

2017

2018

2019

2020

2021

<https://bit.ly/SxB5YA>

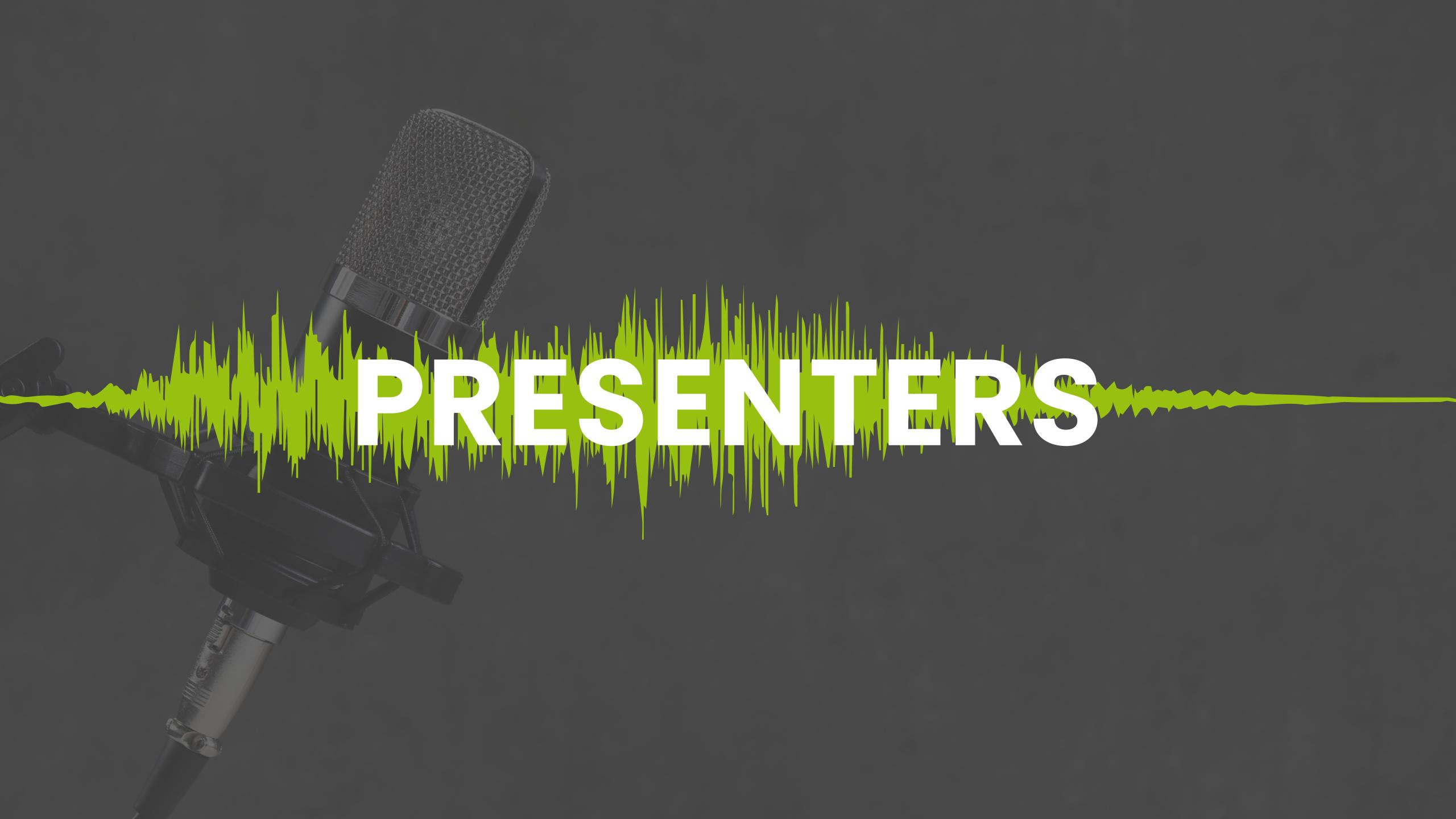


Good Vibrations?

Spotify x Billboard Top 200 Albums 5-Year Analysis Project

Bianca Serrano and Katie Ravenwood



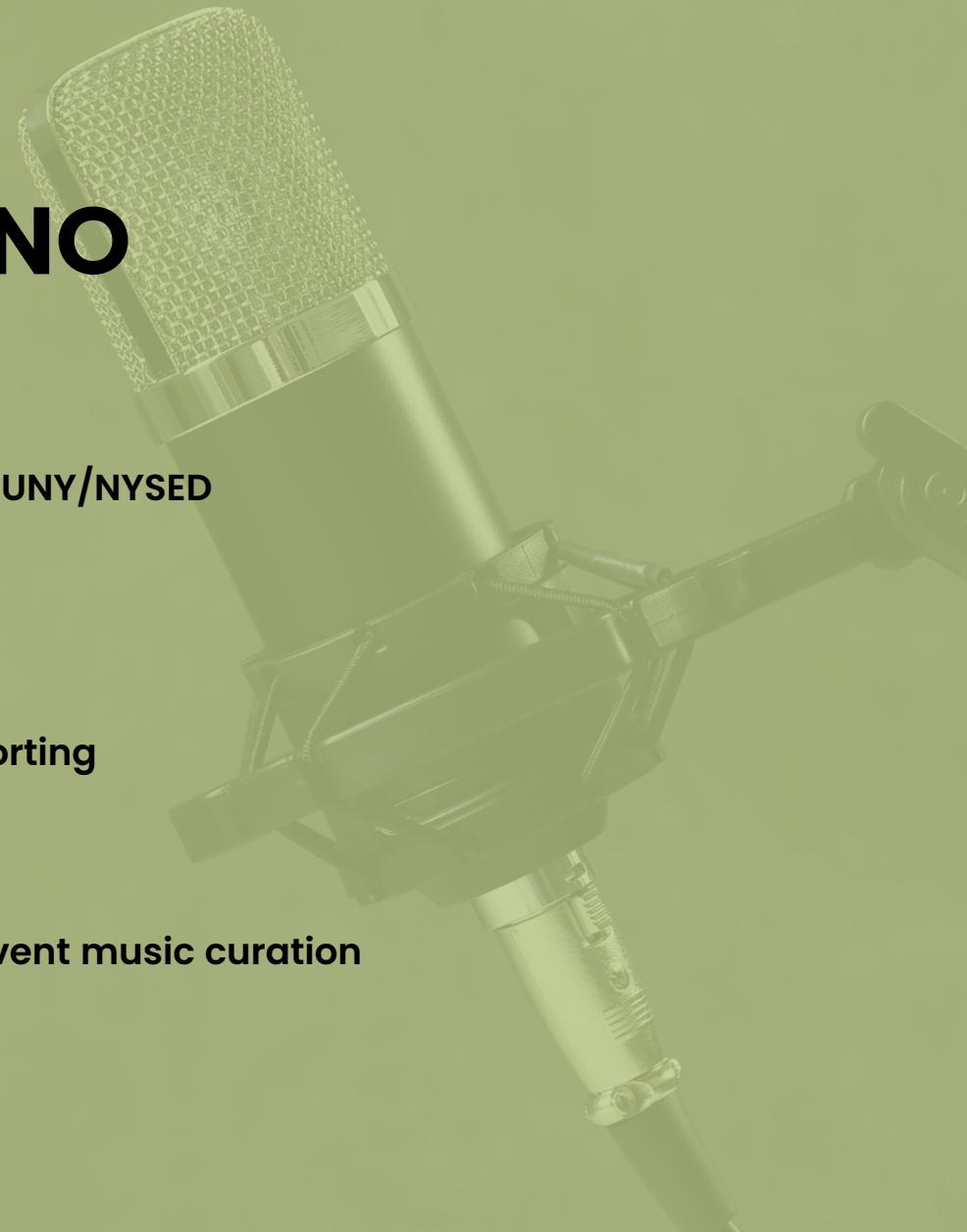


PRESENTERS

PRESENTERS

BIANCA SERRANO

- Educational background in psychology
- Currently employed as a data coordinator for CUNY/NYSED
- 5+ years experience in database management
- Skills in:
 - Data Wrangling
 - Data synthesis for monthly and annual reporting
 - Administration
 - Communication
- Interested in music through a background in event music curation and active listening



PRESENTERS

KATIE RAVENWOOD

- Educational background in music & biochemistry
- Currently employed as a musician & music teacher
- 25 years experience in music performance and music education
- 20 years experience in web development & marketing
- Skills in:
 - Business administration and marketing
 - Contracting and contract negotiation
 - Front end design and development
 - Educational and curriculum design
- Interest in music through a background in music performance and concert promotion



A pair of black over-ear headphones is shown from a slightly elevated angle, facing towards the viewer. A vibrant green soundwave graphic is overlaid across the center of the image, starting from the left ear cup and ending at the right ear cup. The soundwave is composed of numerous vertical spikes of varying heights, creating a textured, digital representation of audio. The background is a solid dark gray.

BACKGROUND

**“If you want to find the secrets of the universe,
think in terms of energy, frequency and vibration.”**

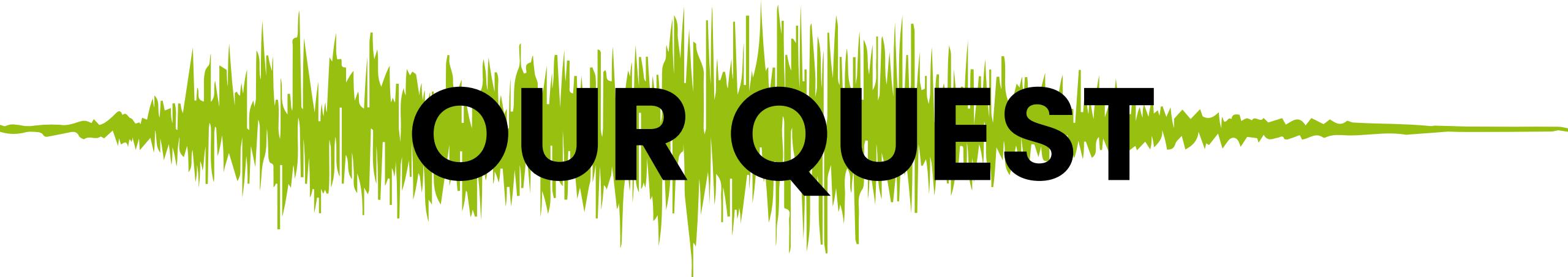
- Attributed to Nicola Tesla

METHODS

```
import java.util.Scanner;
import java.util.ArrayList;
import java.util.List;

public class Main {
    public static void main(String[] args) {
        Scanner keyboard = new Scanner(System.in);
        System.out.print("Enter start city: ");
        String start = keyboard.nextLine();
        System.out.print("Enter goal city: ");
        String goal = keyboard.nextLine();
        ArrayList<String> pathForWiles = new ArrayList<String>();
        ArrayList<String> airlinesvisited = new ArrayList<String>();
        if(cantdead(start, goal, pathForWiles, airlinesvisited)){
            System.out.println("Path to reach " + goal + " is: " + pathForWiles);
        } else{
            System.out.println("Cannot convert cities from " + start + " to " + goal);
        }
    }

    private static boolean cantdead(String current, String goal,
                                    ArrayList<String> pathForWiles, ArrayList<String> airlinesvisited){
        if(current.equals(goal)){
            //base case 1, I have found a path!
            pathForWiles.add(current);
            return true;
        }
        else if(airlinesvisited.contains(current)){
            // base case 2, I have already been here
            // don't go into a cycle
            return false;
        } else{
            // I have not bee
```



OUR QUEST

CHANGES OVER TIME

Have there been any significant changes in music over the past five years?

Are there other significant features that change over time?

Do any of the features have similarities or contrasts?

GENRE PRESENCE

Have there been any changes
in genre presence on the
charts? Which genres are
changing and how?

MOOD

Has mood of the most
popular music changed
over the past five years?
How has it changed?

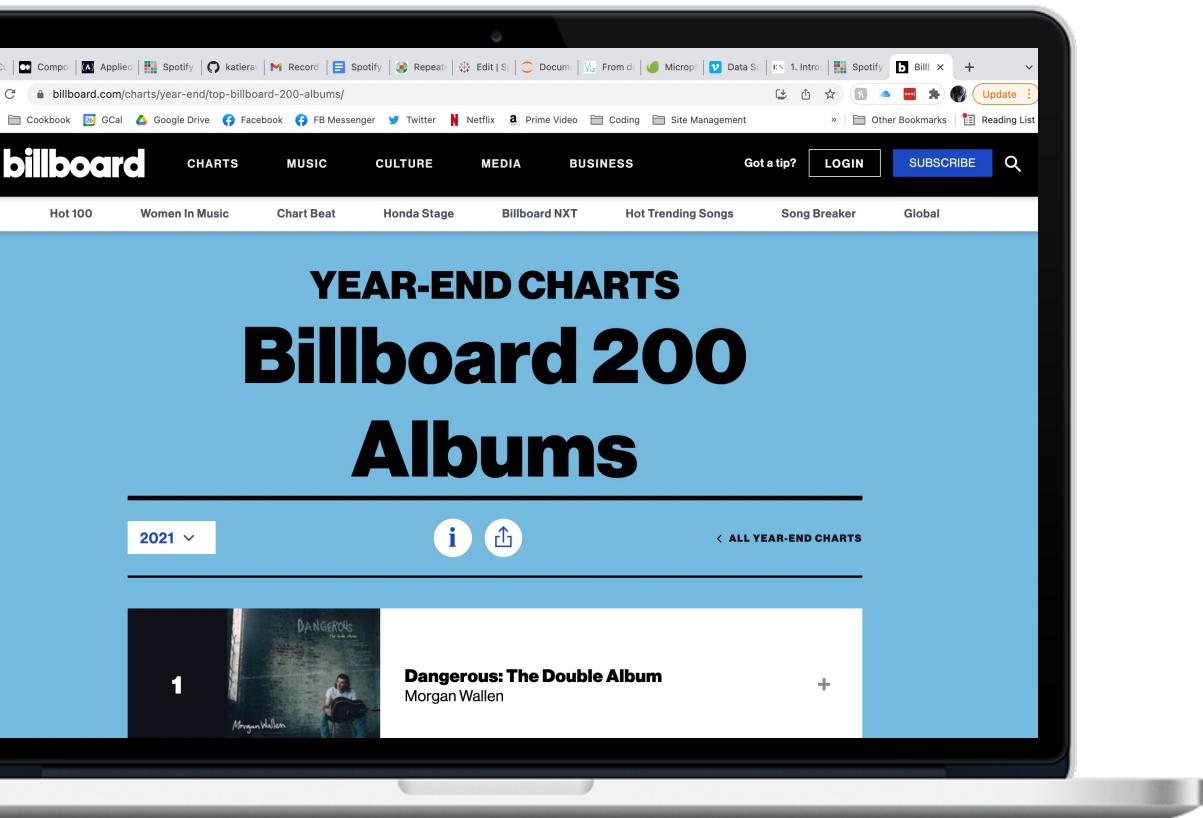
DATASET SELECTION

MUSIC DATA FOR ANALYSIS?

Available datasets were:

- **Too old**
- **Wrong scope**
- **Restricted access**
- **Training datasets with errors**

billboard



Billboard is an American music and entertainment magazine founded in 1894 in Cincinnati, Ohio and now operating in several countries around the world.



Billboard is best known for its weekly charts, including the Billboard 200 albums ranking and the Hot 100 singles chart, as well as its yearly 200 Albums chart.



The Top 200 Album chart lists the year's most popular albums across all genres, ranked by physical and digital album and track sales as well as audio on-demand streaming activity.



Billboard charts are powered by information from MRC Data (formerly Nielsen Music or Nielsen Soundscan.)



HISTORY

Spotify was founded in 2006 in Stockholm, Sweden, by Daniel Ek and Martin Lorentzon.



USERS

365 million people use Spotify at least once a month, and 165 million are paid subscribers.



CATALOG

More than 70 million songs are available on Spotify, and the company has reported that more than 60,000 new tracks are now uploaded to Spotify every day.



API

Data is available to developers through Spotify's API, and includes information about artists, albums, and tracks as well as user data on song popularity and streaming statistics.

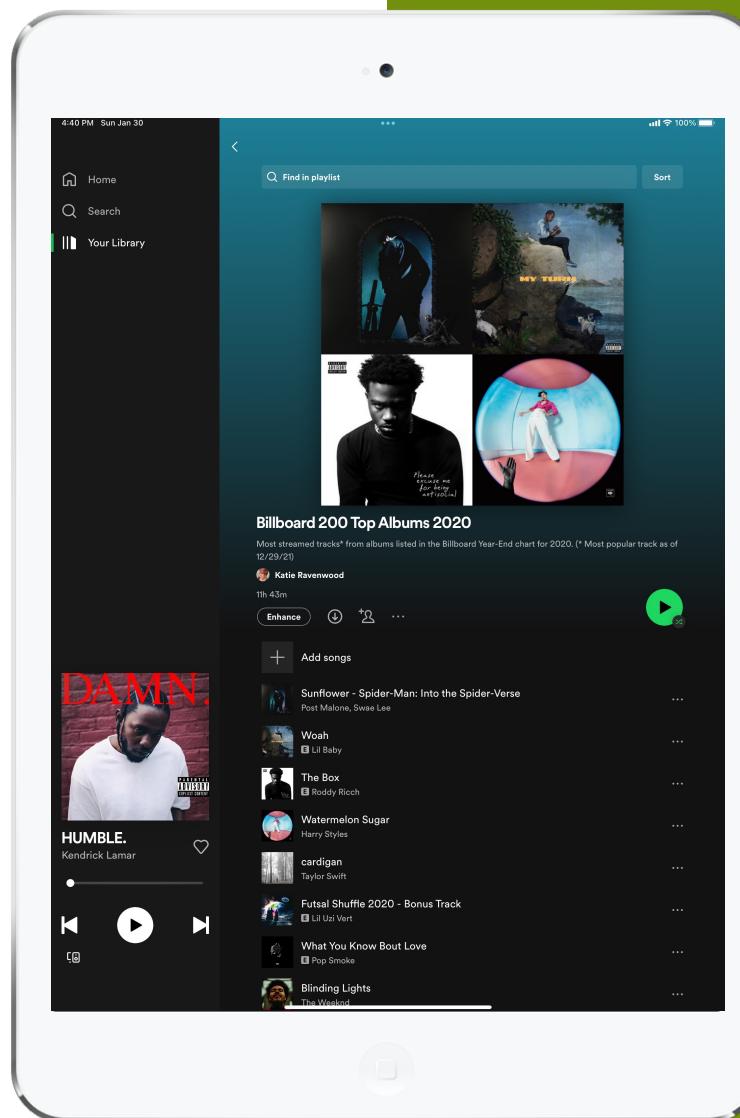


SPOTIFY API

Playlists of each chart were created in Spotify to provide endpoints for accessing chart data.

SPOTIFY PLAYLISTS

- 01. Billboard 200 Top Albums 2017**
- 02. Billboard 200 Top Albums 2018**
- 03. Billboard 200 Top Albums 2019**
- 04. Billboard 200 Top Albums 2020**
- 05. Billboard 200 Top Albums 2021**



SPOTIFY DEVELOPER PLATFORM

The Spotify Developer Platform provides access to calculated audio features of tracks such as danceability, energy, valence, and more. For more advanced use cases, it is possible to read in-depth analysis data about tracks such as the segments, tatusms, bars, beats, pitches, and more.

Spotify Web API endpoints return JSON metadata about music artists, albums, and tracks, directly from the Spotify Data Catalogue.

DATASET CREATION

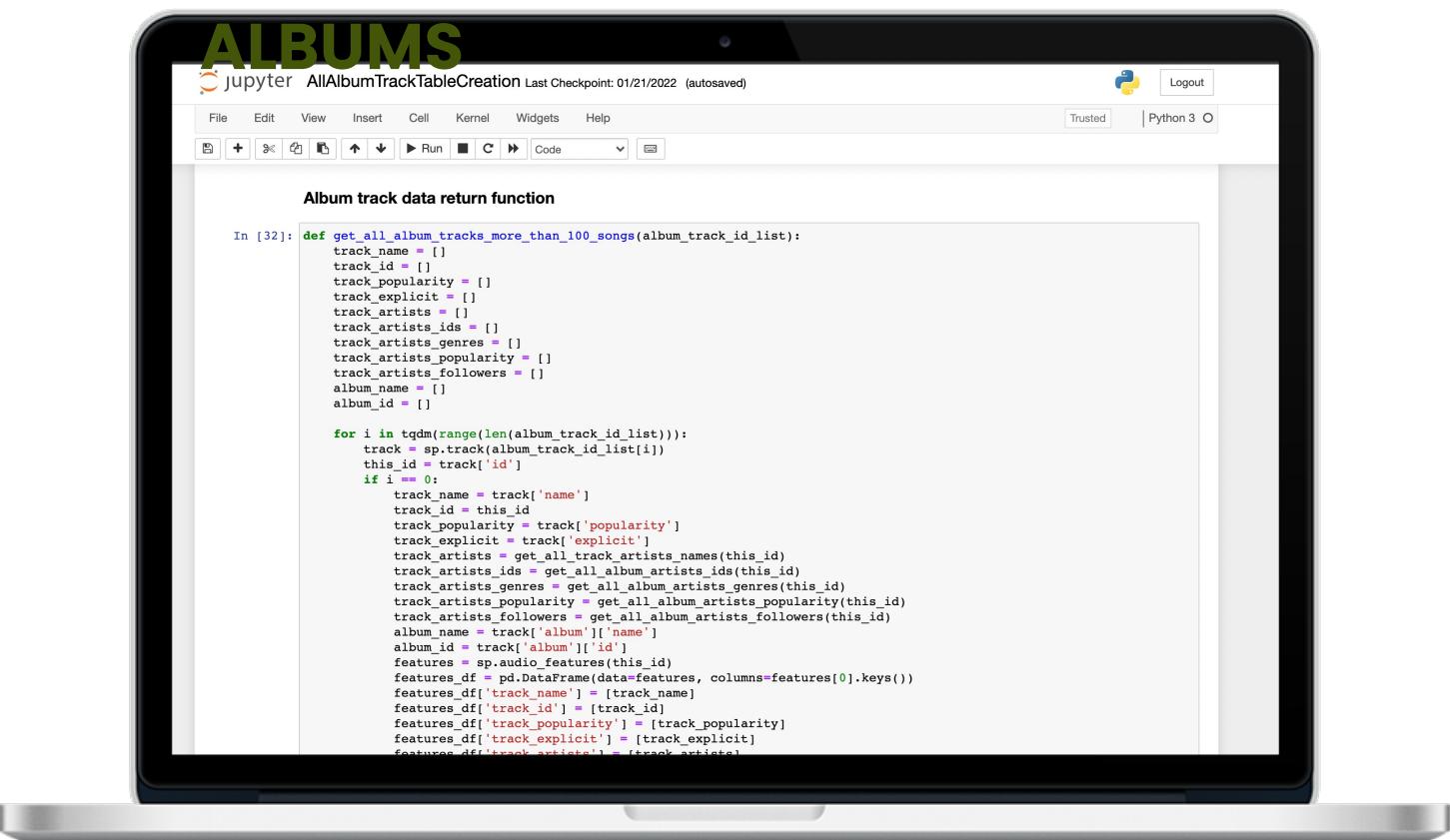
PYTHON & SPOTIPY

Data tables of album and track information were created with Python using the SpotiPy package.

Included variables:

- Billboard album chart name & year
- Album ID, name, release year
- Album artists' names, IDs, popularity, and associated genres
- Track IDs, popularity, and statistics
- Explicit labeling designation
- Audio features

8996 TRACKS FROM 560



The image shows a laptop screen displaying a Jupyter Notebook interface. The title bar of the notebook says "ALBUMS". The notebook has a single cell containing Python code. The code is a function named "get_all_album_tracks_more_than_100_songs" which iterates through a list of album track IDs. For each track, it retrieves the track's name, ID, popularity, explicitness, and artists. It then retrieves the album's name and ID, and the artists' names, IDs, and genres. Finally, it retrieves the track's audio features. The code uses several libraries including tqdm, requests, and pandas.

```
def get_all_album_tracks_more_than_100_songs(album_track_id_list):
    track_name = []
    track_id = []
    track_popularity = []
    track_explicit = []
    track_artists = []
    track_artists_ids = []
    track_artists_genres = []
    track_artists_popularity = []
    track_artists_followers = []
    album_name = []
    album_id = []

    for i in tqdm(range(len(album_track_id_list))):
        track = sp.track(album_track_id_list[i])
        this_id = track['id']
        if i == 0:
            track_name = track['name']
            track_id = this_id
            track_popularity = track['popularity']
            track_explicit = track['explicit']
            track_artists = get_all_track_artists_names(this_id)
            track_artists_ids = get_all_album_artists_ids(this_id)
            track_artists_genres = get_all_album_artists_genres(this_id)
            track_artists_popularity = get_all_album_artists_popularity(this_id)
            track_artists_followers = get_all_album_artists_followers(this_id)
            album_name = track['album'][name]
            album_id = track['album'][id]
            features = sp.audio_features(this_id)
            features_df = pd.DataFrame(data=features, columns=features[0].keys())
            features_df['track_name'] = [track_name]
            features_df['track_id'] = [track_id]
            features_df['track_popularity'] = [track_popularity]
            features_df['track_explicit'] = [track_explicit]
            features_df['track_artists'] = [track_artists]
```

DATA WRANGLING



JOINS

TRACK & ALBUM TABLES

Multiple joins for various charts and analyses



FEATURES

CLEANING & DUMMY CODING

Removing null values and dummy coding genres for analysis



GENRES

UN-NESTING LISTS

Wrangling nested lists of genres linked by artist



DATES

RECODING TO DATETIME

Recoding release dates and chart years to datetime format

RESULTS



he Beatles Chris Stapleton Khalid Kodak Black Juice WRLD
The Beatles AC/DC Logic Elton John DJ Khaled
Twenty One Pilots Blake Shelton Queen Frank Ocean
Lady Gaga Summer Walker Playboi Carti Chris Brown H.E.R.
mas Rhett Ed Sheeran Tory Lanez Five Finger Death Punch
Gunn Adele Michael Jackson NAV Bruno Mars Ba
Halsey Bryson Tiller Mac Miller Smoke Tory Lanez
Eilish DaBaby Trippie Redd Pop Smoke H.E.R.
ingBoy Never Broke Again Migos Imagine Dragons Zac Brown
Lil Uzi Vert Sam Smith Leslie Odom Jr. Lil Bab
Bob Seger Polo G Luke Bryan Emir
Pentatonix Kane Brown Linkin Park Kany
BTS Red Hot Chili Peppers Luke Combs Fut
Shawn Mendes SZA Leslie Odom Jr. Tyler, The Creator
Cole Rod Wave Carrie Underwood Bob Marley & The Wailers
Kendrick Lamar Justin Bieber Ali'i Cravalho
Nirvana Panic! At The Disco Red Hot Chili Peppers
SZA Carrie Underwood Bob Marley & The Wailers
Travis Scott The Notorious B.I.G.

the weeknd

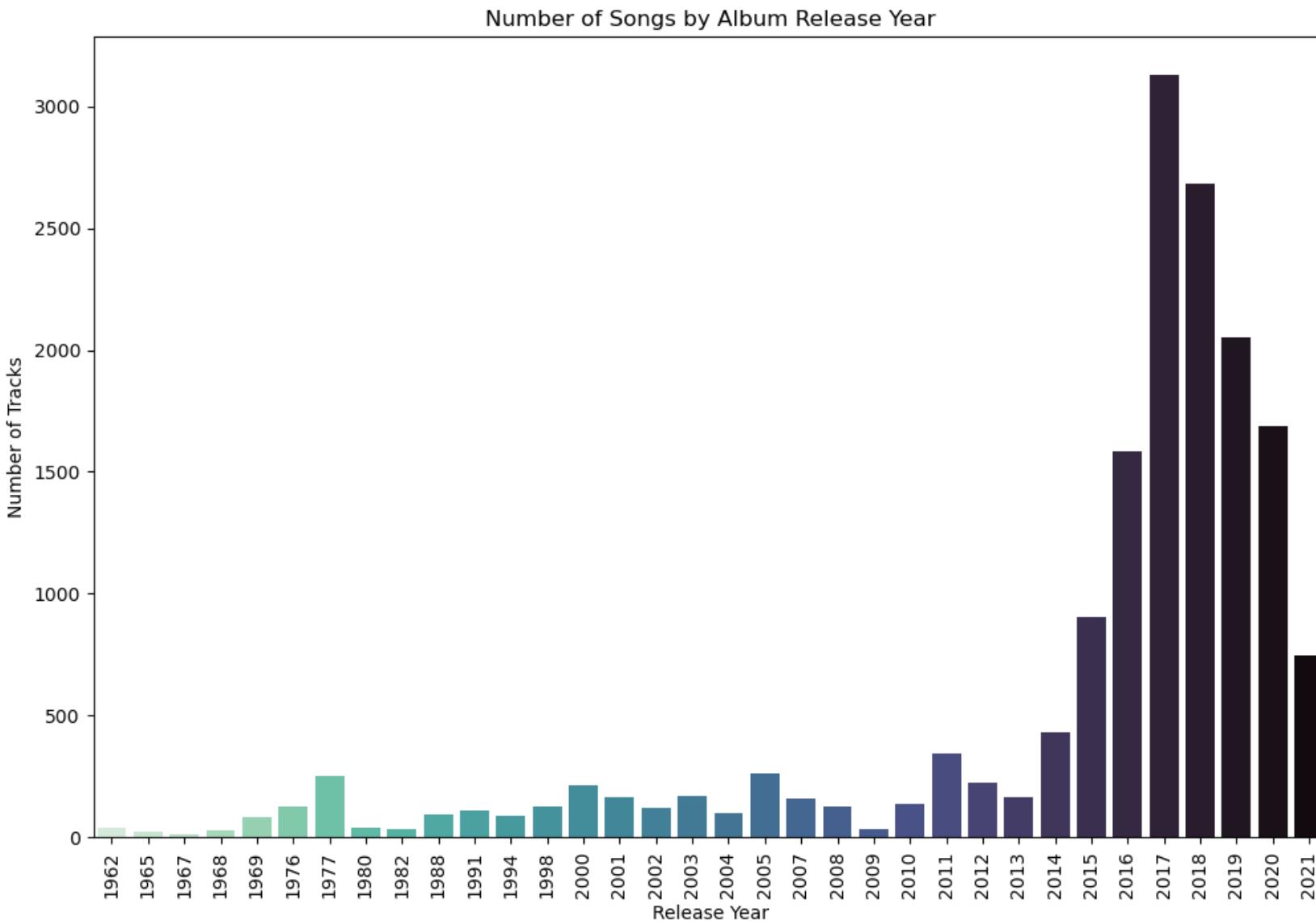
Drake

ost Malone

Kendrick Lamar

Interesting Data Points

- Shape of the tracks dataframe: 16,501 rows and 57 variables
- Most popular song: "good for you" by Olivia Rodrigo and "Woman" by Doja Cat
- Average song length: 3 minutes and 22 seconds
- Longest song: "Like I Never Left" by Future at 25 minutes 33 seconds
- Most frequently appearing album artist: Drake (31 times)
- Appearances by Taylor Swift (15), Queen (7), and Elton John (4)
- Oldest track: "The Christmas Song (Merry Christmas To You)" released in 1962
- Newest track: "Bad Morning" by Young Boy Never Broke Again released on October 5, 2021.



NUMBER OF TRACKS BY RELEASE YEAR

AUDIO FEATURES

ACOUSTICNESS (0 → 1)

A confidence measure of how likely it is that the track is performed acoustically (as opposed to amplified.)



DANCEABILITY (0 → 1)

How suitable a track is for dancing based on elements including tempo and beat strength and overall regularity.



ENERGY (0 → 1)

Measures intensity and activity in a track, influenced by dynamics, loudness, timbre, onset rate, and general entropy.



INSTRUMENTALNESS (0 → 1)

Confidence measure that predicts whether a track contains no vocals (oohs and aahs are considered non-vocal sounds.)



KEY (-1 to 11)

The key the track is in. Integers map to pitches using standard Pitch Class notation.



LIVENESS (0 → 1)

Detects the presence of an audience in the recording. Greater than 0.8 is likely recorded live.



LOUDNESS (-60 to 0)

Average loudness of a track's normalization level in decibels across the entire track.



MODE (1 or 0)

Indicates whether a track is in major (1) or minor (0) mode.



SPEECHINESS (0 → 1)

Detects the presence of spoken words in a track. Closer to 1 means higher spoken word presence.



TEMPO (BPM)

Beats-per-minute measure of the speed or pace of a track derived from beat duration.



TIME SIGNATURE (3 to 7)

Specifies how many beats in each measure or bar of music.



VALENCE (0 → 1)

Measures the musical positiveness of a track. Higher values equal more cheerfulness.

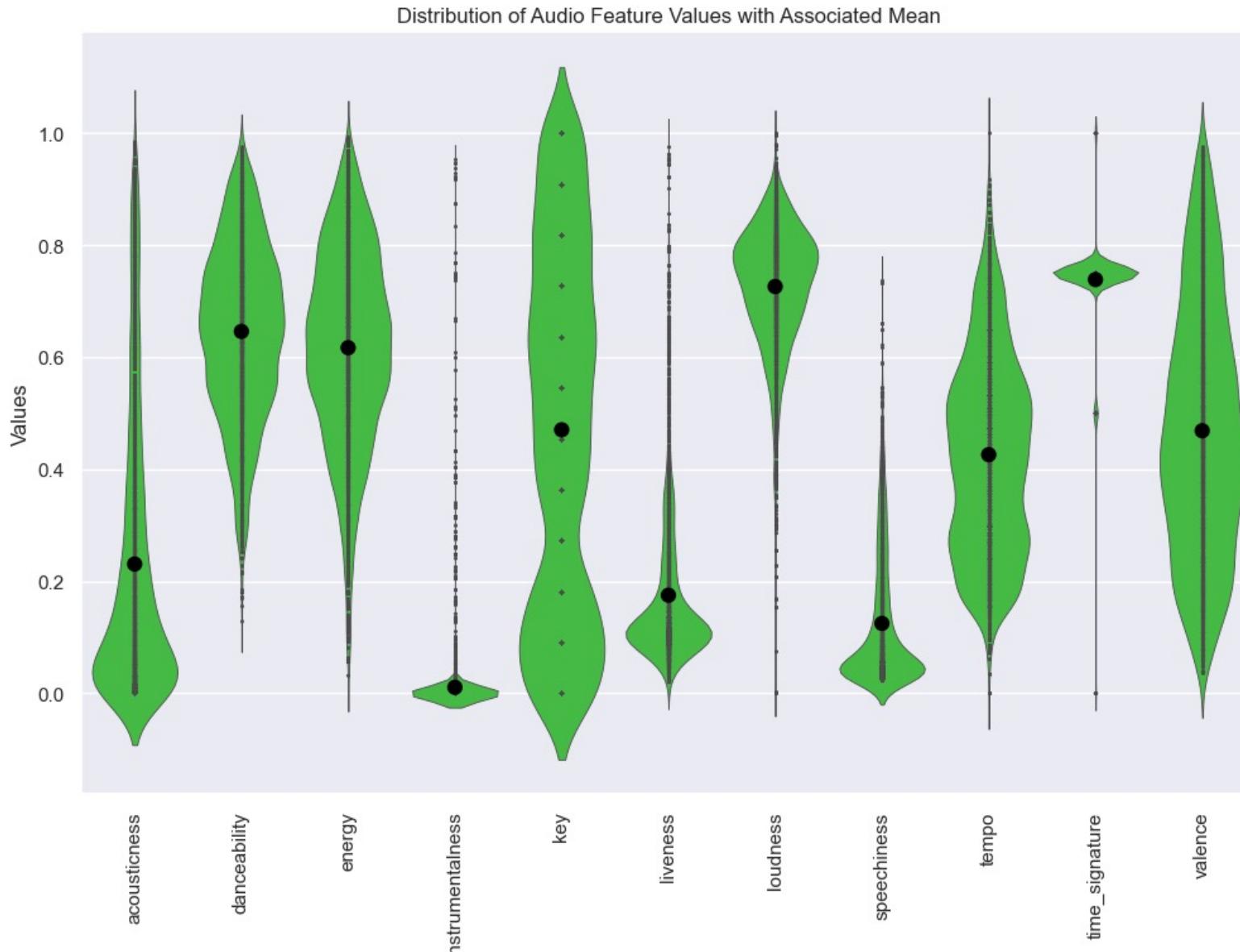


EXPLICIT (T/F)

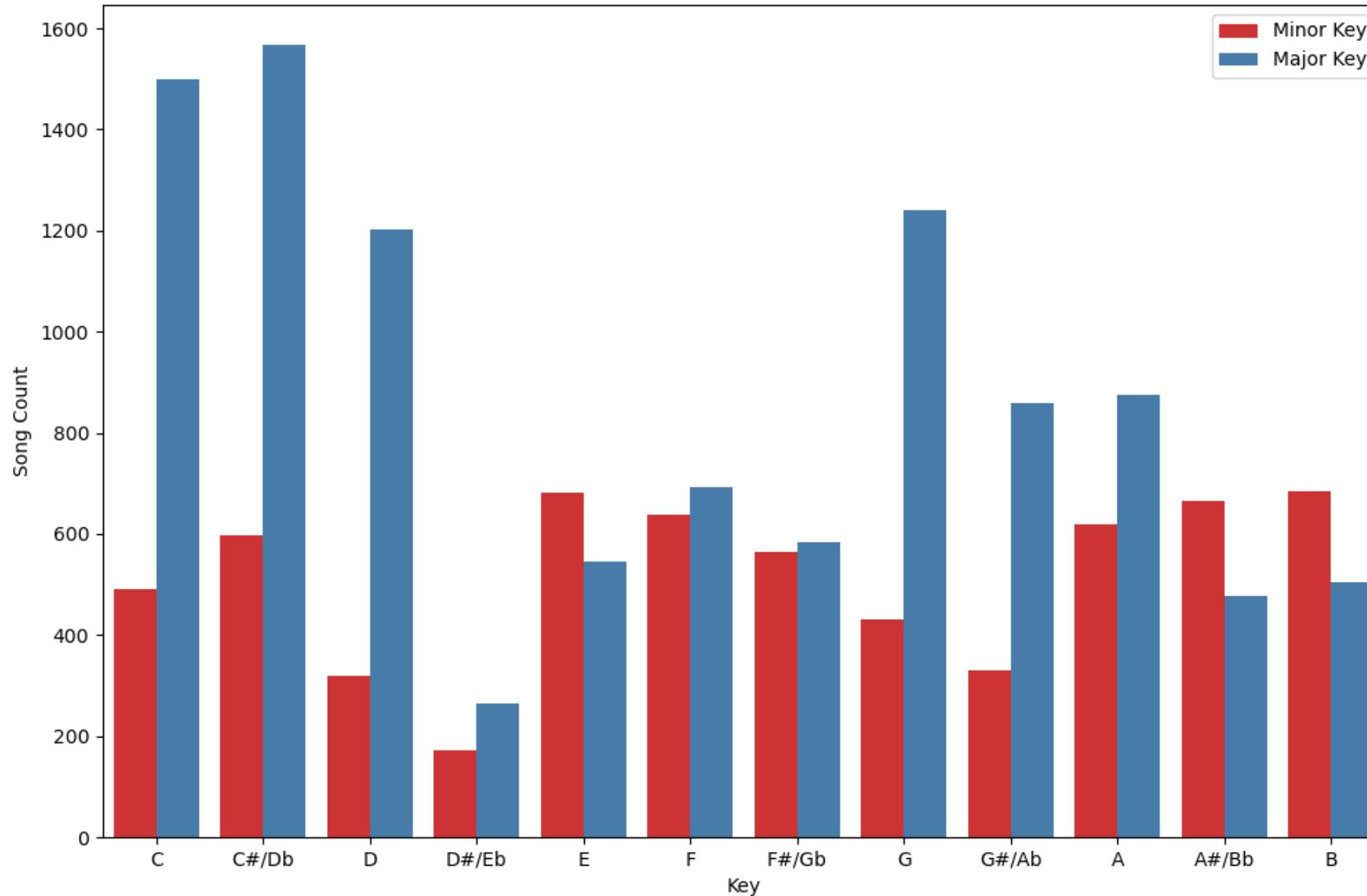
Indicates whether the track has been marked for Parental Advisory labelling. False indicates not marked or no information given.



DISTRIBUTION OF AUDIO FEATURE VALUES



Count of Songs in Major and Minor Keys Across All Album Tracks

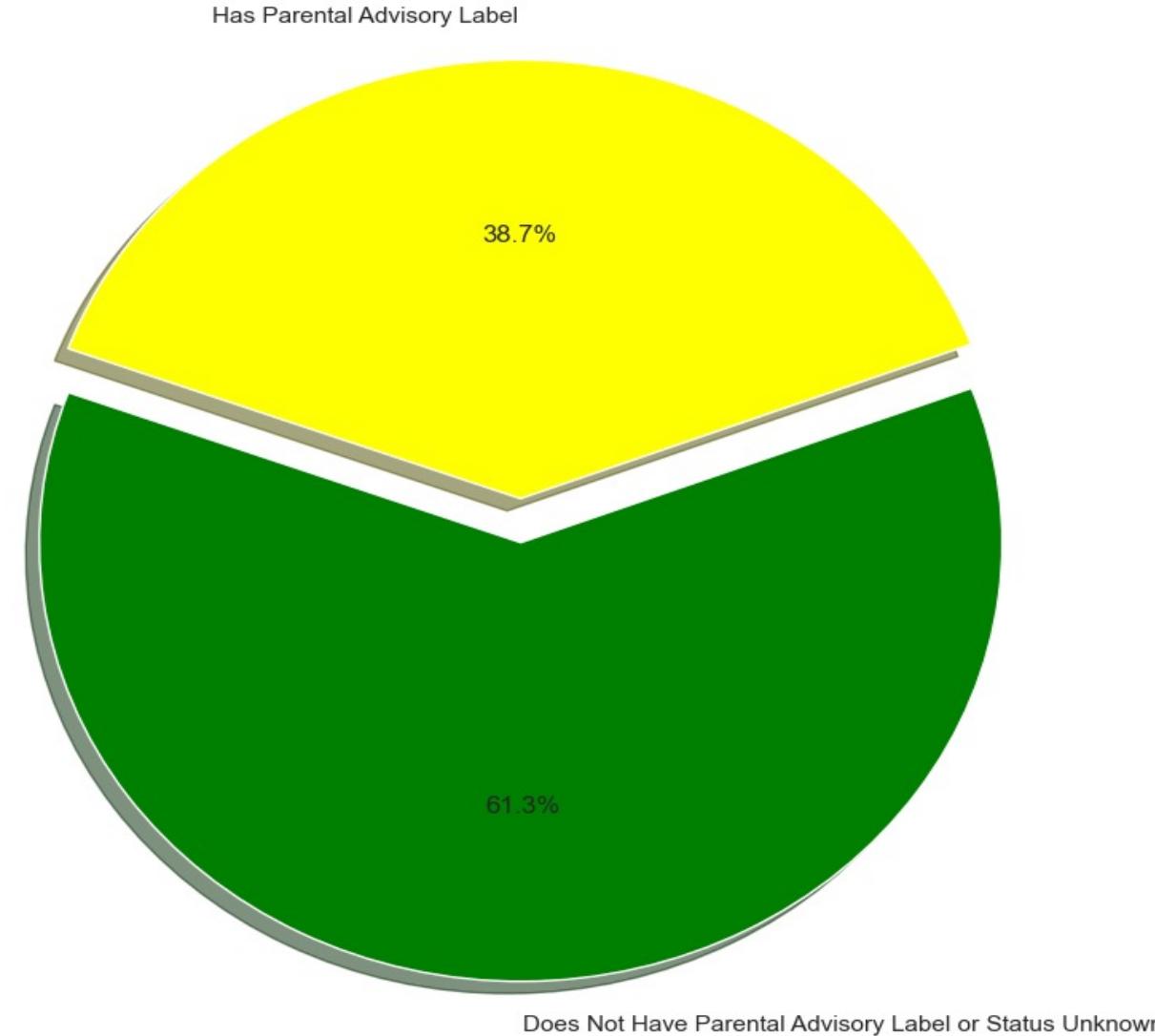


TRACK COUNT BY MAJOR & MINOR KEY

PARENTAL ADVISORY LABELING

- ▶ **Result of Parents Music Resource Center (PMRC) Congressional Hearings in 1985**
- ▶ **Labels are voluntarily applied by record labels or artists**
- ▶ **Album-wide labeling**
- ▶ **No set standard for labeling**
- ▶ **Labeling assessment inconsistent**
- ▶ **Missing data: non explicit, non-labeled & unknown status**

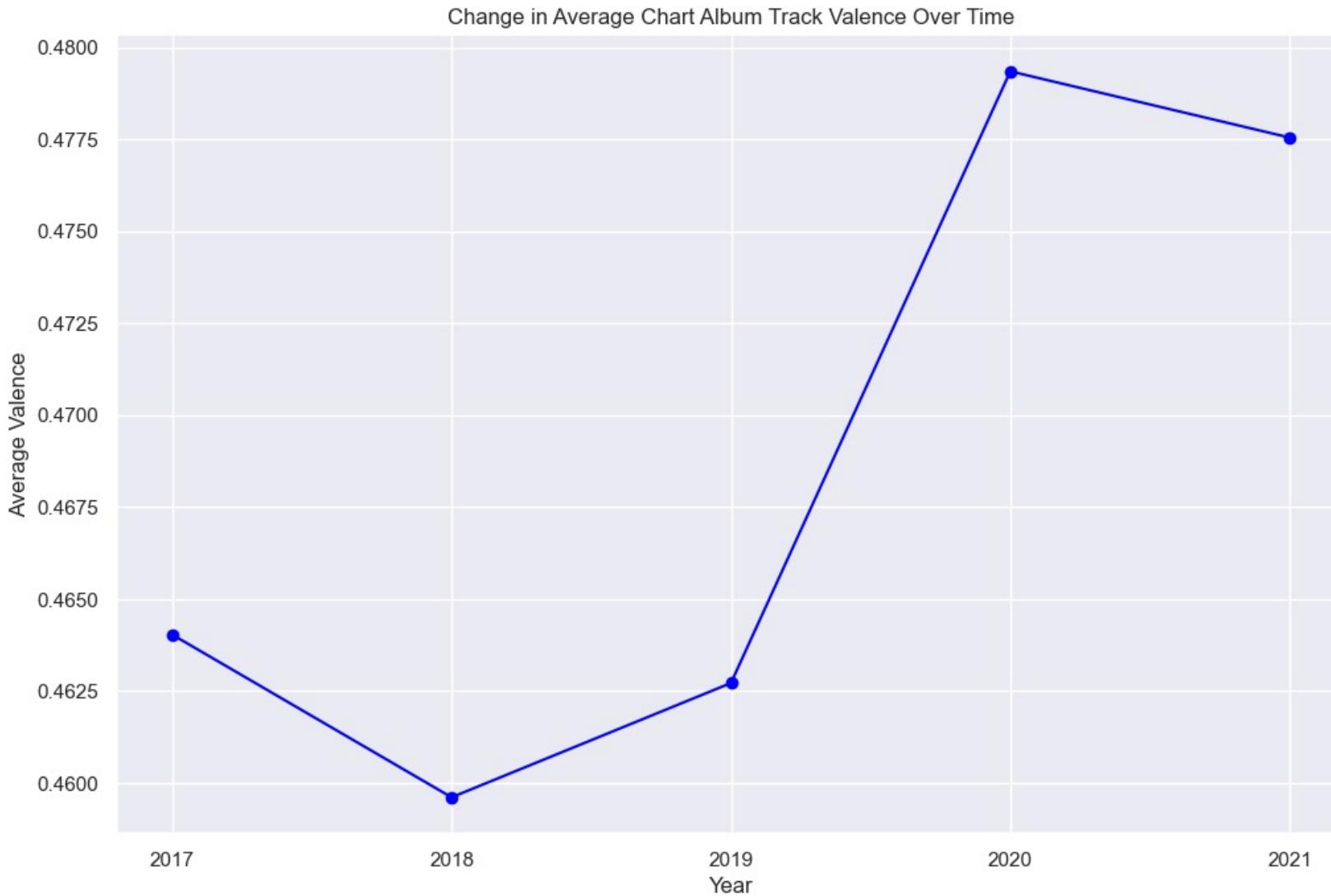
Percentage of Chart Album Tracks with Parental Advisory Label



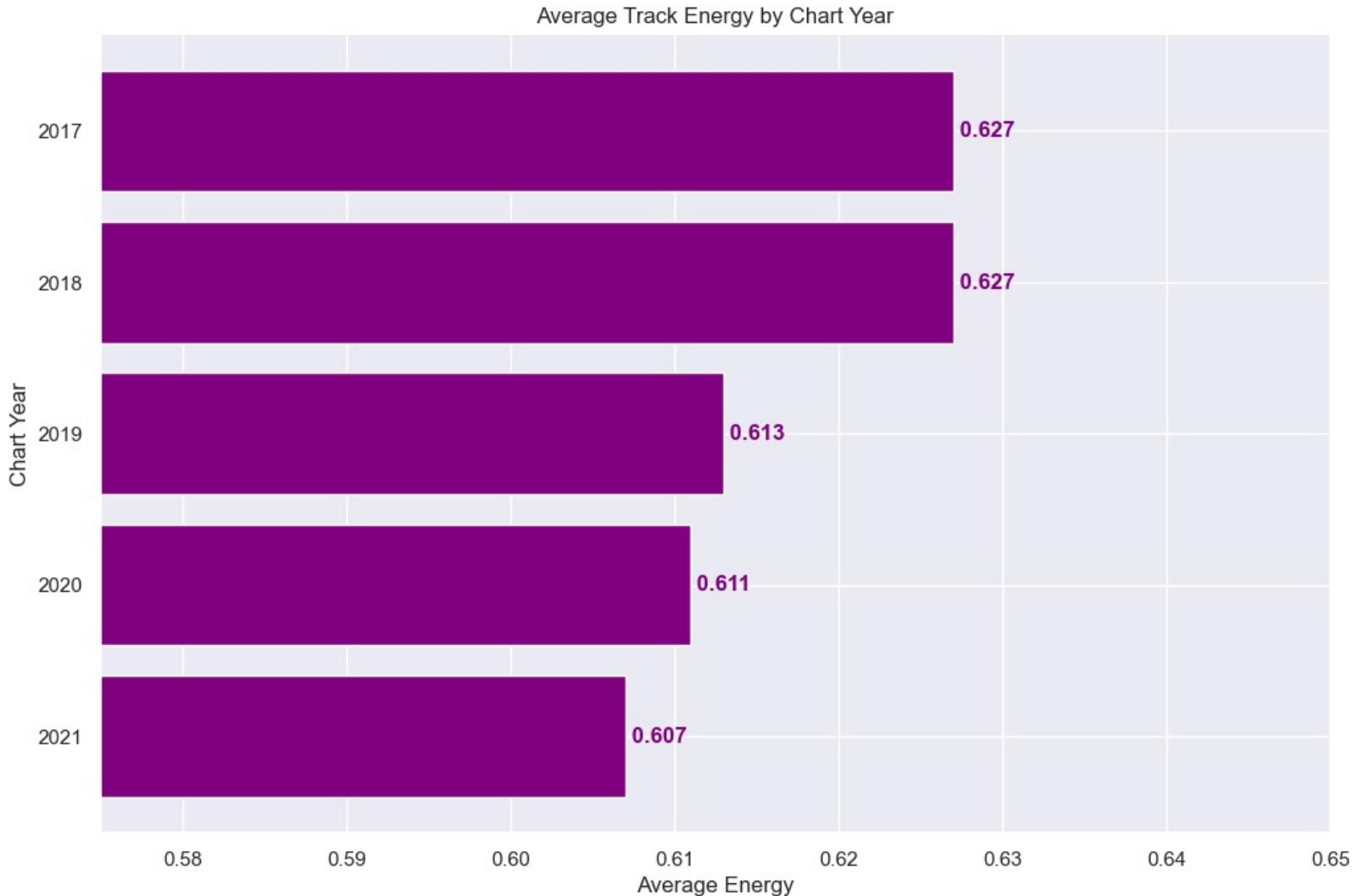
Source: "Parental Advisory Label – RIAA". RIAA, 13 January 2022,
<https://www.riaa.com/resources-learning/parental-advisory-label/>

Does Not Have Parental Advisory Label or Status Unknown

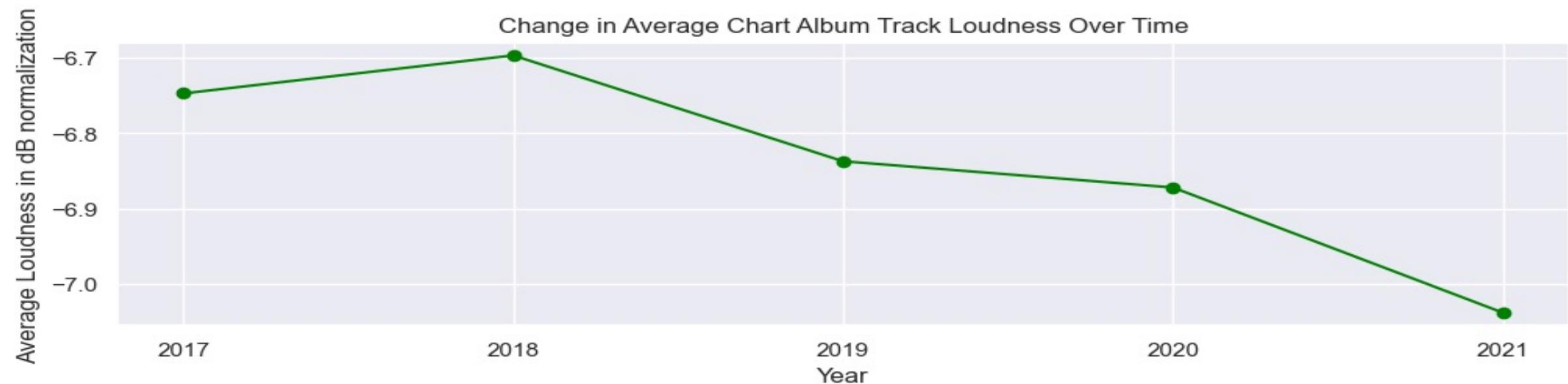
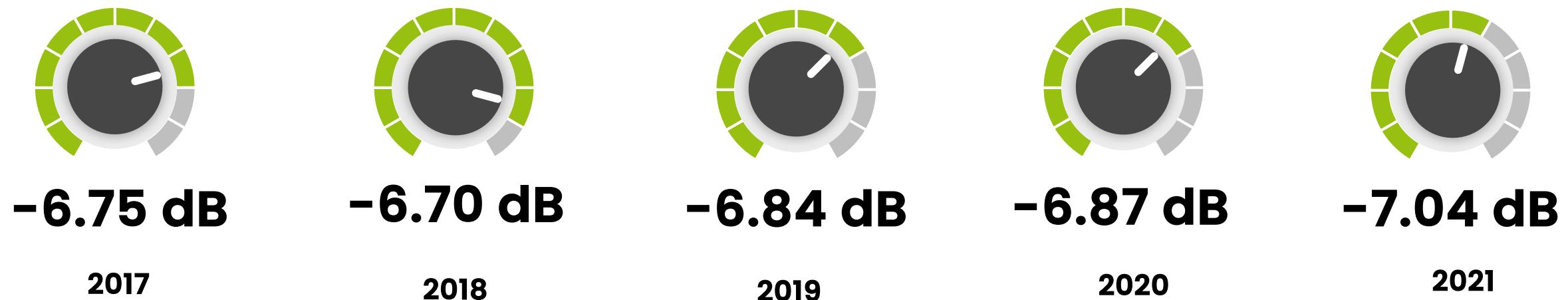
CHANGE IN VALENCE OVER TIME



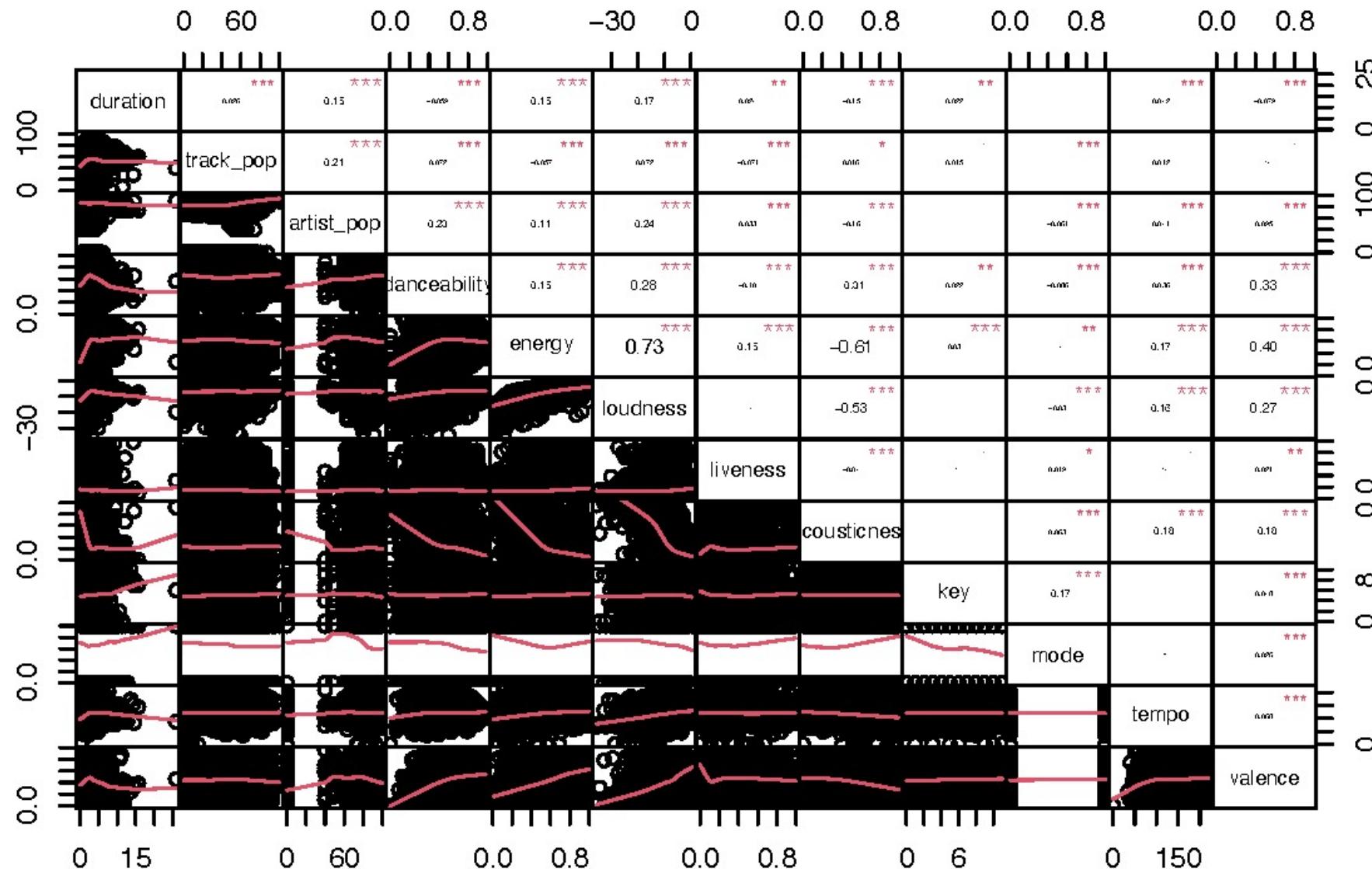
ENERGY



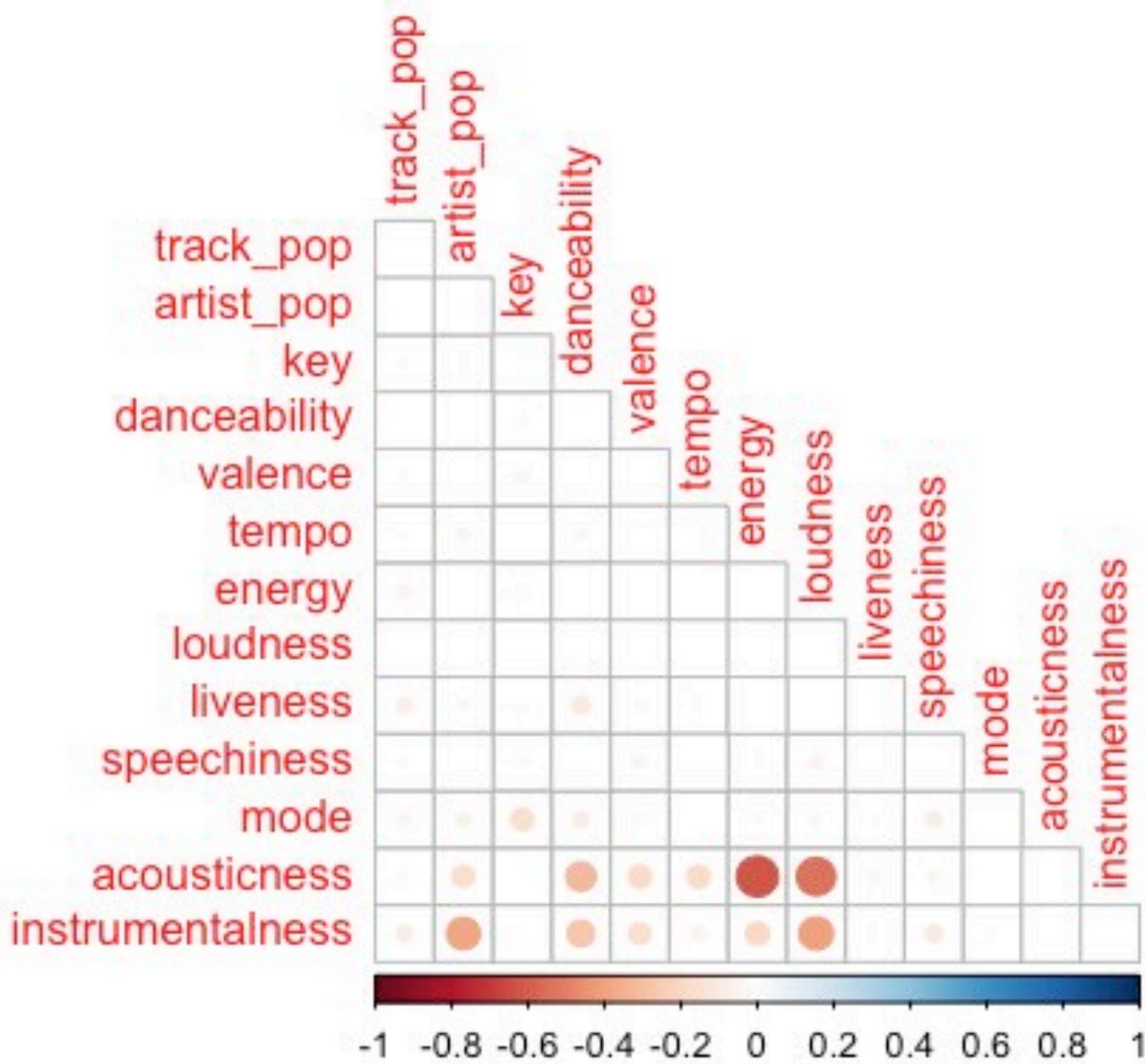
CHANGE IN LOUDNESS OVER TIME

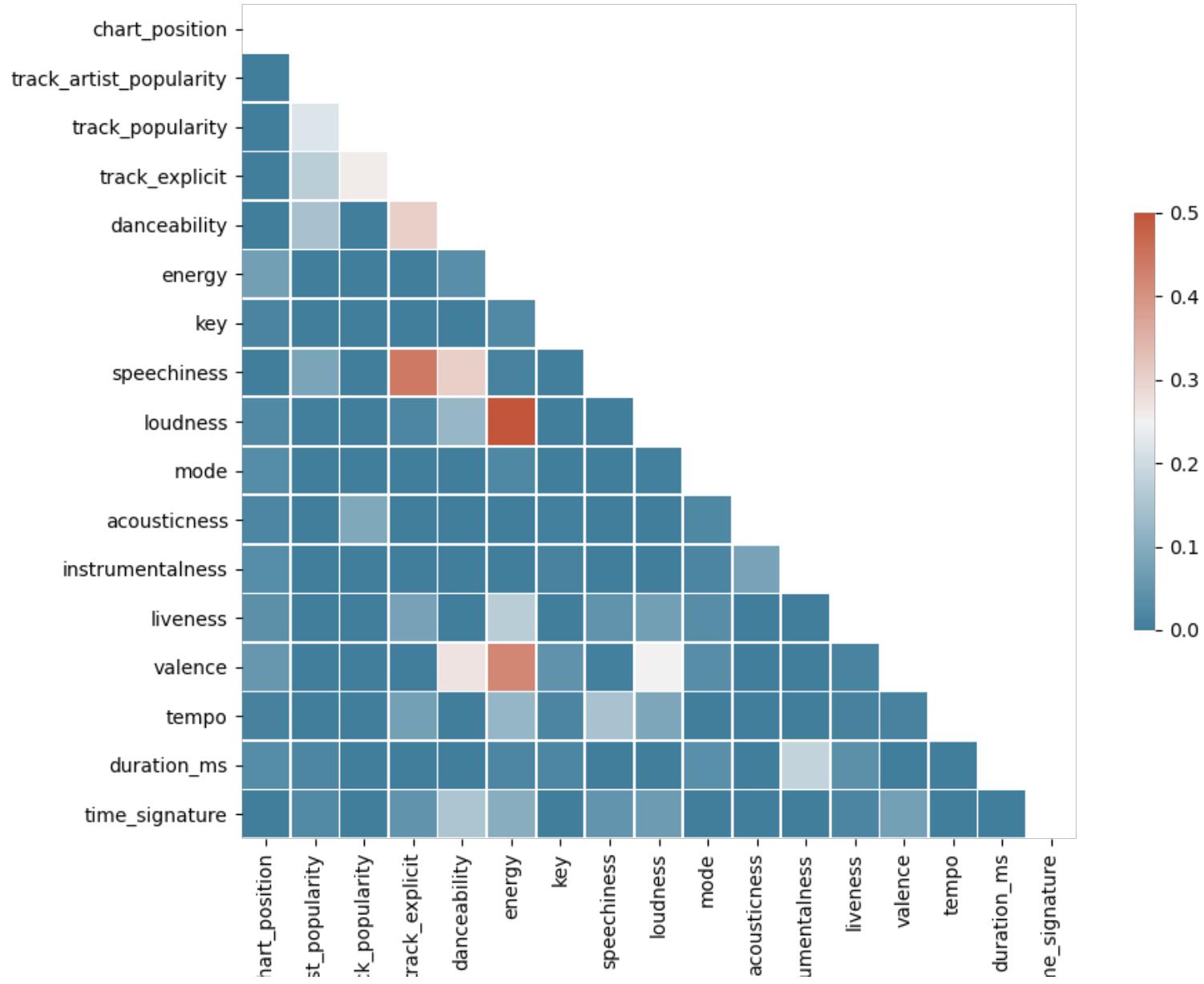


CORRELATION MATRIX OF FEATURES



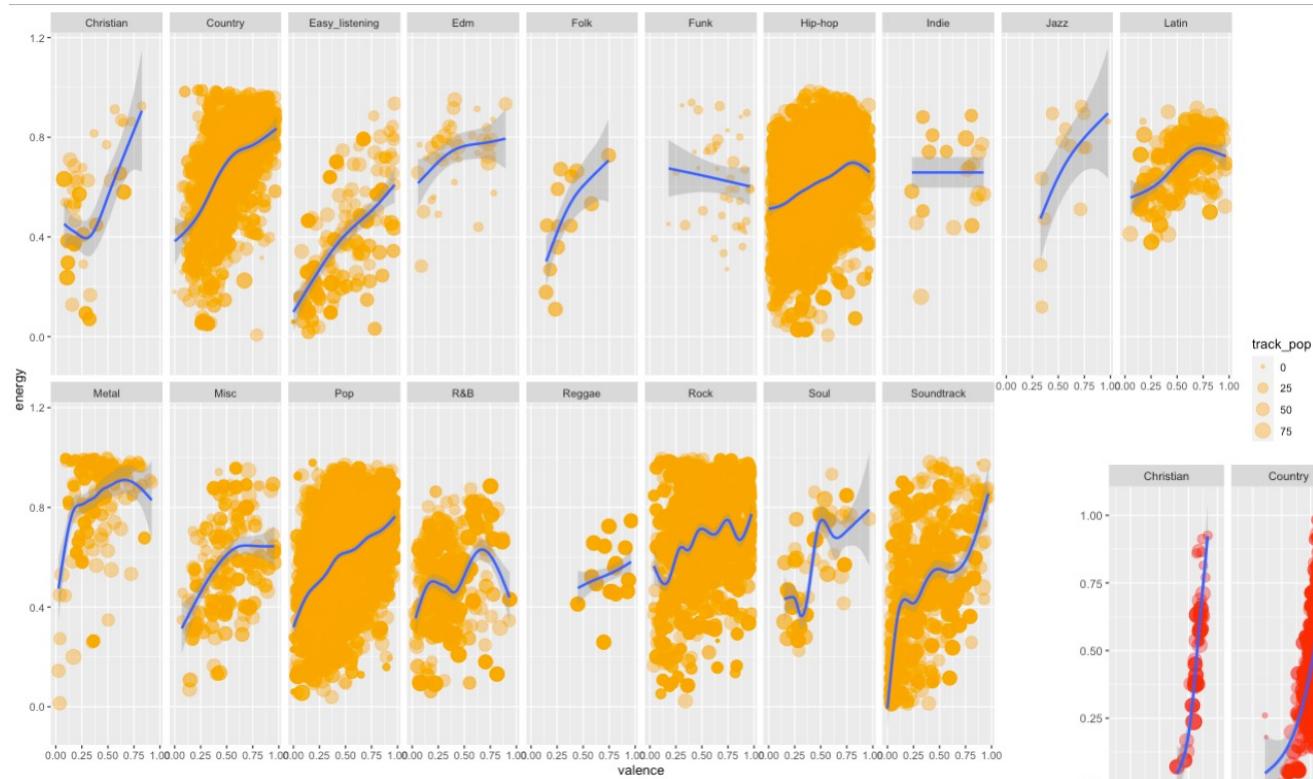
CORRELATION PLOT OF FEATURES





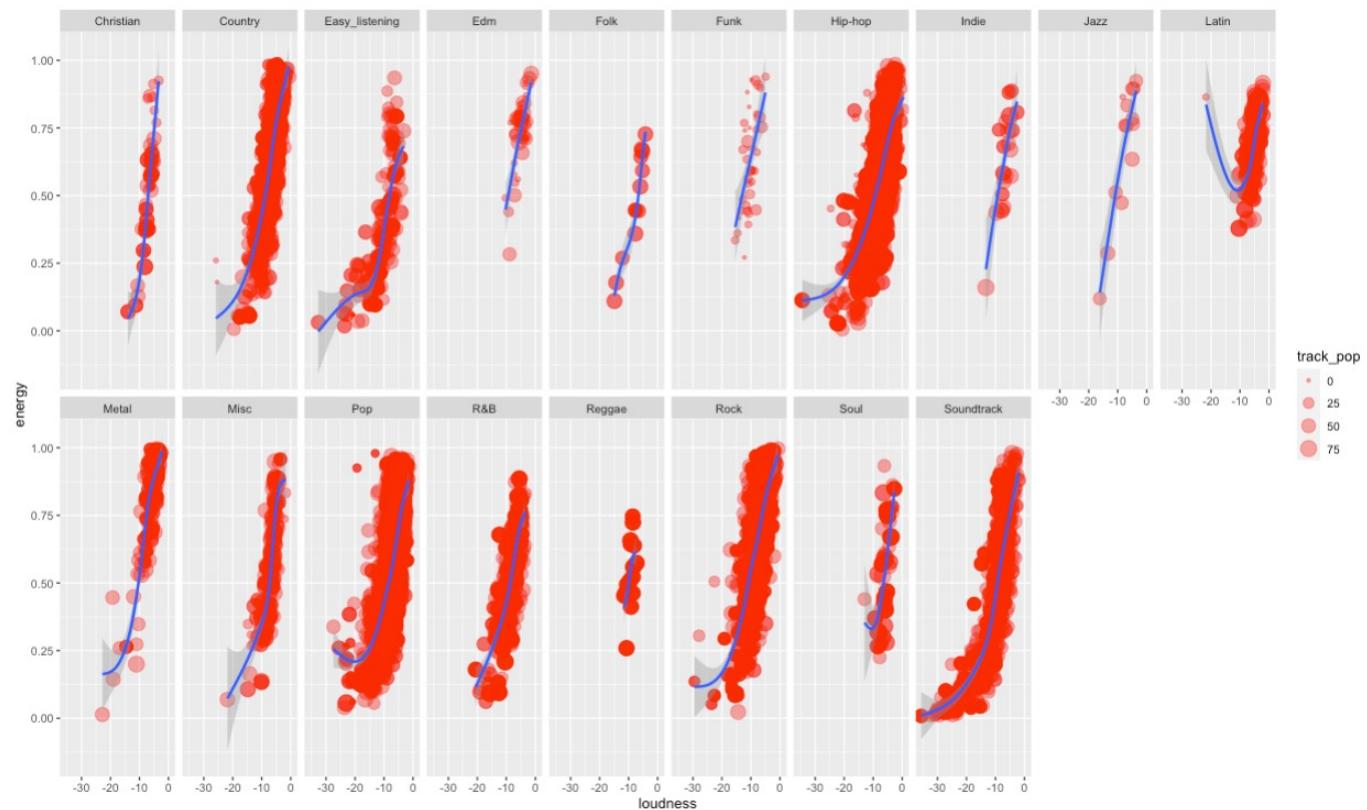
CORRELATION HEAT MAP OF FEATURES

FACTET PLOTS



**VALENCE & ENERGY
SIZED BY TRACK POPULARITY
WRAPPED BY GENRE**

**LOUDNESS & ENERGY
SIZED BY TRACK POPULARITY
WRAPPED IN GENRE**



CORRELATION BETWEEN ENERGY & LOUDNESS

```
> summary(lmLE)
```

Call:

```
lm(formula = energy ~ loudness, data = ACATjoined)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.44630	-0.09052	-0.00265	0.08741	0.84853

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.9094852	0.0024570	370.2	<2e-16 ***
loudness	0.0411684	0.0002972	138.5	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1341 on 16502 degrees of freedom

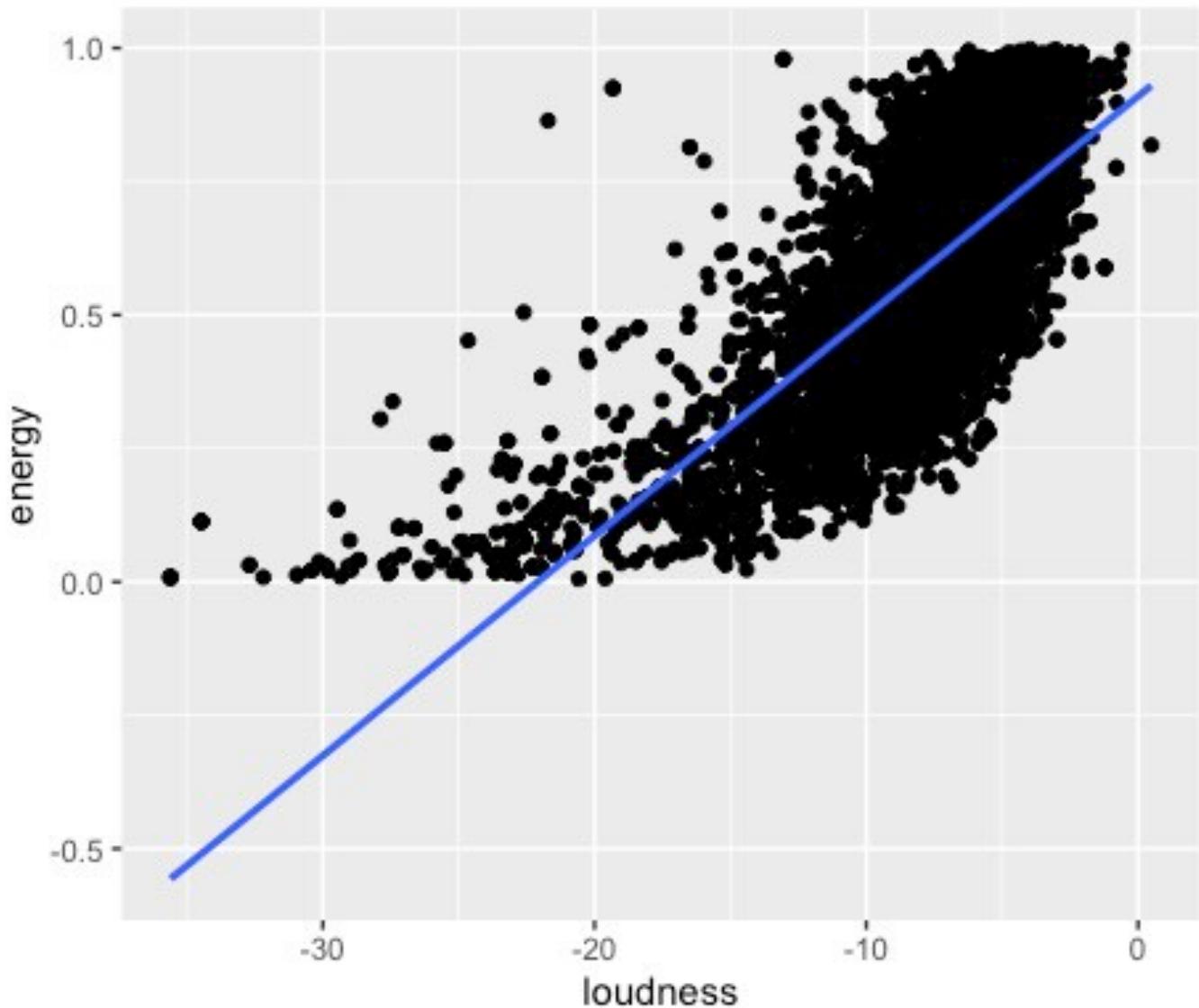
(1 observation deleted due to missingness)

Multiple R-squared: 0.5375, Adjusted R-squared: 0.5375

F-statistic: 1.918e+04 on 1 and 16502 DF, p-value: < 2.2e-16

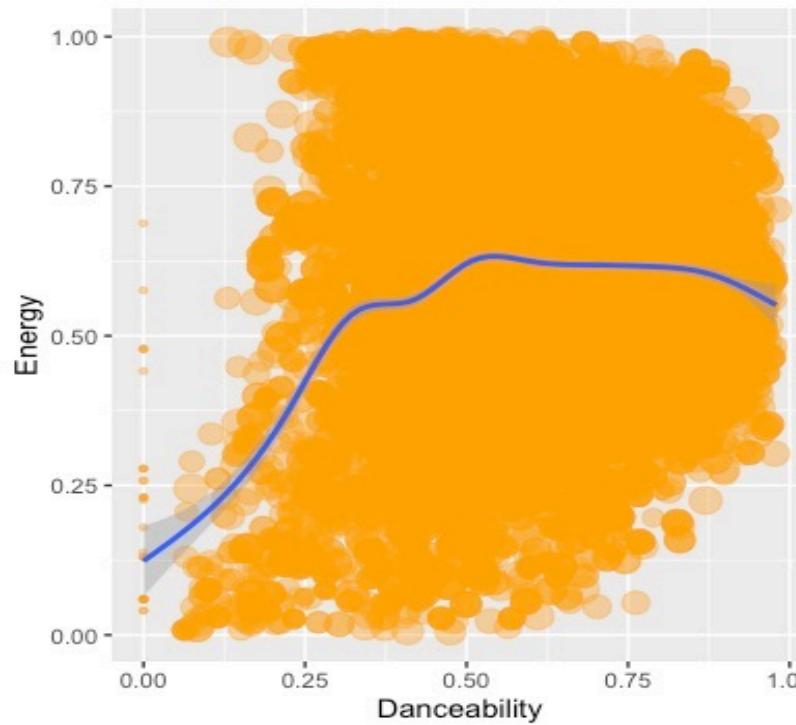
Pearson's product-moment correlation

```
data: ACATjoined$loudness and ACATjoined$energy
t = 138.5, df = 16502, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.7260404 0.7401531
sample estimates:
cor
0.7331757
```



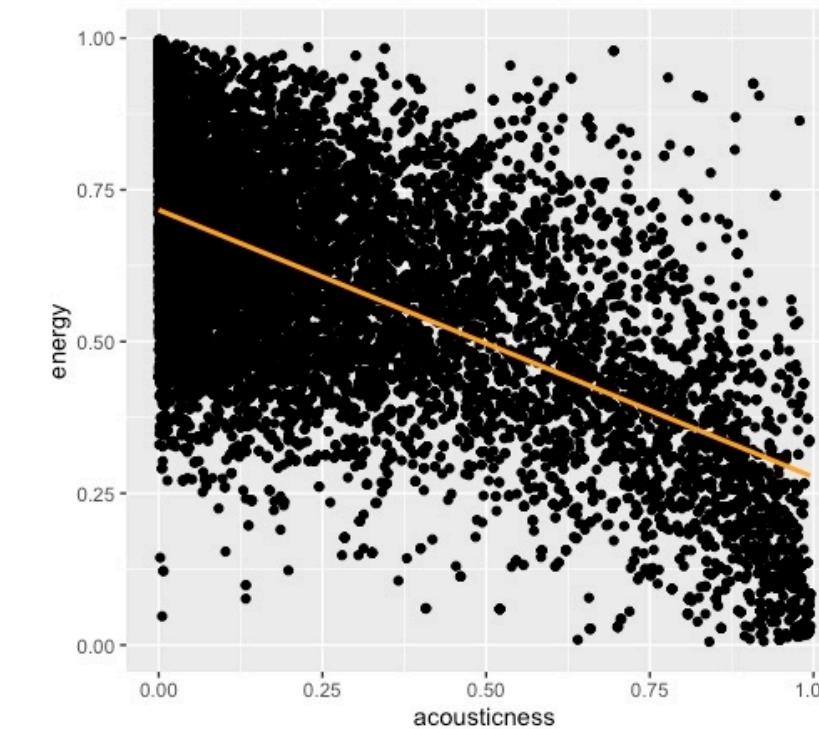
CORRELATIONS

Danceability & Energy



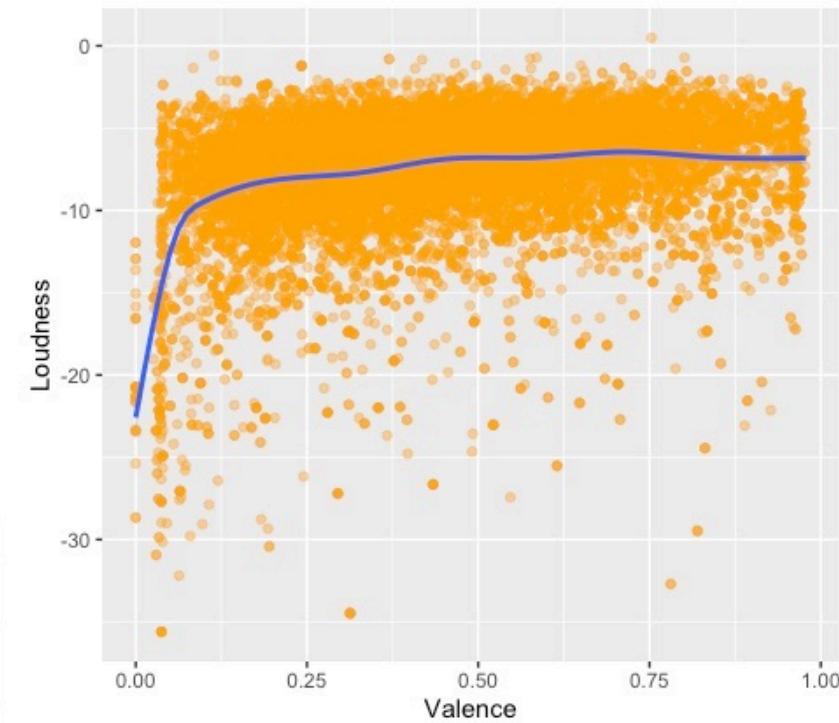
tempo

- 0
- 50
- 100
- 150
- 200

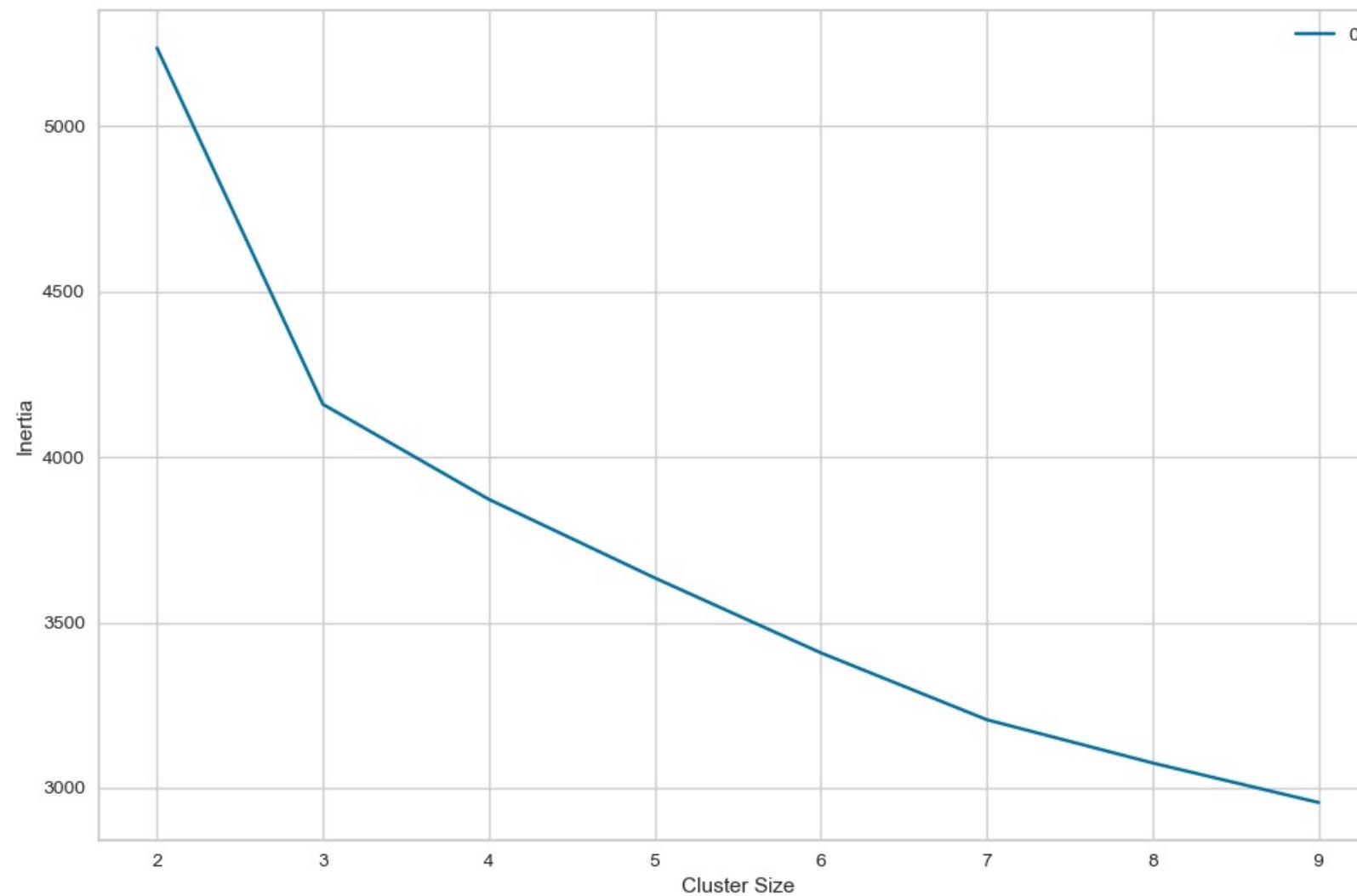


Acousticness & Energy

Valence & Loudness



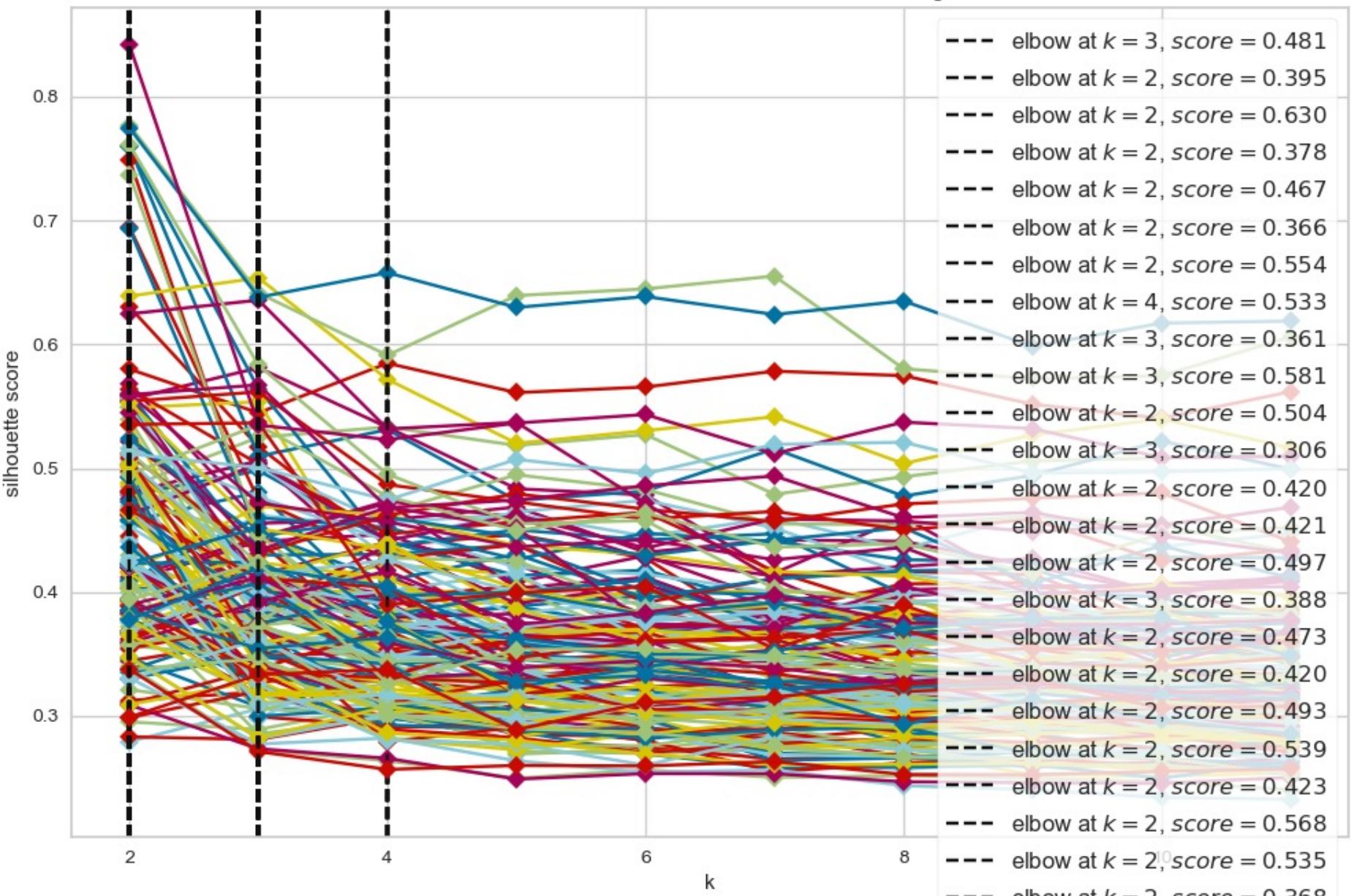
K-MEANS CLUSTERING



IDEAL NUMBER of CLUSTERS: 3

POSSIBLE FEATURE COMBINATIONS

Silhouette Score Elbow for KMeans Clustering



CLUSTERING SCORES

Select most well-clustered group of features

```
In [19]: max_cluster_3 = kmeans_scores.sort_values(["elbow", "score"],  
max_cluster_3
```

Out[19]:

	Unnamed: 0	features	score	elbow
2	2	['instrumentalness', 'liveness', 'speechiness']	0.533339	4
3	3	['instrumentalness', 'key', 'loudness']	0.581160	3
47	47	['instrumentalness', 'time_signature', 'valence']	0.534843	3
14	14	['danceability', 'instrumentalness', 'key']	0.506165	3
27	27	['danceability', 'instrumentalness', 'time_sig...']	0.503687	3
15	15	['instrumentalness', 'loudness', 'valence']	0.503424	3
9	9	['energy', 'instrumentalness', 'time_signature']	0.501529	3
49	49	['instrumentalness', 'loudness', 'time_signatu...']	0.841490	2
37	37	['instrumentalness', 'liveness', 'time_signatu...']	0.777018	2
44	44	['instrumentalness', 'speechiness', 'time_sign...']	0.774674	2
40	40	['instrumentalness', 'liveness', 'loudness']	0.761278	2

BEST CLUSTER GROUP OF 3:

INSTRUMENTALNESS

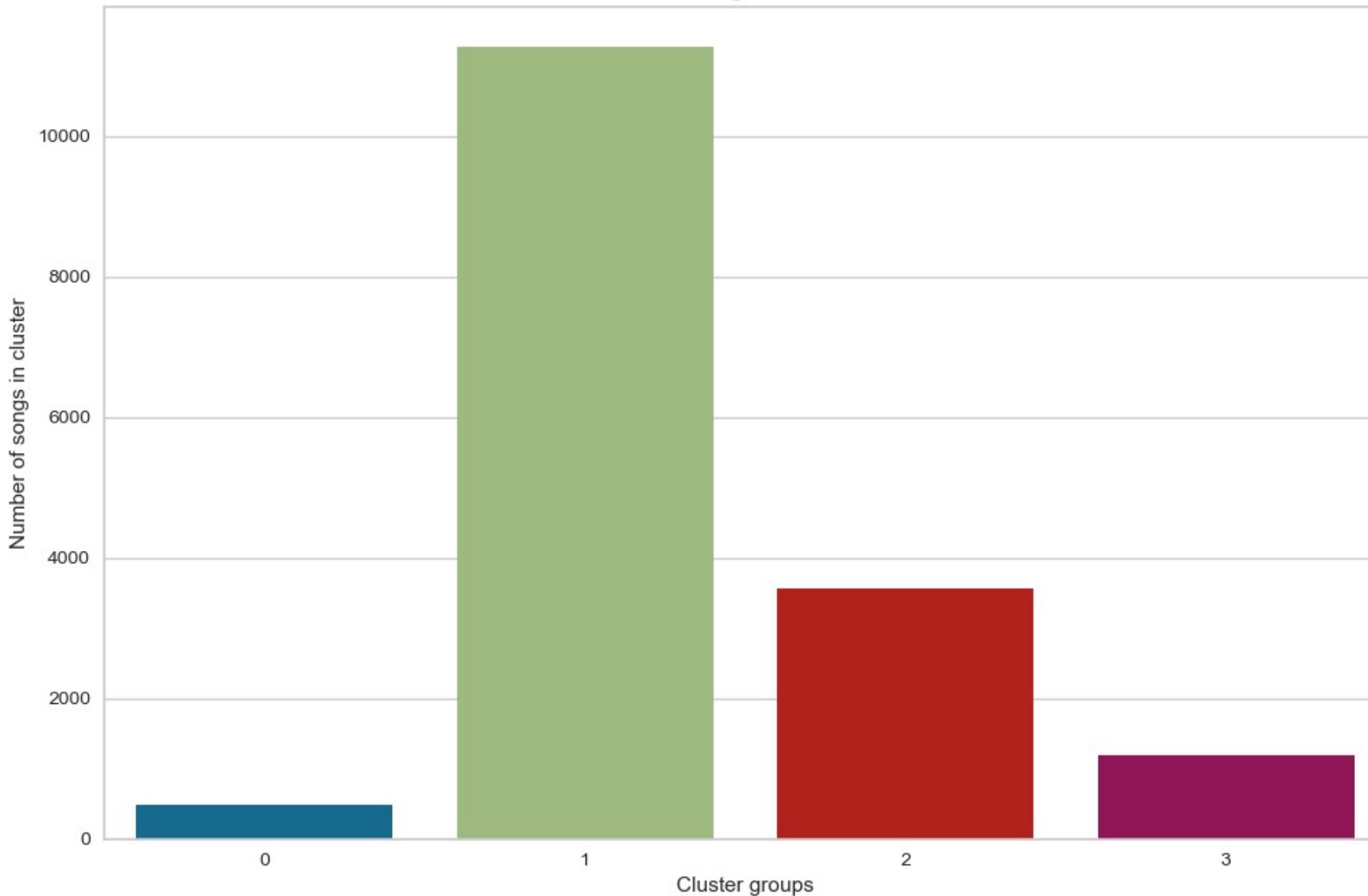
LIVENESS

SPEECHINESS

0.53339

CLUSTERED TRACK GROUPS

Count of number of songs in each KMeans cluster



KMEANS CLUSTERS

1st Cluster:

Mid-high instrumentalness
Low-mid liveness
Low-mid speechiness

2nd Cluster:

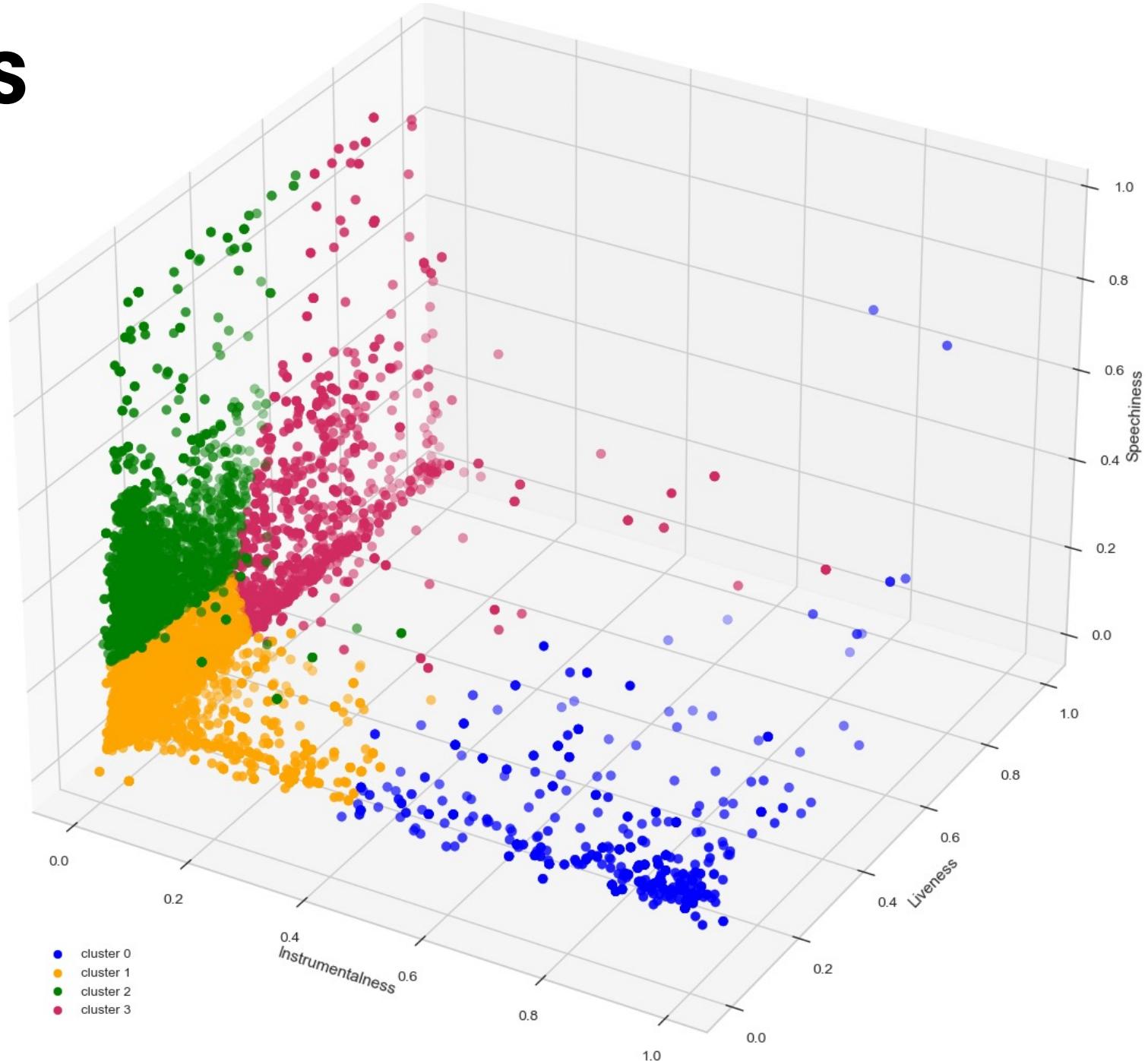
Low-mid instrumentalness
Low-mid liveness
Low-mid speechiness

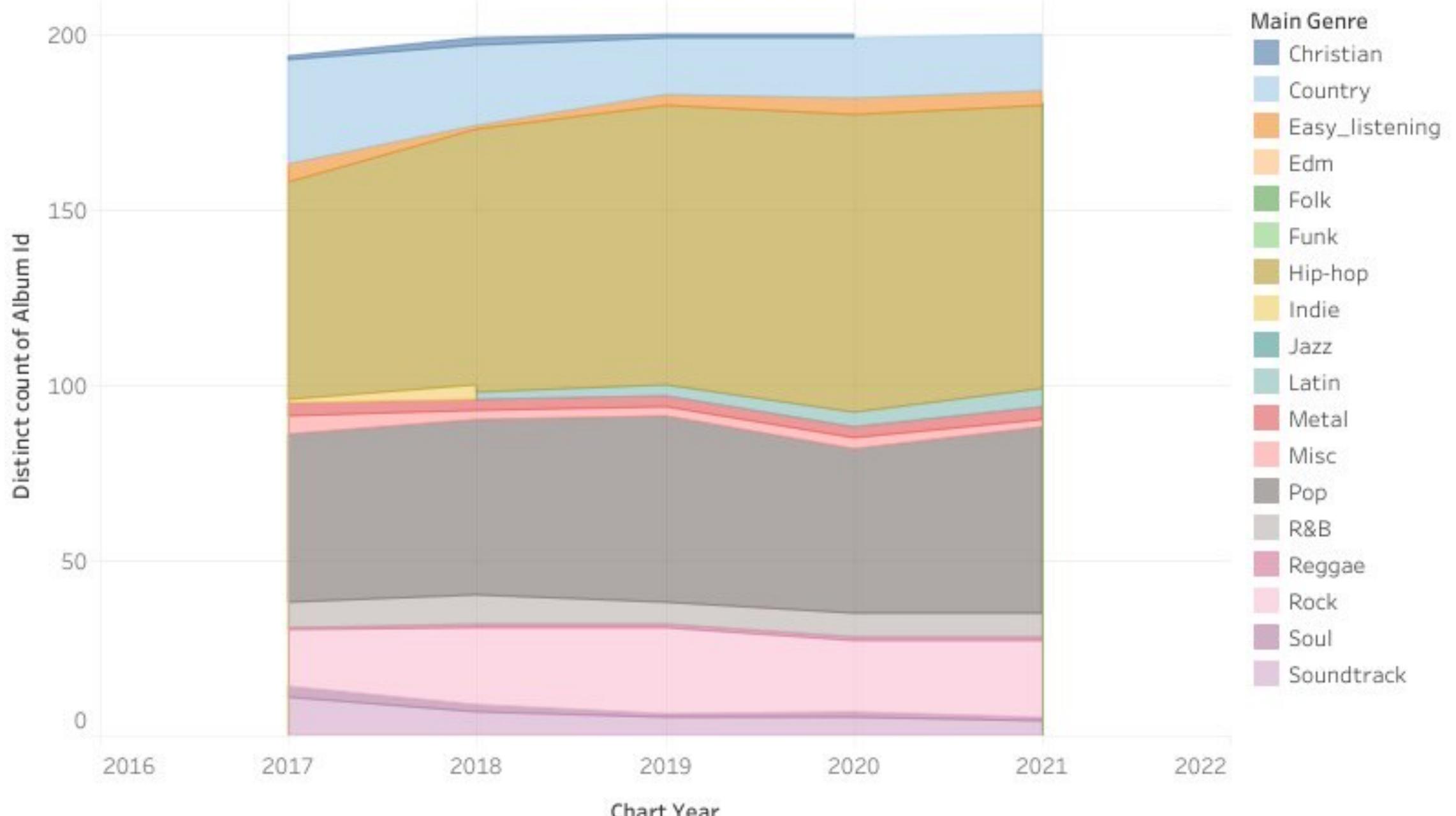
3rd Cluster:

Low instrumentalness
Low-mid liveness
Mid-high speechiness

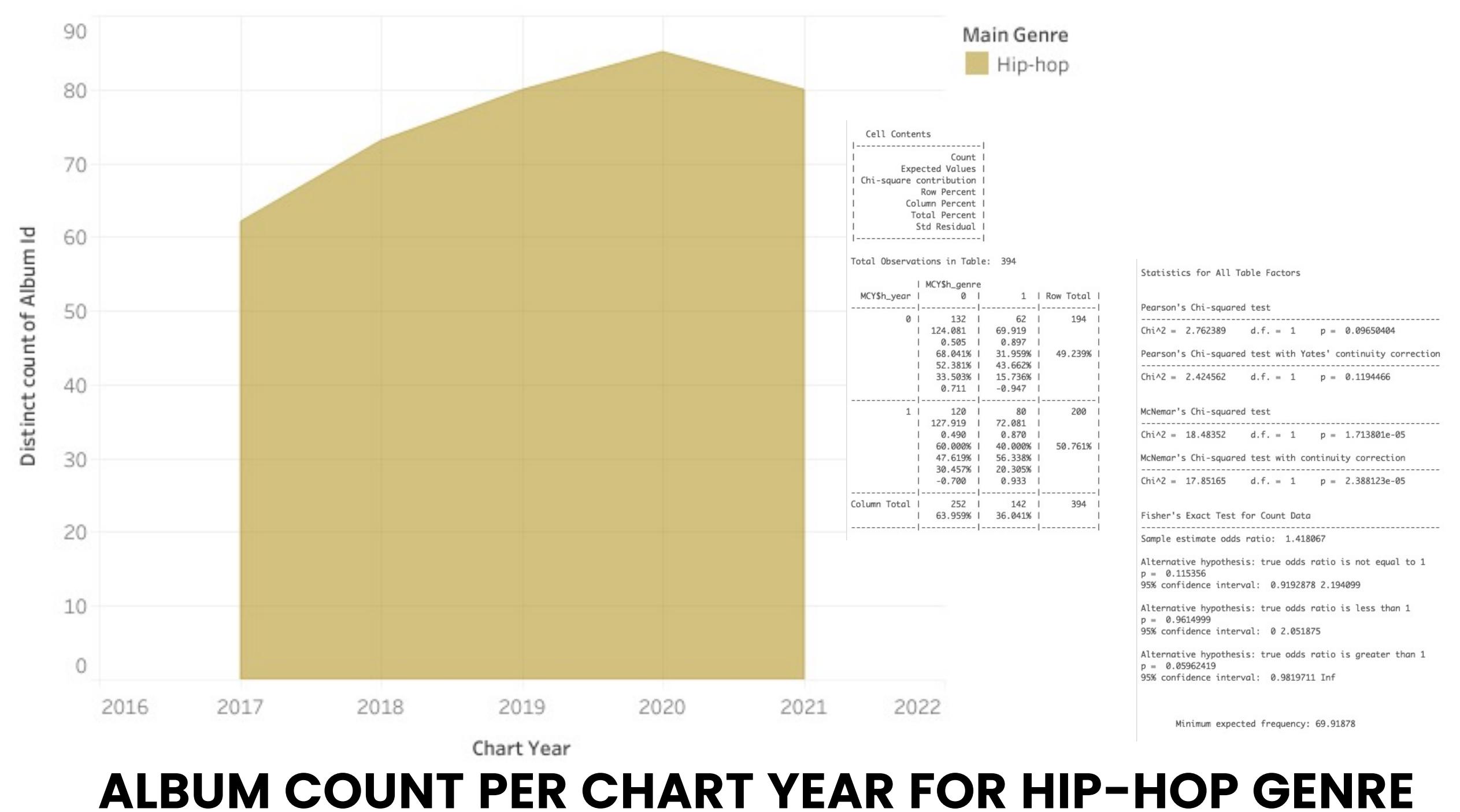
4th Cluster:

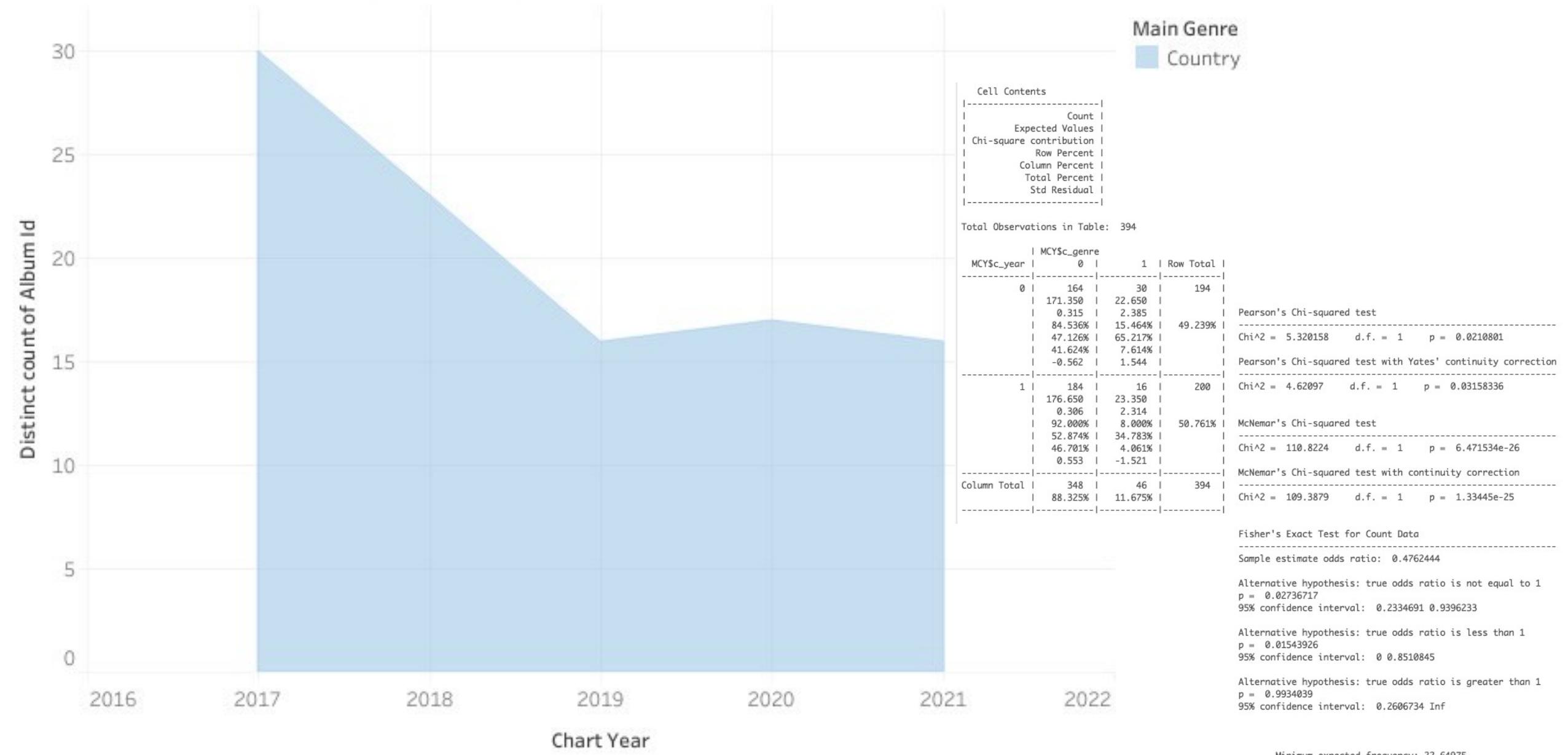
Low instrumentalness
Mid-high liveness
Low-mid speechiness



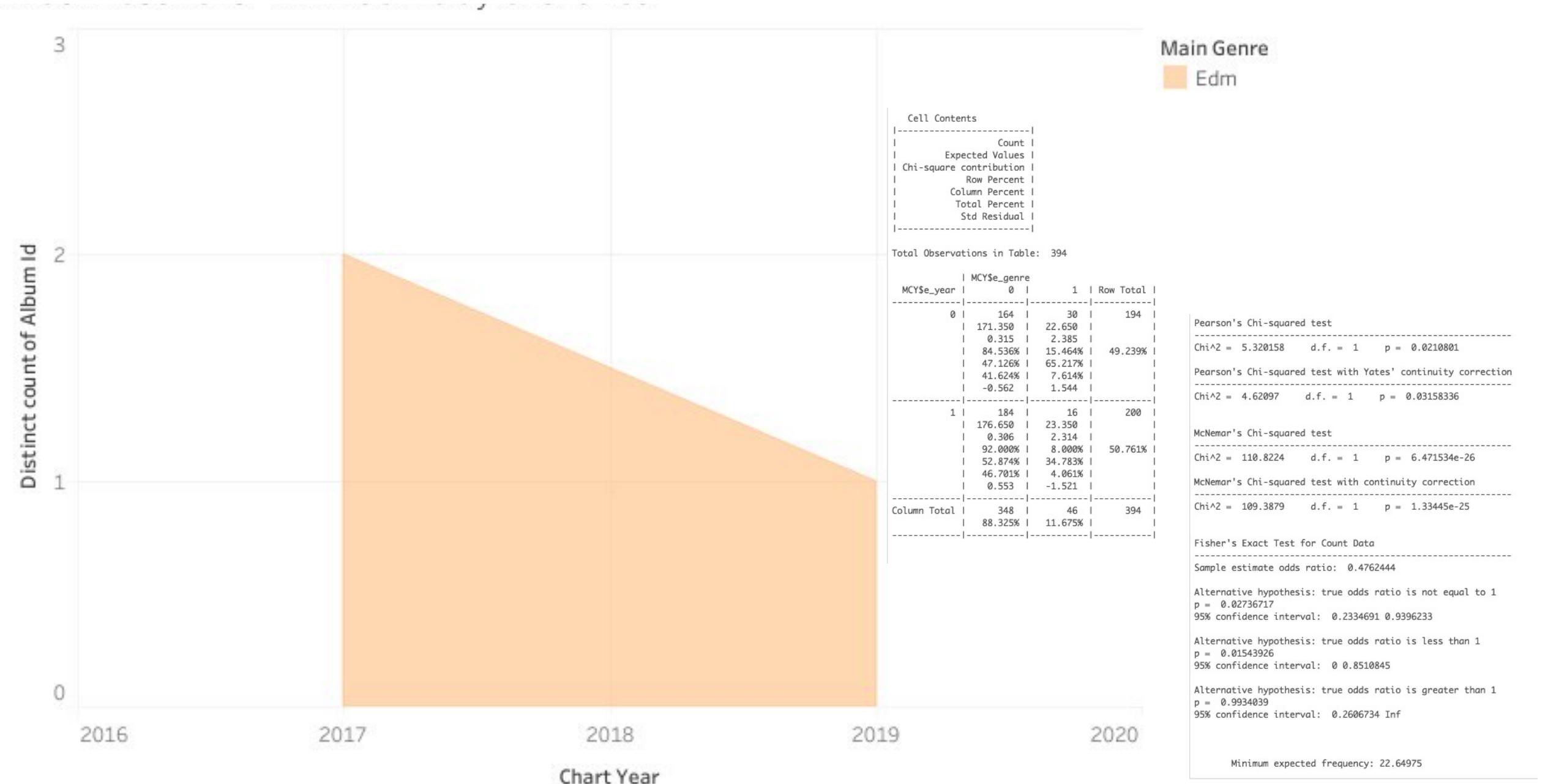


ALBUM COUNT PER CHART YEAR BY GENRE





ALBUM COUNT PER CHART YEAR FOR COUNTRY GENRE



ALBUM COUNT PER CHART YEAR FOR EDM GENRE

GENRE CLASSIFICATION

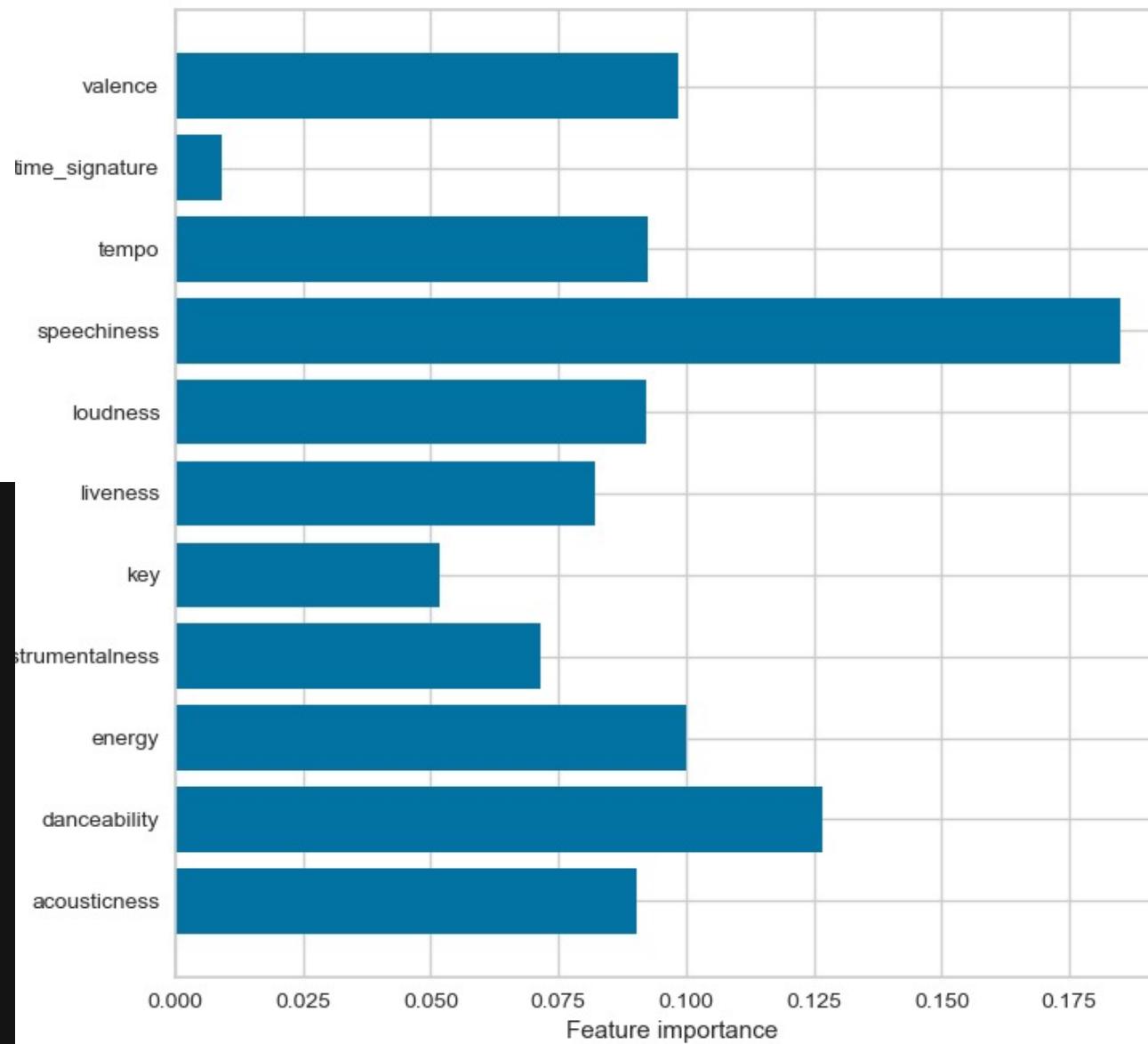
K Nearest Neighbors

	precision	recall	f1-score	support
Christian	0.62	0.57	0.59	14
Country	0.72	0.78	0.75	305
Easy_listening	0.79	0.70	0.74	43
Edm	0.12	0.08	0.10	12
Folk	0.40	1.00	0.57	2
Funk	0.50	0.20	0.29	5
Hip-hop	0.88	0.85	0.87	1259
Indie	1.00	0.80	0.89	5
Jazz	0.00	0.00	0.00	3
Latin	0.47	0.40	0.44	47
Metal	0.87	0.89	0.88	45
Misc	0.74	0.70	0.72	53
Pop	0.76	0.75	0.75	773
R&B	0.65	0.68	0.66	118
Reggae	0.81	1.00	0.90	13
Rock	0.84	0.90	0.87	402
Soul	0.71	0.77	0.74	13
Soundtrack	0.76	0.75	0.75	189
accuracy			0.80	3301
macro avg	0.65	0.66	0.64	3301
weighted avg	0.80	0.80	0.80	3301

Random Forest Classifier

	precision	recall	f1-score	support
Christian	1.00	0.48	0.65	25
Country	0.81	0.80	0.80	450
Easy_listening	0.93	0.63	0.75	62
Edm	0.00	0.00	0.00	13
Folk	1.00	1.00	1.00	8
Funk	0.33	0.12	0.18	8
Hip-hop	0.86	0.94	0.90	1896
Indie	1.00	0.44	0.62	9
Jazz	0.00	0.00	0.00	4
Latin	0.96	0.42	0.59	64
Metal	0.89	0.83	0.86	65
Misc	0.97	0.51	0.67	76
Pop	0.74	0.80	0.77	1142
R&B	0.93	0.65	0.77	176
Reggae	1.00	1.00	1.00	19
Rock	0.85	0.88	0.87	615
Soul	1.00	0.67	0.80	24
Soundtrack	0.94	0.73	0.82	295
accuracy			0.84	4951
macro avg	0.79	0.61	0.67	4951
weighted avg	0.84	0.84	0.83	4951

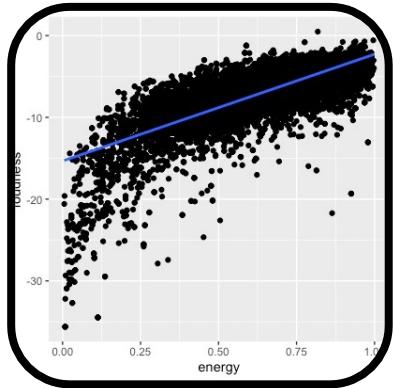
FEATURE INFLUENCE



SUMMARY

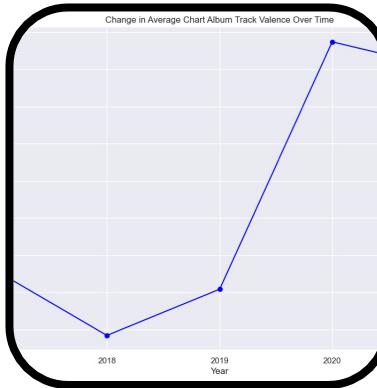


SUMMARY



GENRE PRESENCE

Hip-hop presence on the charts increased while albums in the Country genre decreased.

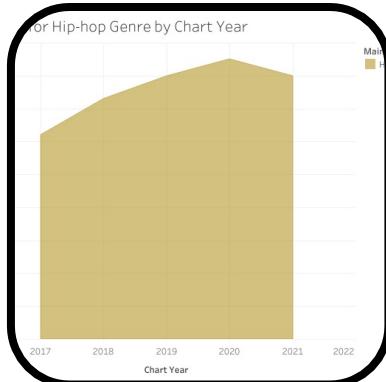


ENERGY CHANGE

Average track energy decreased during the five-year analysis period.

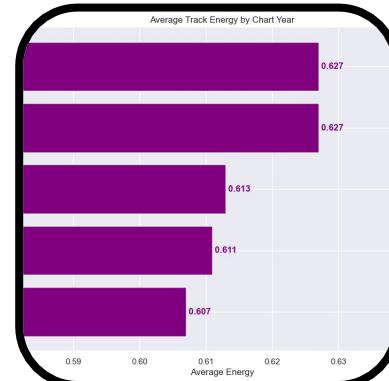
CORRELATIONS

Energy was significantly correlated with loudness and valence.



VALENCE CHANGE

Valence generally increased during the five year analysis period, particularly in 2020.



A dark, atmospheric photograph of a concert. Stage lights in shades of purple, blue, and yellow illuminate the scene from above and behind a band. In the foreground, the silhouettes of a large crowd of spectators are visible, many holding up smartphones to take pictures. The overall mood is energetic and celebratory.

CONCLUSIONS



CONCLUSIONS

IMPORTANCE

Music is a vital part in the human experience and affects each person's life in some way. The details within the music we love are worth exploring and learning about, and they help us find more music that resonates with us.

IMPACT

Analysis of music features helps entertainment, streaming, and social media companies find focus and gain insight on current musical trends. This allows the maximization of profits and revenue.

A photograph of a live concert. The stage is filled with bright, colorful spotlights creating a starburst effect. A band is performing; one member is visible on the left playing a guitar, and another is on the right singing into a microphone. The audience is in the foreground, seen from behind, with many people holding up smartphones to take pictures or record the performance. The overall atmosphere is energetic and vibrant.

QUESTIONS



THANK YOU FOR LISTENING

More information and
analyses available
on our GitHub repo:

<https://bit.ly/SxB5YA>