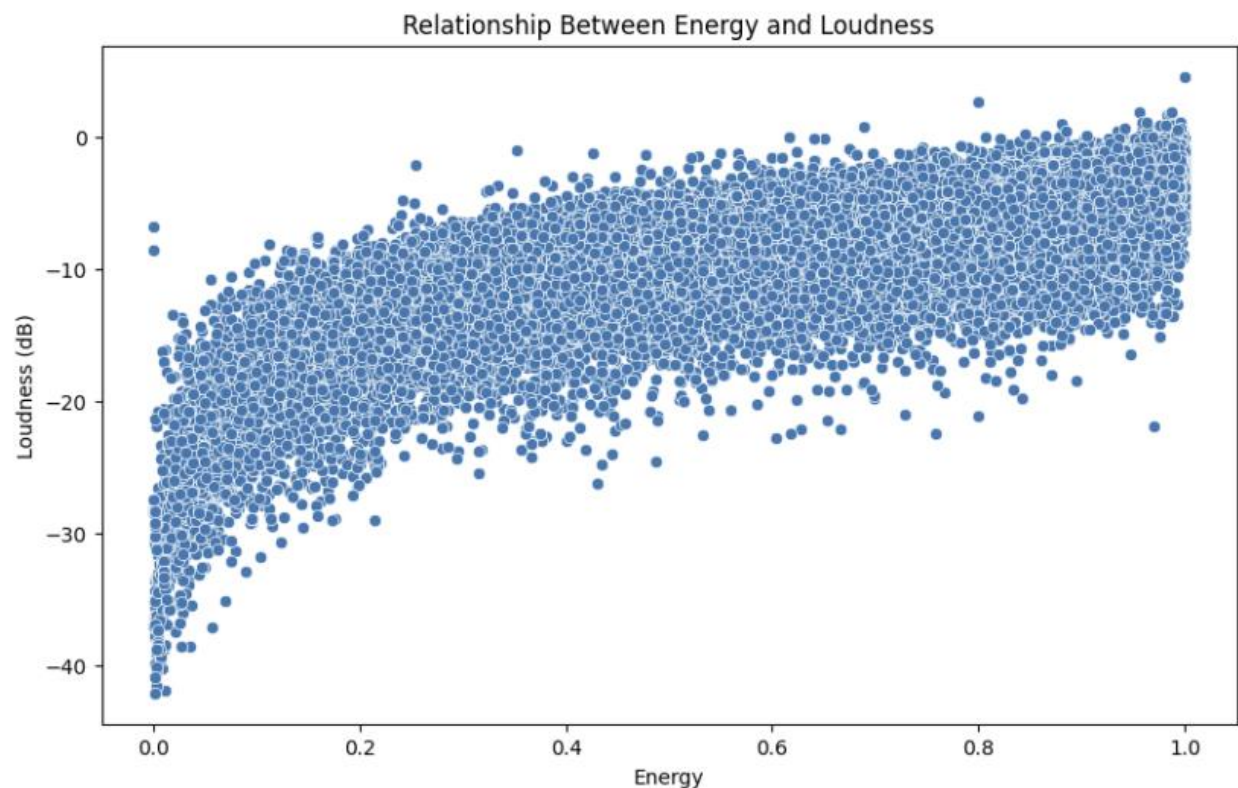


Following the result of the two-tailed test, it became clear that there was not a statistically significant difference between popularity based key mode, as displayed by the p value of 0.127. Therefore, due to the p value being >0.05 , we fail to reject the null hypothesis and major key songs are not significantly more popular than minor key songs.

Question 5:

Energy is believed to largely reflect the “loudness” of a song. Can you substantiate (or refute) that this is the case?

For this analysis, which investigates whether the energy of a song reflects its loudness, I performed a scatterplot to visually examine the relationship between energy and loudness, and calculated the Pearson correlation coefficient to quantify the relationship. The scatterplot allowed visualization of the relationship between energy and loudness to see if there is a linear trend or any pattern. The Pearson correlation coefficient helped in hand with the scatterplot to determine how strongly the energy and loudness metrics are correlated. The results are as follows:



Pearson correlation coefficient between energy and loudness: 0.7748808291850185

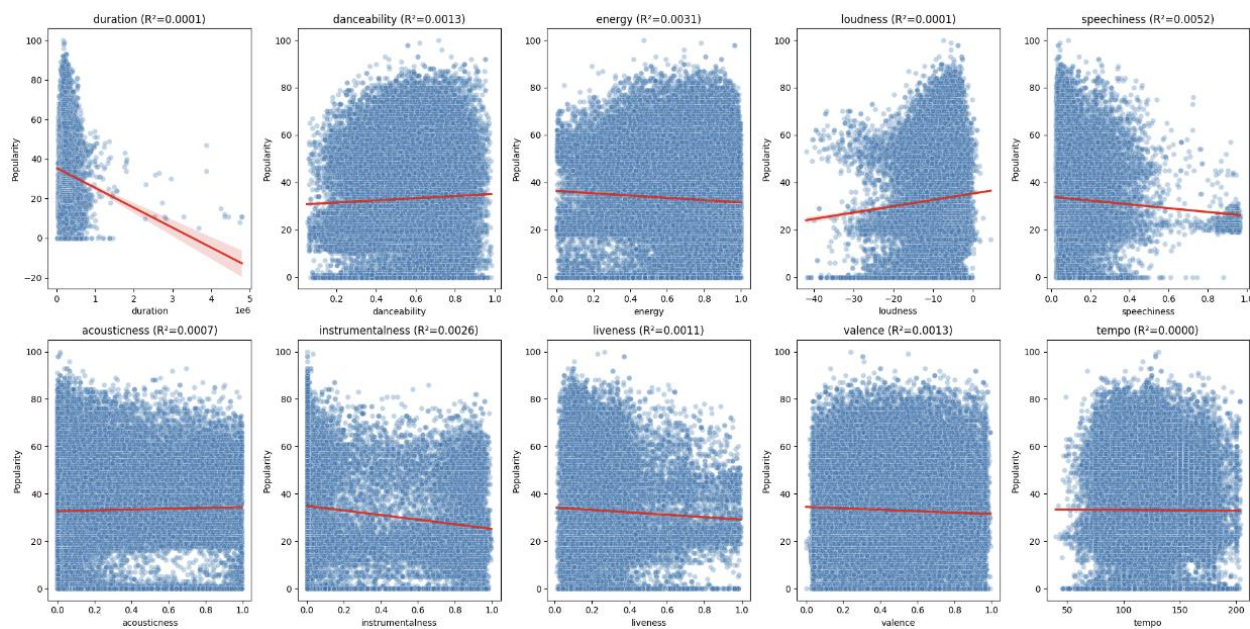
Based on the plot, the scatterplot shows a clear positive relationship between energy and loudness of songs. As the energy increases, the loudness also tends to increase. The plot displays a dense clustering of points, particularly towards higher levels of energy and loudness,

suggesting a strong linear trend in these regions. This analysis is backed with the correlation coefficient of 0.775, which further indicates the strong positive linear relationship between the two variables. Based on the data provided, energy levels largely reflect the loudness of a song.

Question 6:

**Which of the 10 individual (single) song features from question 1 predicts popularity best?
How good is this “best” model?**

For this analysis, where the task was to determine which of the 10 song features from question 1 predicts popularity the best, I employed a regression analysis. I used linear regression analysis for each of the 10 features to receive R^2 values. From the R^2 values, I compared each feature's ability to predict the popularity of a song and identify which feature serves as the best predictor. Understanding the feature that most influences song popularity can guide artists and producers in optimizing song characteristics to enhance their appeal. However, before the linear regression analysis took place, I removed the outliers within the features based on the basis that an outlier is a value that resides outside the parameters of below $Q1 - 1.5IQR$ or above $Q3 + 1.5IQR$. The results are as follows:

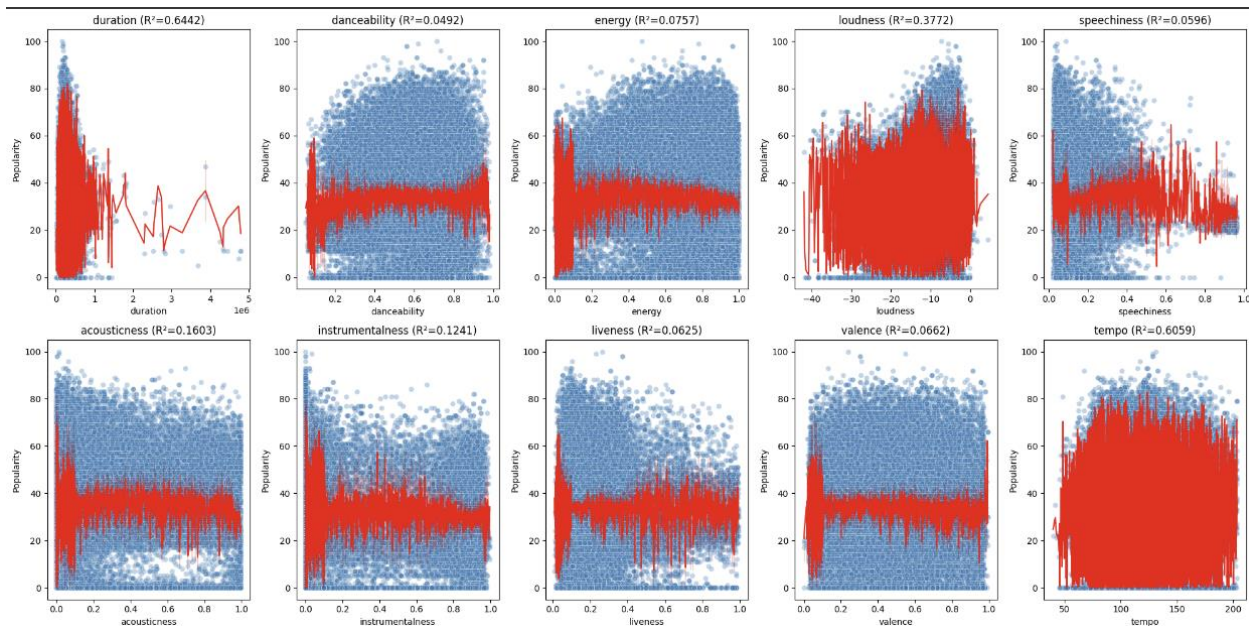



```

R^2 values for each feature predicting popularity:
speechiness: R^2 = 0.0052
energy: R^2 = 0.0031
instrumentalness: R^2 = 0.0026
danceability: R^2 = 0.0013
valence: R^2 = 0.0013
liveness: R^2 = 0.0011
acousticness: R^2 = 0.0007
duration: R^2 = 0.0001
loudness: R^2 = 0.0001
tempo: R^2 = 0.0000

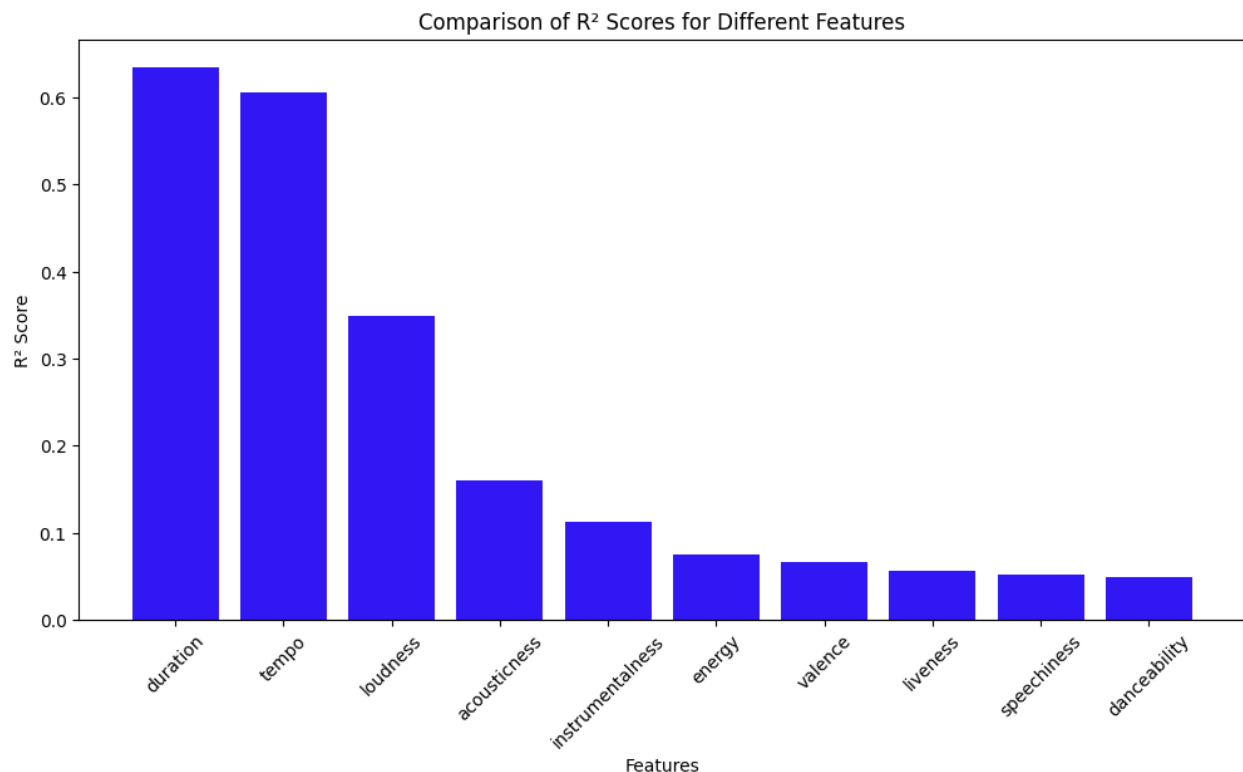
```

The red line depicts the perfect predictor line. Despite all of the features available, no feature had a high R^2 score. Of the features, speechiness appeared to have the largest R^2 score of 0.0052, meaning that speechiness alone explains very little of the variability in popularity. The low scores are influenced by a multitude of outside factors that were not analyzed. Some of these characteristics include marketing efforts, public relations, artist popularity, and cultural/social factors. Not only are those factors contributors, but also the data may not be linear, resulting in low R^2 values for linear regression. To test this, I conducted a Random Forest Regressor because it can handle non-linear relationships better. The results of the regressor are as follows:



```
R^2 values for each feature predicting popularity:  
duration: R^2 = 0.6442  
tempo: R^2 = 0.6059  
loudness: R^2 = 0.3772  
acousticness: R^2 = 0.1603  
instrumentalness: R^2 = 0.1241  
energy: R^2 = 0.0757  
valence: R^2 = 0.0662  
liveness: R^2 = 0.0625  
speechiness: R^2 = 0.0596  
danceability: R^2 = 0.0492
```

The red line in each plot represents the predicted popularity based on the Random Forest model's predictions for each specific feature, plotted against the actual feature values. By overlaying this line on top of the scatterplot of actual data points, you can see where the model predictions align with actual outcomes and where they diverge.

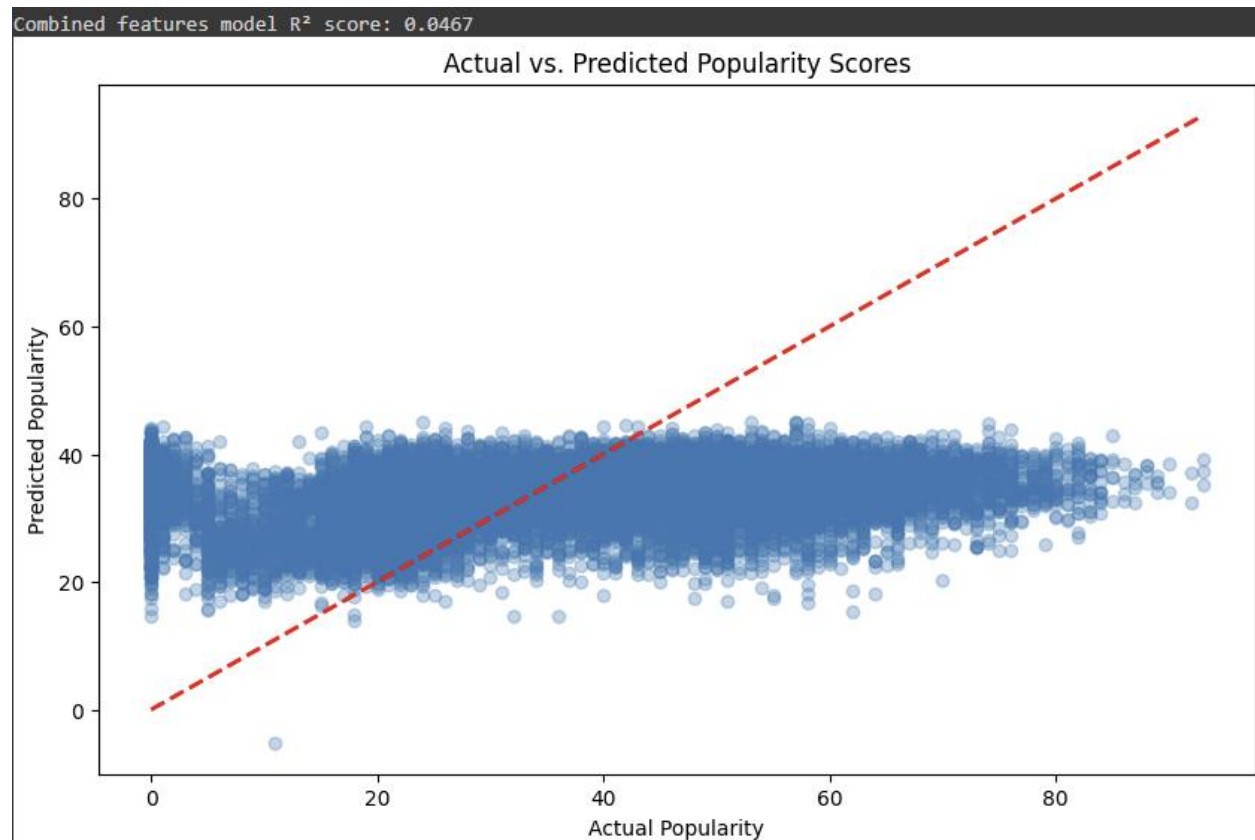


The Random Forest Regressor excelled in comparison to the linear regression previously used. The random forest regressor yielded R² scores ranging from 0.05 to 0.64. The 'best' model was duration, as duration explains 64% of the variance in popularity. Closely following duration was tempo, yielding a score of .61. These R² scores show that the random forest regressor model is a fairly good fit for the data.

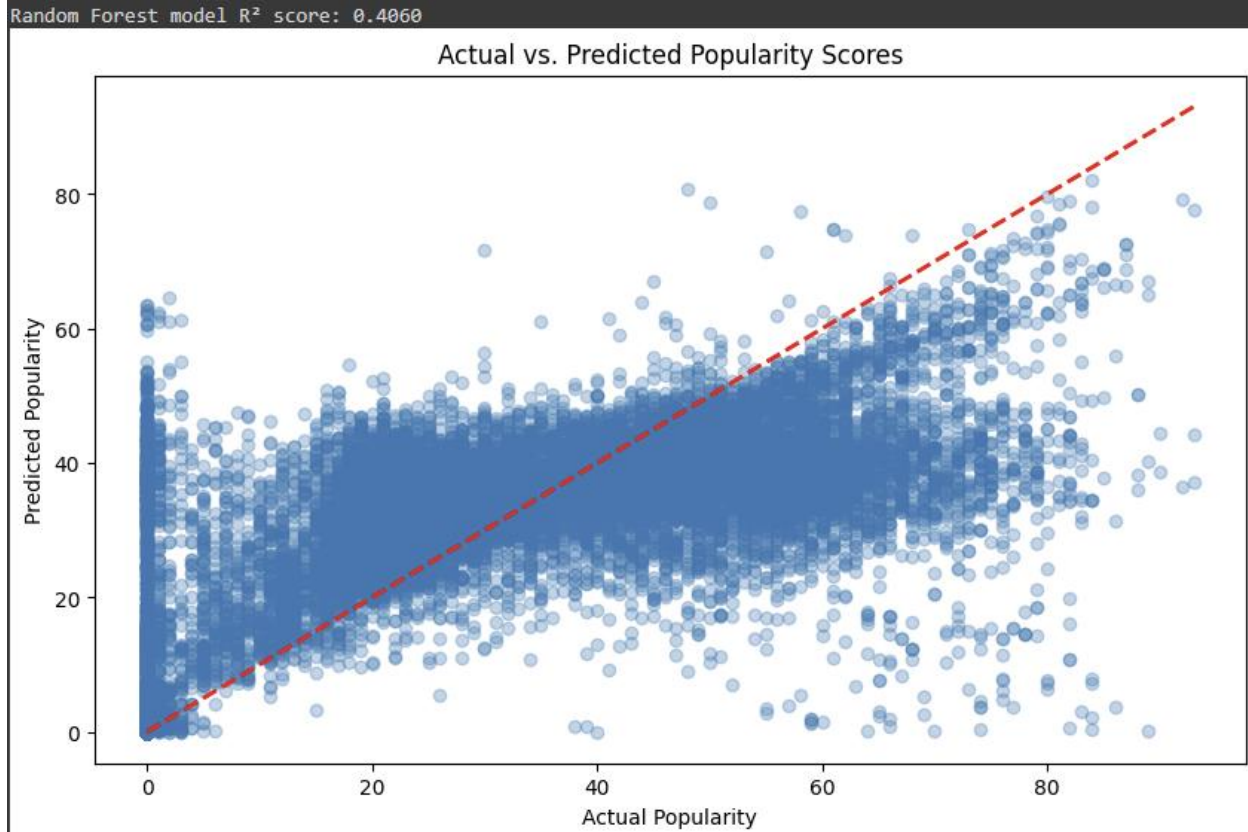
Question 7:

Building a model that uses **all of the song features from question 1, how well can you predict popularity now? How much (if at all) is this model improved compared to the best model in question 6). How do you account for this?**

Building on the findings from Question 6, this analysis tests the hypothesis that a combination of features might explain more variability in popularity due to potential synergistic effects. Multiple linear regression was used to model the interactions. The results are as follows:



The red line in the visual represents a perfect predictor. The use of the multiple linear regression fails to explain much of the variance in the popularity scores, as shown with the R^2 score of 0.0467. This outcome suggests that while the features included might have some predictive power, they are insufficient to robustly predict popularity on their own. This can most likely be because of how complex popularity is, or because the relationship between features and popularity is not linear. In order to test this theory once again, I used a Random Forest Regressor, which can handle non-linearity much better than linear regression. The results are as follows:

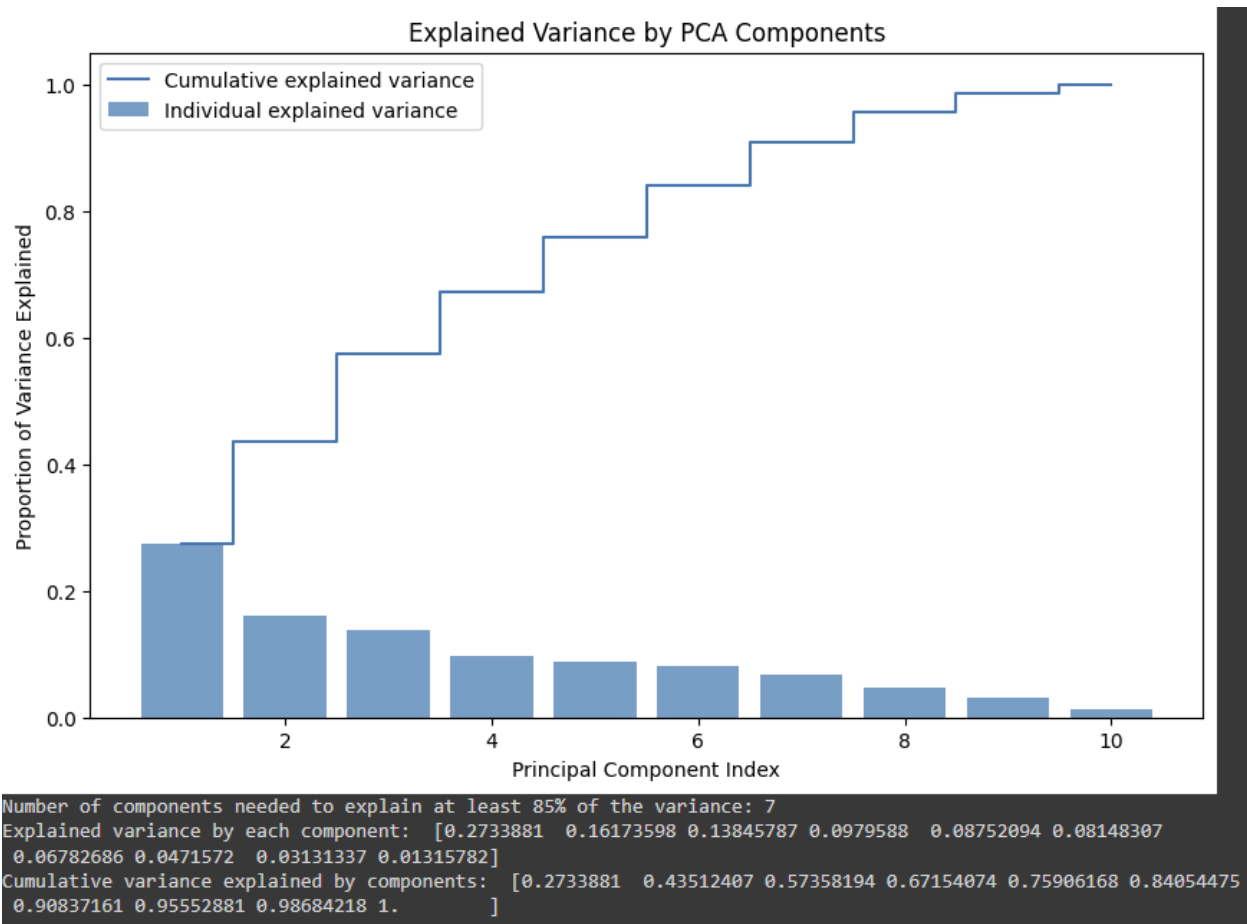


The Random Forest model with combined features demonstrates an R^2 score of 0.4060, which, while showing an improvement over linear regression models by capturing more complex interactions, is still approximately 0.24 points lower than the R^2 scores achieved by the best-performing individual features, such as duration ($R^2 = 0.6442$) and tempo ($R^2 = 0.6059$), from the previous analysis in question 6. This relative underperformance of the combined model can be accounted for by its ability to incorporate multiple features simultaneously, therefore adding complexity by integrating various song aspects and their interactions. The model provides a wider understanding of song popularity, even though it does not reach the high predictive power of the best single-feature models. The combination of features can potentially dilute the impact of any single strong predictor by averaging out its effects with less predictive features, therefore resulting in a lower overall R^2 score than models using only the most predictive features.

Question 8:

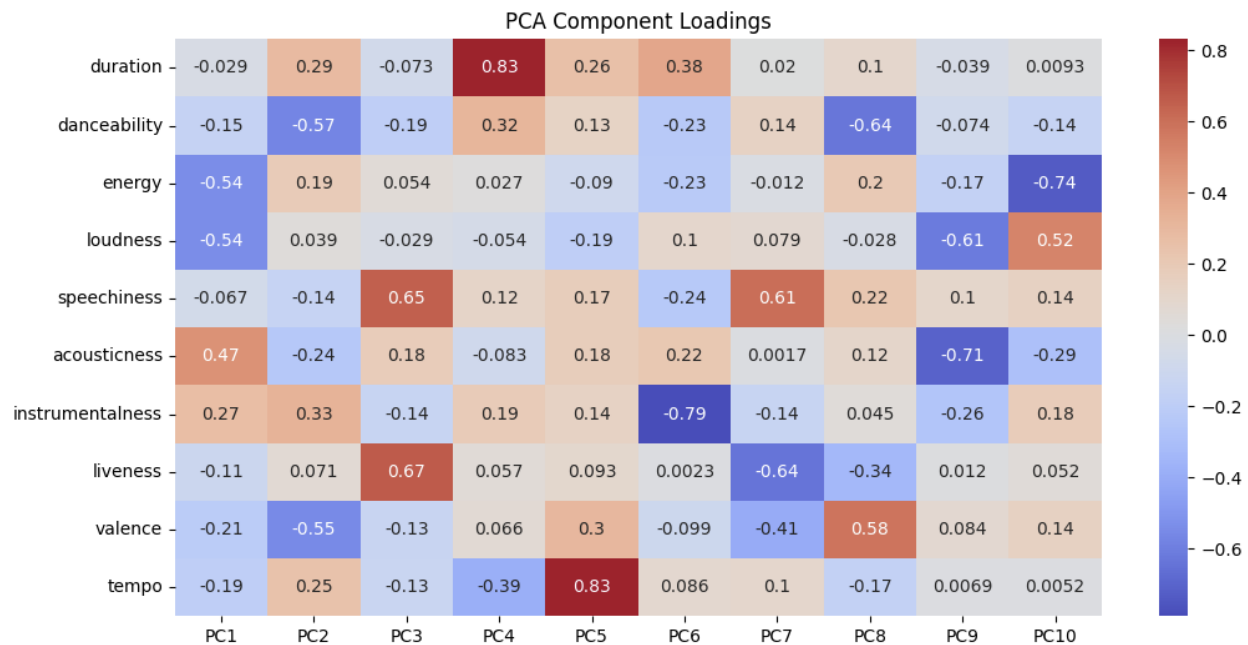
When considering the 10 song features above, how many meaningful principal components can you extract? What proportion of the variance do these principal components account for?

For this analysis, I aimed to explore the effectiveness of Principal Component Analysis (PCA) in reducing the dimensionality of the dataset consisting of ten song features—duration, danceability, energy, loudness, speechiness, acousticness, instrumentality, liveness, valence, and tempo. The goal was to determine how many meaningful principal components can be extracted and what proportion of the variance these components account for. PCA is a powerful tool for dimensionality reduction, used to simplify the data while retaining as much information as possible. It transforms the original variables into a new set of variables, which are orthogonal (uncorrelated) and ranked according to the variance they capture from the data. I performed PCA to extract the principal components and then looked into the proportion of variance explained by each component. The results are as follows:



The PCA results indicate that the first three principal components account for 60% of the total variance in the dataset (0.273, 0.162, 0.138). To capture at least 85% of the total variance, we

found that seven principal components are required. This reduction from ten original features to seven principal components allows for a more efficient representation of the dataset without significant loss of information, enabling more effective subsequent analyses. The one issue with the PCA is that their interpretation is lost. In order to try and maintain an understanding of each component, I loaded them onto a heatmap as follows:



From the heat map, the interpretation of the components are:

PC1: High values might indicate songs that are quieter and more acoustic, while low values suggest songs that are loud and energetic

PC2: High values indicate less danceable, less joyful music with more instrumental elements.

PC3: High values suggest songs that likely feature live performance aspects and more speech

PC4: Highly influenced by the duration of the songs and inversely by tempo, suggesting a dimension where longer songs tend to have slower tempos

PC5: Contrasts fast-tempo songs against acoustic characteristics, indicating that songs with a higher score are faster and less acoustic

PC6: High values could indicate longer songs that are less instrumental, potentially pointing towards more vocal or lyrical content

PC7: High values suggest a greater emphasis on speechiness in the track, whereas low values could indicate a focus on live performance attributes

PC8: High values indicate songs that are emotionally positive but less danceable, which could characterize certain genres like ballads or slower, more melodic music.

PC9: Higher values might be associated with songs that are less acoustic and quieter, potentially indicating more subdued or softer music styles.

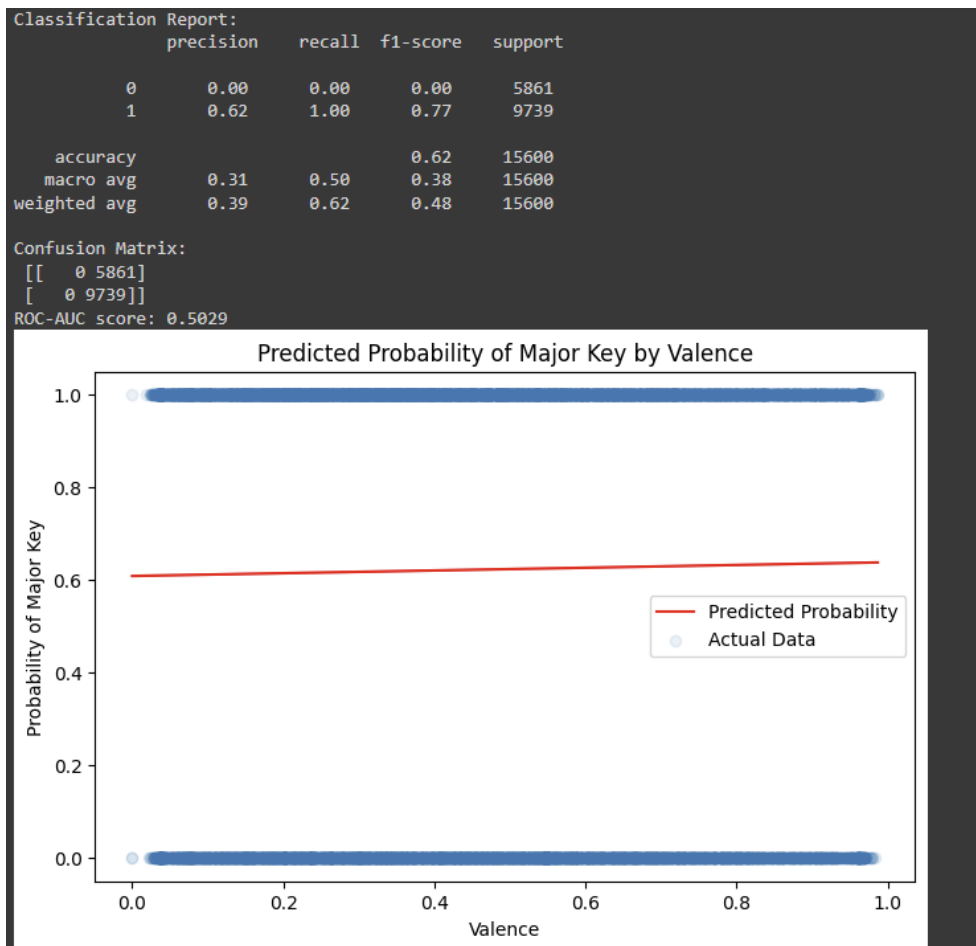
PC10: Songs with high values may be louder but have less energetic content.

This interpretation can help with understanding each component, leading to a more beneficial finding of results.

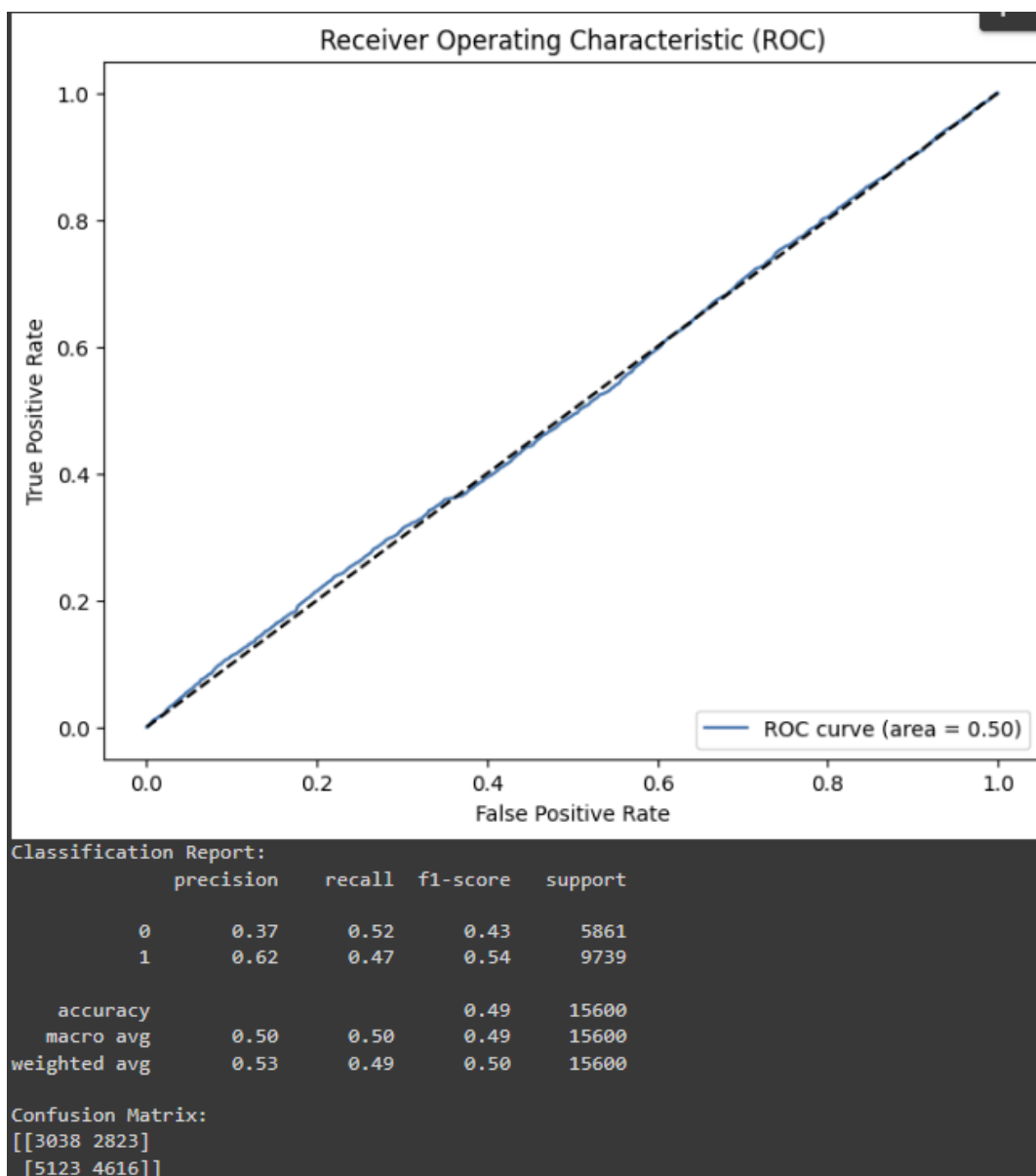
Question 9:

Can you predict whether a song is in major or minor key from valence? If so, how good is this prediction? If not, is there a better predictor?

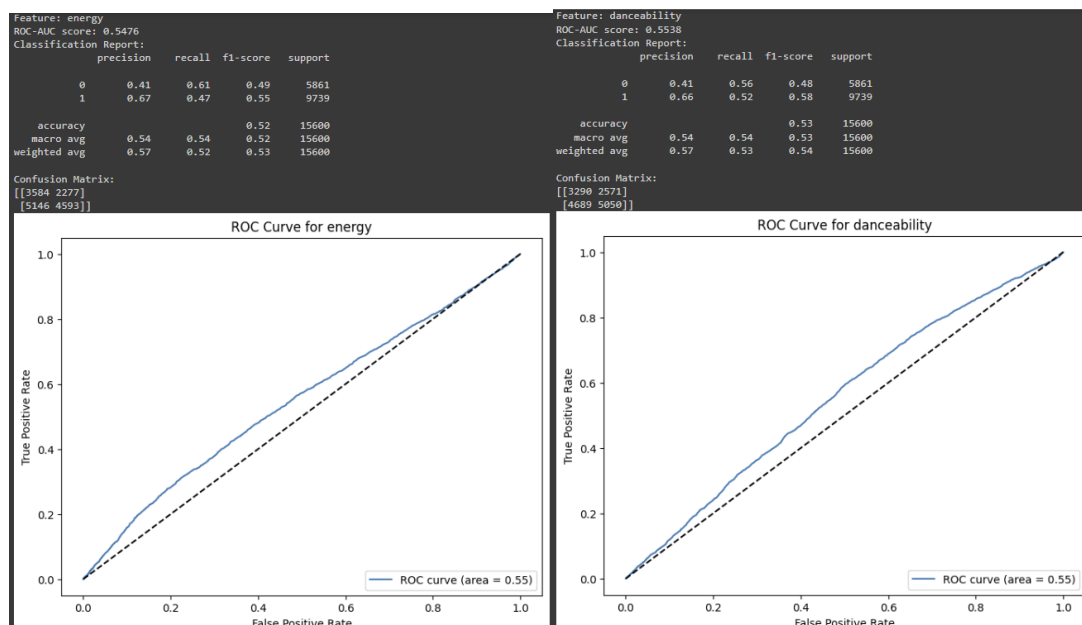
For this analysis, I aimed to determine if song valence (a measure of musical positiveness) can predict whether a song is in a major or minor key. Understanding this relationship can provide insights into how the emotional content of music correlates with musical theory concepts like key modes. In order to do this, I used a logistic regression to model the relationship between valence and key note. The results are as follows:

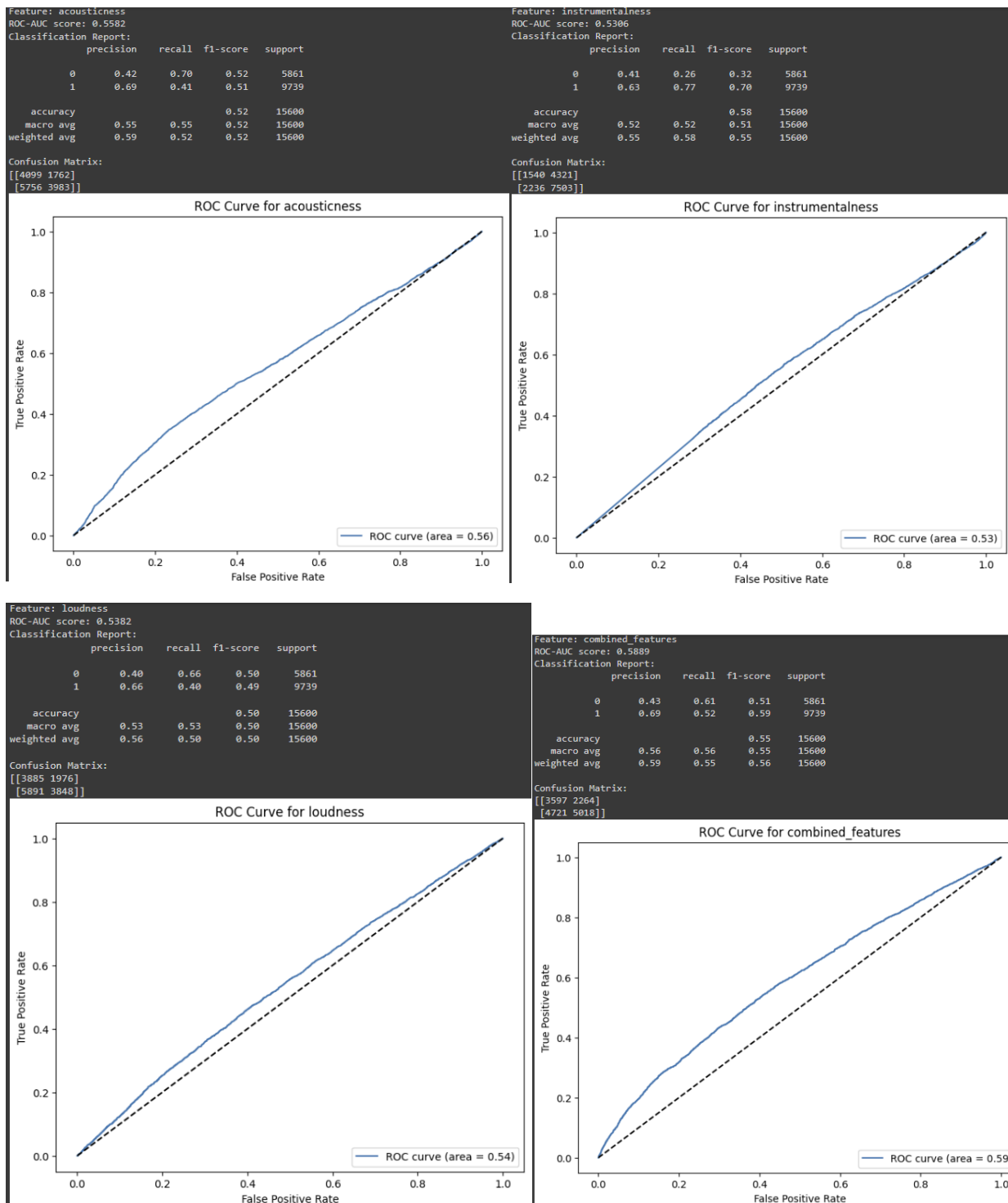


The model does has a precision of 0.62, implying that when the model predicts a song is in a major key, it is correct 62% of the time. At the surface level, that seems like the model is working better than randomly guessing, however the major key recall is 1.00 and all of the minor key classification reports are 0. The confusion matrix predicted all songs as being in a major key (1), as shown by the zeros in the first column (actual minor) and zeros in the first row of the second column (predicted minor). And the ROC-AUC score of 0.5029 indicates a model with no discriminative ability whatsoever between major and minor keys, essentially performing no better than random guessing. The model may have an issue due to the imbalance of major and minor keys within the data set. In order to investigate if this was the case, I used a resampling technique called SMOTE (Synthetic Minority Over-sampling Technique). This will synthesize new minority class instances rather than duplicating existing ones until it is balanced. The results are as follows:



The model's performance, as captured by the ROC-AUC score, was found to be 0.51, indicating a marginal improvement over random guessing, but still showing limited predictive accuracy. For songs in a minor key (labelled as 0 in the confusion matrix), the model achieved a precision of 0.38 and a recall of 0.52, resulting in an F1-score of 0.44. This suggests that while the model is relatively moderate at identifying true minor key songs, it lacks precision and falsely identifies major key songs as minor key songs quite frequently. For songs in a major key (labelled as 1), the precision was somewhat higher at 0.62 with a recall of 0.48, leading to an F1-score of 0.54. This indicates that while the model is more precise when predicting major key songs, it fails to identify just over half of them correctly. Based on this information, it appears that it is feasible but not very effective to predict the songs key from valence by itself. I wanted to find a better predictor, so I took a look at the features that may elude to a better model. Features I selected were: energy, danceability, acousticness, instrumentalness, and loudness. I followed the same method as previous and added a combined feature predictor (a combination of the features I hand selected), and received the following results:



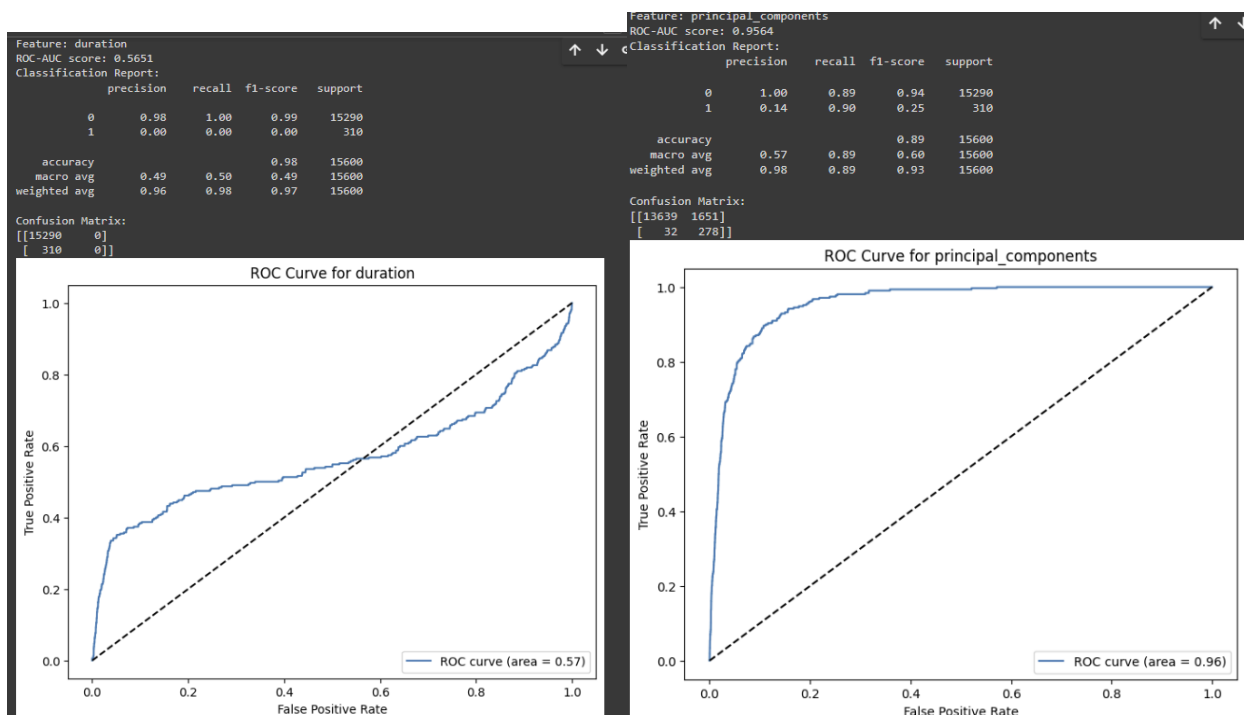


As observed from the results, all of the selected features yielded higher ROC-AUC values than valence, thus indicating that they are all better predictors. However, the "best" predictor from the data analysis was the combined feature predictor, with the ROC-AUC of 0.59.

Question 10:

Which is a better predictor of whether a song is classical music – duration or the principal components you extracted in question 8?

For this analysis, I aimed to determine if song duration or principal components (extracted from various song features in question 8) can predict whether a song is classified as classical music. Understanding this relationship can provide insights into how different musical characteristics contribute to the classification of music genres, particularly classical music. In order to do this, I used logistic regression models to assess the predictive power of both duration and principal components. The results are as follows:



Based on the analysis, principal components are a significantly better predictor of whether a song is classical music compared to using duration alone. The ROC-AUC score for the principal components model was 0.9564, which is much higher than the 0.5651 for the duration model. The classification report for the principal components model shows a substantial improvement in recall for the classical music class (0.90) compared to the duration model (0.00). Additionally, the principal components model has a much higher precision for the classical music class (0.14) than the duration model (0.00), although it is still quite low. The ROC curve for the principal components model illustrates a far superior performance, with an area under the curve close to 1, indicating excellent discriminative ability. In contrast, the ROC curve for the duration model is much closer to the diagonal line, indicating performance close to random guessing. These results suggest that the combination of various song features captured by principal components provides

a more comprehensive understanding and significantly better predictive power for classifying classical music compared to using duration alone.

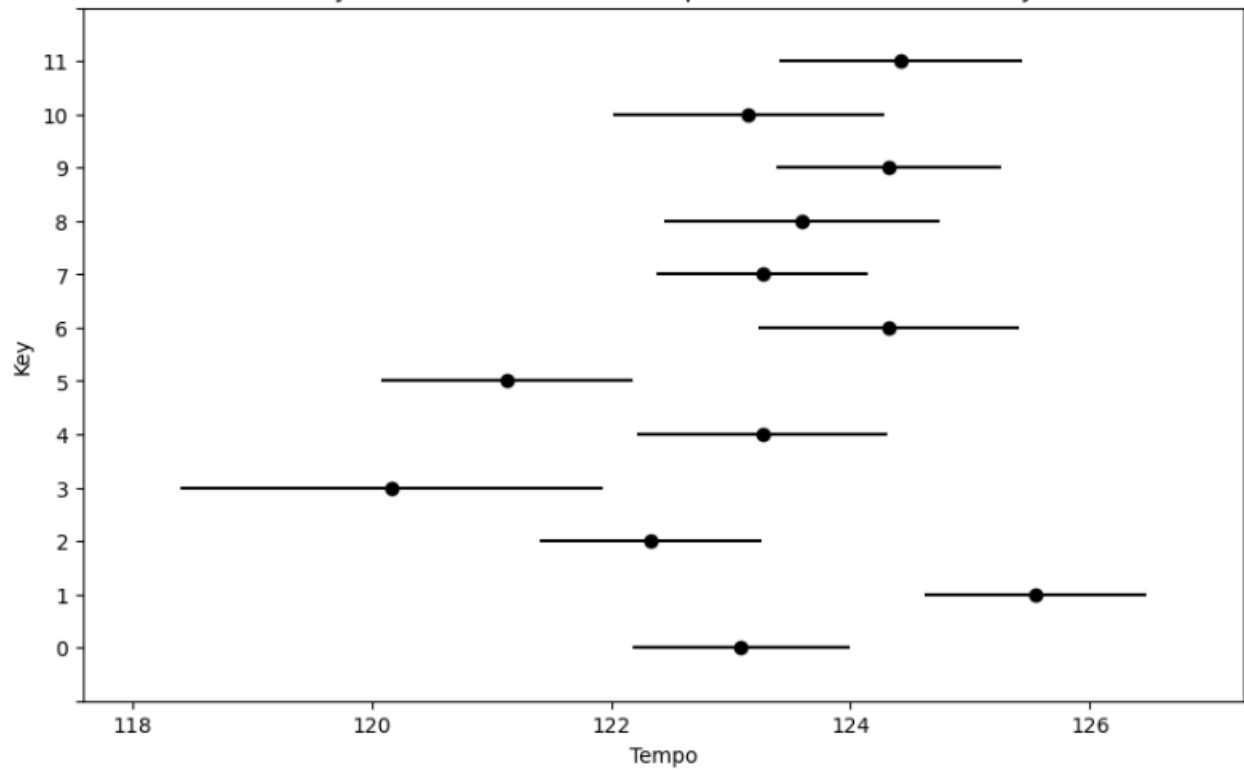
Extra Credit:

Tell us something interesting about this dataset that is not trivial and not already part of an answer (implied or explicitly) to these enumerated questions

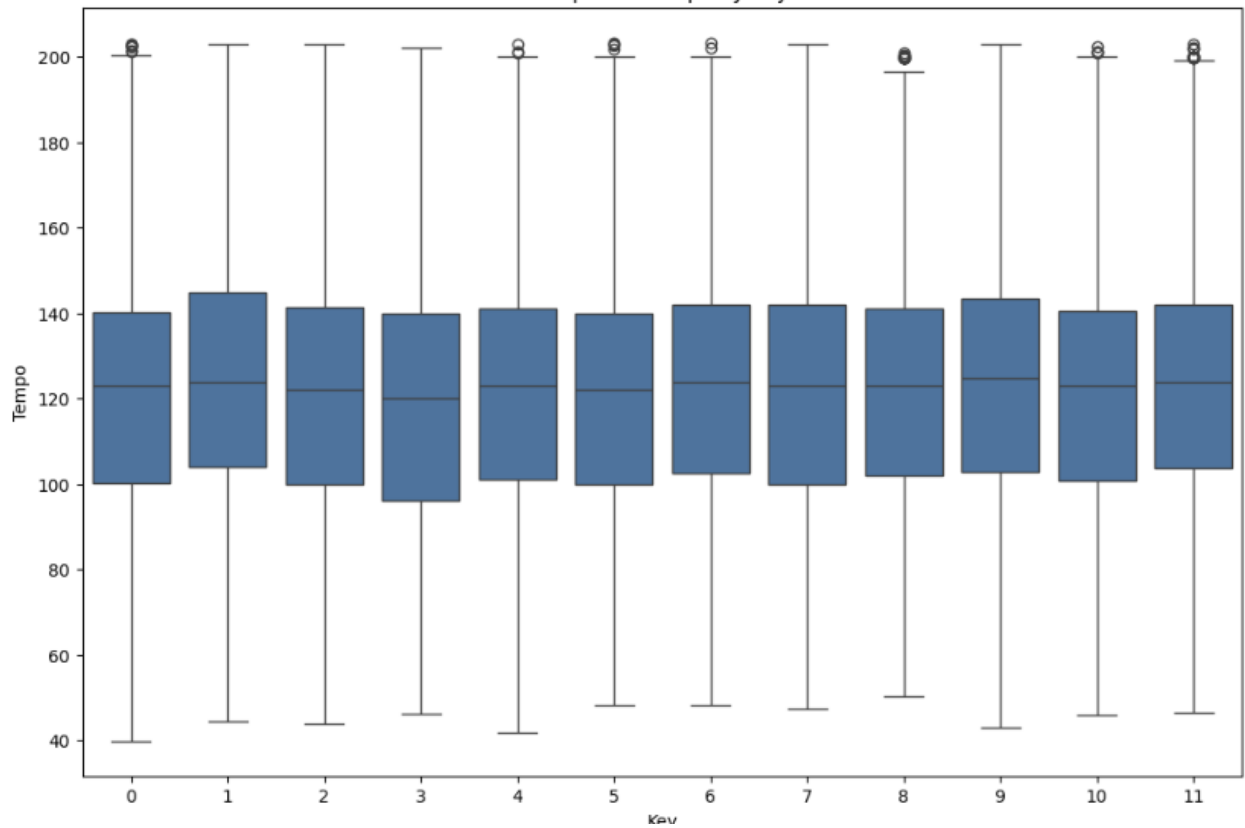
For this analysis, I aimed to determine whether there are statistically significant differences in the average tempo of songs between different musical keys. Understanding these differences can provide insights into how key selection might influence the tempo of a composition, which is valuable for both music theory and practical composition. To achieve this, I used an Analysis of Variance (ANOVA) test followed by Tukey's Honestly Significant Difference (HSD) test for post-hoc analysis. ANOVA is useful in this context as it allows us to determine whether there are any statistically significant differences in the means of multiple groups (in this case, the different keys), and Tukey's HSD test helps to identify which specific pairs of keys differ. The results are as follows:

ANOVA result: F-statistic = 8.80289349652358, p-value = 7.841299214564952e-16							4	7	0.0012	1.0	-1.9251	1.9275	False
Multiple Comparison of Means - Tukey HSD, FWER=0.05							4	8	0.3273	1.0	-1.8687	2.5234	False
=====							4	9	1.0602	0.8449	-0.9202	3.0406	False
group1	group2	meandiff	p-adj	lower	upper	reject	4	10	-0.1144	1.0	-2.2894	2.0606	False
0	1	2.4653	0.0006	0.6432	4.2874	True	4	11	1.1569	0.7977	-0.9014	3.2153	False
0	2	-0.7561	0.9714	-2.5782	1.066	False	5	6	3.1906	0.0001	1.0576	5.3235	True
0	3	-2.9188	0.0211	-5.6191	-0.2186	True	5	7	2.1348	0.0158	0.2046	4.0649	True
0	4	0.1774	1.0	-1.7721	2.1269	False	5	8	2.4609	0.0136	0.2615	4.6603	True
0	5	-1.9562	0.0493	-3.9095	-0.0029	True	5	9	3.1938	0.0	1.2096	5.1779	True
0	6	1.2344	0.678	-0.7601	3.2289	False	5	10	2.0192	0.1001	-0.1592	4.1976	False
0	7	0.1786	1.0	-1.5974	1.9545	False	5	11	3.2905	0.0	1.2285	5.3524	True
0	8	0.5047	0.9987	-1.5607	2.5701	False	6	7	-1.0558	0.8448	-3.0276	0.916	False
0	9	1.2376	0.5461	-0.5969	3.0721	False	6	8	-0.7296	0.996	-2.9657	1.5065	False
0	10	0.063	1.0	-1.98	2.1061	False	6	9	0.0032	1.0	-2.0215	2.028	False
0	11	1.3343	0.496	-0.5841	3.2527	False	6	10	-1.1714	0.8556	-3.3868	1.0441	False
1	2	-3.2214	0.0	-5.0643	-1.3785	True	6	11	0.0999	1.0	-2.0011	2.201	False
1	3	-5.3841	0.0	-8.0985	-2.6698	True	7	8	0.3262	1.0	-1.7174	2.3697	False
1	4	-2.2879	0.0081	-4.2568	-0.319	True	7	9	1.059	0.7518	-0.7508	2.8688	False
1	5	-4.4215	0.0	-6.3942	-2.4487	True	7	10	-0.1156	1.0	-2.1365	1.9054	False
1	6	-1.2309	0.6951	-3.2444	0.7826	False	7	11	1.1557	0.6982	-0.7391	3.0505	False
1	7	-2.2867	0.0019	-4.084	-0.4894	True	8	9	0.7329	0.9927	-1.3618	2.8275	False
1	8	-1.9605	0.0883	-4.0443	0.1232	False	8	10	-0.4417	1.0	-2.7212	1.8378	False
1	9	-1.2277	0.577	-3.0828	0.6275	False	8	11	0.8296	0.9848	-1.3389	2.998	False
1	10	-2.4023	0.0078	-4.4639	-0.3406	True	9	10	-1.1746	0.7887	-3.2472	0.898	False
1	11	-1.131	0.7551	-3.0692	0.8072	False	9	11	0.0967	1.0	-1.8531	2.0465	False
2	3	-2.1627	0.2772	-4.8771	0.5516	False	10	11	1.2713	0.7372	-0.8759	3.4185	False
2	4	0.9335	0.9266	-1.0354	2.9024	False							
2	5	-1.2001	0.7018	-3.1728	0.7726	False							
2	6	1.9905	0.0561	-0.023	4.004	False							
2	7	0.9347	0.8689	-0.8626	2.732	False							
2	8	1.2608	0.709	-0.823	3.3446	False							
2	9	1.9937	0.0226	0.1385	3.8489	True							
2	10	0.8191	0.9795	-1.2425	2.8807	False							
2	11	2.0904	0.0217	0.1522	4.0286	True							
3	4	3.0962	0.016	0.2948	5.8977	True							
3	5	0.9627	0.9938	-1.8414	3.7667	False							
3	6	4.1532	0.0001	1.3202	6.9862	True							
3	7	3.0974	0.0089	0.4138	5.781	True							
3	8	3.4236	0.0059	0.5402	6.3069	True							
3	9	4.1564	0.0	1.4337	6.8791	True							
3	10	2.9818	0.033	0.1145	5.8492	True							
3	11	4.2531	0.0	1.4732	7.033	True							
4	5	-2.1336	0.0406	-4.2245	-0.0427	True							
4	6	1.057	0.9013	-1.0725	3.1864	False							

Tukey HSD Post-Hoc Test for Tempo Differences Between Keys



Boxplot of Tempo by Key



Based on the analysis, there are statistically significant differences in the average tempo of songs between different keys. The ANOVA test resulted in an F-statistic of 8.8028 with a p-value of $7.84e-16$, indicating that the variations in tempo across keys are not due to random chance. The post-hoc analysis using Tukey's HSD test further identified which specific keys have significant differences in average tempo. The results showed that certain key pairs have significantly different average tempos, as indicated by the 'True' values in the reject column of the Tukey HSD output. These significant differences highlight that some keys are associated with faster or slower tempos compared to others. Some of these include: keys C and E, C# and A#, D and F#, D# and E, and F and G. The Tukey HSD plot visually illustrates these significant differences, with some keys showing clear separation in average tempo. Additionally, the boxplot of tempo by key provides a visual representation of the distribution of tempos across different keys, supporting the statistical findings. These results suggest that the key of a song does influence its average tempo, and certain keys tend to be associated with faster or slower tempos.