

2017

2018

2019

2020

2021

<https://bit.ly/SxB5YA>

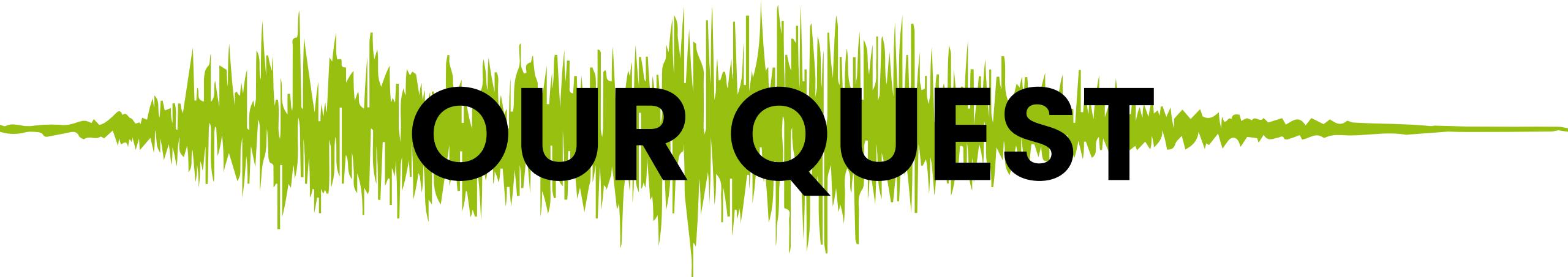


Good Vibrations?

Spotify x Billboard Top 200 Albums 5-Year Analysis Project

Bianca Serrano and Katie Ravenwood





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 CREATION

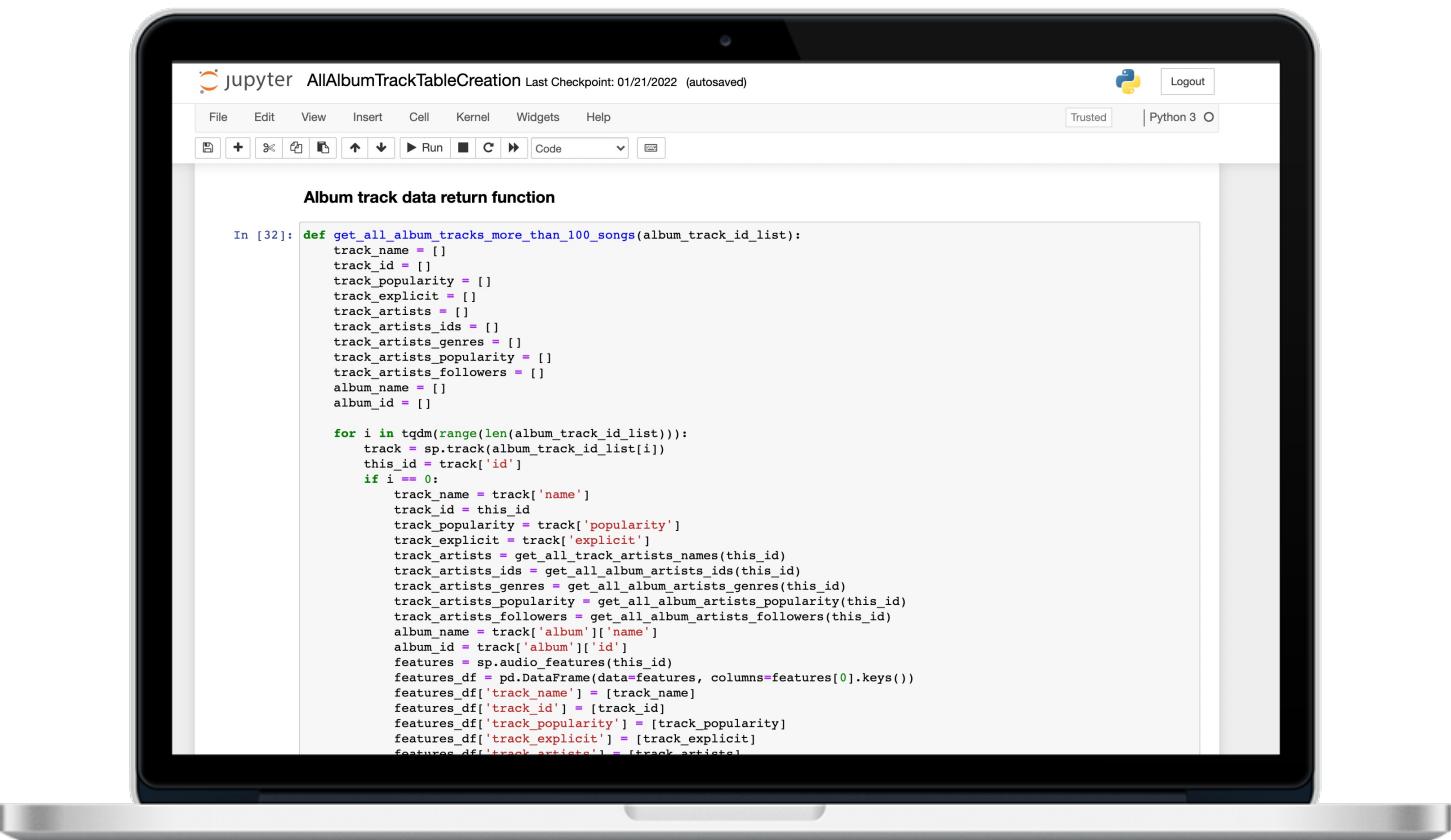
PYTHON & SPOTIPY

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 ALBUMS



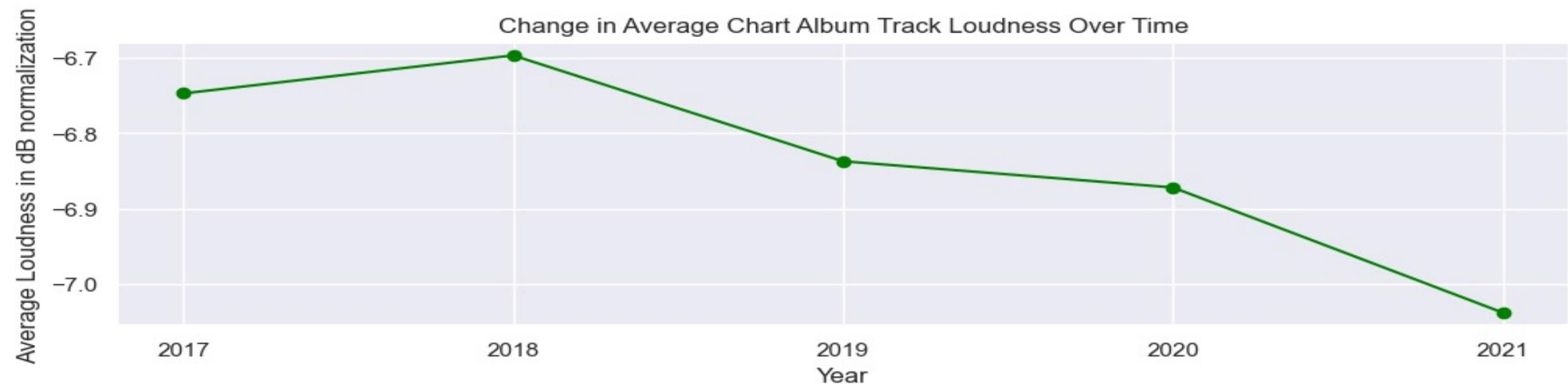
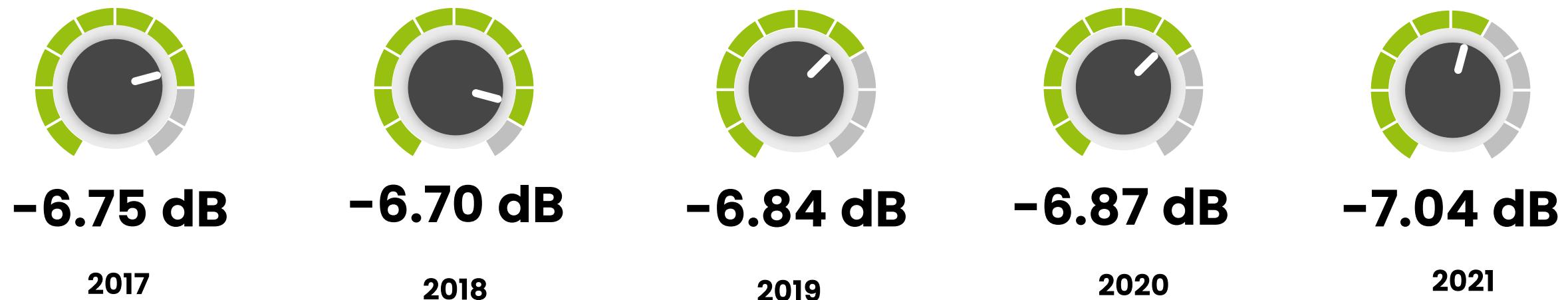
```
jupyter AllAlbumTrackTableCreation Last Checkpoint: 01/21/2022 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
Logout Trusted Python 3 O

Album track data return function

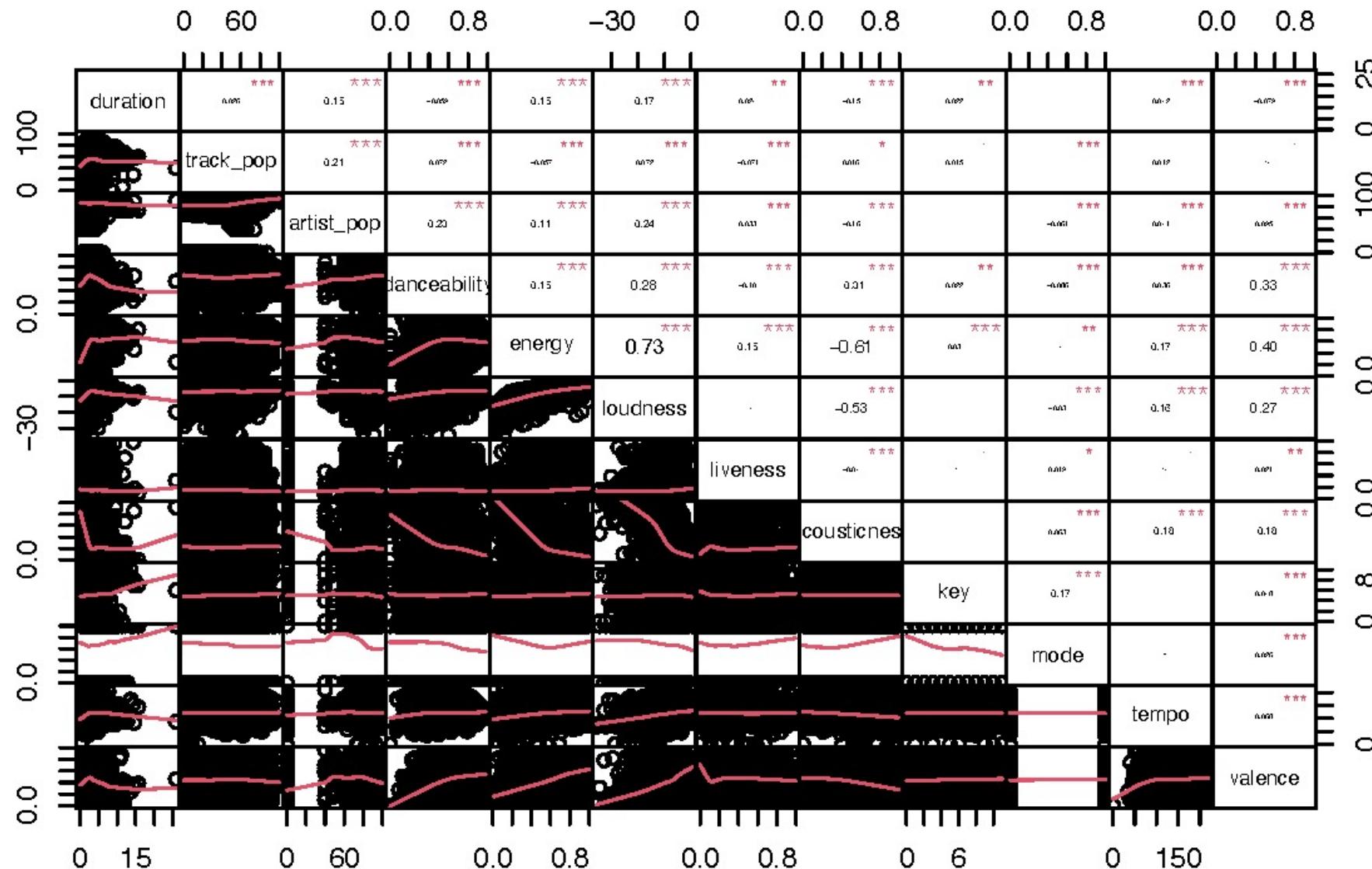
In [32]: def get_all_album_tracks_more_than_100_songs(album_track_id_list):
    track_name = []
    track_id = []
    track_popularity = []
    track_explicit = []
    track_artists = []
    track_artists_ids = []
    track_artists_genres = []
    track_artists_popularity = []
    track_artists_followers = []
    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]
```

CHANGE IN LOUDNESS OVER TIME



CORRELATION MATRIX OF FEATURES



KMEANS CLUSTERS

1st Cluster:

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

2nd Cluster:

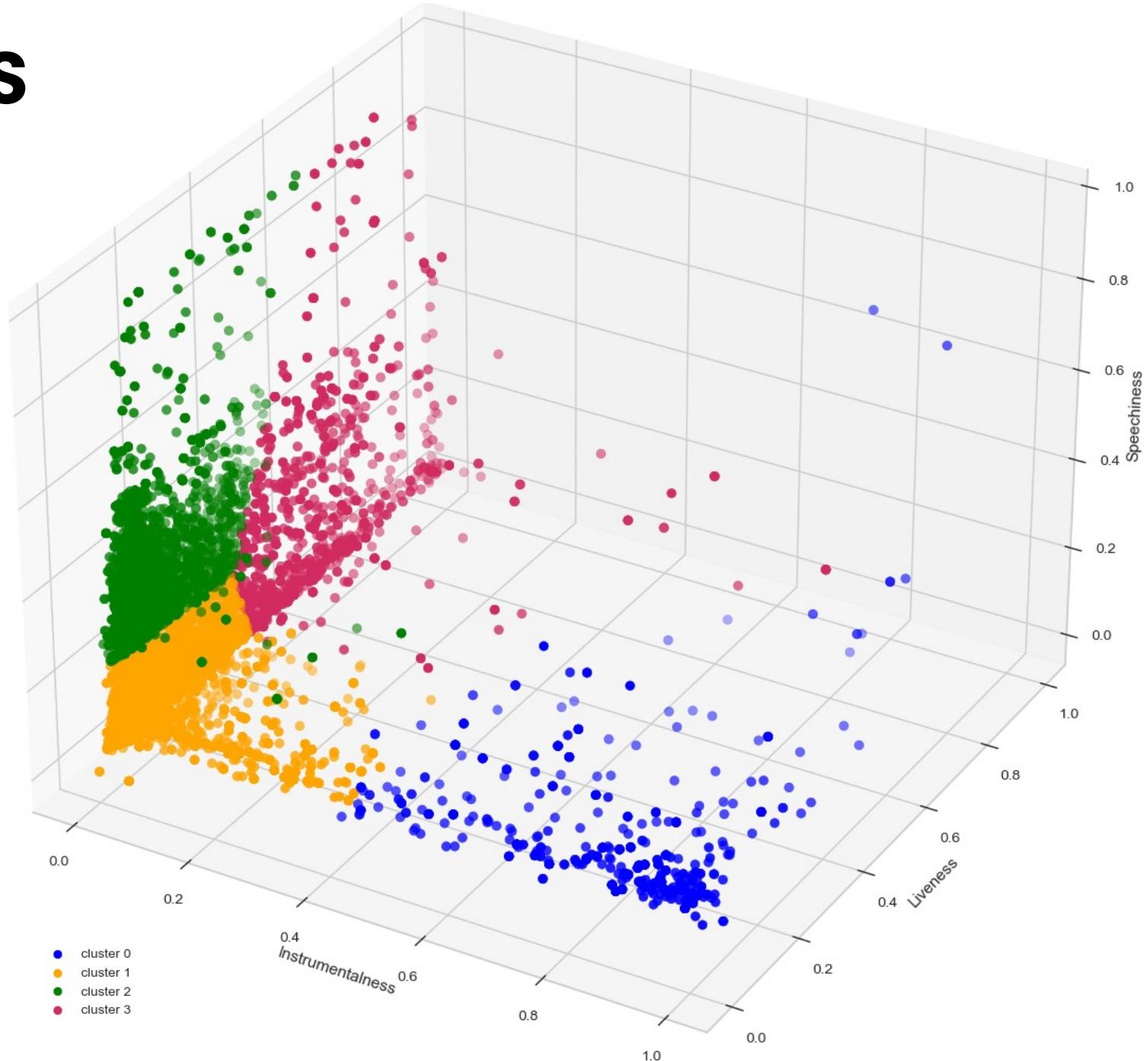
Low-mid instrumentalness
Low-mid liveness
Low-mid speechiness

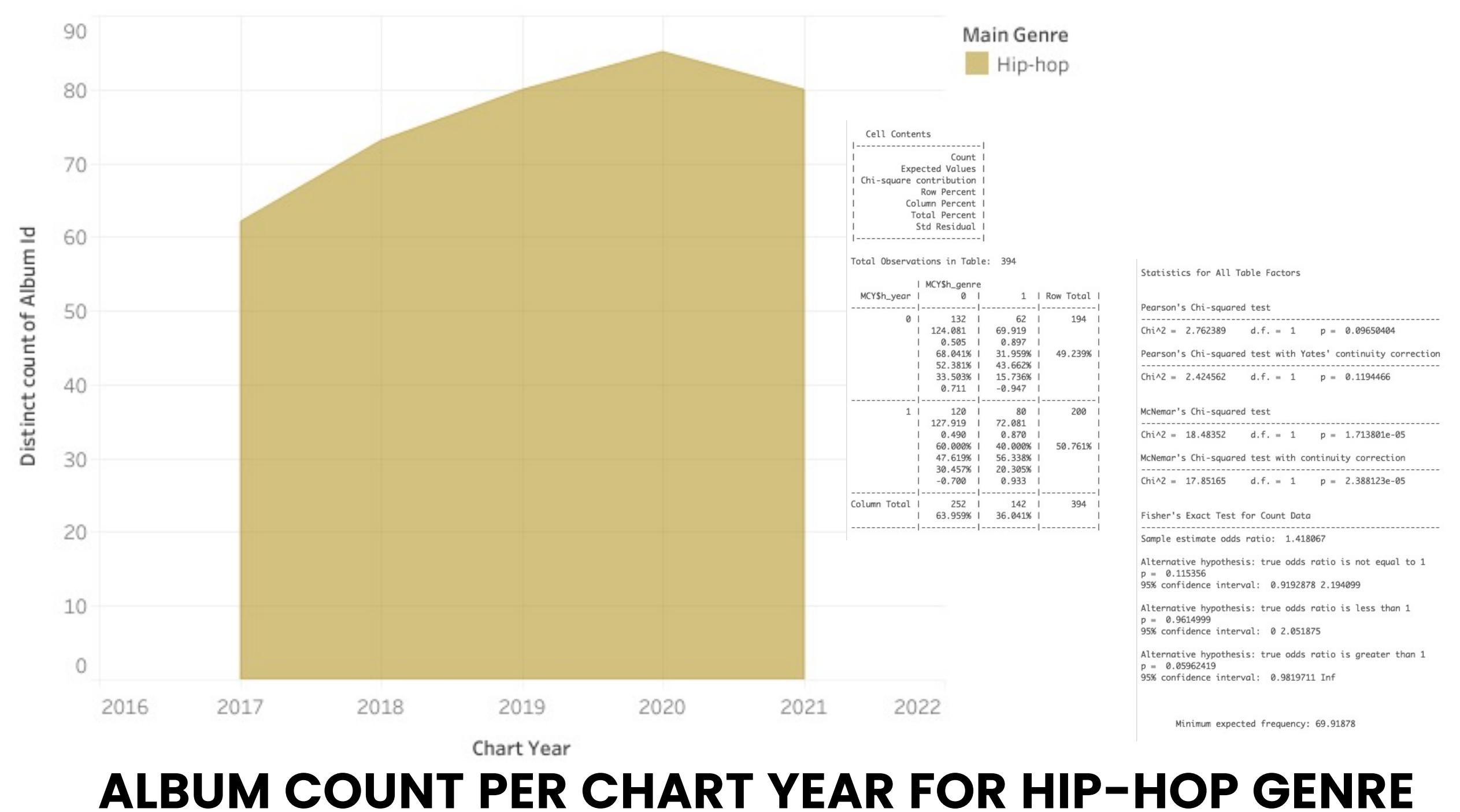
3rd Cluster:

Low instrumentalness
Low-mid liveness
Mid-high speechiness

4th Cluster:

Low instrumentalness
Mid-high liveness
Low-mid speechiness





GENRE CLASSIFICATION

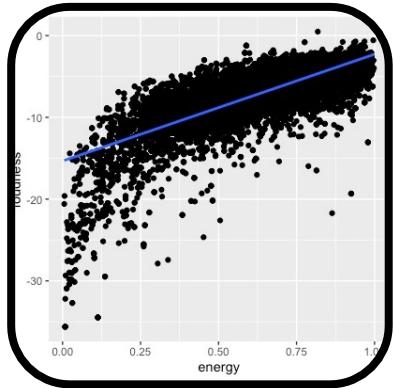
K Nearest Neighbors

	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

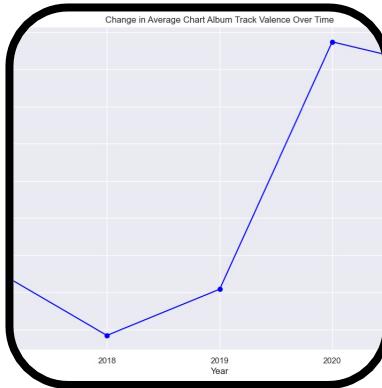
	precision	recall	f1-score	support
Christian	1.00	0.48	0.65	25
Country	0.81	0.80	0.80	450
Easy_listening	0.93	0.63	0.75	62
Edm	0.00	0.00	0.00	13
Folk	1.00	1.00	1.00	8
Funk	0.33	0.12	0.18	8
Hip-hop	0.86	0.94	0.90	1896
Indie	1.00	0.44	0.62	9
Jazz	0.00	0.00	0.00	4
Latin	0.96	0.42	0.59	64
Metal	0.89	0.83	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

SUMMARY



GENRE PRESENCE

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

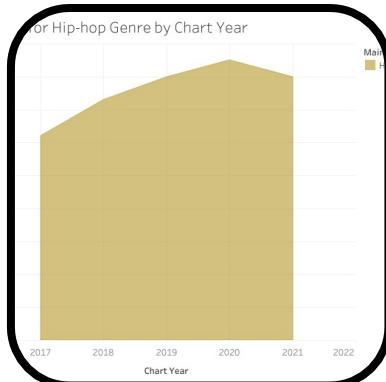


ENERGY CHANGE

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.

