more_exploration

February 23, 2022

1 Executive Summary

1. Across all available metrics of success, Spotify-created playlists are much more successful than nearly all user-created playlists.
2. We consider nine metrics of playlist success, including measures of total streams and total users over the past day, week, month and two months. Amongst both the user-created and Spotify-created playlists, these measures of success are very highly correlated with one another, thus suggesting any choice of these success metrics would likely be sufficient.
6. Across both Spotify-created and user-created playlists, the number of tracks is the most important predictor of short-term and long-term playlist success, with longer playlists tending to be more successful.
7. For Spotify-created playlists, having greater musical diversity (as measured by tracks per album and tracks per artist) is associated with increased success. This effect is present for the user-created playlists, but is less pronounced.
8. Greater musical diversity is more important for a playlist's long-term staying power than for short-term success, with one of the musical diversity metrics being the second most important predictor (behind total tracks) for the long-term success metric.
10. For Spotify-created playlists, the primary genres of 'dance & house' and 'pop' are associated with both short-term and long-term success.
11. The most success-inducing primary genres for Spotify-created and user-created vary considerably. For user-created playlists, "Children's", "Latin" and "Traditional" increase both short- and long-term success.

2 Outline

1. Explore massive differences in playlist success between the Spotify-created and user-created playlists.
2. Establish the substantial correlation between each of the measures of success across both Spotify-created and user-created playlists will next be discussed. A principal components analysis will be performed.
3. Briefly present the distribution of the predictors of success, including both the numeric (number of tracks, tracks per artist and tracks per album) as well as the categorical (the mood and genre variables).

4. Fit gradient boosting models to predict measures of both short-term and long-term success, stratifying by Spotify-created and user-created playlists. Use SHAP values to assess the inferential implactions of the models.

5. The Appendix contains a number of other subanalyses and assumption checks.

3 Introduction

The data under consideration for these analyses consists of 403,366 distinct playlists, with 314,899 distinct playlist owners. Of the 314,899 unique playlist owners, 261,040 (83%) have exactly one playlist in the data. Of the owners with more than one playlist, Spotify itself has the most, with 399. The data is composed of only playlists from US owners, and thus extrapolating any of these analyses to other countries is likely unwarranted or should be done with great caution. Each playlist is categorized by its top three genres and top three moods. There are 26 genres and 27 moods under consideration.

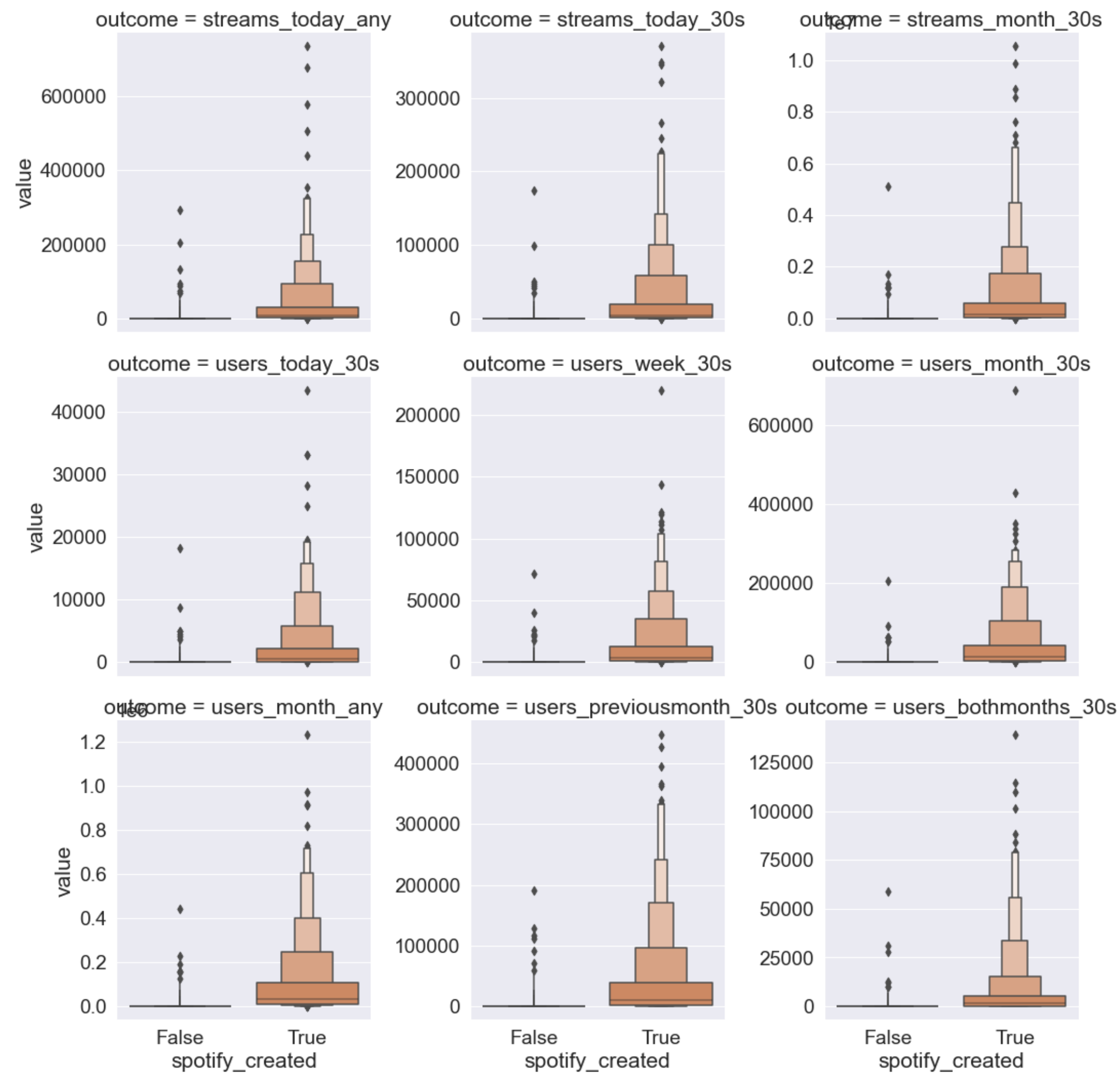
There are a number of potential measures of playlist success included in this dataset. Specifically, we have (1) the number of streams of any length from the playlist today, (2) the number of streams greater than 30 seconds today, (3) the number of active users today, where an active user is defined as having a stream > 30 seconds, (4) the number of active users in the past week, (5) the number of active users in the past month, (6) the number of users who had a stream from this playlist for any length of time in the past month, (7) the number of active users in the previous month, (8) the total number of > 30 second streams in the past month, (9) the number of users who were active this month and the previous month. The data also contains a variable signifying the number of > 30 second streams by the playlist owner in the past month, but this analysis will focus on broader measures of success, and thus we will not analyze this variable. Finally, the data also includes the number of users who skipped more than 90% of their total streams today who also used this playlist. This variable does not directly measure the proportion of skips for each playlist, but instead measures simply whether users who skip often listen to this playlist. Thus variable will thus also not be included in these analyses.

Some of the potential predictors of stream success include: (1) the number of tracks in the playlist, (2) the number of tracks that were added to the playlist today, (3) The number of unique artists in the playlist, (4) the number of unique albums in the playlist, (5-7) the first, second and third most common genre found in the playlist, (8-10) the first, second and third most common mood found in the playlist. While number of tracks, artists and albums are each measures of playlist size, the most direct measure is number of tracks, so we will thus use it directly in the analyses. We will use number of tracks per artist and number of tracks per album as inverse measures of musical diversity. The dataset also contains unstructured tokens associated with each playlist. Due to the time constraints for these analyses and the unstructured nature of the data, these tokens will not be considered in the analyses.

4 Comparing Spotify-Created and User-Created Playlists

There are two Spotify-created playlists that constitute extreme outliers across each of the potential outcome variables. The first is a pop, Dance & House, Indie Rock playlist with 100 tracks and has tokens ‘top’, ‘tracks’, ‘currently’, ‘spotify’. The second is a pop, R&B, Dance & House with 51 tracks and has tokens ‘top’, and ‘hits’. These two playlists have more than three times as many streams today as their nearest competitor and more than four times as many > 30 second streams in the past month as the nearest competitor. For the purposes of plotting, these playlists will be removed.

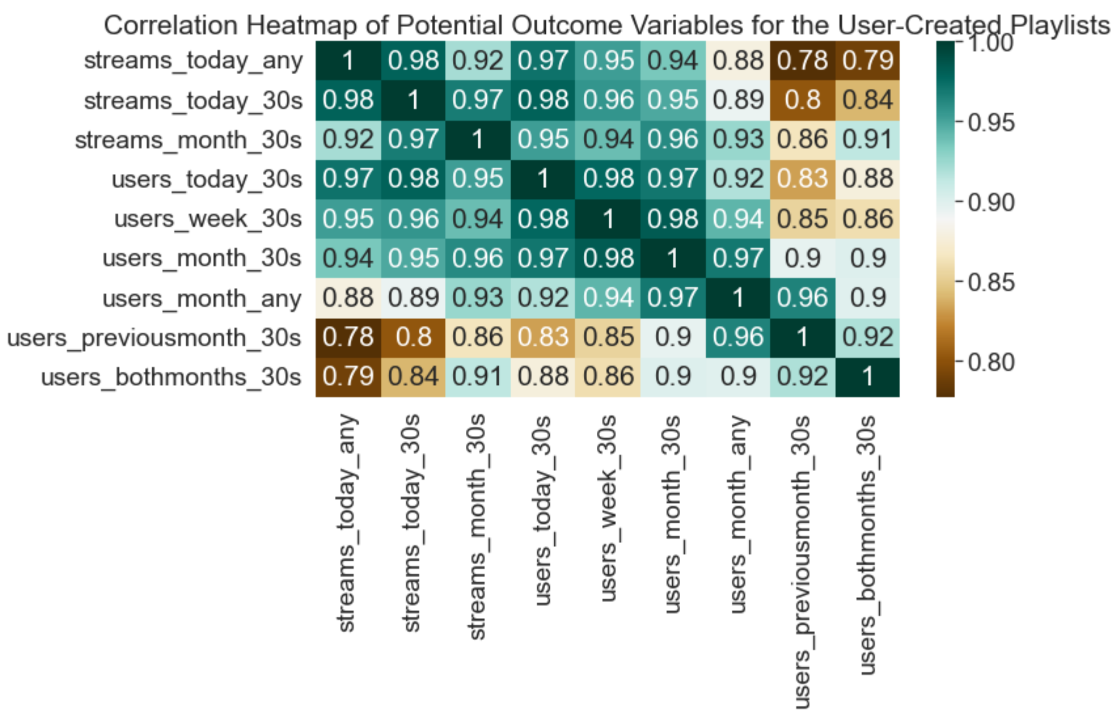
Even after removing the two most successful Spotify-created playlists, there is still a wide gulf between the Spotify-created and user-created playlists. To illustrate this, consider the boxen plot below, which shows the distribution of each of the potential outcome variables stratified by whether the playlist is creator type. We observe that the Spotify-created playlists are much more successful than the vast majority of user-created playlists. However, there are a few user-created playlists that can rival an average Spotify-created playlist.



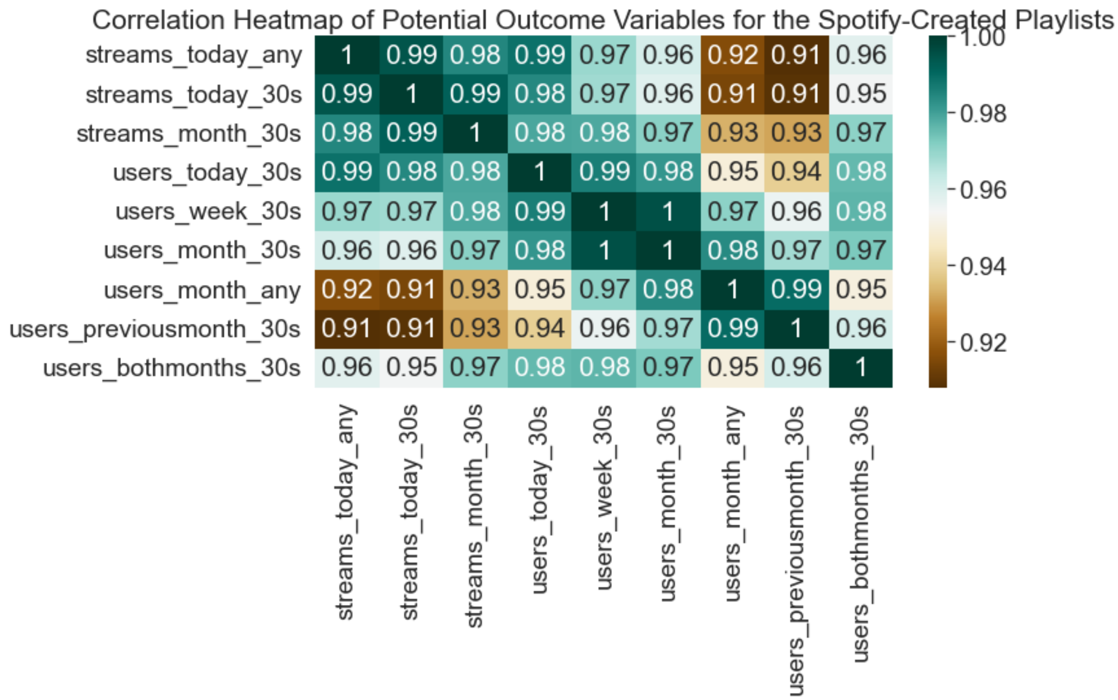
5 Exploration of Potential Outcomes

Because there is such an enormous difference between Spotify-created and user-created playlists, we will continue our data exploration stratifying by whether the playlist is Spotify-created.

To begin understanding the relationship between the potential outcomes of interest, we present a heatmap of the Pearson correlation between each of the outcomes. We create one such heatmap for the user-created playlists and another for the Spotify-created playlists. Amongst the user-created playlists, we see that the minimum correlation between any of the outcomes is 0.78, thus signifying a great deal of correlation between our potential outcomes. While all potential outcomes are highly correlated, the mostly weakly correlated outcomes are the outcomes related to playlists' longer-term success (i.e. the monthly average users in the given month and the previous month) with the more recent measures of success (i.e. the number of total and >30 second streams today and the number of active users today and in the past week). Amongst the Spotify-created playlists, the potential outcomes are even more highly correlated, with the smallest correlation being 0.91. The Spotify-created playlists exhibit a similar general pattern as the user-created in that the weakest correlation between the outcomes is between the more long-term measures of success and the measures of success in the more recent past. However, the degree of correlation is still immense between monthly active users over the past two months and the number of streams that occurred today, thus indicating that Spotify playlists tend to have considerable 'staying power'. Of course, if more successful playlists in the past are algorithmically pushed to users, then this could become a self-fulfilling prophecy rather than a true indication of how 'intrinsically good' the playlist is.



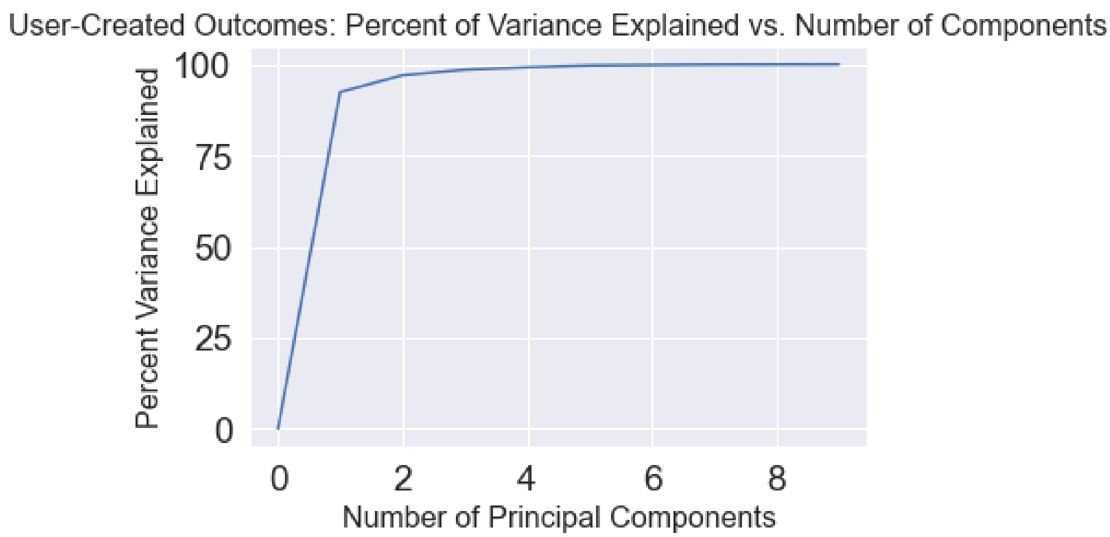
	streams_today_any	streams_today_30s	streams_month_30s	users_today_30s	users_week_30s	users_month_30s	users_month_any	users_previousmonth_30s	users_bothmonths_30s
streams_today_any	1	0.98	0.92	0.97	0.95	0.94	0.88	0.78	0.79
streams_today_30s	0.98	1	0.97	0.98	0.96	0.95	0.89	0.8	0.84
streams_month_30s	0.92	0.97	1	0.95	0.94	0.96	0.93	0.86	0.91
users_today_30s	0.97	0.98	0.95	1	0.98	0.97	0.92	0.83	0.88
users_week_30s	0.95	0.96	0.94	0.98	1	0.98	0.94	0.85	0.86
users_month_30s	0.94	0.95	0.96	0.97	0.98	1	0.97	0.9	0.9
users_month_any	0.88	0.89	0.93	0.92	0.94	0.97	1	0.96	0.9
users_previousmonth_30s	0.78	0.8	0.86	0.83	0.85	0.9	0.96	1	0.92
users_bothmonths_30s	0.79	0.84	0.91	0.88	0.86	0.9	0.9	0.92	1



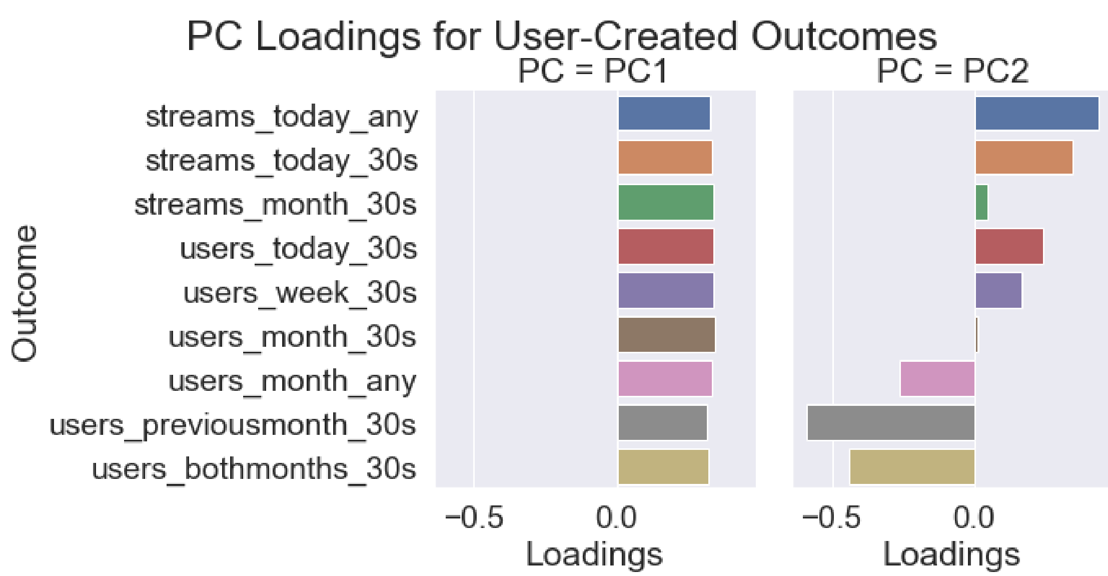
5.1 Principal Components Analysis of Potential Outcomes

We will continue our exploration of the relationship between the potential outcomes by performing a factor analysis. We observed that the correlation matrix for both the Spotify-created and user-created playlists is quite high. We would thus next like to address the question of to what degree can this information be reduced to one or a few factors, i.e. how much of the information encoded in these 9 potential outcomes is redundant. To help answer this question, we will employ a principal components analysis (PCA) on the user-created outcomes data. The results for the Spotify-created outcomes data are similar and can be found in the Appendix.

From the plot below, we see that the first principal component explains approximately 92% of the variation across the 9 potential outcomes, thus indicating that there is a great deal of redundancy in these measured outcomes. The second principal component explains another approximately 5% of variance. We can thus explain 97% of the variability of the 9 outcomes with 2 numbers (i.e. components).



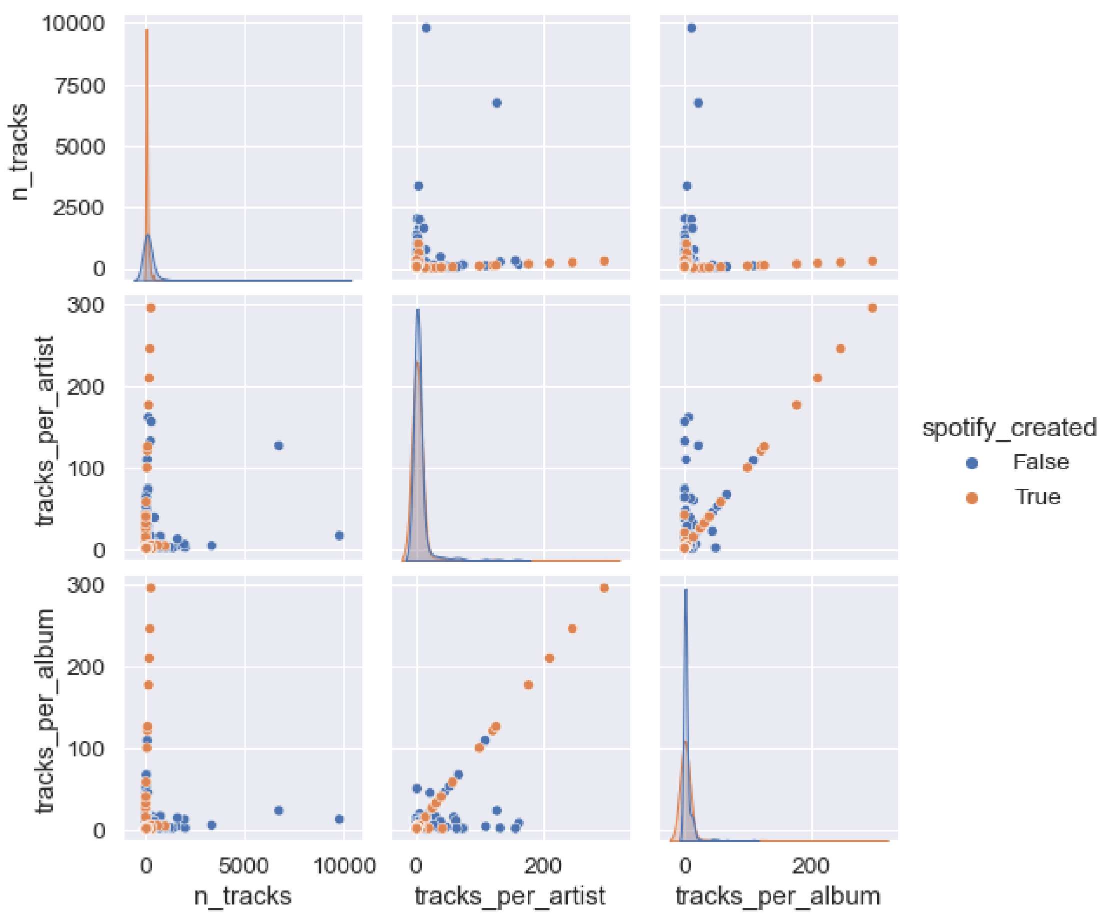
Let's next consider the loadings of each of the first two principal components. In the plot below, we see that each of the 9 outcomes are weighted very equally for the first PC. This is to be expected given the high level of correlation amongst the predictors. Looking at the second PC, we see that the number of users in the previous month and the number of users in past two months have a sizeable negative weighting while the total streams today variables have noticeable positive weighting. This indicates that there are some differences in the behavior of the "total users over the course of months" variables and the "total streams today" variables. We, however, should be careful not to overemphasize these differences, given the high correlation between the variables and the fact that the first PC explainds 92% of variability across the outcomes. Nevertheless, when we proceed to modeling playlist success, we will consider two outcomes: total users who had a stream in both months and total users who streamed today.



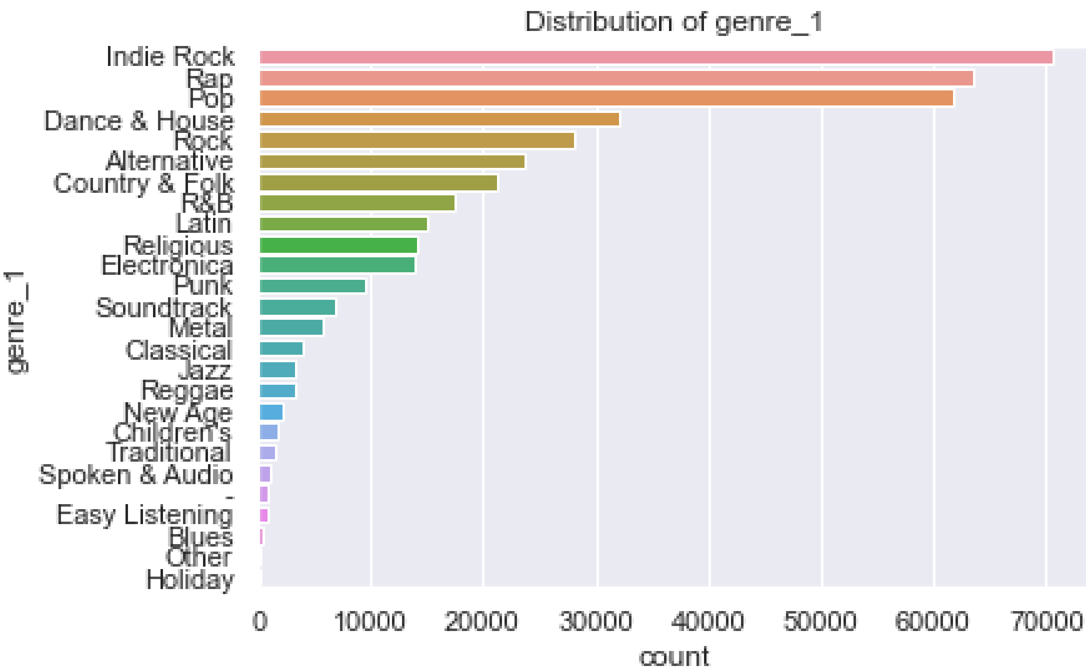
5.2 Exploration of Predictors of Interest

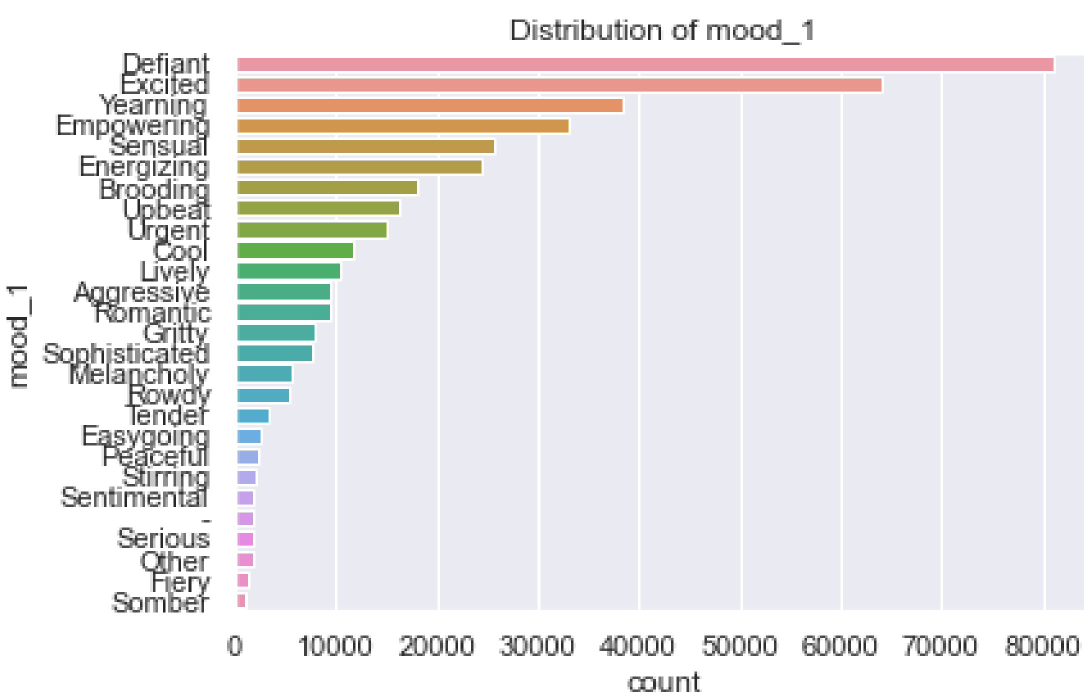
Note that for the plots below, we drew a random sample of user-created playlists for these plots in order for them not to visually swamp the Spotify-created playlists. Because the number of artists and albums in a playlist will tend to increase as the total tracks in a playlist increase, we define tracks per artist and tracks per album as scaled measures of the number of artists and albums that account for the overall size of the playlist. Also, note that the distribution of number of tracks, artists and albums does not vary between Spotify-created and user-created to the degree that the outcome variables do. However, we still observe some difference between user-created and Spotify-created playlists, namely that user-created playlists tend to have more tracks, artists and albums, with some user-created playlists having many more of each.

Let's consider the correlation of the new scaled measures of artists and albums as well as whether we observe differences between the user-created and Spotify-created playlists. Notice that the distribution of tracks per artist and tracks per album appear quite similar for the user-created and Spotify-created playlists. Spotify-created playlist appear to exhibit a little bit more artist and album diversity than the user-created playlists, but the difference is miniscule compared to the outcome variables.



We now turn our attention to describing the categorical predictors: genre and mood.





6 Predicting Playlist Success

We now turn our attention to predicting playlist success as measured by (1) the number of users who had a stream of greater than 30 seconds today (i.e. 'users_today_30s') and (2) the number of users who had a stream of greater than 30 seconds during the current month and the previous month (i.e. 'users_bothmonths_30s'). We choose 'users_today_30s' because amongst the user-created playlists, it very highly correlated with streams_today_any, streams_today_30s, streams_month_30s, users_week_30s, and users_month_30s (with the minimum correlation being 0.95). It is also less correlated with users_previousmonth_30s (0.83) and users_bothmonths_30s (0.88). We choose 'users_bothmonths_30s' because (1) it is less corelated than many of the other outcomes and (2) it is a good measure of a playlist's longer-term staying power. We build the models for both the Spotify-created and user-created playlist separately.

The modeling process that we employ is as follows. We train a lightgbm gradient boosting model using using tenfold cross validation. We employ early stopping so as to avoid overfitting and prevent wasting time and compute resources. The Huber loss function is used in order to reduce the effect that large outliers will have on the model. Based on the best number of trees to build as determined by cross validation, we then refit the model using all the data and fit the given number of trees. Because of the nature in which the lightgbm package handles categorical data, we do not have to one-hot encode the categorical genre or mood data, instead only having to input these as categorical columns in pandas. Note that all hyperparameters are left equal to their default values.

After fitting the models, we will then use SHAP values to help draw inference from the models. In particular, we will show plots of the most important predictors ranked by their mean absolute SHAP value. These plots will also highlight the directional effect of the predictor (i.e. the plots will showcase whether the given variable tends to be positively or negatively related to the outcome). We will also show plots for the categorical genre and mood variables that highlight which genres and moods tend to be related to higher and lower levels of playlist success.

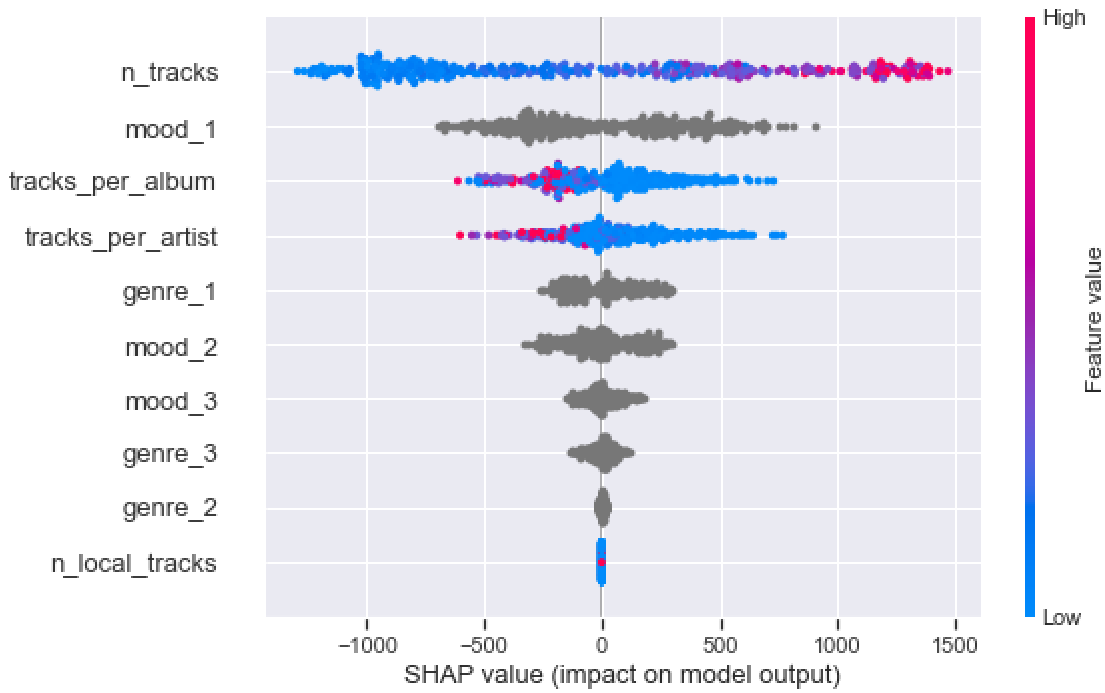
Deriving measures of inferential uncertainty from gradient boosting models can be challenging. Thus, we also fit a robust M-estimation linear regression with a Huber loss function. These modeling results are presented in the Appendix. For the academically curious, we also present the results for an Ordinary Least Squares (OLS) model to show the importance of handling outliers in a robust fashion.

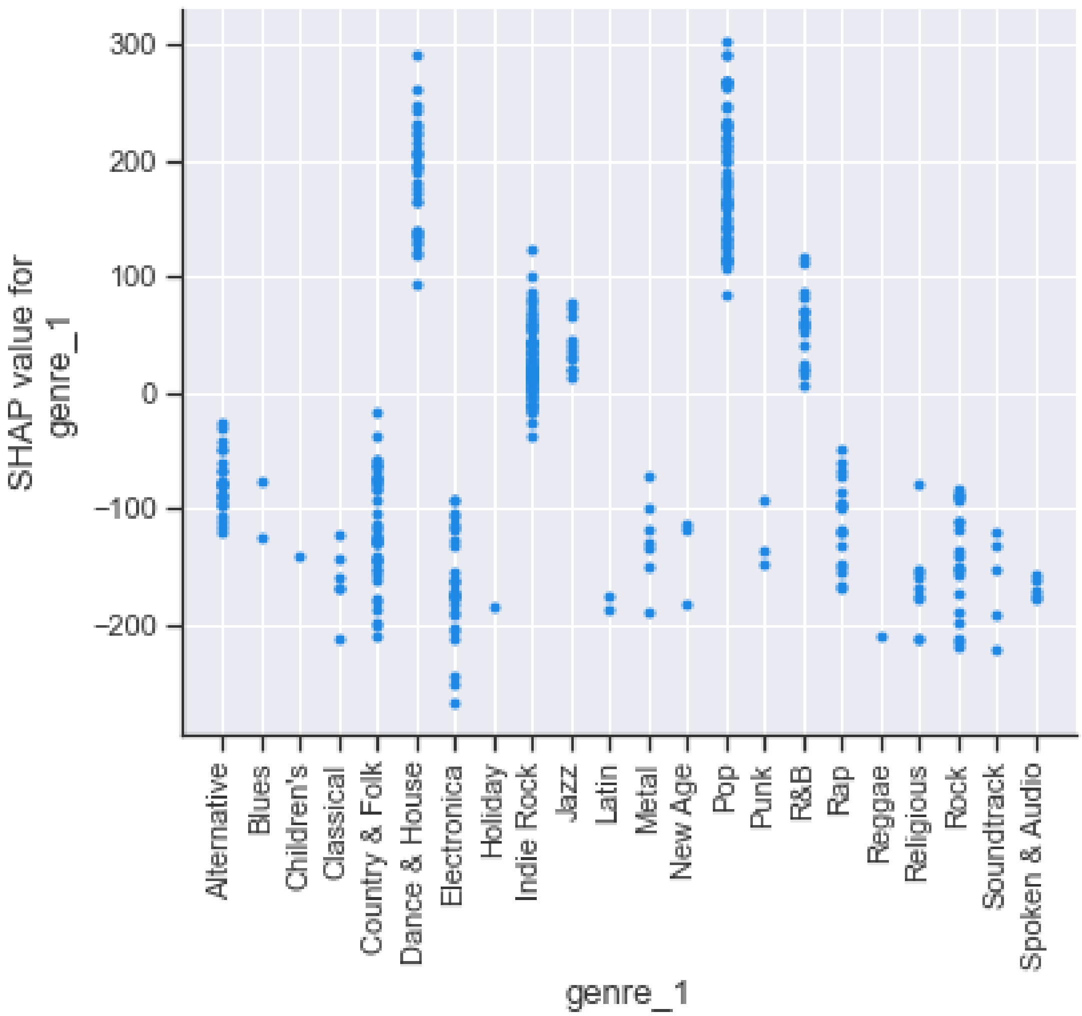
6.1 Spotify-Created Playlists: Modeling Results for Number of Users Today

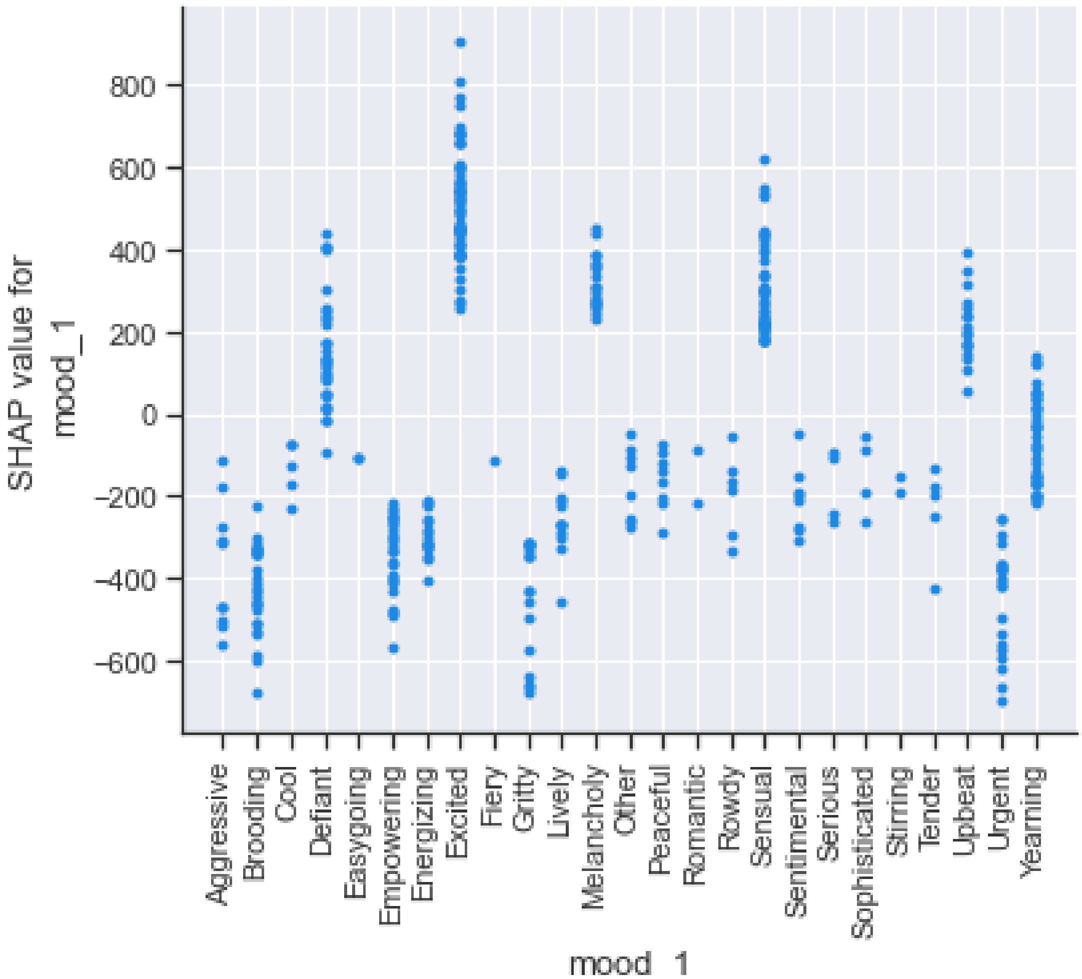
From the first plot below, we see that amongst the Spotify-created playlists, the most important variable predicting success is the number of tracks, followed by the primary mood, tracks per album and tracks per artist. Also, notice that large values of tracks (colored in pink) tend to be associated with larger SHAP values. Thus, we find that playlists with more tracks tend to have more users today. On the other hand, notice that for tracks per album and tracks per artist, the large values tend to be associated with negative SHAP values. Thus, large tracks per artist and tracks per album tend to be associated with reduced playlist success. Said another way, increasing the musical diversity of a playlist tends to increase playlist success as measured by number of users today.

The first plot does not easily allow us to understand individual levels of the categorical predictors of mood and genre. We thus plot SHAP values for each level of the primary genre and primary mood below. From the plot for primary genre, we can see that 'Dance and House' and 'Pop' music tends to be positively related to playlist success, while 'Country and Folk', 'Electronica' and 'Rock' tend to negatively related to success (again, as measured by number of users today).

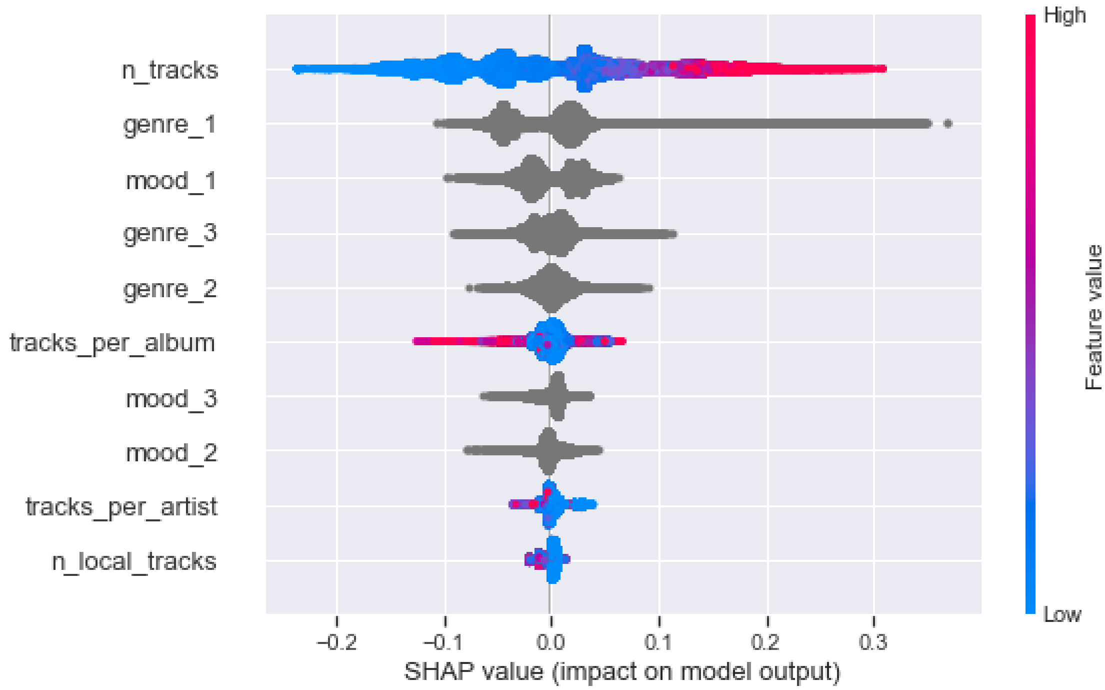
Note that due to a relatively limited sample size for Spotify playlists, all results should be taken with some skepticism. The results here should be thought of as more hypothesis-generating than hypothesis-confirming. Experimentation would be warranted to assess whether these observational-data-based results hold. With that said, the positive effect of both total tracks and

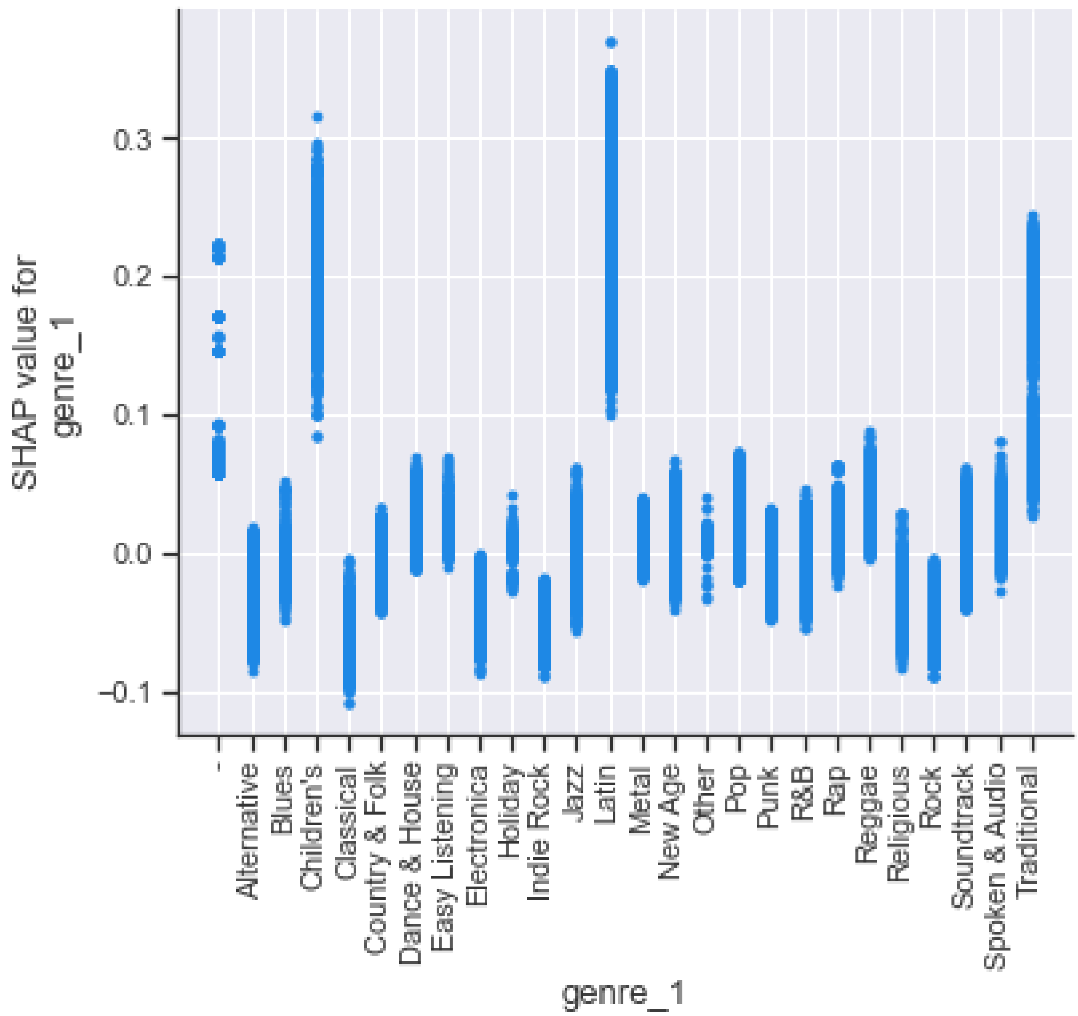



SHAP value for genre_1
300
200
100
0
−100
−200
Alternative
Blues
Children's
Classical
Country & Folk
Dance & House
Electronica
Holiday
Indie Rock
Jazz
Latin
Metal
New Age
Pop
Punk
R&B
Rap
Reggae
Religious
Rock
Soundtrack
Spoken & Audio
genre_1

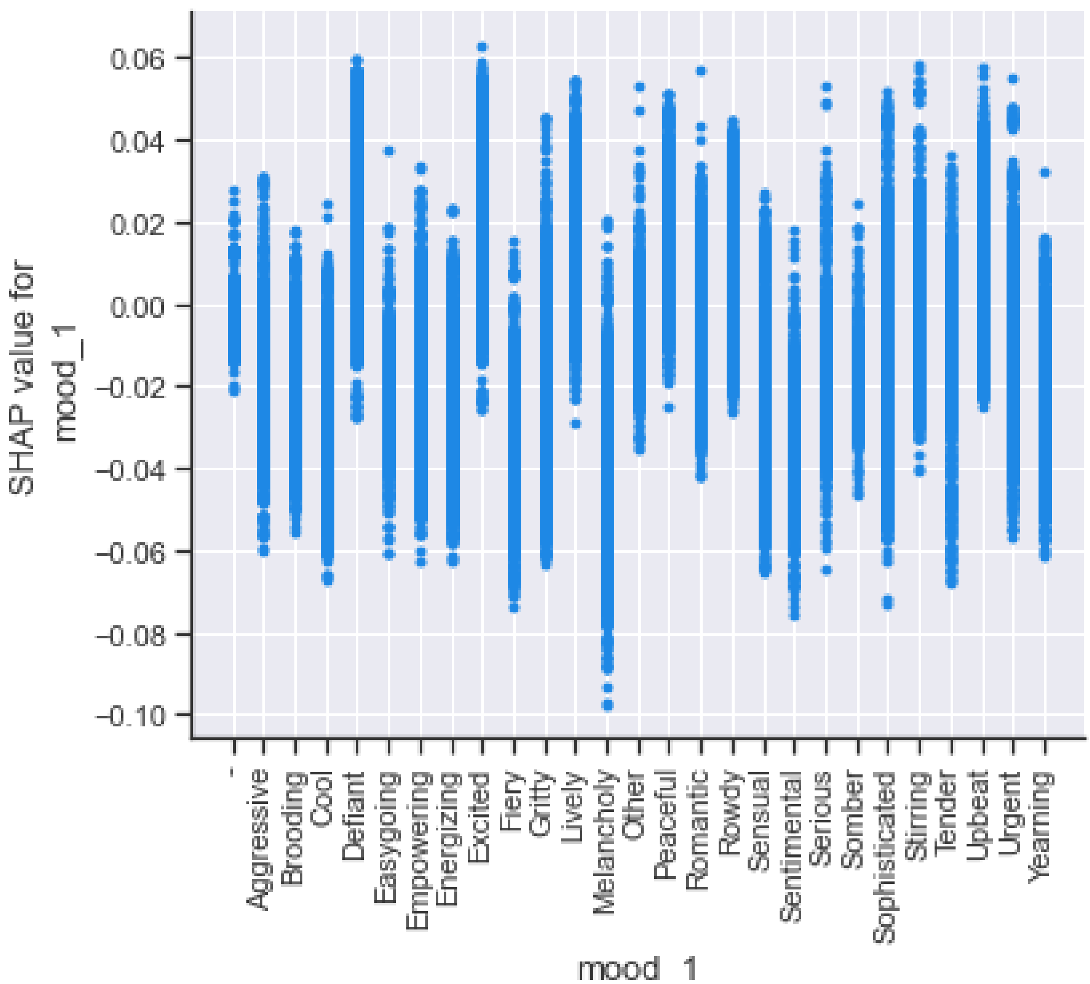

SHAP value for mood_1
800
600
400
200
0
−200
−400
−600
Aggressive
Brooding
Cool
Defiant
Easygoing
Empowering
Energizing
Excited
Fiery
Gritty
Lively
Melancholy
Other
Peaceful
Romantic
Rowdy
Sensual
Sentimental
Serious
Sophisticated
Stirring
Tender
Upbeat
Urgent
Yearning
mood_1

6.2 User-Created Playlists: Modeling Results for Number of Users Today



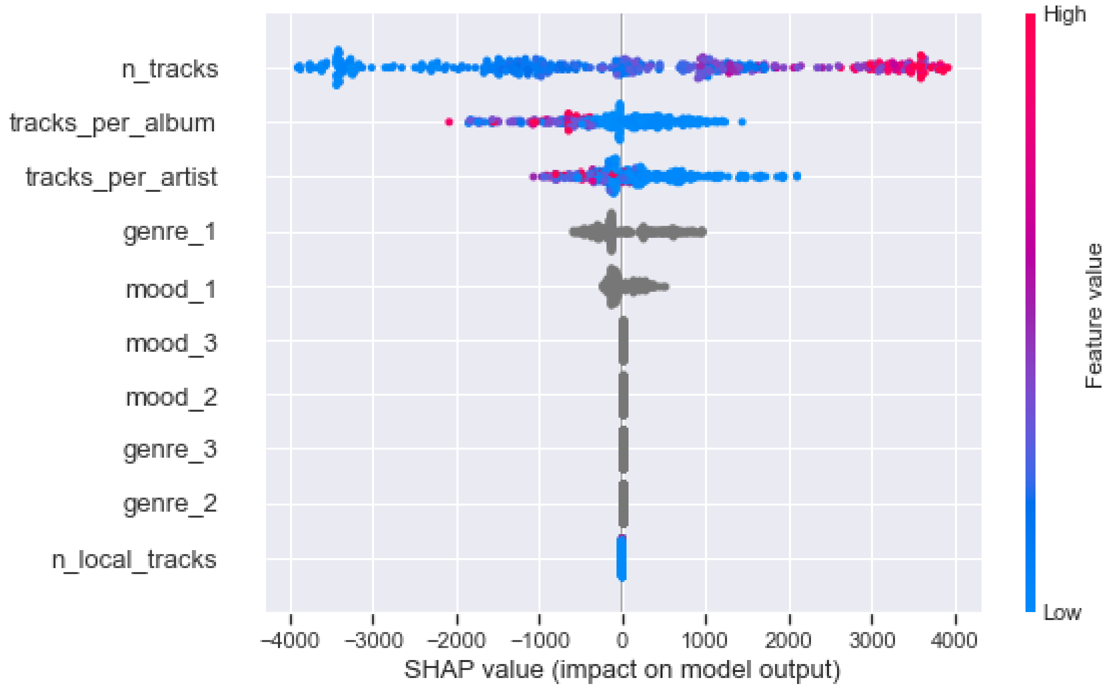


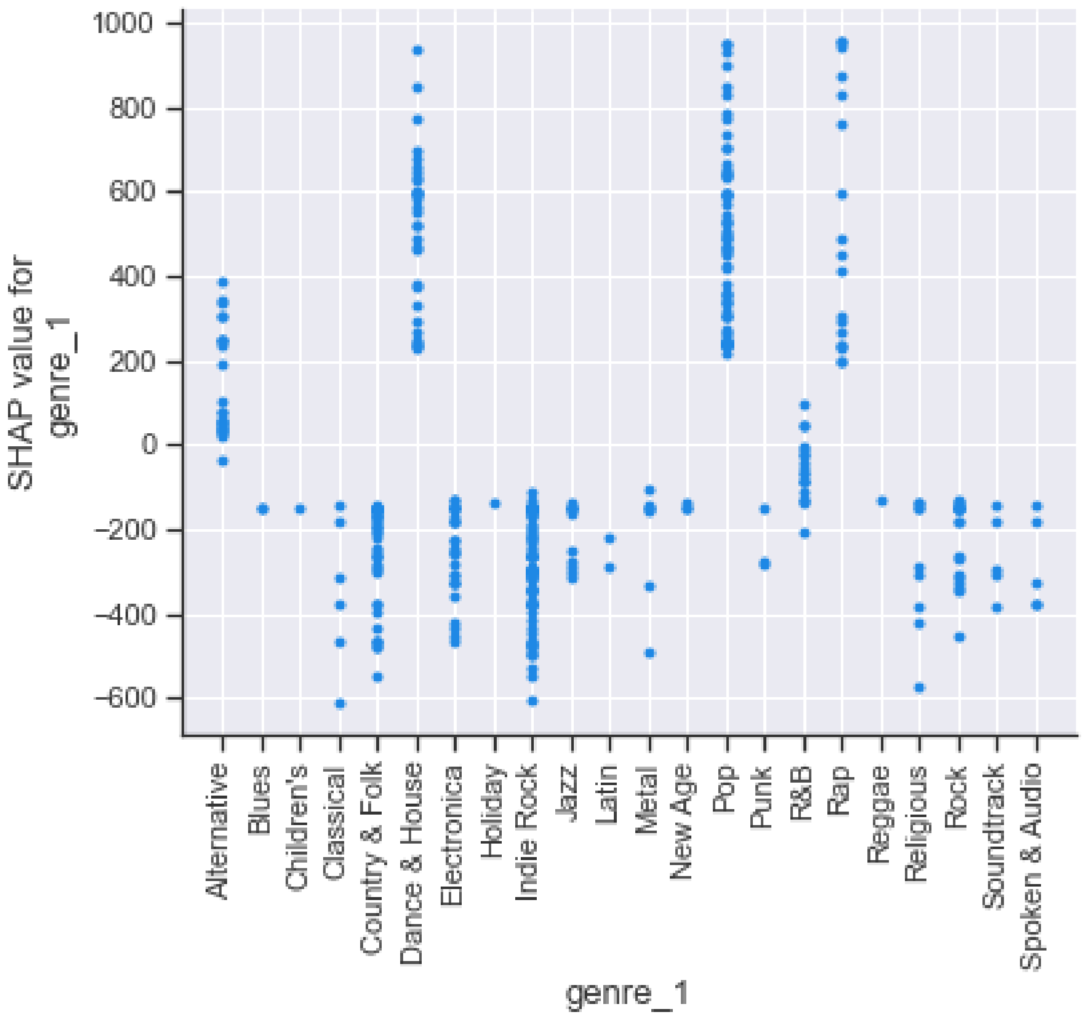
SHAP value for genre_1
0.3
0.2
0.1
0.0
−0.1
-
Alternative
Blues
Children's
Classical
Country & Folk
Dance & House
Easy Listening
Electronica
Holiday
Indie Rock
Jazz
Latin
Metal
New Age
Other
Pop
Punk
R&B
Rap
Reggae
Religious
Rock
Soundtrack
Spoken & Audio
Traditional
genre_1

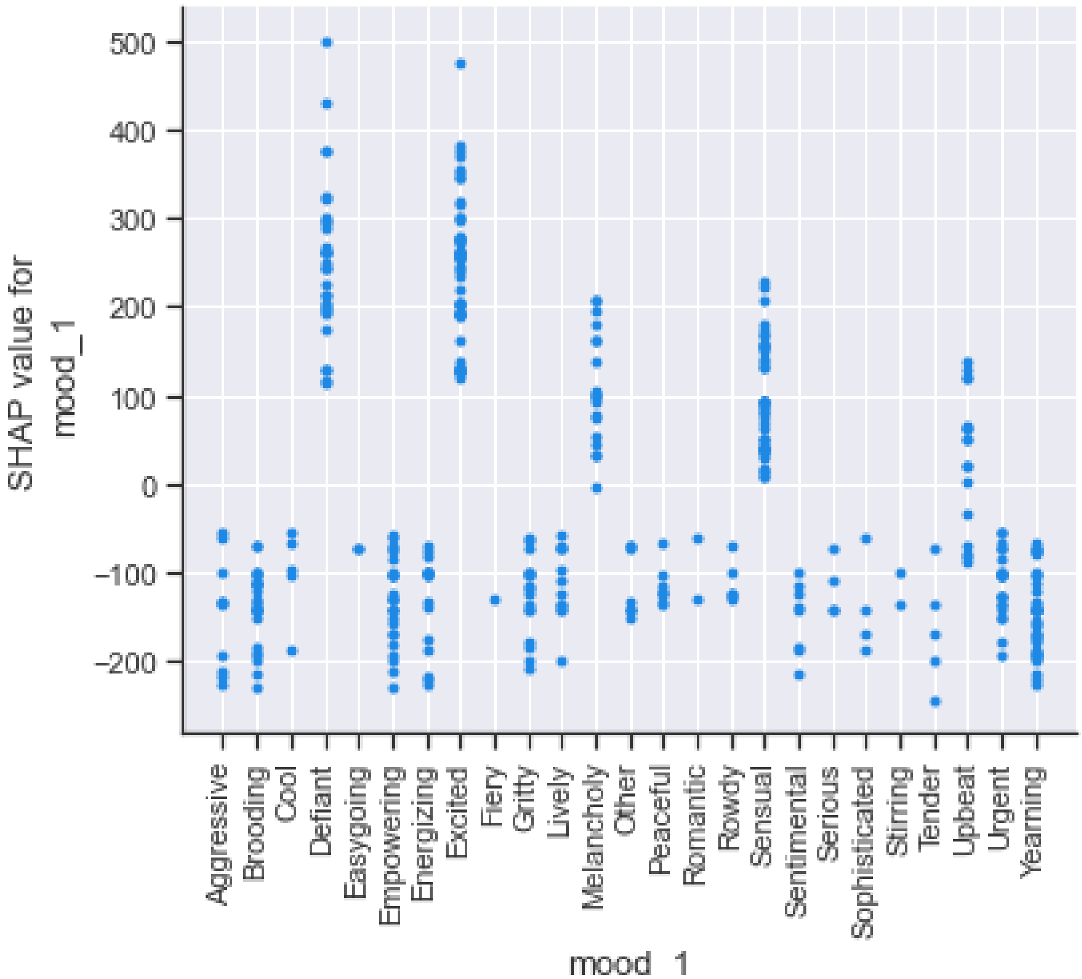

SHAP value for mood_1
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08
−0.10
-
Aggressive
Brooding
Cool
Defiant
Easygoing
Empowering
Energizing
Excited
Fiery
Gritty
Lively
Melancholy
Other
Peaceful
Romantic
Rowdy
Sensual
Sentimental
Serious
Somber
Sophisticated
Stirring
Tender
Upbeat
Urgent
Yearning
mood_1

7 Appendix

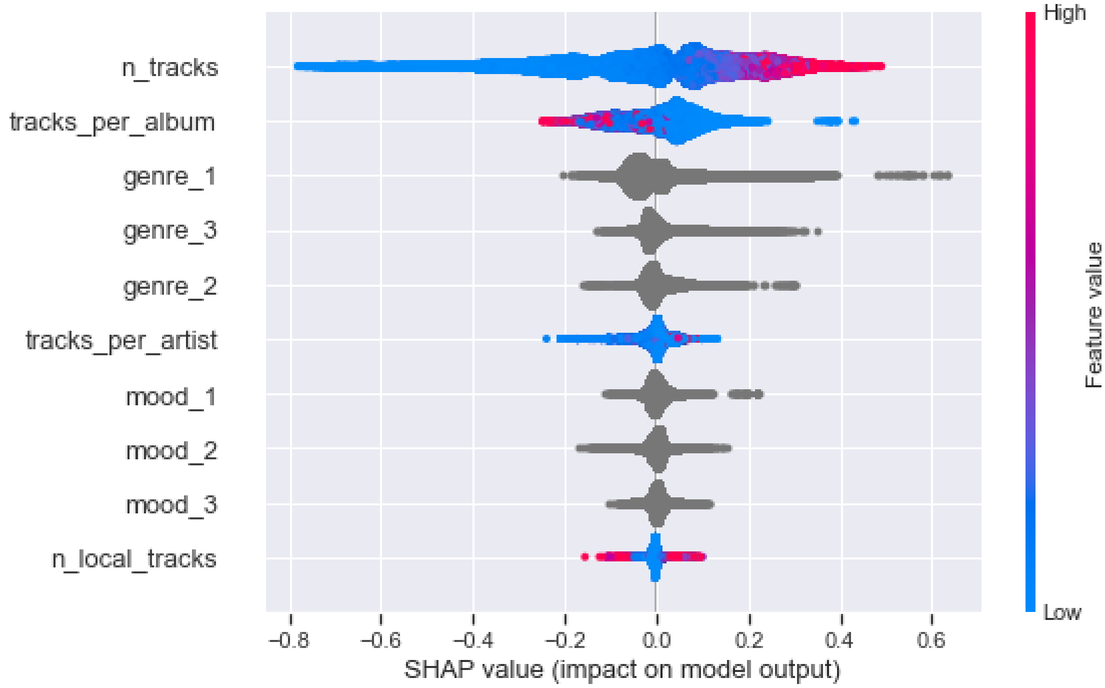
7.1 Spotify-Owned Playlists: Modeling Results for Number of Users in Current and Previous Month

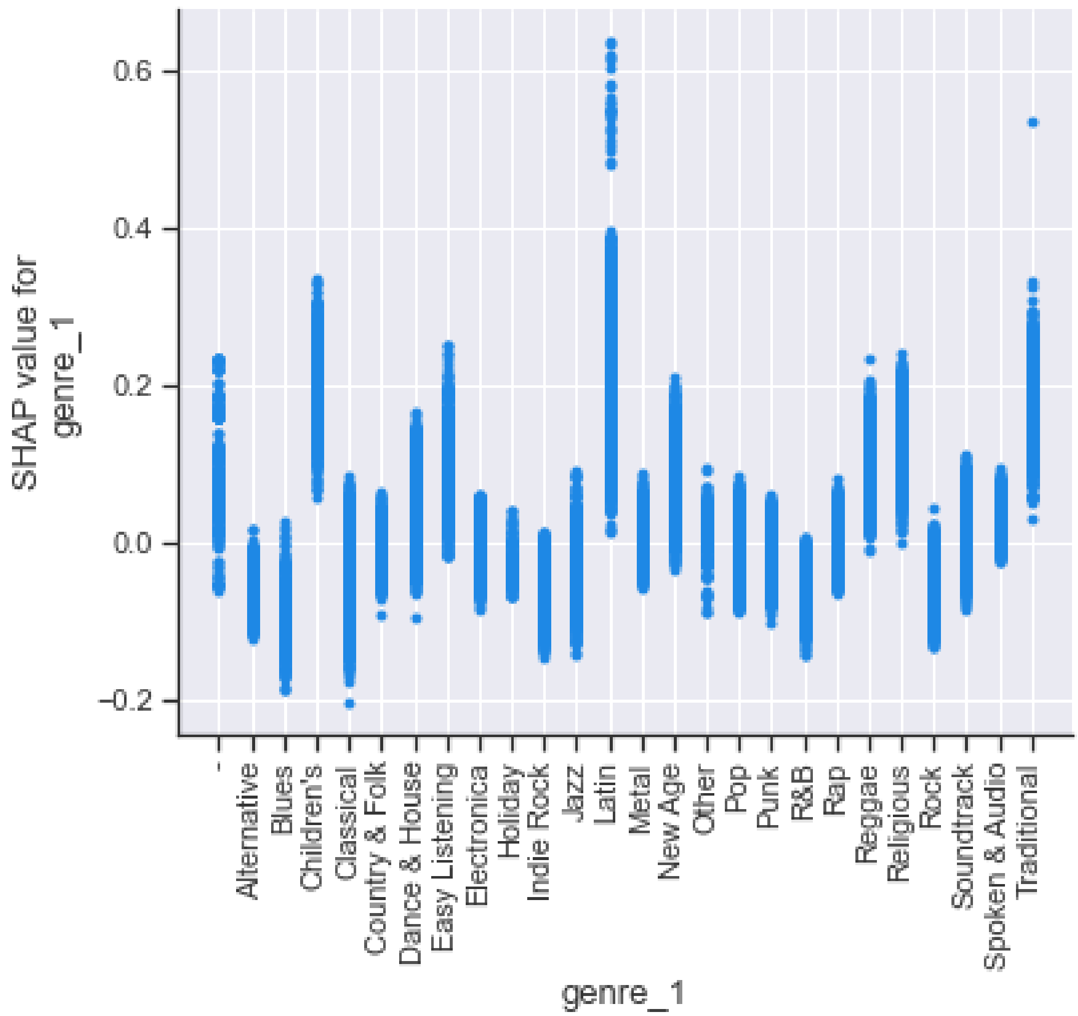



SHAP value for genre_1
1000
800
600
400
200
0
−200
−400
−600
Alternative
Blues
Children's
Classical
Country & Folk
Dance & House
Electronica
Holiday
Indie Rock
Jazz
Latin
Metal
New Age
Pop
Punk
R&B
Rap
Reggae
Religious
Rock
Soundtrack
Spoken & Audio
genre_1

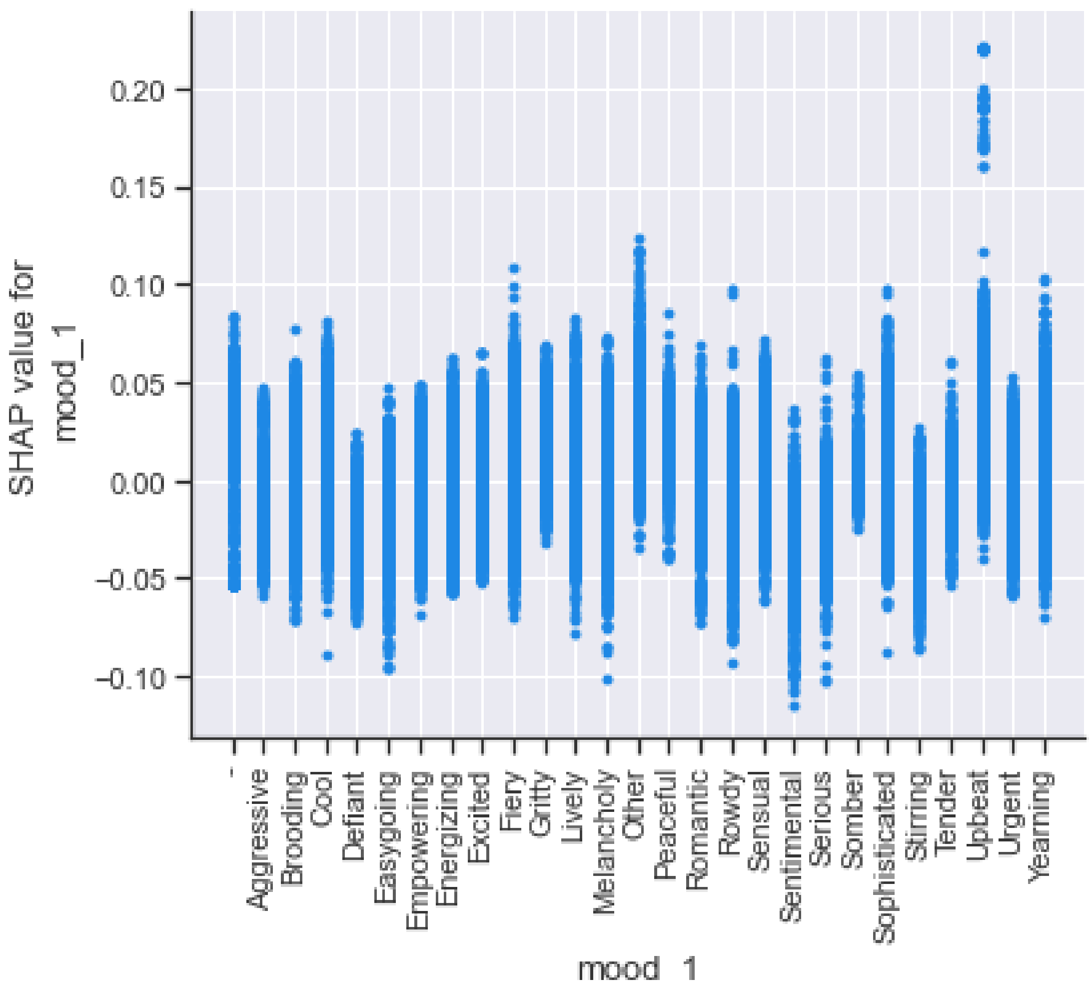

SHAP value for mood_1
500
400
300
200
100
0
−100
−200
Aggressive
Brooding
Cool
Defiant
Easygoing
Empowering
Energizing
Excited
Fiery
Gritty
Lively
Melancholy
Other
Peaceful
Romantic
Rowdy
Sensual
Sentimental
Serious
Sophisticated
Stirring
Tender
Upbeat
Urgent
Yearning
mood_1

7.2 User-Created Playlists: Modeling Results for Number of Users in Current and Previous Month Amongst





SHAP value for genre_1
0.6
0.4
0.2
0.0
-0.2
-
Alternative
Blues
Children's
Classical
Country & Folk
Dance & House
Easy Listening
Electronica
Holiday
Indie Rock
Jazz
Latin
Metal
New Age
Other
Pop
Punk
R&B
Rap
Reggae
Religious
Rock
Soundtrack
Spoken & Audio
Traditional
genre_1



7.3 Assumptions:

1. This is a random sample of playlists. If this a biased sample, then any generalizations that we make from the data is likely to be meaningfully inaccurate.
2. Spotify treats each user-created playlist equally in terms of promotion. For examples, if the Spotify algorithms were promoting some genres above others at the time this data was collected, then we are unlikely to get a good read on how genre affects listenership.
3. (a) Spotify treats its own playlists differently than the non-Spotify playlists. If this assumption is correct, then it is likely that Spotify playlists are not particularly comparable to non-Spotify playlists.

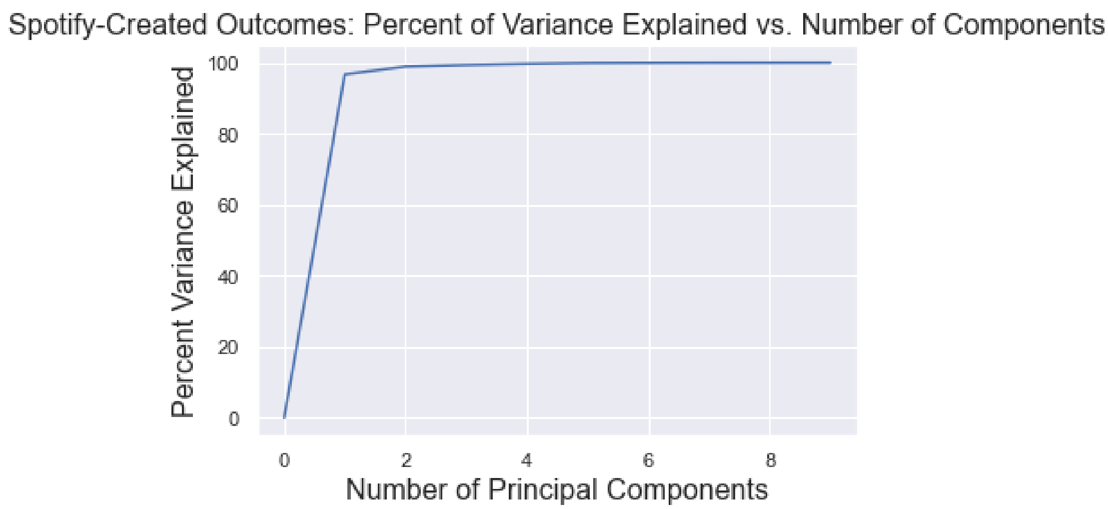
   (b) Spotify treats its own playlists equally with each other. Thus, an analysis with only Spotify playlists should be okay.
4. Each playlist included in the dataset has existed for at least two months. This ensures that the monthly average users in the previous month variable is not biased by how long the

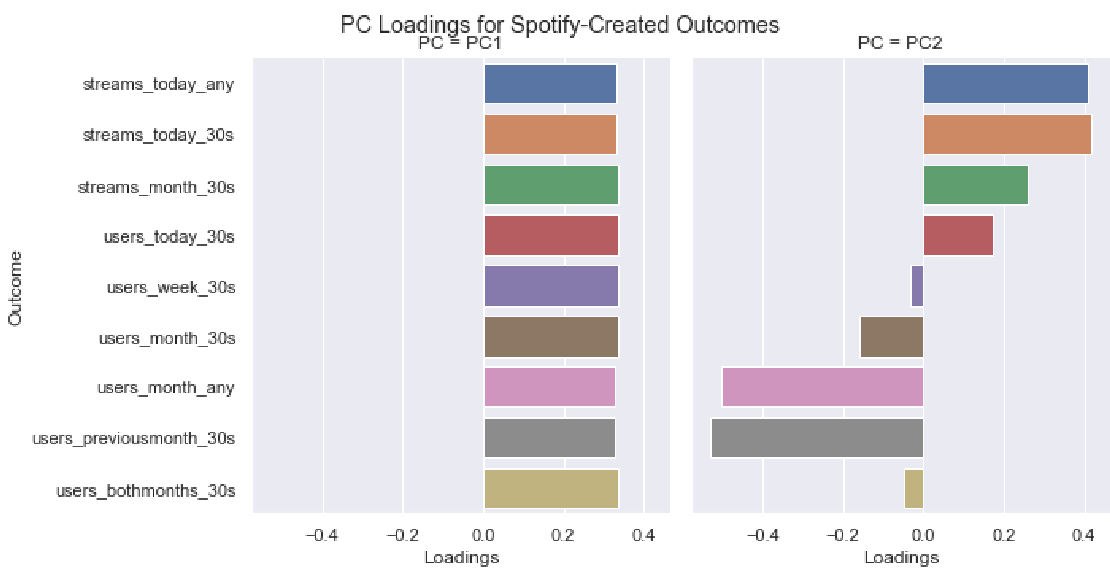
playlist has existed.

5. The Spotify algorithms do not amplify small variations in success. If playlist A was slightly more successful than Playlist B two months ago under 'fair' algorithmic treatment, then the algorithms will not amplify playlist A over playlist B, and thus widen the gulf between the success of the two playlists. In other words, there is a fair marketplace for the playlists to compete, where success does not necessarily beget success simply due to the algorithms.

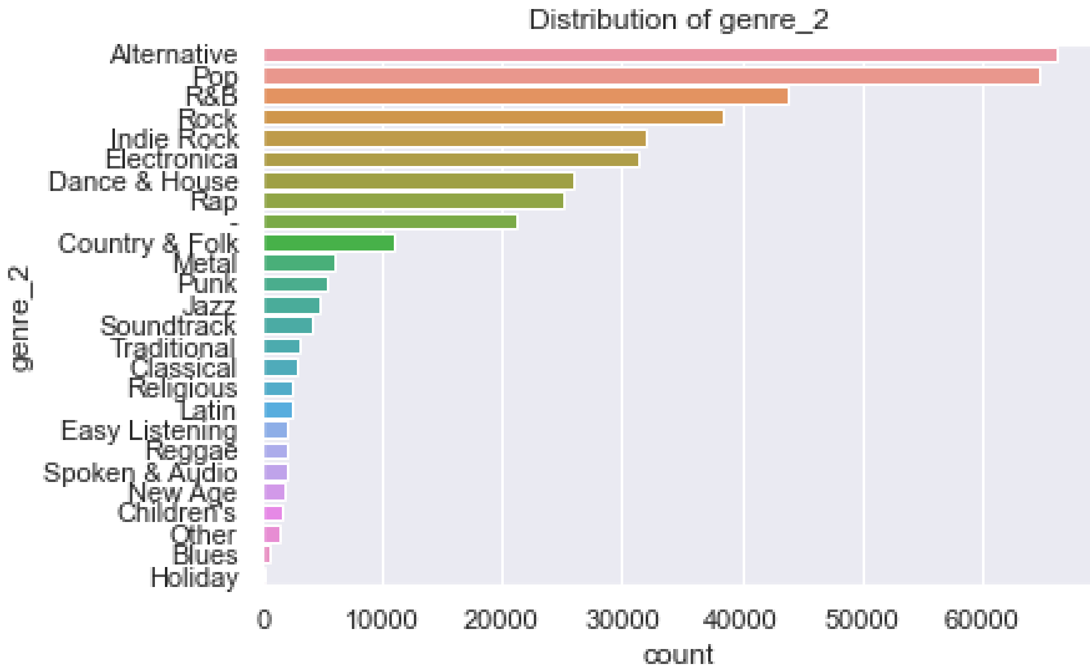
6. For the categorical variables, genre_1-genre_3 and mood_1-mood_3, when the value is '-' this is not a missing value, but is instead imparting the information that the given playlist does not easily fit into the predefined genre and mood types.

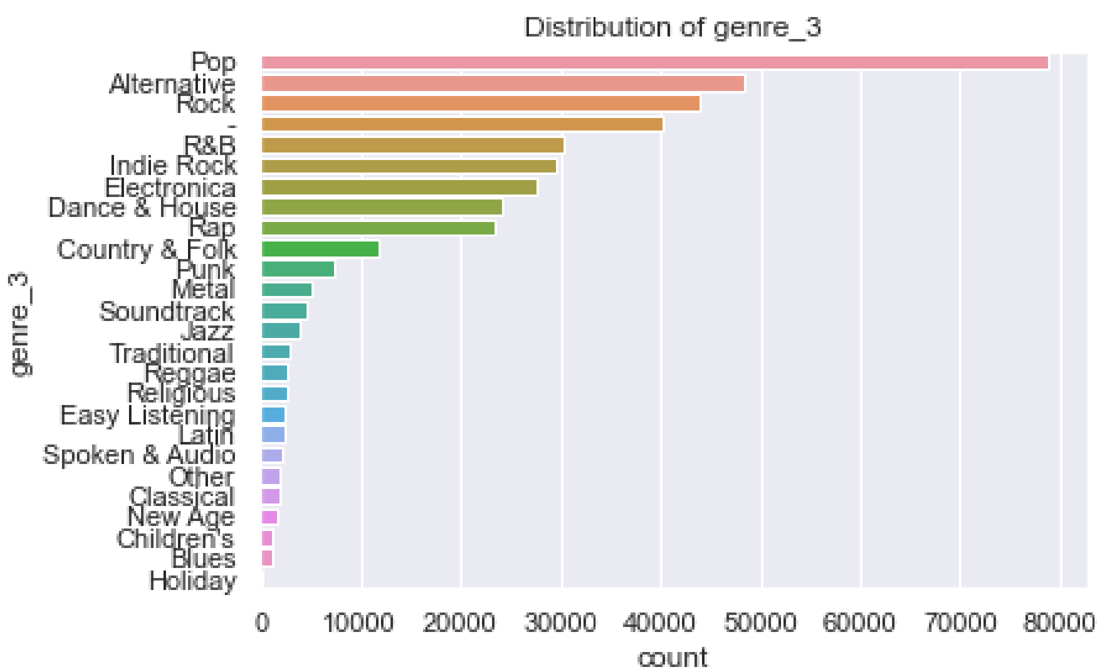
7.4 PCA Analysis of Outcomes for Spotify-Created Playlists

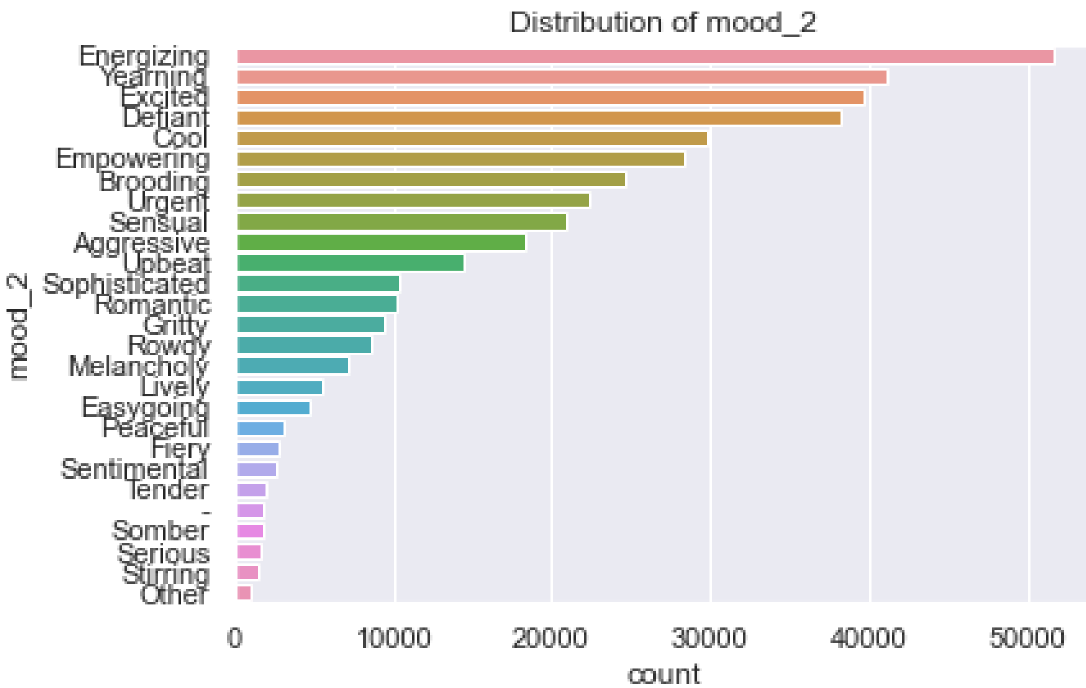


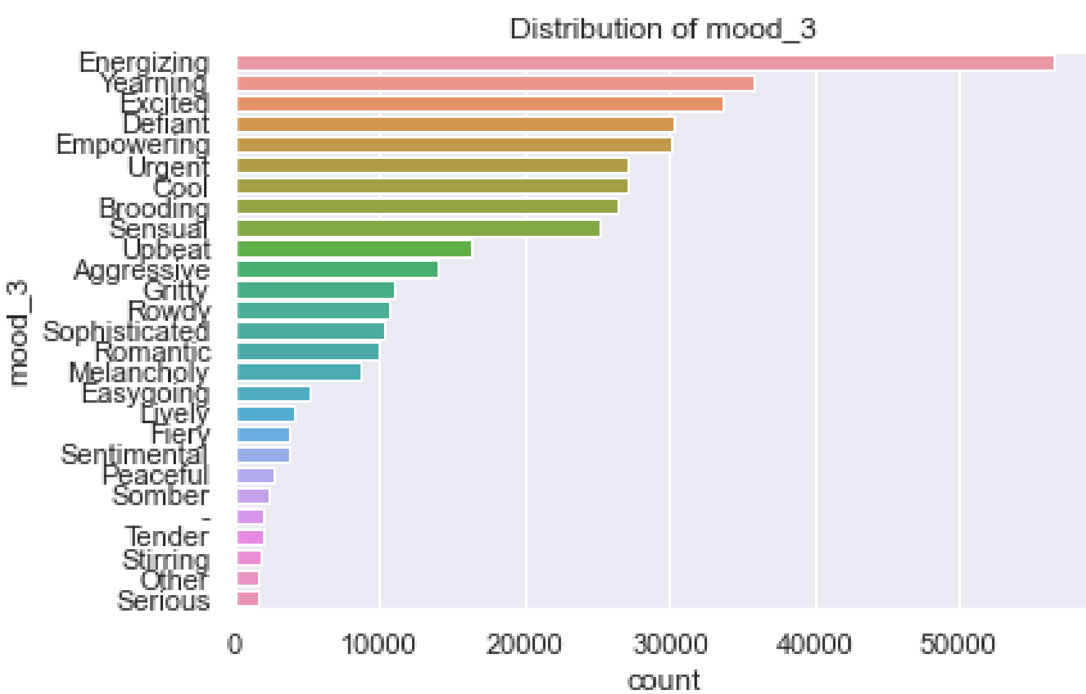


7.5 Plots of Distribution of Secondary and Tertiary Genre and Mood






Distribution of mood_2
Energizing
Yearning
Excited
Defiant
Cool
Empowering
Brooding
Urgent
Sensual
Aggressive
Upbeat
Sophisticated
Romantic
Gritty
Rowdy
Melancholy
Lively
Easygoing
Peaceful
Fiery
Sentimental
Tender
-
Somber
Serious
Stirring
Other
mood_2
0
10000
20000
30000
40000
50000
count


Distribution of mood_3
Energizing
Yearning
Excited
Defiant
Empowering
Urgent
Cool
Brooding
Sensual
Upbeat
Aggressive
Gritty
Rowdy
Sophisticated
Romantic
Melancholy
Easygoing
Lively
Fiery
Sentimental
Peaceful
Somber
-
Tender
Stirring
Other
Serious
mood_3
0
10000
20000
30000
40000
50000
count

7.6 Robust Linear Regression Model Results

7.6.1 RLM Model for Spotify-Created Playlists, Outcome = Users Today

```
                           Results: Robust linear model
================================================================================
==
Model:                      RLM                      Df Residuals:          351
Dependent Variable:         users_today_30s          Norm:
HuberT
Date:                       2022-02-23 19:06         Scale Est.:            mad
No. Observations:           399                      Cov. Type:             H1
Df Model:                   47                       Scale:
1289.6
--------------------------------------------------------------------------------
--
                             Coef.     Std.Err.    z    P>|z|    [0.025
0.975]
--------------------------------------------------------------------------------
--
Intercept                    592.5850  210.9870  2.8086 0.0050   179.0581
1006.1119
genre_1[T.Alternative]       321.4248  387.0639  0.8304 0.4063  -437.2066
1080.0561
genre_1[T.Blues]            1730.9971 1164.6920  1.4862 0.1372  -551.7574
4013.7515
genre_1[T.Children's]       -445.0210 1603.4611 -0.2775 0.7814 -3587.7470
2697.7050
genre_1[T.Classical]        1301.9412 1010.0657  1.2890 0.1974  -677.7513
3281.6336
genre_1[T.Country & Folk]   -169.5201  365.2716 -0.4641 0.6426  -885.4392
546.3991
genre_1[T.Dance & House]     660.8945  377.2358  1.7519 0.0798   -78.4740
1400.2630
genre_1[T.Easy Listening]     -0.0000    0.0000 -0.5453 0.5855    -0.0000
0.0000
genre_1[T.Electronica]      -603.0586  405.2874 -1.4880 0.1368 -1397.4072
191.2900
genre_1[T.Holiday]          -199.2876 1612.9078 -0.1236 0.9017 -3360.5288
2961.9536
genre_1[T.Indie Rock]        266.7984  283.7900  0.9401 0.3472  -289.4198
823.0166
genre_1[T.Jazz]             -360.3224  556.3045 -0.6477 0.5172 -1450.6592
730.0144
genre_1[T.Latin]              -5.5204 1156.6357 -0.0048 0.9962 -2272.4847
2261.4439
genre_1[T.Metal]            -147.9796  714.6698 -0.2071 0.8360 -1548.7068
1252.7475
```

```
genre_1[T.New Age]          -570.1787 1000.4309 -0.5699 0.5687 -2530.9872
1390.6299
genre_1[T.Other]               0.0000    0.0000  1.7542 0.0794    -0.0000
0.0000
genre_1[T.Pop]               695.8092  296.2827  2.3485 0.0189   115.1058
1276.5127
genre_1[T.Punk]               17.1923  961.5249  0.0179 0.9857 -1867.3619
1901.7465
genre_1[T.R&B]               365.5038  452.5511  0.8077 0.4193  -521.4800
1252.4876
genre_1[T.Rap]               296.8161  501.3839  0.5920 0.5539  -685.8783
1279.5104
genre_1[T.Reggae]            198.2037 1805.6950  0.1098 0.9126 -3340.8934
3737.3008
genre_1[T.Religious]        -931.6740  613.0447 -1.5197 0.1286 -2133.2195
269.8714
genre_1[T.Rock]             -516.6896  418.6339 -1.2342 0.2171 -1337.1970
303.8178
genre_1[T.Soundtrack]      -1068.7898  966.9534 -1.1053 0.2690 -2963.9836
826.4040
genre_1[T.Spoken & Audio]   -244.9543 1205.7641 -0.2032 0.8390 -2608.2086
2118.3000
genre_1[T.Traditional]        -0.0000    0.0000 -0.3063 0.7594    -0.0000
0.0000
mood_1[T.Aggressive]        -254.2457  631.9588 -0.4023 0.6875 -1492.8622
984.3709
mood_1[T.Brooding]          -437.9987  363.1691 -1.2060 0.2278 -1149.7970
273.7997
mood_1[T.Cool]              -221.0991  893.3308 -0.2475 0.8045 -1971.9953
1529.7971
mood_1[T.Defiant]            140.3623  381.5733  0.3679 0.7130  -607.5076
888.2322
mood_1[T.Easygoing]         -968.6157 1166.1635 -0.8306 0.4062 -3254.2542
1317.0227
mood_1[T.Empowering]        -348.6438  365.1943 -0.9547 0.3397 -1064.4115
367.1239
mood_1[T.Energizing]        -157.2315  453.6704 -0.3466 0.7289 -1046.4091
731.9462
mood_1[T.Excited]            700.3149  298.4466  2.3465 0.0189   115.3704
1285.2594
mood_1[T.Fiery]            -1175.7843 1605.0526 -0.7326 0.4638 -4321.6296
1970.0611
mood_1[T.Gritty]            -759.8694  448.2712 -1.6951 0.0901 -1638.4648
118.7260
mood_1[T.Lively]            -206.9745  525.4376 -0.3939 0.6936 -1236.8133
822.8642
mood_1[T.Melancholy]         346.0206  418.6357  0.8265 0.4085  -474.4904
1166.5315
```

```
mood_1[T.Other]            -790.2208  997.5676 -0.7921 0.4283 -2745.4174
1164.9759
mood_1[T.Peaceful]          594.4401  615.4594  0.9658 0.3341  -611.8382
1800.7184
mood_1[T.Romantic]          167.0885 1472.1465  0.1135 0.9096 -2718.2655
3052.4426
mood_1[T.Rowdy]            -126.0988  682.2825 -0.1848 0.8534 -1463.3478
1211.1502
mood_1[T.Sensual]           372.3687  345.6100  1.0774 0.2813  -305.0144
1049.7519
mood_1[T.Sentimental]       951.1252  614.1904  1.5486 0.1215  -252.6658
2154.9162
mood_1[T.Serious]          -504.0052  932.1720 -0.5407 0.5887 -2331.0287
1323.0182
mood_1[T.Somber]             -0.0000    0.0000 -0.8718 0.3833    -0.0000
0.0000
mood_1[T.Sophisticated]    3097.9696  835.1919  3.7093 0.0002  1461.0235
4734.9158
mood_1[T.Stirring]         -778.8670 1276.0579 -0.6104 0.5416 -3279.8946
1722.1606
mood_1[T.Tender]            989.2884  948.0818  1.0435 0.2967  -868.9177
2847.4945
mood_1[T.Upbeat]            752.1248  445.0139  1.6901 0.0910  -120.0865
1624.3361
mood_1[T.Urgent]           -625.8362  399.0035 -1.5685 0.1168 -1407.8687
156.1962
mood_1[T.Yearning]         -163.0275  328.3333 -0.4965 0.6195  -806.5490
480.4940
n_tracks                      4.3197    1.0171  4.2470 0.0000     2.3261
6.3132
tracks_per_album             23.7652    3.3748  7.0419 0.0000    17.1507
30.3798
================================================================================
==
```

```
<statsmodels.robust.robust_linear_model.RLMResultsWrapper at 0x14418c34040>
```

7.6.2 RLM Model of User-Created Playlists, Outcome = Users Today

Note that for the following model outputs, the baseline level for both the genre and mood categorical predictors is "-", which I am taking to mean there is no easily discernible genre or mood detectable for the given playlist.

```
                      Results: Robust linear model
=========================================================================
Model:                   RLM                    Df Residuals:      402913
Dependent Variable:      users_today_30s        Norm:              HuberT
```

Date:	2022-02-23 19:07	Scale Est.:	mad
No. Observations:	402967	Cov. Type:	H1
Df Model:	53	Scale:	0.73545

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	0.4775	0.0191	25.0304	0.0000	0.4401	0.5149
genre_1[T.Alternative]	-0.0373	0.0280	-1.3342	0.1821	-0.0922	0.0175
genre_1[T.Blues]	-0.0499	0.0392	-1.2741	0.2026	-0.1268	0.0269
genre_1[T.Children's]	0.1435	0.0312	4.6023	0.0000	0.0824	0.2046
genre_1[T.Classical]	-0.0907	0.0293	-3.0895	0.0020	-0.1482	-0.0331
genre_1[T.Country & Folk]	0.0057	0.0281	0.2024	0.8396	-0.0495	0.0609
genre_1[T.Dance & House]	0.0254	0.0280	0.9064	0.3648	-0.0295	0.0802
genre_1[T.Easy Listening]	0.0205	0.0338	0.6068	0.5440	-0.0457	0.0867
genre_1[T.Electronica]	-0.0404	0.0282	-1.4323	0.1520	-0.0956	0.0149
genre_1[T.Holiday]	0.0271	0.0542	0.5004	0.6168	-0.0792	0.1335
genre_1[T.Indie Rock]	-0.0407	0.0279	-1.4625	0.1436	-0.0953	0.0139
genre_1[T.Jazz]	-0.0234	0.0296	-0.7878	0.4308	-0.0815	0.0347
genre_1[T.Latin]	0.2287	0.0282	8.1150	0.0000	0.1734	0.2839
genre_1[T.Metal]	0.0101	0.0289	0.3495	0.7267	-0.0465	0.0667
genre_1[T.New Age]	-0.0051	0.0302	-0.1702	0.8649	-0.0643	0.0540
genre_1[T.Other]	-0.0759	0.0519	-1.4627	0.1436	-0.1775	0.0258
genre_1[T.Pop]	0.0018	0.0278	0.0654	0.9478	-0.0527	0.0564
genre_1[T.Punk]	-0.0318	0.0284	-1.1182	0.2635	-0.0875	0.0239
genre_1[T.R&B]	-0.0094	0.0281	-0.3351	0.7375	-0.0646	0.0457
genre_1[T.Rap]	0.0399	0.0279	1.4292	0.1529	-0.0148	0.0946
genre_1[T.Reggae]	0.0639	0.0297	2.1478	0.0317	0.0056	0.1221
genre_1[T.Religious]	-0.0631	0.0282	-2.2376	0.0253	-0.1183	-0.0078
genre_1[T.Rock]	-0.0452	0.0280	-1.6176	0.1058	-0.1000	0.0096
genre_1[T.Soundtrack]	-0.0039	0.0286	-0.1350	0.8926	-0.0599	0.0522
genre_1[T.Spoken & Audio]	-0.0047	0.0332	-0.1404	0.8883	-0.0698	0.0605
genre_1[T.Traditional]	0.0993	0.0314	3.1655	0.0015	0.0378	0.1609
mood_1[T.Aggressive]	-0.0045	0.0212	-0.2115	0.8325	-0.0460	0.0370
mood_1[T.Brooding]	-0.0226	0.0207	-1.0891	0.2761	-0.0632	0.0180
mood_1[T.Cool]	-0.0174	0.0210	-0.8280	0.4077	-0.0586	0.0238
mood_1[T.Defiant]	0.0532	0.0204	2.6097	0.0091	0.0133	0.0932
mood_1[T.Easygoing]	-0.0553	0.0233	-2.3711	0.0177	-0.1009	-0.0096
mood_1[T.Empowering]	0.0055	0.0205	0.2668	0.7896	-0.0347	0.0456
mood_1[T.Energizing]	-0.0176	0.0206	-0.8570	0.3914	-0.0579	0.0227
mood_1[T.Excited]	0.0387	0.0204	1.8972	0.0578	-0.0013	0.0786
mood_1[T.Fiery]	-0.0221	0.0256	-0.8632	0.3880	-0.0723	0.0281
mood_1[T.Gritty]	-0.0355	0.0213	-1.6634	0.0962	-0.0772	0.0063
mood_1[T.Lively]	0.0377	0.0211	1.7841	0.0744	-0.0037	0.0790
mood_1[T.Melancholy]	-0.0545	0.0217	-2.5092	0.0121	-0.0970	-0.0119
mood_1[T.Other]	0.0040	0.0254	0.1591	0.8736	-0.0457	0.0538
mood_1[T.Peaceful]	-0.0040	0.0236	-0.1716	0.8638	-0.0503	0.0422
mood_1[T.Romantic]	-0.0045	0.0211	-0.2133	0.8311	-0.0458	0.0368
mood_1[T.Rowdy]	0.0042	0.0218	0.1908	0.8487	-0.0386	0.0469

```
mood_1[T.Sensual]         -0.0135   0.0206 -0.6539 0.5132 -0.0538  0.0269
mood_1[T.Sentimental]     -0.0833   0.0246 -3.3923 0.0007 -0.1315 -0.0352
mood_1[T.Serious]         -0.0143   0.0250 -0.5703 0.5685 -0.0633  0.0348
mood_1[T.Somber]          -0.0623   0.0280 -2.2258 0.0260 -0.1171 -0.0074
mood_1[T.Sophisticated]    0.0033   0.0213  0.1543 0.8774 -0.0384  0.0450
mood_1[T.Stirring]         0.0200   0.0244  0.8224 0.4109 -0.0277  0.0678
mood_1[T.Tender]           0.0062   0.0229  0.2732 0.7847 -0.0386  0.0510
mood_1[T.Upbeat]           0.0372   0.0209  1.7755 0.0758 -0.0039  0.0783
mood_1[T.Urgent]           0.0178   0.0208  0.8549 0.3926 -0.0230  0.0586
mood_1[T.Yearning]        -0.0143   0.0205 -0.6987 0.4847 -0.0544  0.0258
n_tracks                   0.0001   0.0000 53.0255 0.0000  0.0001  0.0001
tracks_per_album          -0.0000   0.0000 -1.6380 0.1014 -0.0000  0.0000
=========================================================================
```

```
<statsmodels.robust.robust_linear_model.RLMResultsWrapper at 0x144187b5040>
```

7.6.3 OLS Model of User Created Playlists, Outcome = Users Today

This is a model of number of users today for the non-Spotify playlists. Take note of the incredibly large standard errors on our estimates when we do not employ a loss function that is robust to outliers.

```
                    Results: Ordinary least squares
=======================================================================
Model:              OLS               Adj. R-squared:     0.000
Dependent Variable: users_today_30s   AIC:                4140974.1152
Date:               2022-02-23 19:07  BIC:                4141563.0721
No. Observations:   402967            Log-Likelihood:     -2.0704e+06
Df Model:           53                F-statistic:        1.180
Df Residuals:       402913            Prob (F-statistic): 0.173
R-squared:          0.000             Scale:              1699.7
-----------------------------------------------------------------------
                          Coef.  Std.Err.    t    P>|t|   [0.025 0.975]
-----------------------------------------------------------------------
Intercept                  1.2035   1.3420  0.8968 0.3698 -1.4268 3.8338
genre_1[T.Alternative]    -0.5632   1.9695 -0.2860 0.7749 -4.4233 3.2969
genre_1[T.Blues]          -0.4999   2.7576 -0.1813 0.8562 -5.9047 4.9049
genre_1[T.Children's]      1.1479   2.1936  0.5233 0.6008 -3.1515 5.4473
genre_1[T.Classical]      -1.1414   2.0647 -0.5528 0.5804 -5.1882 2.9054
genre_1[T.Country & Folk] -0.6222   1.9803 -0.3142 0.7534 -4.5036 3.2591
genre_1[T.Dance & House]  -0.0923   1.9686 -0.0469 0.9626 -3.9507 3.7661
genre_1[T.Easy Listening] -0.6473   2.3767 -0.2724 0.7854 -5.3056 4.0110
genre_1[T.Electronica]    -0.8544   1.9821 -0.4310 0.6664 -4.7392 3.0305
genre_1[T.Holiday]        -0.5885   3.8160 -0.1542 0.8774 -8.0678 6.8909
genre_1[T.Indie Rock]     -0.7986   1.9597 -0.4075 0.6836 -4.6396 3.0424
genre_1[T.Jazz]           -0.4959   2.0855 -0.2378 0.8120 -4.5834 3.5916
genre_1[T.Latin]           0.5322   1.9823  0.2685 0.7883 -3.3531 4.4176
```

```
genre_1[T.Metal]          -0.5730   2.0306 -0.2822 0.7778 -4.5529 3.4069
genre_1[T.New Age]        -0.9145   2.1236 -0.4306 0.6667 -5.0767 3.2477
genre_1[T.Other]          -0.8709   3.6496 -0.2386 0.8114 -8.0240 6.2821
genre_1[T.Pop]            -0.2399   1.9576 -0.1225 0.9025 -4.0766 3.5969
genre_1[T.Punk]           -0.8336   2.0002 -0.4168 0.6769 -4.7540 3.0868
genre_1[T.R&B]            -0.7183   1.9798 -0.3628 0.7167 -4.5986 3.1620
genre_1[T.Rap]            -0.3503   1.9636 -0.1784 0.8584 -4.1988 3.4983
genre_1[T.Reggae]         -0.3277   2.0920 -0.1566 0.8755 -4.4279 3.7725
genre_1[T.Religious]      -0.7106   1.9828 -0.3584 0.7201 -4.5968 3.1756
genre_1[T.Rock]           -0.0104   1.9670 -0.0053 0.9958 -3.8656 3.8448
genre_1[T.Soundtrack]      0.3482   2.0116  0.1731 0.8626 -3.5945 4.2910
genre_1[T.Spoken & Audio] -0.5224   2.3385 -0.2234 0.8232 -5.1059 4.0610
genre_1[T.Traditional]     0.6116   2.2080  0.2770 0.7818 -3.7159 4.9392
mood_1[T.Aggressive]       0.5344   1.4906  0.3585 0.7199 -2.3870 3.4559
mood_1[T.Brooding]         0.0589   1.4573  0.0404 0.9677 -2.7973 2.9151
mood_1[T.Cool]            -0.0634   1.4776 -0.0429 0.9658 -2.9594 2.8326
mood_1[T.Defiant]          0.3676   1.4347  0.2562 0.7978 -2.4445 3.1796
mood_1[T.Easygoing]       -0.2866   1.6398 -0.1748 0.8612 -3.5005 2.9273
mood_1[T.Empowering]       0.2961   1.4410  0.2055 0.8372 -2.5281 3.1203
mood_1[T.Energizing]       0.4055   1.4465  0.2803 0.7792 -2.4297 3.2407
mood_1[T.Excited]          0.7369   1.4337  0.5140 0.6072 -2.0730 3.5468
mood_1[T.Fiery]            0.0122   1.8012  0.0068 0.9946 -3.5181 3.5426
mood_1[T.Gritty]           0.0013   1.4995  0.0009 0.9993 -2.9376 2.9402
mood_1[T.Lively]           0.2418   1.4850  0.1628 0.8706 -2.6686 3.1523
mood_1[T.Melancholy]       0.1516   1.5269  0.0993 0.9209 -2.8410 3.1442
mood_1[T.Other]            0.0283   1.7856  0.0158 0.9874 -3.4714 3.5279
mood_1[T.Peaceful]         0.1250   1.6594  0.0753 0.9399 -3.1273 3.3773
mood_1[T.Romantic]         0.9962   1.4824  0.6720 0.5016 -1.9093 3.9017
mood_1[T.Rowdy]            0.4852   1.5336  0.3164 0.7517 -2.5206 3.4909
mood_1[T.Sensual]          0.5650   1.4482  0.3901 0.6965 -2.2735 3.4034
mood_1[T.Sentimental]     -0.1200   1.7282 -0.0694 0.9446 -3.5073 3.2673
mood_1[T.Serious]          1.8377   1.7595  1.0444 0.2963 -1.6109 5.2863
mood_1[T.Somber]           0.0115   1.9681  0.0058 0.9953 -3.8459 3.8689
mood_1[T.Sophisticated]    0.1403   1.4974  0.0937 0.9253 -2.7945 3.0751
mood_1[T.Stirring]         0.1790   1.7150  0.1044 0.9169 -3.1823 3.5403
mood_1[T.Tender]           0.3622   1.6079  0.2252 0.8218 -2.7892 3.5136
mood_1[T.Upbeat]           0.7530   1.4738  0.5109 0.6094 -2.1356 3.6417
mood_1[T.Urgent]           0.2116   1.4655  0.1444 0.8852 -2.6608 3.0839
mood_1[T.Yearning]         0.1184   1.4399  0.0822 0.9345 -2.7037 2.9404
n_tracks                   0.0003   0.0001  2.5696 0.0102  0.0001 0.0005
tracks_per_artist         -0.0005   0.0007 -0.7858 0.4320 -0.0018 0.0008
------------------------------------------------------------------------
Omnibus:           2142164.888  Durbin-Watson:      1.999
Prob(Omnibus):     0.000        Jarque-Bera (JB):   180443014585193.719
Skew:              268.776      Prob(JB):           0.000
Kurtosis:          103668.698   Condition No.:      108308
========================================================================
* The condition number is large (1e+05). This might indicate
```

strong multicollinearity or other numerical problems.