Final 210 Project

Sean Villoresi and Ellie Kang

Introduction

Music is an essential part of culture, creativity, and history. Specific songs and types of music can have great significance to groups of people and individuals alike. In America today, the music industry is both highly regarded and hotly debated. One successful song can launch an artist to the top of the charts, etching them into modern history. The importance of music and potential significance of a single song motivates the question - what makes a song successful?

Data

Wanting to explore this question in our project, we found a dataset1 on Kaggle with data on Spotify streams, Youtube views, and various song characteristics. There are 28 columns, and 20,719 observations. The data was collected on Februrary 7th, 2023 by extracting the data from Youtube and Spotify. Our goal is to determine and develop the best model for predicting the success of a song based on the number of streams. We chose to use streams (as opposed to Youtube views) as our outcome variable because of inconsistencies within the data when it comes to music videos. Some music videos were not from the artist's channel (unofficial), and we wanted to test this variable as a predictor.

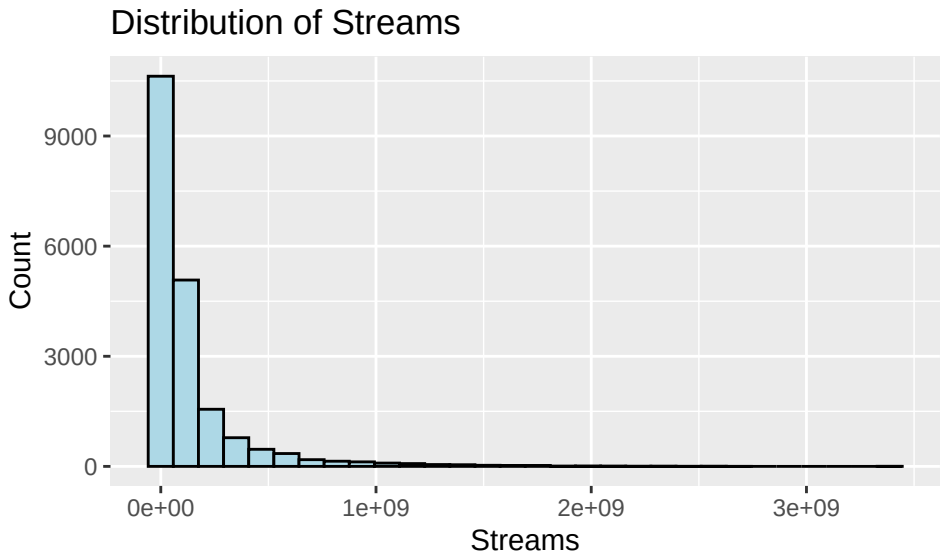
Variables: We will be using streams as the model's outcome variable. We chose the following variables as potential predictors based on their relevance to the listening experience of a song (as opposed to a more descriptive variable such as Description) *Stream*: number of streams of the song on Spotify *Energy*: a measure from 0.0 to 1.0 representing a perceptual measure (dynamic range, loudness, timbre, onset rate, general entropy) of intensity and activity. *Key*: the key the track is in measured in ntegers representing pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C /D , 2 = D. If no key was detected, the value is -1. *Loudness*: the overall loudness of a track in decibels (dB) *Speechiness*: a measure from 0.0 to 1.0 representing the presence of spoken words in a track. *Acousticness*: a from 0.0 to 1.0 of whether the track is acoustic. *Instrumentalness*: a measure from 0.0 to 1.0 that predicts whether a track contains no vocals. *Liveness*: a measure from 0.0 to 1.0 that detects the presence of an audience in the recording. *Valence*: a measure from 0.0 to 1.0 describing the musical positiveness conveyed

by a track. *Tempo*: the overall estimated tempo of a track in beats per minute (BPM). *Duration_ms*: the duration of the track in milliseconds. *Official_video*: boolean value that indicates if the video found is the official video of the song.

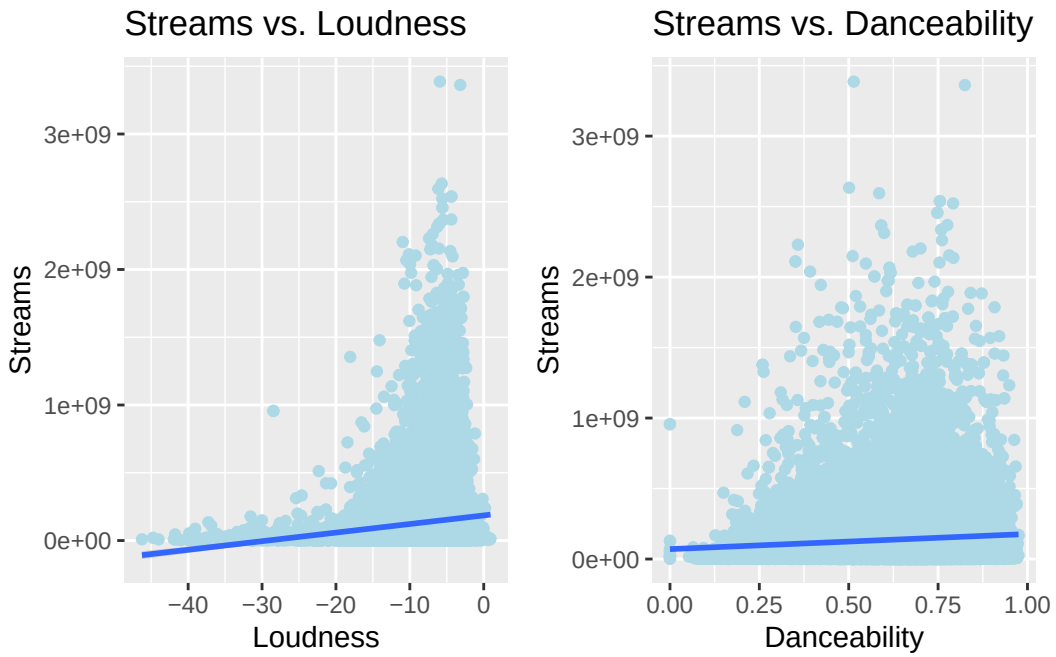
We felt that the dataset had a sufficient number of both quantitative and categorical variables to test predictability. Thus, we chose not to create additional predictors. However, in our data cleaning process, we removed any observations with missing values for streams, danceability, and licensed. After removing missing values for these variables, there were no remaining observations with missing data for relevant variables as listed above.

We hypothesized that Danceability and Loudness would be significant predictors of song success measured in number of streams. We used domain knowledge and the understanding that songs that top the charts tend to be the ones that are catchy, and both danceability and loudness contribute to this. We use this model in our results section to determine if our final model is better than this model.

Exporatory Data Analysis



The distribution of the response variable Stream is skewed to the right with most of the data having lower values for Streams. This is understandable because a handful of songs are extremely popular and these outliers cause the distribution visualization to have most of the data concentrated on the left.



We visualized the relationship between Streams and all the predictors in the dataset. Danceability and Loudness showed slight positive relationships with the response variable Stream. The remaining visualizations are in the Appendix.

Table 1: VIF values

	x
Danceability	1.614911
Energy	3.447032
Loudness	3.240748
Speechiness	1.090299
Acousticness	1.915335
Instrumentalness	1.569425
Liveness	1.066099
Valence	1.529906
Tempo	1.063285
Duration_ms	1.014723
official_video	1.030153

The results of looking at Variance Inflation Factor are values in a range between 1 and 4. Because all the values were below 4, they did not demonstrate much multicollinearity. Thus, we chose to keep all the initial predictors when performing variable selection and did not include any interaction terms.

Table 2: Summary Statistics of Response and Predictors of Interest

	Stream	Danceability	Loudness
Min.	6574	0.0000000	-46.251000
1st Qu.	17769857	0.5190000	-8.785500
Median	49737033	0.6390000	-6.518000
Mean	136930194	0.6209205	-7.641455
3rd Qu.	139092991	0.7420000	-4.931000
Max.	3386520288	0.9750000	0.920000

Methods

We will be fitting a linear model to predict streams because Stream is a non-binary numerical variable. Additionally, we are using it to represent a measure of success, and we want to determine what factors have a linear relationship with streams to determine correlations between predictors and success. Based on our EDA, we chose not to use any interaction terms.

Variable Selection

To start our modeling process, we determined which variables we would use as our "baseline", as described in our introduction above. From these, we decided the first thing we needed to do was determine if any of the variables were seen as non important/non essential, as we want to avoid overcomplicating our model. To start, we performed two variable selection methods, by using a forward and backwards stepwise method starting at our linear model for all terms, and we then proceeded to use a lasso method as well. Between these two methods, we are fairly confident that we can determine the best variables to use.

Step-wise Selection

From our stepwise method, the only variable it seemed to remove was Tempo, and even with that the change in AIC with or without tempo was fairly minimal, so it was from this point that we decided to also use lasso in order to get another perspective on the matter, knowing that stepwise functions can be very influenced by the starting model and its order of variables

Lasso Model

Table 3: Lasso Coefficients

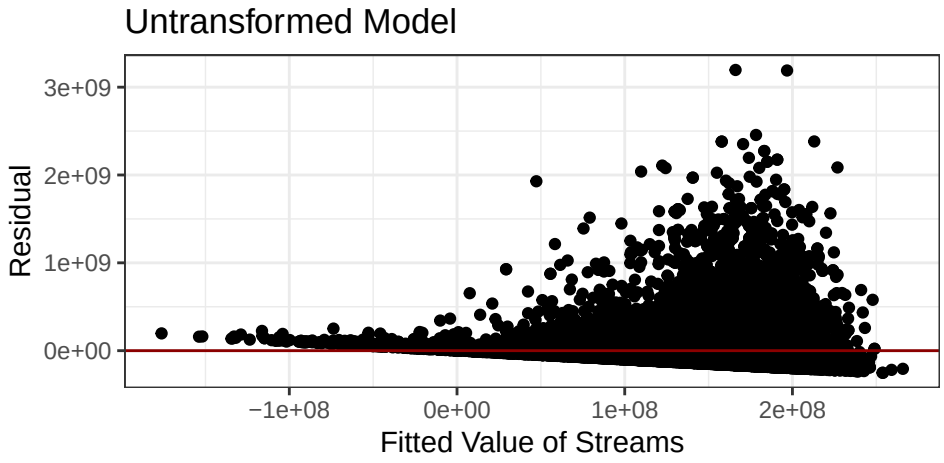
	s0
(Intercept)	0.000000e+00
Danceability	5.449040e+07
Energy	-1.247357e+08
factor(Key)1	1.208188e+07
factor(Key)2	-8.149839e+06
factor(Key)3	-2.303635e+06
factor(Key)4	0.000000e+00
factor(Key)5	0.000000e+00
factor(Key)6	1.873367e+06
factor(Key)7	-1.007666e+07
factor(Key)8	2.576629e+06
factor(Key)9	-1.538578e+07
factor(Key)10	-2.077360e+06
factor(Key)11	6.101606e+06
Loudness	6.595251e+06
Speechiness	-6.511330e+07
Acousticness	-7.939672e+07
Instrumentalness	-3.787604e+07
Liveness	-4.133460e+07
Valence	-4.981840e+07
Tempo	-3.708147e+04
Duration_ms	-2.220704e+01
official_videoTRUE	4.918297e+07

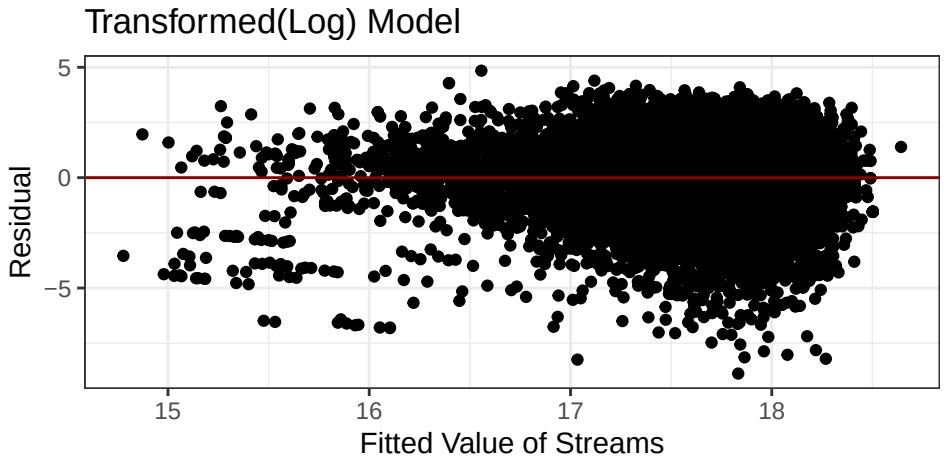
With lasso however, we came to the same conclusion, as our lasso kept essentially every variable to a fairly significant coefficient. As such, we have decided to move on using all of the variables that we had started with at the beginning as predictor variables

Linearity Assumptions and Checks for Transformations

With our variables chosen, we move on to now looking at whether our base model satisfies our assumptions required for a linear mode. We also compared our results to a transformed model where we take the log of our outcome variable Streams.

Residual Models

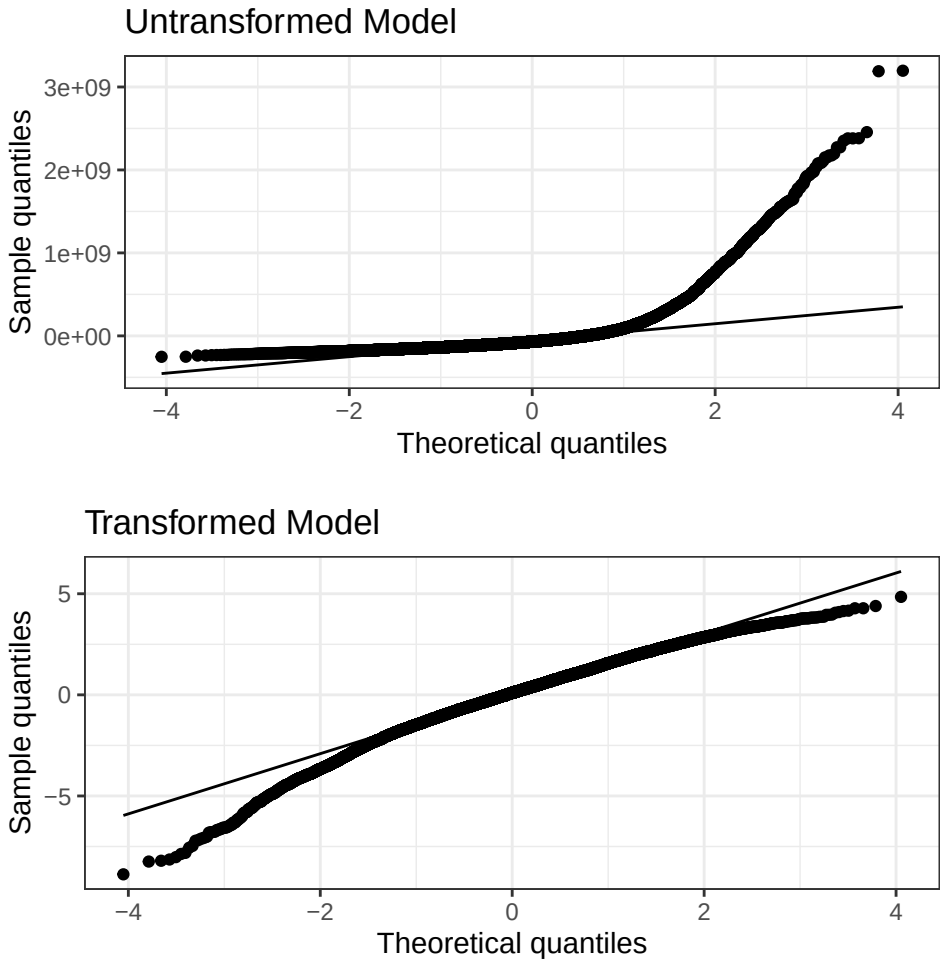




Looking at the visualizations above, we can see that the transformed model gives us a much better spread on the residual split around our red line then our untransformed model. As such, the residuals appear roughly symmetrical along the horizontal axis for our transformed plot, so we feel it safe to assume approximate linearity, specifically for our transformed model.

As it relates to constant variance, we believe that our fitted values for our transformed model seem to satisfy this condition. Other than a few outliers in our negative residual side on the right, overall we seem to see fairly constant trends with how spread out our data is.

QQ Plots



Now looking at our qq plots, we see our trend continue, where our untransformed model performs quite bad as can be seen above, while our transformed model hangs much closer to our standardized line, making it a better fit. Here, we feel safe to assume normality for our transformed plot, as other then some slight deviation towards the tails, our data points hang tight to the normal line.

For our independence assumption, we believe that our data set satisfies this condition. Our data was collected all on the same day, and the streams and/or variables associated with each song should not be impacted by those of other songs.

After looking at our two graphs above, we believe that our transformed model will provide a better measurement of our data, and provide better predicitons of a songs streams.

Results

Our final model is log(Stream) ~ Danceability + Energy + factor(Key) + Loudness + Speechiness + Acousticness + Instrumentalness + Liveness + Valence + Tempo + Duration_ms + official_video

```
Call:
lm(formula = log(Stream) ~ Danceability + Energy + factor(Key) +
    Loudness + Speechiness + Acousticness + Instrumentalness +
    Liveness + Valence + Tempo + Duration_ms + official_video,
    data = music)

Residuals:
    Min      1Q  Median      3Q     Max
-8.8753 -0.9352  0.0965  1.0739  4.8471

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.894e+01  1.299e-01 145.746  < 2e-16 ***
Danceability        2.433e-01  8.726e-02   2.788   0.0053 **
Energy             -1.088e+00  9.862e-02 -11.029  < 2e-16 ***
factor(Key)1       -2.984e-03  4.889e-02  -0.061   0.9513
factor(Key)2       -7.841e-02  4.975e-02  -1.576   0.1150
factor(Key)3        1.597e-02  7.173e-02   0.223   0.8238
factor(Key)4        5.302e-03  5.400e-02   0.098   0.9218
factor(Key)5       -1.527e-02  5.186e-02  -0.294   0.7684
factor(Key)6        2.348e-02  5.487e-02   0.428   0.6688
factor(Key)7       -5.105e-02  4.831e-02  -1.057   0.2907
factor(Key)8        1.707e-02  5.428e-02   0.314   0.7532
factor(Key)9       -7.646e-02  4.992e-02  -1.532   0.1257
factor(Key)10      -8.253e-02  5.508e-02  -1.498   0.1341
factor(Key)11      -2.742e-02  5.258e-02  -0.522   0.6020
Loudness            6.579e-02  4.423e-03  14.877  < 2e-16 ***
Speechiness        -2.041e+00  1.125e-01 -18.147  < 2e-16 ***
Acousticness       -5.650e-01  5.492e-02 -10.289  < 2e-16 ***
Instrumentalness   -4.900e-01  7.376e-02  -6.643 3.16e-11 ***
Liveness           -3.855e-01  7.100e-02  -5.429 5.73e-08 ***
Valence            -2.931e-01  5.728e-02  -5.117 3.13e-07 ***
Tempo               1.109e-03  3.951e-04   2.807   0.0050 **
Duration_ms         1.507e-07  9.032e-08   1.669   0.0951 .
official_videoTRUE  2.820e-01  2.771e-02  10.174  < 2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.591 on 19668 degrees of freedom
Multiple R-squared:  0.06768,   Adjusted R-squared:  0.06664
F-statistic:  64.9 on 22 and 19668 DF,  p-value: < 2.2e-16


Linear Regression

19691 samples
    2 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 17723, 17722, 17722, 17721, 17722, 17722, ...
Resampling results:

  RMSE       Rsquared    MAE
  243735229  0.01524967  141514771

Tuning parameter 'intercept' was held constant at a value of TRUE


Linear Regression

19691 samples
   12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 17720, 17721, 17722, 17723, 17721, 17723, ...
Resampling results:

  RMSE       Rsquared    MAE
  240763712  0.03747003  139822822

Tuning parameter 'intercept' was held constant at a value of TRUE


Linear Regression

19691 samples
   12 predictor
```

```
No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 17721, 17723, 17721, 17722, 17723, 17723, ...
Resampling results:

  RMSE     Rsquared    MAE
  1.59216  0.06613585  1.238856

Tuning parameter 'intercept' was held constant at a value of TRUE
```

Table 4: Model Fit Statistics

Model	R_squared	RMSE
Hypothesis_Model	0.0152497	2.437352e+08
All_Model	0.0374700	2.407637e+08
Transform_Model	0.0661358	1.592160e+00

Looking at our final model, we'll just look at a couple of our important hypothesized variables.

For Danceability, we have a coefficient of .2433. In this context, this means that for every 1 unit increase in the measure of danceability, we would predict an increase of .2433 in the log of the number of Streams for a given song holding all else constant. Looking at our p-value for our Danceability as well, our p-value compared to an alpha level of .05 is less then it, and as such danceability appears to be a strong predictor of streams.

For Loudness, we have a coefficient of .06579. In this context, this means that for every 1 unit increase in the measure of loudness, we would predict an increase of .2433 in the log of the number of Streams for a given song holding all else constant. Looking at our p-value for our loudness as well, our p-value compared to an alpha level of .05 is less then it, and as such loudness appears to be a strong predictor of streams.

Based on our model fit statistics, our final model with the log-transformed response variable and 12 predictors has the highest r-squared value of 0.0661358 and the lowest RMSE value of 1.592160. In addition, the un-transformed model with 12 predictors performed better than our hypothesized model using only Danceability and Loudness. Thus, we find that the model with the most predictive power included all 12 potential predictors and also required a transformation. We were correct in hypothesizing that Danceability and Loudness would be significant predictors. The final model tells us that the danceability, energy, key, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and duration of a song in addition to whether the music video was official, are all significant in predicting the success of a song measured in number of streams.

Appendix