Song Popularity on Spotify

Rose Dennis
March 13, 2020

TABLE OF CONTENTS

Problem Statement

How do we glean insights on song popularity, what types of listeners are listening, and could we manipulate audio features to have an impact on the popularity of a song?

Why is this important?

Increase Popularity

A certain combination of audio features produces a certain type of popular song

Increase Listeners

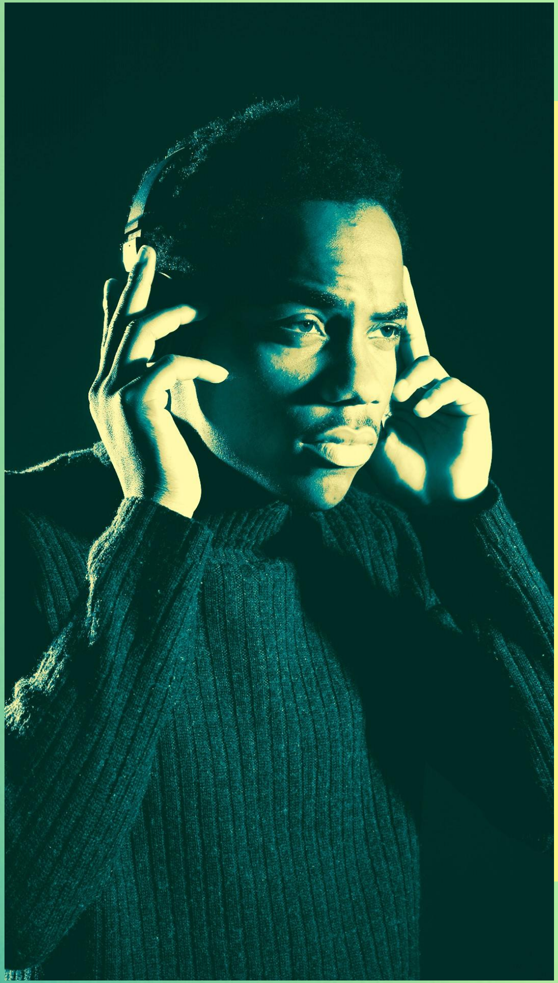
If a song is popular in multiple genres, you increase the amount of listeners for that song

Increase Revenue

More listeners for your song leads to more revenue for the artist

Executive Summary

- Having one or more featured artist really contributed to the popularity of a song
- Contrary to prior beliefs, there was no strong linear trend between any audio feature and song popularity, even when filtering by genre was present
- Provided evidence for more intuitive predispositions



The Data

- 16, 864 songs → 12,527 songs (no duplicates)
- 22 variables
- 6 ‘categories’: country, alternative, pop, rap, club, hits
- Feature Engineered ‘featured_artist’ column
- Target is ‘track_pop’ on a range from 0-100

Acquiring the Data

track_name	track_ids	speechiness	category
Oh Cacchio	1yjyFN62WDYfjH7FaiVHL8	0.3420	rap
Blu Skyes	2cFHJw9MFRzt3Bo3s8aPFU	0.1160	pop
Miss You 2	0j5QKkgh58UmjjONmDxYhE	0.0421	pop
Solitude	31AuJaUuJV7LI0grZeJM1y	0.0843	alternative

...One...hour...latahh..

- Spotify client ID and client secret
- Spotify API queries (Feb, 27, 2020)
  - playlists have a max of 100 songs
  - some artists didn't have a genre
  - skipped music videos
  - popularity score is time dependent
  - artist popularity derived from track popularity

Only pulled playlists with over 20,000 followers

1 function to get playlist IDs, 1 function to get tracks and create df

The Constraints

- Could not pull every full playlist
- Only had 4 well-defined genres
- Did not have a time filter
- Only pulled playlists with 20,000 followers or more

Alternative

SUGAR by BROCKHAMPTON
Score: 90

Pop

The Box by Roddy Ricch
Score: 100

10,000 Hours by Dan and Shay

Rap

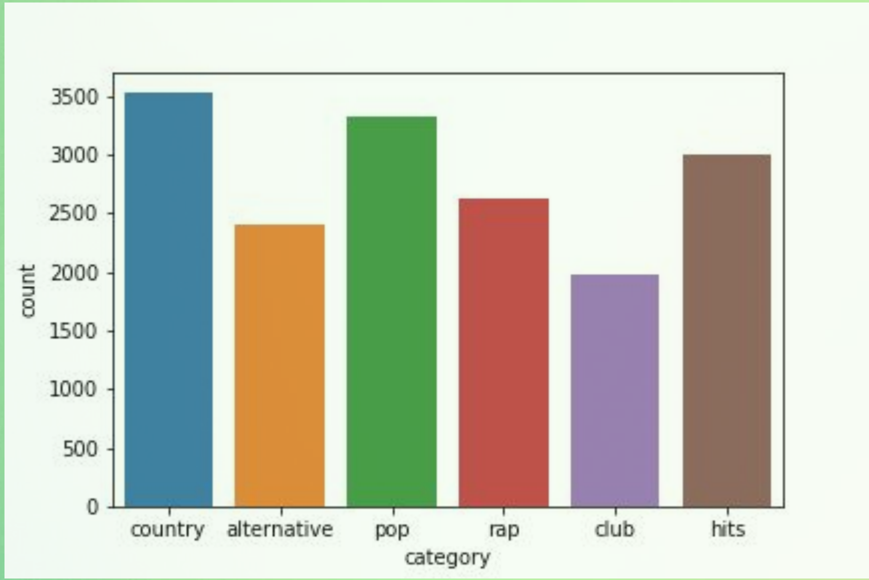
Life is Good by Future
Score: 96

Hits

Someone you loved by Lewis Capaldi
Score: 93

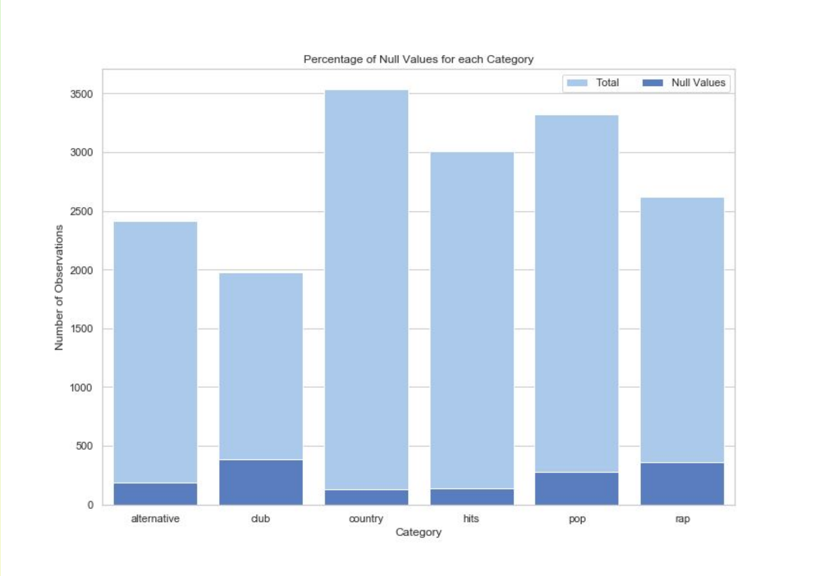
Exploratory Analysis

Distribution of Genres (categories)



* * This **includes** duplicates

Missing Values within each Genre



Duplicates

ID

ID

	track_ids	track_name	playlist_id	playlist_name	category	duplicate_counts_ids
0	3GJ4hzg4lrGwU51Y3VARbF	Speechless	37i9dQZF1DX8WMG8VPSOJC	Country Kind of Love	country	2
99	7q7jyVU0f0hnod8tsaUmxg	Speechless	37i9dQZF1DWYnwbYQ5HnZU	Country Gold	country	3
218	3GJ4hzg4lrGwU51Y3VARbF	Speechless	37i9dQZF1DXdfhOsjRMISB	Country Drive	country	2
1728	7q7jyVU0f0hnod8tsaUmxg	Speechless	37i9dQZF1DX13ZzXoot6Jc	Country Favourites	country	3
3198	7q7jyVU0f0hnod8tsaUmxg	Speechless	37i9dQZF1DX6P1Nsk3wSZX	Hot Country - Top Tracks of 2018	country	3

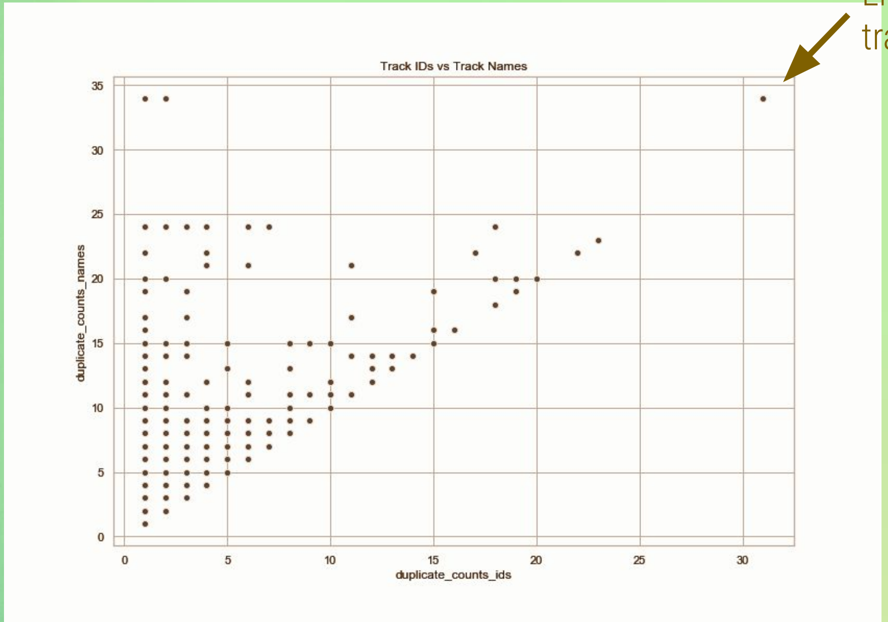
Duplicate Track Name: Multiple Track Names in Dataframe

Duplicate Artist Name: Multiple Artist Names in Dataframe

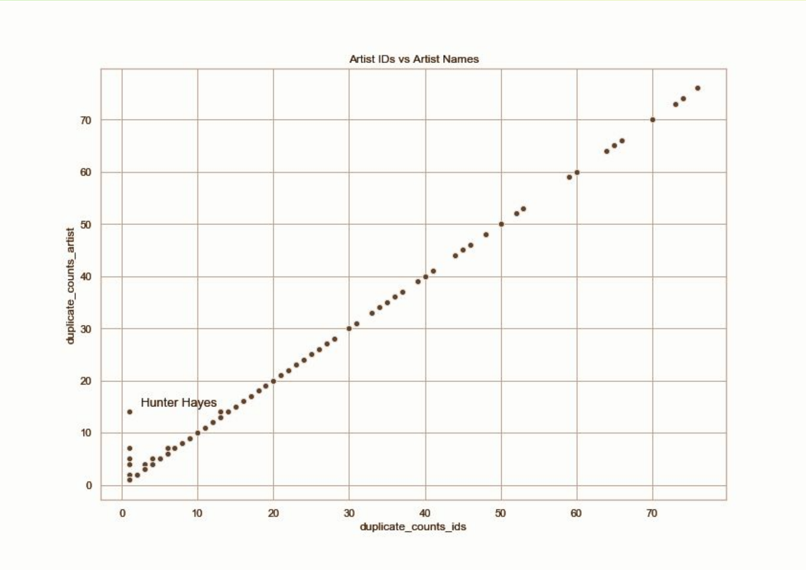
	artist	artist_ids	category
count	16864	16864	16864
unique	6111	6122	6
top	Luke Bryan	0BvkDsjlUla7X0k6CSWh1l	country
freq	76	76	3533

Track Duplicates

Don’t Start Now by Dua Lipa: 31 track ID’s, 34 track names



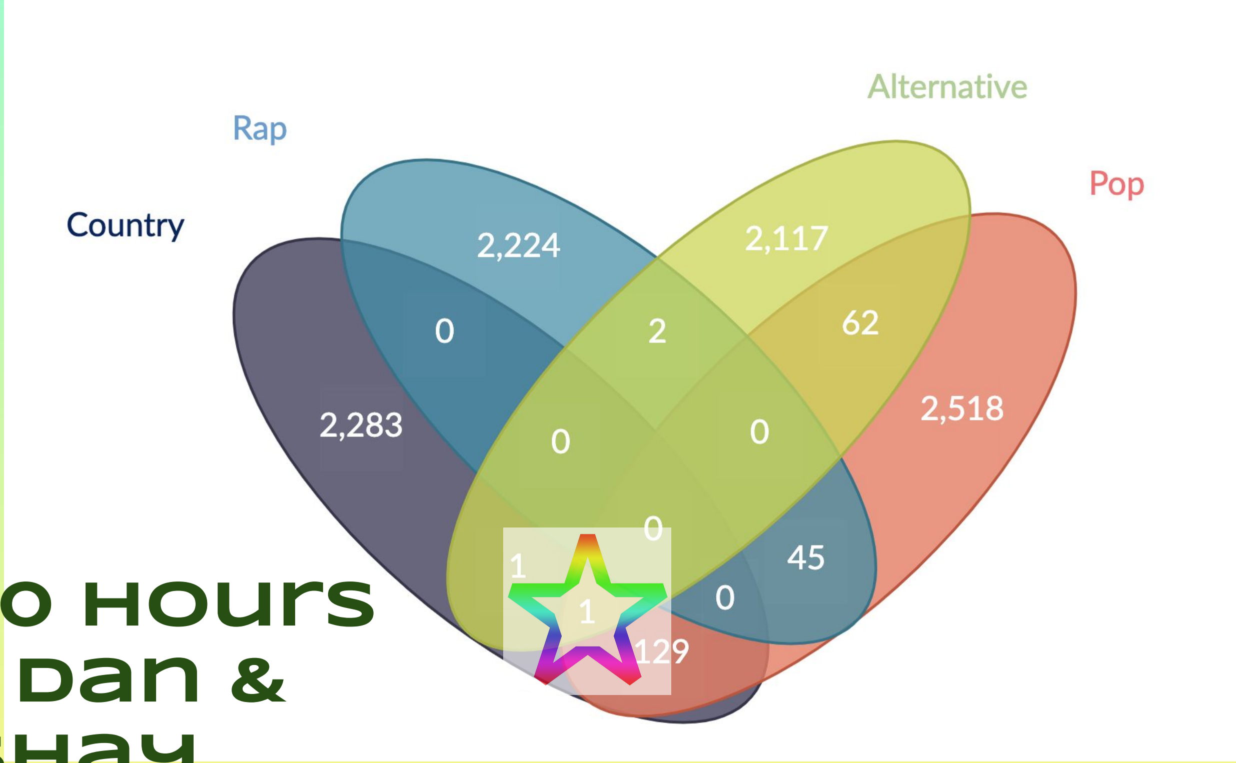
Artist Duplicates



* * This **includes** duplicates

Goal: Have the song overlap into all genres
Reality: Only 1 song appeared in three distinct genre playlists

More Genres → More Money $$$



10,000 Hours by Dan & Shay

Audio Features

Danceability	How suitable a track is for dancing based off tempo, beat strength, etc (0-1)
Energy	Perceptual measure of activity and intensity. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy (0-1)
Time Signature	How many beats in each bar (1-5)
Speechiness	The presence of spoken word (0-1)
Acousticness	A confidence measure of whether the track is acoustic (0-1)
Liveness	A confidence measure of detecting an audience (0-1)
Tempo	Beats per minute, speed/pace of a song

Audio Features

Danceability: 0.62

Energy: 0.73

Time Signature: 3.98

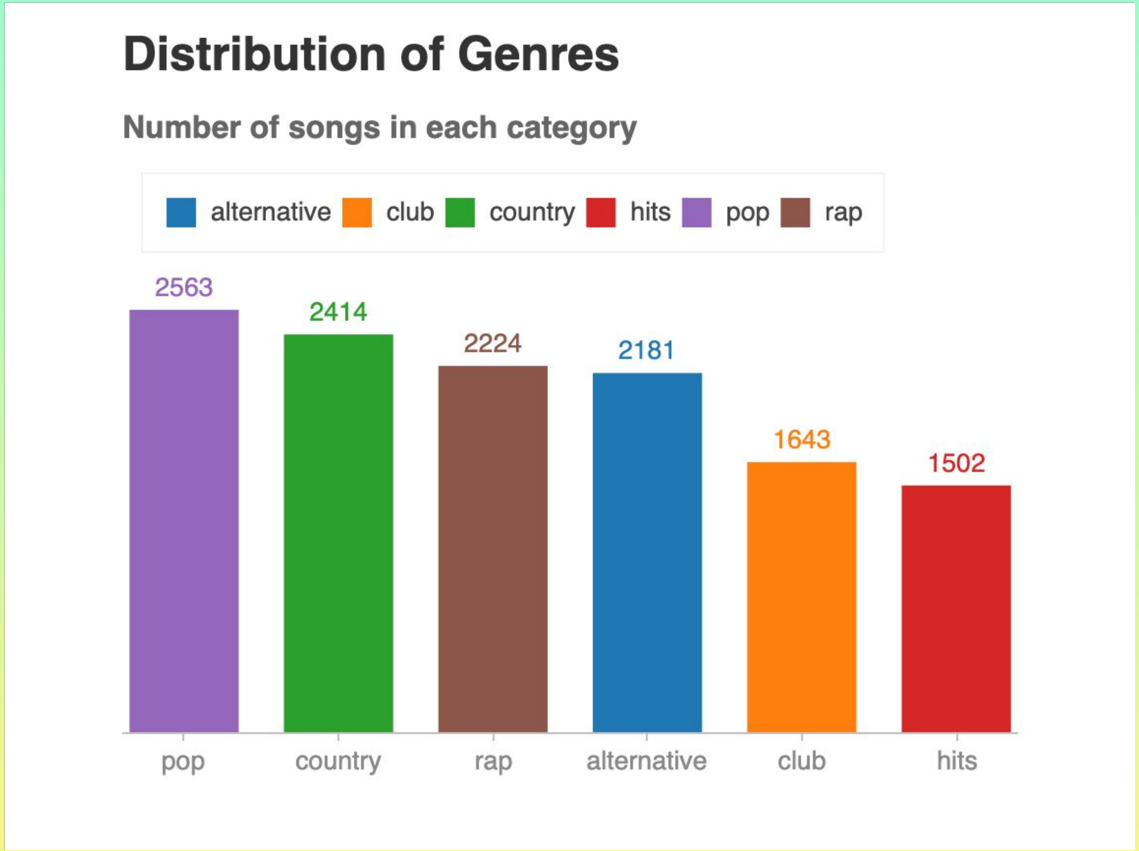
Audio Features

Speechiness: 0.07

Acousticness: 0.17

Liveness: 0.17

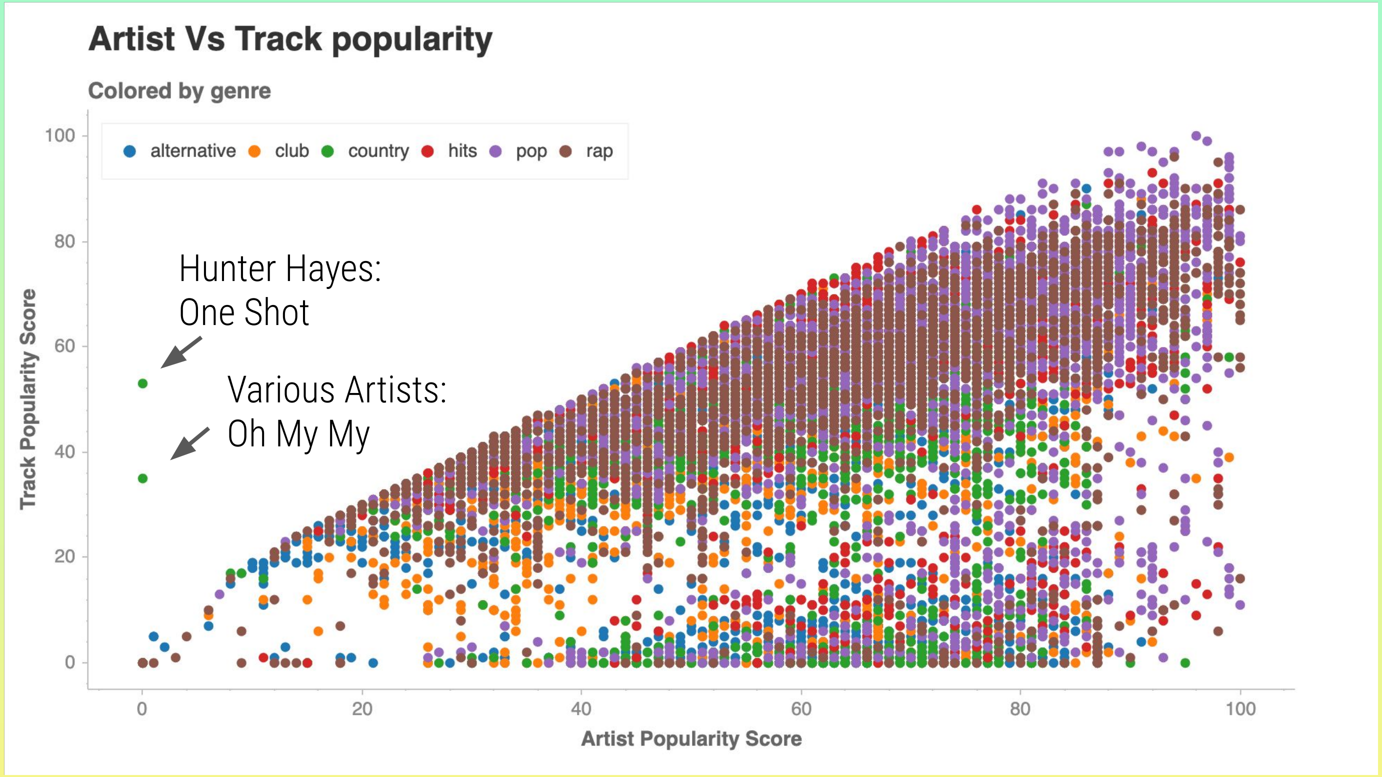
Exploratory Analysis WITHOUT Duplicates



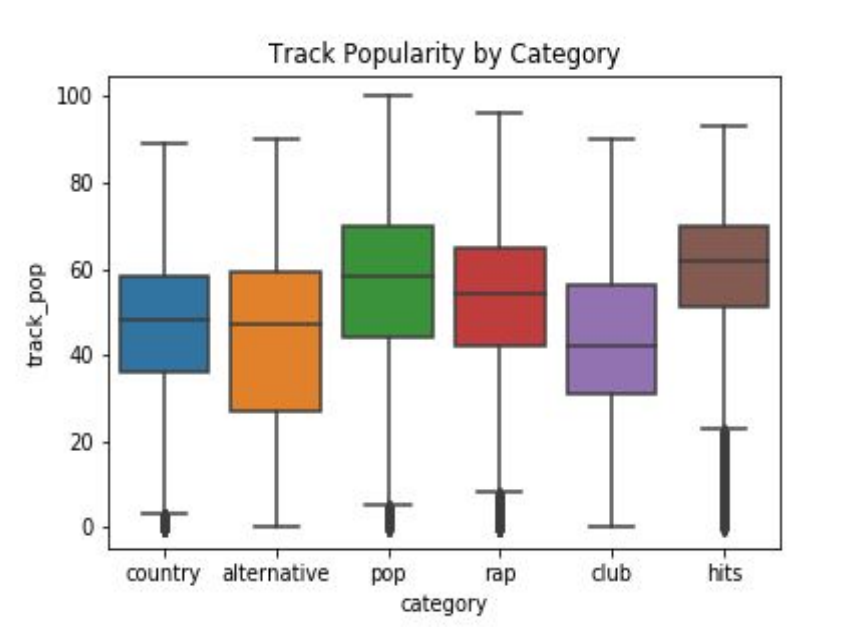
Remember: Artist Popularity is derived from Track Popularity

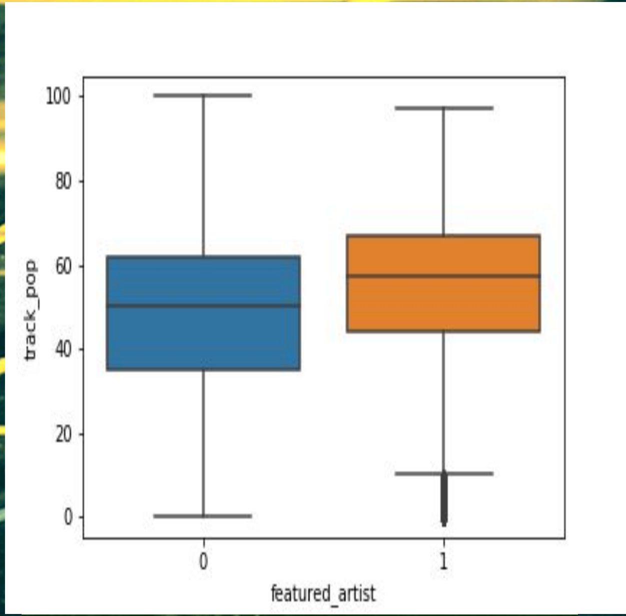
Filtered by Genre (category)

Two very apparent outliers

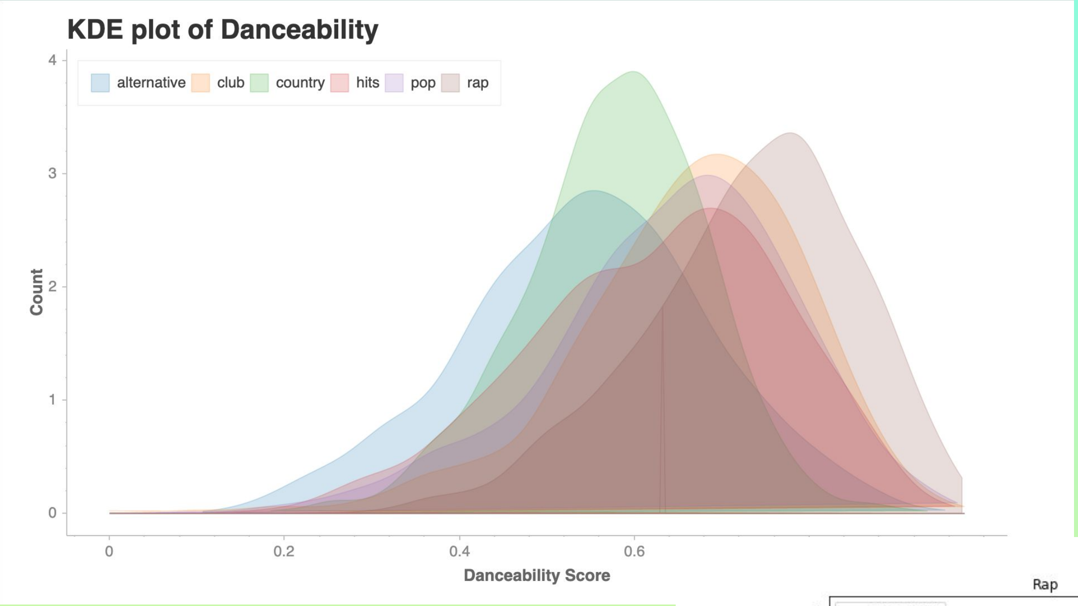


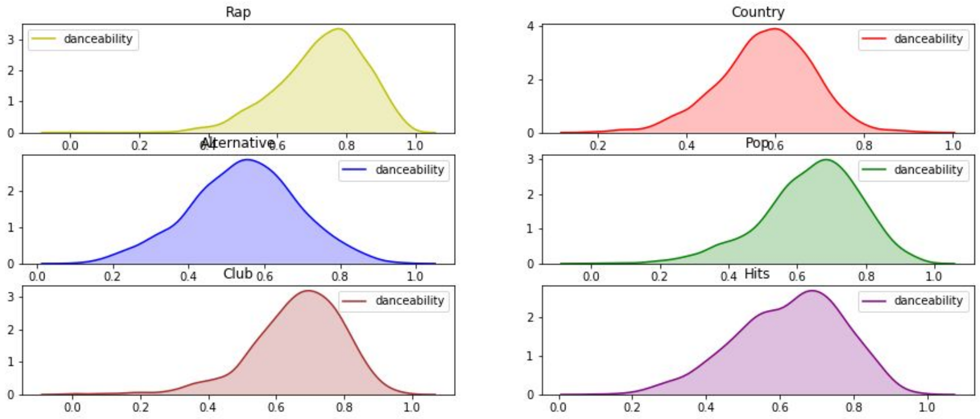
Track Popularity





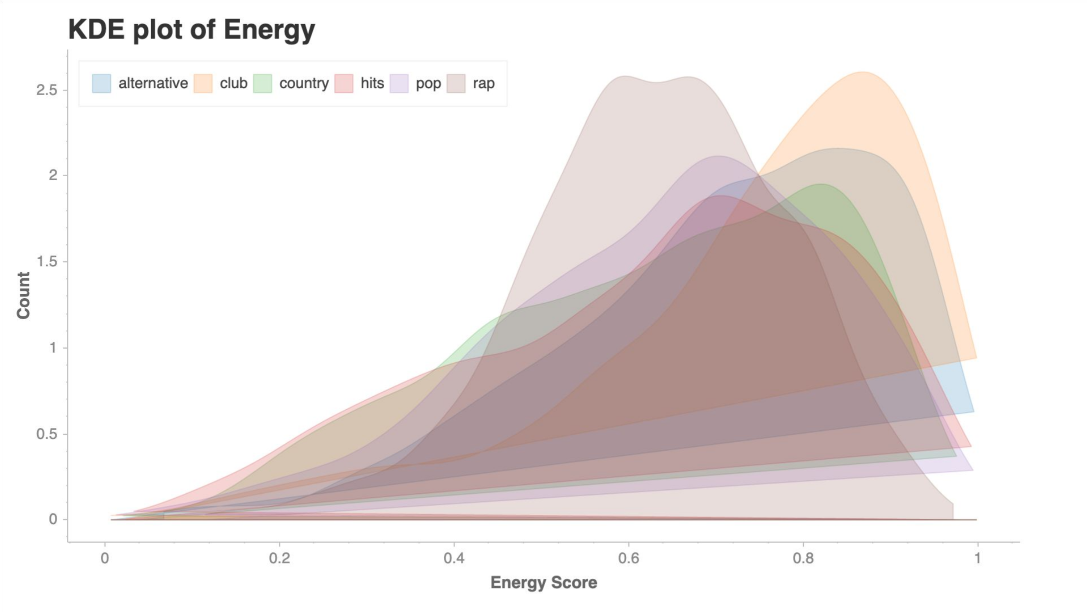
KDE plots

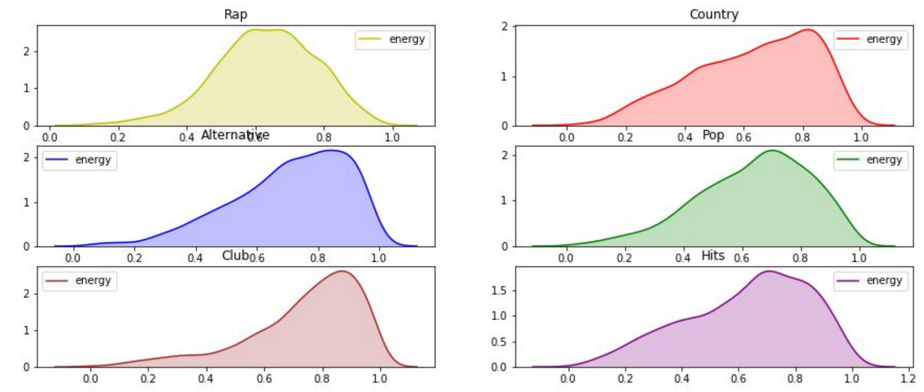




Look at labels, not colors

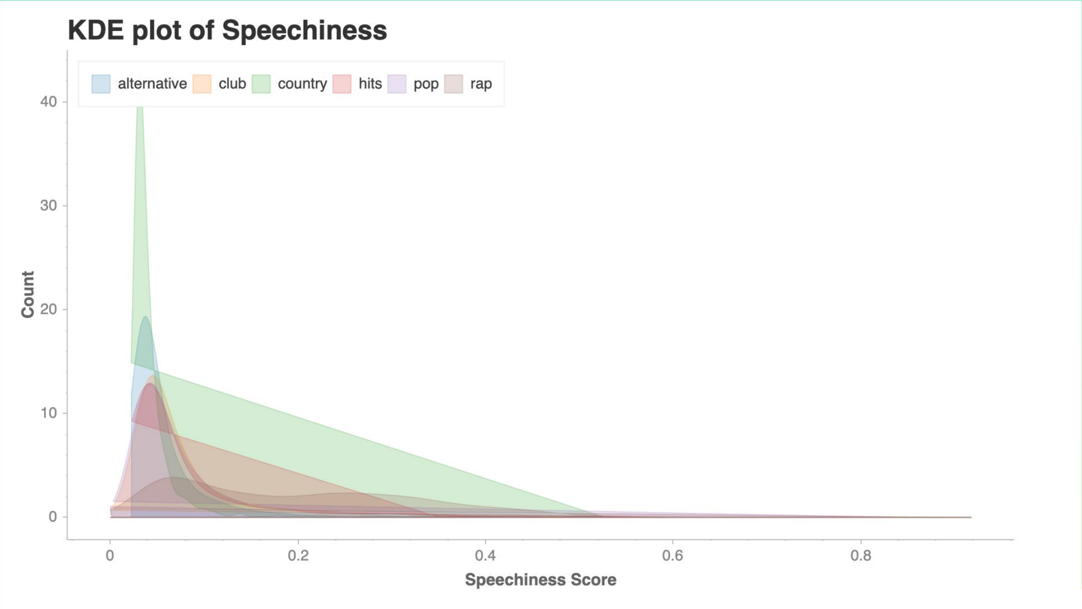
KDE plots





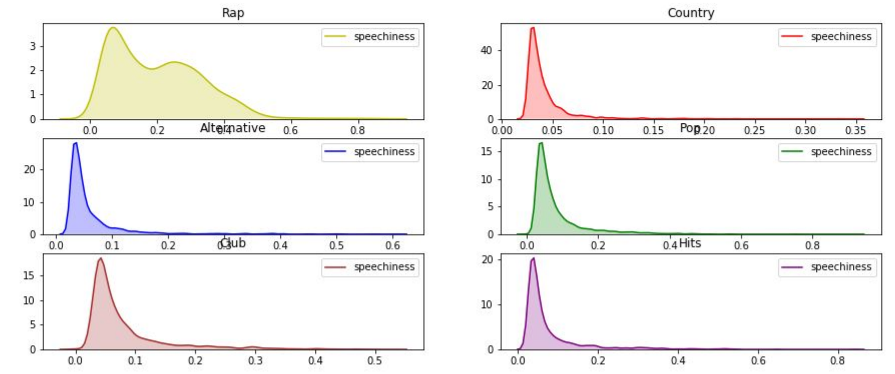
Unintuitive-- look at country vs rap
Look at labels, not colors

KDE plots

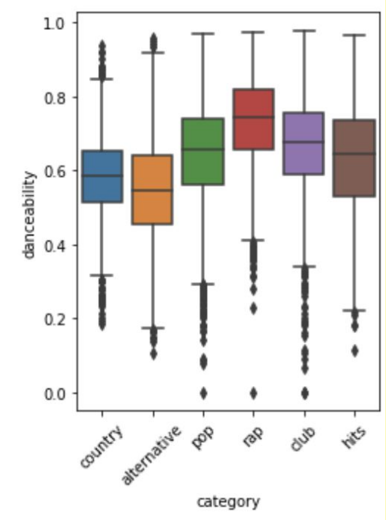


Most genres are right skewed

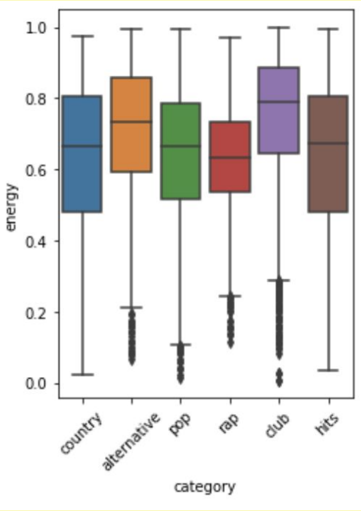
Look at labels, not colors



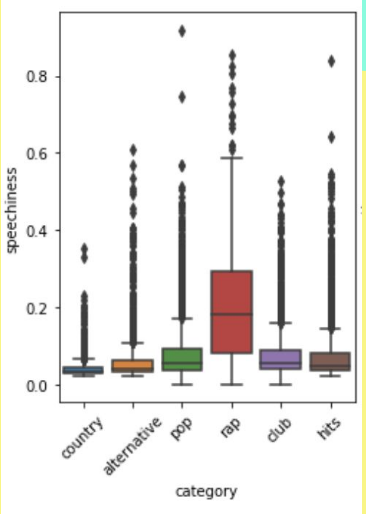
Boxplots



Dance

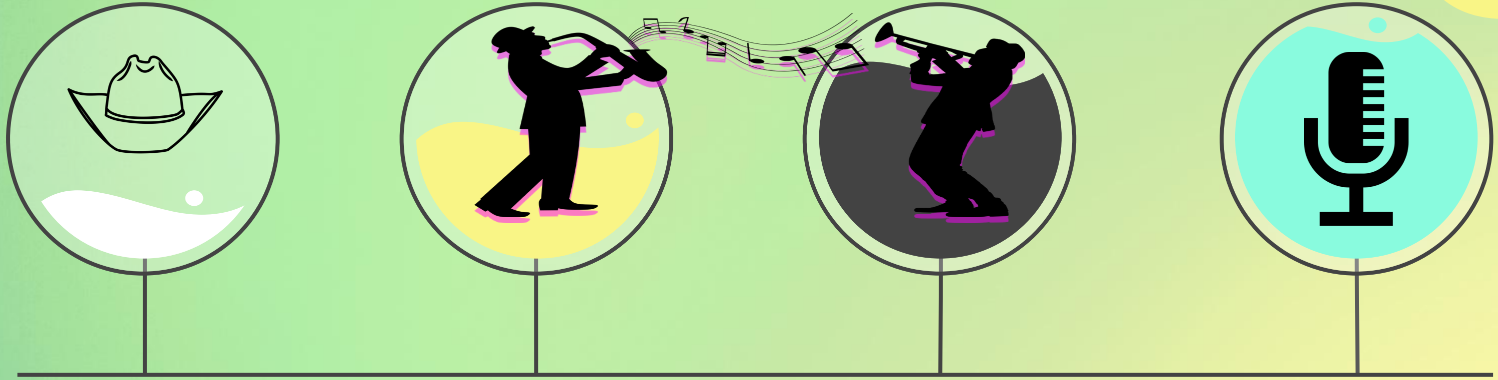


Energy



Speech

Correlation



Country

- \+ artist pop
- \+ featured artist
- \+ danceability
- \- acousticness

Alternative

- \+ artist pop
- \+ loudness
- \+ danceability
- \- tempo

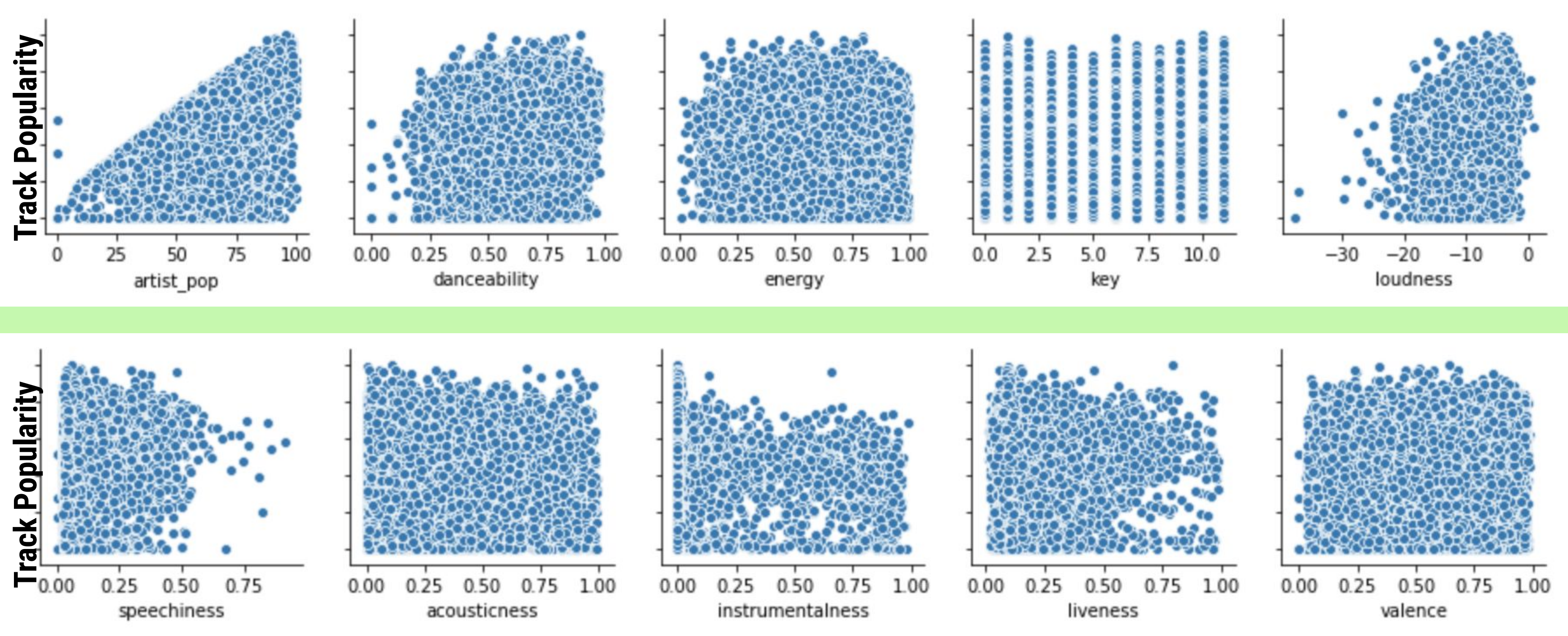
Rap

- \+ artist pop
- \+ loudness
- \+ featured artist
- \- instrumentalness

Pop

- \+ artist pop
- \+ danceability
- \+ loudness/featured artist
- \- instrumentalness

A Bad Omen...


Track Popularity
0
25
50
75
100
artist_pop
0.00
0.25
0.50
0.75
1.00
danceability
energy
0.0
2.5
5.0
7.5
10.0
key
−30
−20
−10
0
loudness
Track Popularity
speechiness
acousticness
instrumentalness
liveness
valence

Taylor Swift > Future
95 to 94

Modeling

Hid the ‘artist popularity’ variable because it is derived from track popularity

Track popularity is time dependent

Fall Out Boy < Panic! At The Disco
85 to 88

Saweetie = City Girls
77 tied

Sam Smith > Maggie Rogers
92 to 76

Lizzo < Ariana Grande
87 to 95

Justin Bieber > Beyoncé
99 to 90

Cardi B > Nicki Minaj
91 to 95

Models Run

- Linear Regression
- Logistic Regression
- K Nearest Neighbors
- Ridge Regression
- Lasso Regression
- Bagged Decision Trees
- Random Forests

Linear Regression

Root Mean Squared Error:
19.88

R Squared:
0.14

Lasso Regression

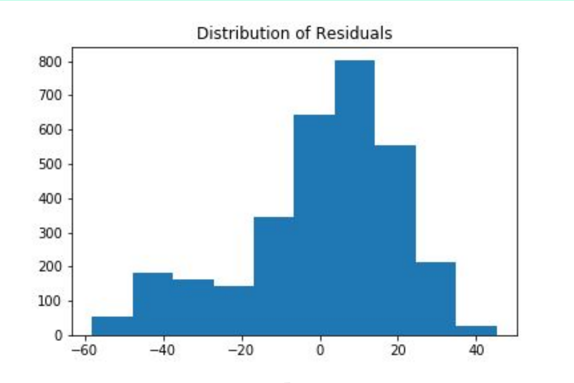
Root Mean Squared Error:
19.94

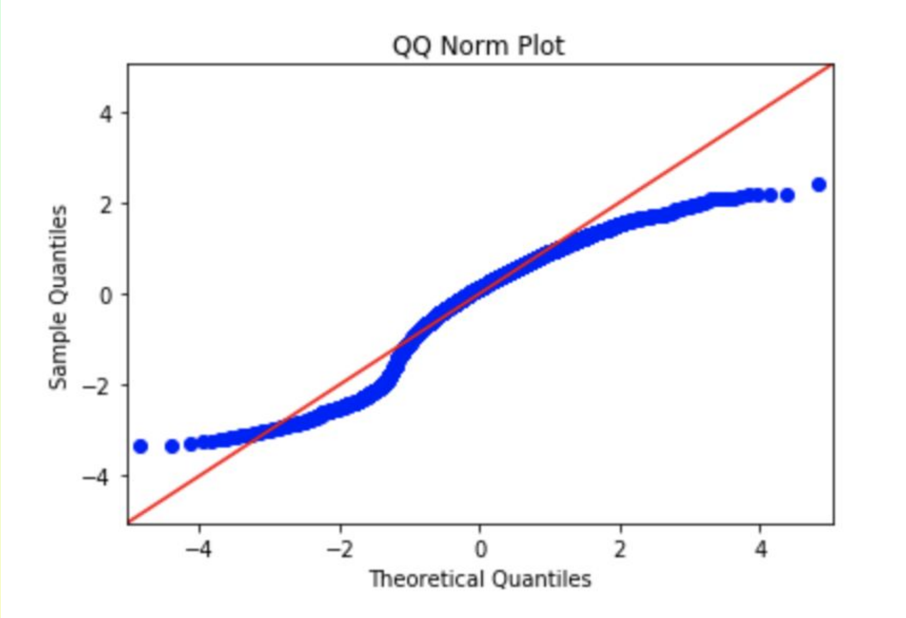
R Squared:
0.13

Ridge Regression

Root Mean Squared Error:
19.94

R Squared:
0.13


Distribution of Residuals
800
700
600
500
400
300
200
100
0
−60
−40
−20
0
20
40


QQ Norm Plot
4
2
0
−2
−4
Sample Quantiles
−4
−2
0
2
4
Theoretical Quantiles

	coef	std err	t	P>|t|	[0.025	0.975]
const	48.5928	3.850	12.622	0.000	41.046	56.139
featured_artist	4.0349	0.496	8.129	0.000	3.062	5.008
danceability	12.0320	1.761	6.834	0.000	8.581	15.483
energy	-5.0423	1.922	-2.623	0.009	-8.810	-1.275
key	0.0335	0.058	0.582	0.561	-0.079	0.146
loudness	0.8097	0.101	7.991	0.000	0.611	1.008
mode	-0.2896	0.458	-0.632	0.527	-1.188	0.609
speechiness	7.1720	2.582	2.778	0.005	2.111	12.233
acousticness	-0.9143	1.036	-0.882	0.378	-2.946	1.117
instrumentalness	-11.9163	1.181	-10.087	0.000	-14.232	-9.601
liveness	-1.0183	1.513	-0.673	0.501	-3.983	1.947
valence	-3.9807	1.042	-3.822	0.000	-6.023	-1.939
tempo	-0.0018	0.008	-0.238	0.812	-0.017	0.013
duration_ms	1.259e-06	4e-06	0.315	0.753	-6.58e-06	9.1e-06
time_signature	-0.2810	0.749	-0.375	0.708	-1.750	1.188
category_club	-2.4146	0.843	-2.864	0.004	-4.067	-0.762
category_country	0.3729	0.709	0.526	0.599	-1.017	1.763
category_hits	11.1593	0.802	13.913	0.000	9.587	12.732
category_pop	8.8633	0.718	12.342	0.000	7.456	10.271
category_rap	3.7228	0.855	4.353	0.000	2.046	5.399

Interpretation: At an alpha level of 0.05, we say that a one unit increase in featured artist, increases the track popularity score by about 4 points.

Next steps: See if that holds true when you continue to add on featured artists


Neural Networks

Fully Connected Feed Forward

No Regularization

Root Mean Squared Error:
10.48
R Squared:
0.76

L2 Regularization

Root Mean Squared Error:
10.59
R Squared:
0.75

50 % Dropout

Root Mean Squared Error:
10.22
R Squared:
0.77

L2 and Dropout

Root Mean Squared Error:
9.37
R Squared:
0.81

Conclusions

Did not have high predictive power for interpretable models

'Normal' Linear Regression model performed best, other than neural networks

Featured Artists seem to be most important

Future Work

More Data!

Include not-so-popular songs

Featured Artists

Is track popularity proportional to number of featured artists?

Fine Tuning

Perform more in depth tuning of parameters

Recommender System

Input a song name, recommend 5 songs

Resoucres:

- https://slidesgo.com/theme/music-app-pitch-deck
- https://spotipy.readthedocs.io/en/latest/#installation
- https://developer.spotify.com/documentation/web-api/
- https://developer.spotify.com/
- https://developer.spotify.com/documentation/web-api/reference/tracks/get-several-audio-features/
- http://pansentient.com/2009/09/spotify-song-popularity/
- Software engineer, friend, Ada
- Color guru, friend, Clara

Thank You!

CREDITS

This is where you give credit to the ones who are part of this project.

Did you like the resources on this template? Get them for free at our other websites.

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Images created by Freepik
- Author introduction slide photo created by Freepik
- Text & Image slide photo created by Freepik.com (and Hadley)