Popularity Determined by Audio Features in a Song

Abstract

Music has always been an important focus of my life. I was in band all throughout middle and high school. I am constantly learning about new music theory and instruments. What better way to integrate my passion with songs then by doing analysis on a musical dataset? The purpose of this dataset is to create a model that will predict the popularity of a song given, its acousticness, danceability, duration in milliseconds, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time signature, and valence. I used a variety of multi linear regression model with different coefficients to determine which model would best predict popularity.

Introduction

This study summarizes all the primary statistical modeling and analysis results associated with what Spotify determines is the popularity of a song. This analysis will also look at whether acousticness, danceability, duration in milliseconds, energy, instrumentalness, key, liveness, loudness, mode, speechiness, tempo, time signature, and valence influences popularity. Of course any musician wants their song to be popular and heard throughout the world so by narrowing down a model future musicians may pass their song through and potentially determine if their song will be popular. I have no hypothesis on which predictor variable will be the most significant.

Description of Data

I found this dataset on Kaggle and it was mined using the Spotify API. The audio features are

Feature	Description
Duration ms	The duration of the track in milliseconds.
Key	The estimated overall key of the track.
Mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived.
Time signature	An estimated overall time signature of a track
Acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
Danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.

Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.
Instrumentalness	Predicts whether a track contains no vocals.
Liveness	Detects the presence of an audience in the recording.
Loudness	The overall loudness of a track in decibels (dB).
Speechiness	Speechiness detects the presence of spoken words in a track.
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)
Popularity	How popular it is with 100 being max.

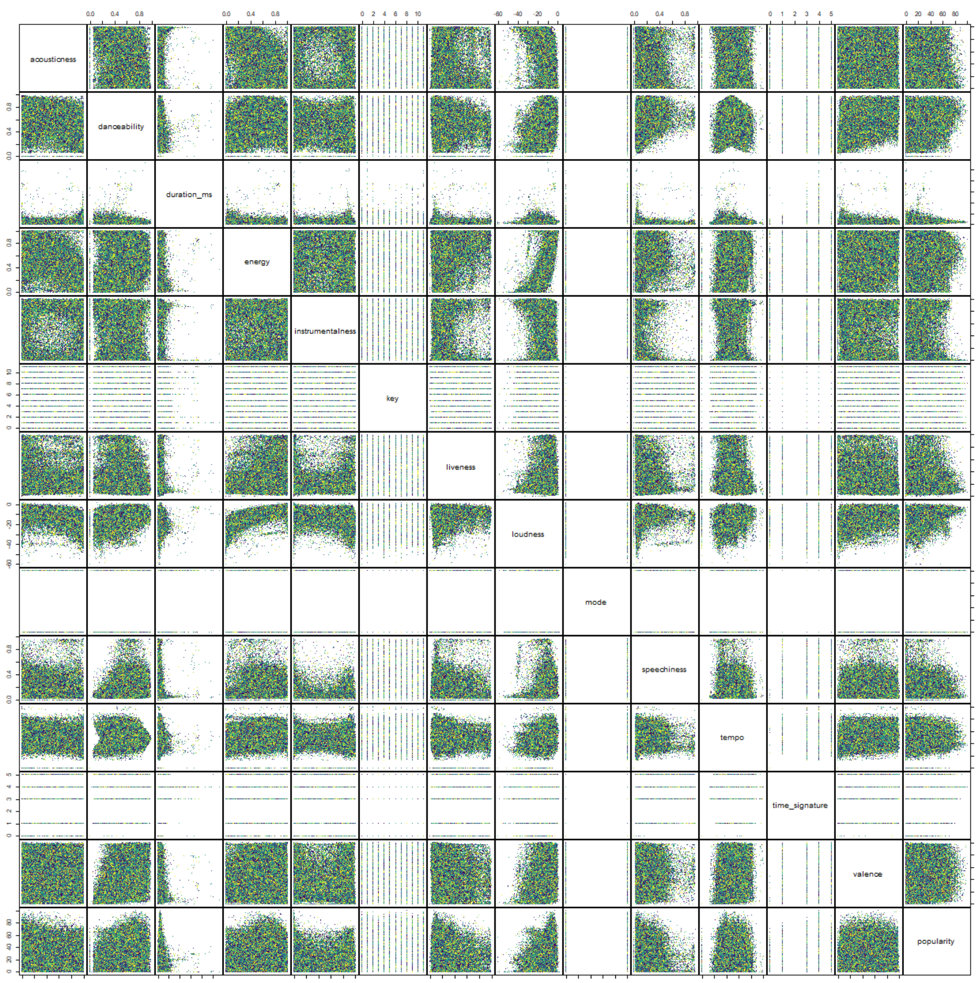
DESCRIPTION OF STATISTICAL METHODS AND RESULTS

Looking at this chart of the basic data.

Feature	**Min**	**Max**	**Median**	**Mean**
Acousticness	0.0000	0.9960	0.2030	0.3425
Danceability	0.0000	0.9960	0.6050	0.5815
Duration_Ms	3203.0000	5610020.0000	201901.0000	212633.0000
Energy	0.0000	1.0000	0.6030	0.5692
Instrumentalness	0.0000	1.0000	0.0001	0.2240
Key	0.0000	11.0000	5.0000	5.2320
Liveness	0.0000	0.9990	0.1240	0.1949
Loudness	-60.0000	1.8060	-7.9790	-9.9740

Mode	0.0000	1.0000	1.0000	0.6077
Speechiness	0.0000	0.9660	0.0559	0.1120
Tempo	0.0000	249.9800	120.0300	119.4700
Time_Signature	0.0000	5.0000	4.0000	3.8790
Valence	0.0000	1.0000	0.4200	0.4396
Popularity	0.0000	100.0000	22.0000	24.2100

As we can see there are a wide variety of values for each variable with the mean for most being generally in the middle except for liveness, instrumentalness, and acousticness. This suggests not a lot of songs are instrumental based.



This matrix plot shows there is not much of a relationship between any variable except maybe a slight upward quadratic between energy and loudness.

Acousticness	Danceability	Duration Ms	Energy	Instrumentalness	Key	Liveness	Loudness	Mode	Speechiness	Tempo	Time Signature	Valence
-0.1165	0.1311	-0.0008	0.1225	-0.2164	0.0027	-0.0312	0.2441	-0.0091	-0.0002	0.0371	0.0649	0.0143

The correlation between popularity and all the variables is not very high. The strongest positive relationships are loudness, danceability, energy. The strongest negative relationship are instrumentalness, acousticness, liveness.

```
Call:
lm(formula = popularity ~ ., data = spotifyCSVnums)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.868 -15.095  -1.858  12.737  74.193 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       3.584e+01  6.334e-01  56.580  < 2e-16 ***
acousticness      7.491e-01  2.253e-01   3.324 0.000887 ***
danceability      5.451e+00  3.622e-01  15.049  < 2e-16 ***
duration_ms      -8.402e-07  4.308e-07  -1.951 0.051118 .  
energy           -6.612e+00  3.880e-01 -17.039  < 2e-16 ***
instrumentalness -7.003e+00  1.762e-01 -39.741  < 2e-16 ***
key              -8.465e-03  1.472e-02  -0.575 0.565146    
liveness         -2.878e+00  3.308e-01  -8.700  < 2e-16 ***
loudness          7.896e-01  1.485e-02  53.180  < 2e-16 ***
mode             -1.717e-01  1.092e-01  -1.573 0.115796    
speechiness      -5.963e+00  4.499e-01 -13.255  < 2e-16 ***
tempo            -2.704e-03  1.793e-03  -1.508 0.131587    
time_signature    8.088e-01  1.046e-01   7.729 1.09e-14 ***
valence          -7.067e+00  2.397e-01 -29.481  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.84 on 130649 degrees of freedom
Multiple R-squared:  0.08705,	Adjusted R-squared:  0.08696 
F-statistic: 958.3 on 13 and 130649 DF,  p-value: < 2.2e-16
```

Our full linear model shows an adjusted R-squared of 0.8696 which means that about 87% of the variation can be explained by this model that includes all the predictors. Some ways we could improve this model are perhaps removing the insignificant variables.

```
Call:
lm(formula = popularity ~ ., data = lm2)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.910 -15.084  -1.874  12.729  74.124 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       35.1270     0.5745  61.145  < 2e-16 ***
acousticness       0.7627     0.2246   3.396 0.000684 ***
danceability       5.5607     0.3603  15.433  < 2e-16 ***
energy            -6.6158     0.3871 -17.090  < 2e-16 ***
instrumentalness  -6.9978     0.1761 -39.740  < 2e-16 ***
liveness          -2.8486     0.3303  -8.625  < 2e-16 ***
loudness           0.7867     0.0148  53.160  < 2e-16 ***
speechiness       -5.9053     0.4480 -13.182  < 2e-16 ***
time_signature     0.7962     0.1045   7.623 2.49e-14 ***
valence           -7.0609     0.2377 -29.704  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.84 on 130653 degrees of freedom
Multiple R-squared:  0.08699,	Adjusted R-squared:  0.08693 
F-statistic:  1383 on 9 and 130653 DF,  p-value: < 2.2e-16
```

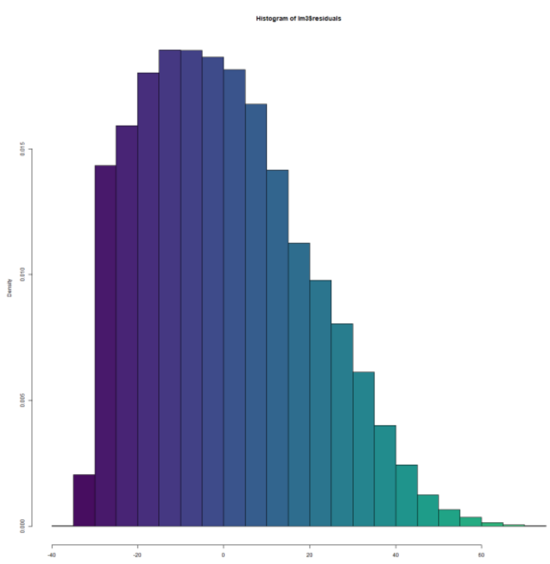
However, it seems that doing so worsened the model. Another attempt at improving the model could be to remove the variables with the lowest absolute correlation values.

```
Call:
lm(formula = popularity ~ ., data = lm2)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.966 -15.101  -1.879  12.792  72.817 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      36.163708   0.618575  58.463  < 2e-16 ***
acousticness      0.764781   0.225424   3.393 0.000692 ***
danceability      4.426231   0.352321  12.563  < 2e-16 ***
energy           -7.004398   0.387121 -18.094  < 2e-16 ***
instrumentalness -6.534236   0.172752 -37.824  < 2e-16 ***
liveness         -3.395957   0.328539 -10.337  < 2e-16 ***
loudness          0.820825   0.014654  56.013  < 2e-16 ***
mode             -0.105333   0.107456  -0.980 0.326969    
tempo            -0.003800   0.001793  -2.120 0.034006 *  
time_signature    0.796744   0.104625   7.615 2.65e-14 ***
valence          -6.968325   0.238427 -29.226  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.85 on 130652 degrees of freedom
Multiple R-squared:  0.08582,	Adjusted R-squared:  0.08575 
F-statistic:  1226 on 10 and 130652 DF,  p-value: < 2.2e-16
```

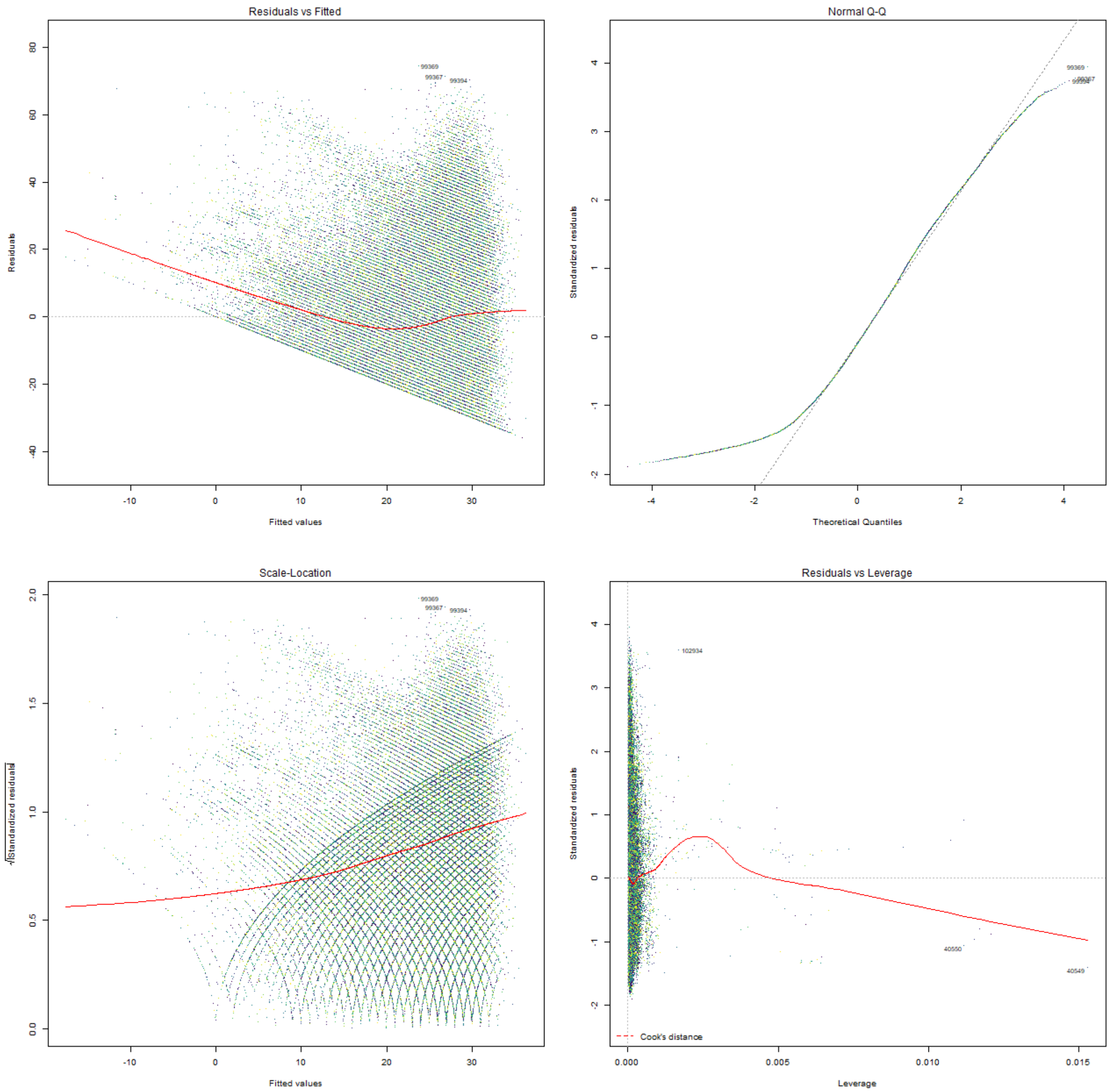
This also seems to not have helped with our model. Trying a step by step removal of all individual variables I have found that removing key led to the only increase in the model.

```
Call:
lm(formula = popularity ~ ., data = lm3)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.841 -15.095  -1.857  12.740  74.233 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       3.579e+01  6.281e-01  56.982  < 2e-16 ***
acousticness      7.477e-01  2.253e-01   3.318 0.000906 ***
danceability      5.453e+00  3.622e-01  15.054  < 2e-16 ***
duration_ms      -8.414e-07  4.308e-07  -1.953 0.050803 .  
energy           -6.615e+00  3.880e-01 -17.051  < 2e-16 ***
instrumentalness -7.002e+00  1.762e-01 -39.737  < 2e-16 ***
liveness         -2.878e+00  3.308e-01  -8.702  < 2e-16 ***
loudness          7.896e-01  1.485e-02  53.185  < 2e-16 ***
mode             -1.606e-01  1.075e-01  -1.495 0.134987    
speechiness      -5.962e+00  4.499e-01 -13.252  < 2e-16 ***
tempo            -2.703e-03  1.793e-03  -1.507 0.131699    
time_signature    8.088e-01  1.046e-01   7.729 1.09e-14 ***
valence          -7.072e+00  2.395e-01 -29.523  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.84 on 130650 degrees of freedom
Multiple R-squared:  0.08705,	Adjusted R-squared:  0.08697 
F-statistic:  1038 on 12 and 130650 DF,  p-value: < 2.2e-16
```

With this model we can figure out the residuals and histogram of residuals.



This histogram is most definitely not normal with a right skew. This shows that our residuals are predicting too high.



Looking at these plots we can see the residuals follow the QQ plot but are very weird otherwise.

CONCLUSION

With the model we have reduced to. We can predict with 86% the variation in popularity given 12 audio features. The model itself is significant even though some of the variables with in it are not. Removing key was the only predictor variable that helped improve the model's prediction. Performing individual linear models with each variable show that any one model is not good enough to predict popularity. This is most likely owed to humans' taste in different music. With a large enough set almost, any music can be popular with wildly different audio features. If I had more variables such as times stream, amount of time in top played, genre, production cost, maybe even amount of words in song. I could potentially make a model that is even more accurate than now.