Spotify: Top Hits Analysis

Cody Allison
James Beckmeyer
Hannah Weber

Introduction

What makes a hit a hit?:

Spotify may have some info on that

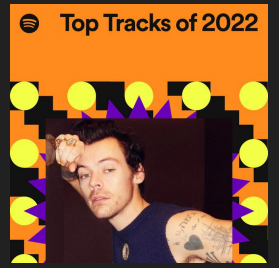
Questions to be answered:

- What traits are present in top hits from the 2010s, and did they change over the decade?
- Was the Covid outbreak a cause for large shift in trends?

Where We Got Our Data

kaggle™

Spotify®



Spotify 2010 - 2019 Top 100.csv (102.76 kB)

Detail Compact Column

About this file

Spotify's top 100 hits of 2010-2019 in a csv file

title	artist	top genre	year released	added	bpm
Song title	Song's artist	Genre of the song	Year song was released	Date song was added to Spotify's top hits playlist	Beats Per of the so
946 unique values	Taylor Swift 2% Drake 2% Other (964) 96%	dance pop 36% pop 6% Other (585) 58%	1975 2021		
STARSTRUKK (feat. Katy Perry)	3OH!3	dance pop	2009		
My First Kiss (feat. Ke$ha)	3OH!3	dance pop	2010		
I Need A Dollar	Aloe Blacc	pop soul	2010		
Airplanes (feat. Hayley Williams of Paramore)	B.o.B	atl hip hop	2010		
Nothin' on You (feat. Bruno Mars)	B.o.B	atl hip hop	2010		
Magic (feat. Rivers Cuomo)	B.o.B	atl hip hop	2010		

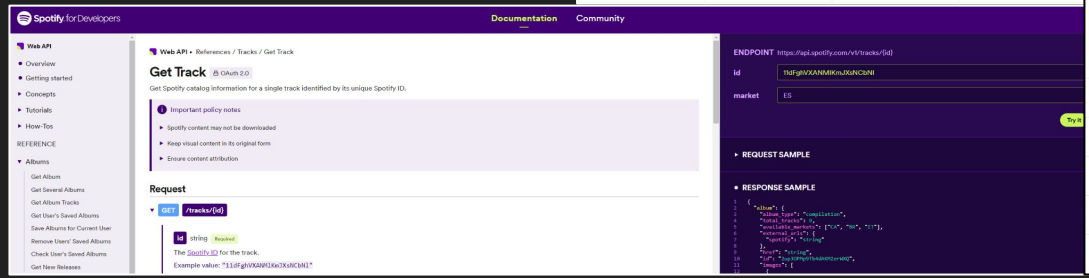
spotify_top50_2021.csv (6.57 kB)

Detail Compact Column

About this file

The dataset contains 50 songs and 14 characteristic variables of them.

id	artist_name	track_name	track_id	popularity	dance
Position of the song in the list	Name of artist	Name of track	Unique ID for the track in spotify	The higher the value the more popular the song is	The high easier it song
1 50	Olivia Rodrigo 8% Doja Cat 8% Other (42) 84%	50 unique values	50 unique values	67 95	0.38
1	Olivia Rodrigo	drivers license	5wANPM4fQCJwkGd4rN57mH	92	0.561
2	Lil Nas X	MONTERO (Call Me By Your Name)	1SC5rEoYDGUK4NfG82494W	90	0.593
3	The Kid LAROI	STAY (with Justin Bieber)	5PjdY0CKGZdEuoNab3yDmX	92	0.591
4	Olivia Rodrigo	good 4 u	4ZtFanR9U6ndgddUvNcjcG	95	0.563
5	Dua Lipa	Levitating (feat. DaBaby)	5nujrmhLynf4yMoMtj8AQF	89	0.702



API Calls Used

-Spotify Authorization Token

-Search, Get Playlist, Get Track Features

```python
def search_song_data(song_name):
    spotify_base_url= "https://api.spotify.com/v1"
    type='track'
    limit='1'

    headers = {'Authorization': 'Bearer {}'.format(get_spotify_token())}
    params = {'type':type,
              'limit':limit}

    search_endpoint = f"/search?query={song_name}"
    query_url = ''.join([spotify_base_url,search_endpoint])
    response = requests.get(query_url,headers=headers, params=params)
    results = response.json()
    return results

#Search for 'As It Was' by Harry Styles
results = search_song_data('As It Was')
track_id = results['tracks']['items'][0]['id']
#track_id returns as '4LRPiXqCikLlN15c3yImP7'
```

```python
def get_spotify_token():
    auth_url = 'https://accounts.spotify.com/api/token'
    data = {
    'grant_type': 'client_credentials',
    'client_id': client_id,
    'client_secret': client_secret}
    auth_response = requests.post(auth_url, data=data)
    access_token = auth_response.json().get('access_token')
    return access_token

get_spotify_token()
```

```python
def get_playlist_songs(playlist_id):
    spotify_base_url= "https://api.spotify.com/v1"

    headers = {
    'Authorization': 'Bearer {}'.format(get_spotify_token())}

    playlists_endpoint = f"/playlists/{playlist_id}"
    playlist_url = ''.join([spotify_base_url,playlists_endpoint])
    response = requests.get(playlist_url,headers=headers)
    playlist = response.json()
    return playlist
```

Example Get Track Features Output

```python
def get_track_features(trackID):

    track_url = f"https://api.spotify.com/v1/audio-features/{trackID}"

    headers = {
    'Authorization': 'Bearer {}'.format(get_spotify_token())}

    response = requests.get(track_url, headers=headers)
    song_features = response.json()
    return song_features
```


```
{'danceability': 0.824,
 'energy': 0.587,
 'key': 6,
 'loudness': -6.401,
 'mode': 0,
 'speechiness': 0.0937,
 'acousticness': 0.69,
 'instrumentalness': 0.000105,
 'liveness': 0.149,
 'valence': 0.514,
 'tempo': 98.029,
 'type': 'audio_features',
 'id': '2N8m6CYs74qQO4mjVcXO30',
 'uri': 'spotify:track:2N8m6CYs74qQO4mjVcXO30',
 'track_href': 'https://api.spotify.com/v1/tracks/2N8m6CYs74qQO4mjVcXO30',
 'analysis_url': 'https://api.spotify.com/v1/audio-analysis/2N8m6CYs74qQO4mjVcXO30',
 'duration_ms': 209438,
 'time_signature': 4}
```


Features

- **Danceability**: How suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
- **Energy**: Represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. A value of 0.0 is least Energy and 1.0 is most Energy
- **Valence**: The musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).Valence is a measure from 0.0 to 1.0
- **Acousticness**: Acousticness is a confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- **Speechiness**: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

Cleaning

Null Removal

```
tracks_2010s_df = tracks_2010s_df.dropna(how='any')
```

Column Manipulation

Rename Columns

```
tracks_df_21 =tracks_df_21.rename(columns = {'duration_ms': 'duration (ms)','tempo':'bpm',
                                             'artist_name': 'Artist', 'track_name': 'Song Title', 'track_id': 'Song ID'})

tracks_df_21['duration (secs)'] = tracks_df_21['duration (ms)']/1000
tracks_df_21['top year'] = '2021'
tracks_df_21
```

Insert New Columns

danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	bpm	duration (ms)	time_signature	duration (secs)	top year
0.561	0.431	10	-8.810	1	0.0578	0.76800	0.000014	0.1060	0.1370	143.875	242013	4	242.013	2021
0.593	0.503	8	-6.725	0	0.2200	0.29300	0.000000	0.4050	0.7100	178.781	137704	4	137.704	2021
0.591	0.764	1	-5.484	1	0.0483	0.03830	0.000000	0.1030	0.4780	169.928	141806	4	141.806	2021
0.563	0.664	9	-5.044	1	0.1540	0.33500	0.000000	0.0849	0.6880	166.928	178147	4	178.147	2021
0.702	0.825	6	-3.787	0	0.0601	0.00883	0.000000	0.0674	0.9150	102.977	203064	4	203.064	2021
0.677	0.696	0	-6.181	1	0.1190	0.32100	0.000000	0.4200	0.4640	90.030	198082	4	198.082	2021

Concat

```
#combine and reduce columns to relevant data
all_years_df = pd.concat([tracks_2010s, tracks_2020s])
all_years_df = all_years_df[['Song Title','Artist','bpm','energy','danceability'
all_years_df
```

	Song Title	Artist	bpm	energy	danceability	dB	valence	duration (secs)
0	STARSTRUKK (feat. Katy Perry)	3OH!3	140.000	81.0	61.0	-6.000	23.0	203.000
1	My First Kiss (feat. Ke$ha)	3OH!3	138.000	89.0	68.0	-4.000	83.0	192.000
2	I Need A Dollar	Aloe Blacc	95.000	48.0	84.0	-7.000	96.0	243.000
3	Airplanes (feat. Hayley Williams of Paramore)	B.o.B	93.000	87.0	66.0	-4.000	38.0	180.000
4	Nothin' on You (feat. Bruno Mars)	B.o.B	104.000	85.0	69.0	-6.000	74.0	268.000
...	...	...	...	...	...	...	...	...
45	Late Night Talking	Harry Styles	114.996	72.8	71.4	-4.595	90.1	177.955
46	Until I Found You	Stephen Sanchez	101.358	50.8	53.9	-6.050	22.7	177.720
47	Lo Siento BB:/ (with Bad Bunny & Julieta Venegas)	Tainy	169.888	70.3	63.9	-6.330	13.8	207.301
48	Yonaguni	Bad Bunny	179.951	64.8	64.4	-4.601	44.0	206.710
49	Tarot	Bad Bunny	114.011	68.4	79.5	-3.971	41.9	237.895

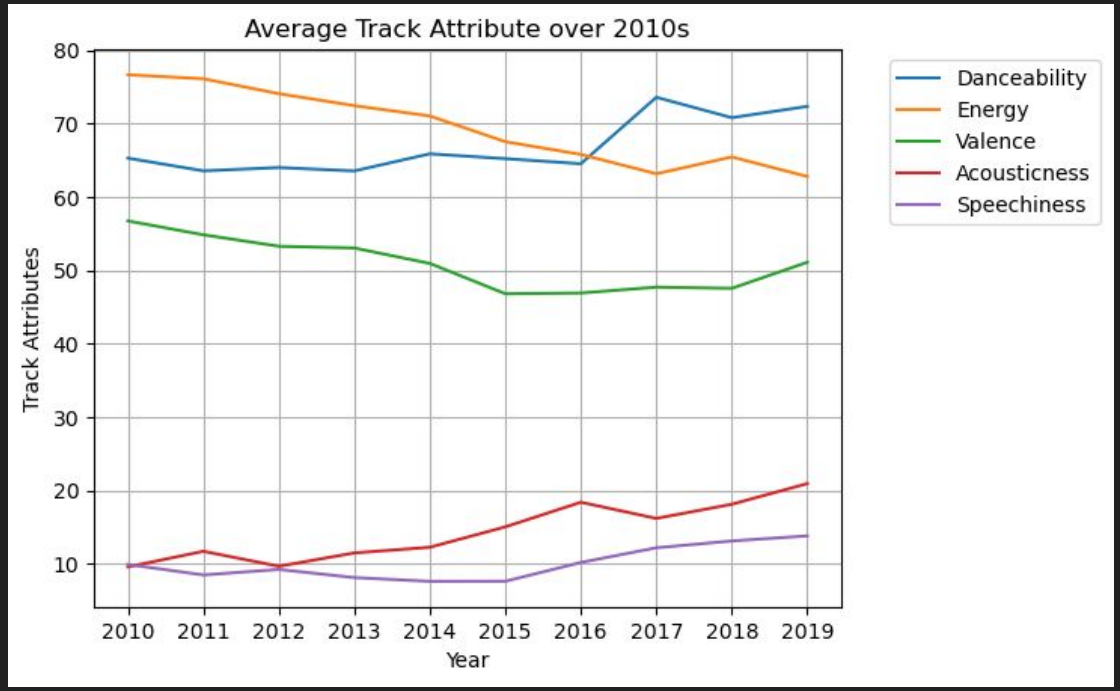
1150 rows × 13 columns

10 years with 100 top hits
3 years with 50 top hits

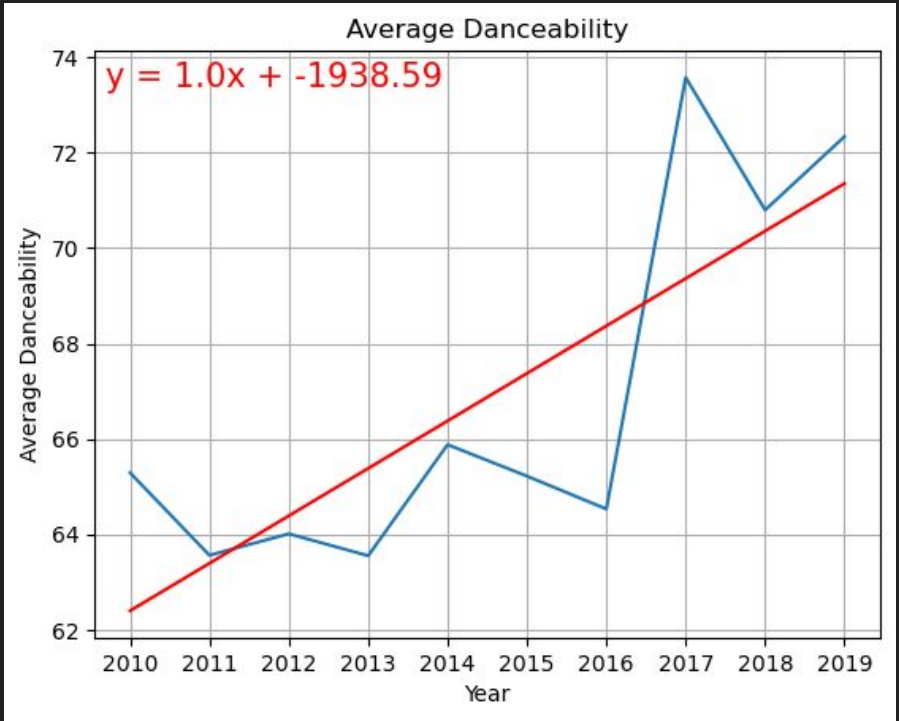
Question 1: What traits are present in top hits from the 2010s, and did they change over the decade?

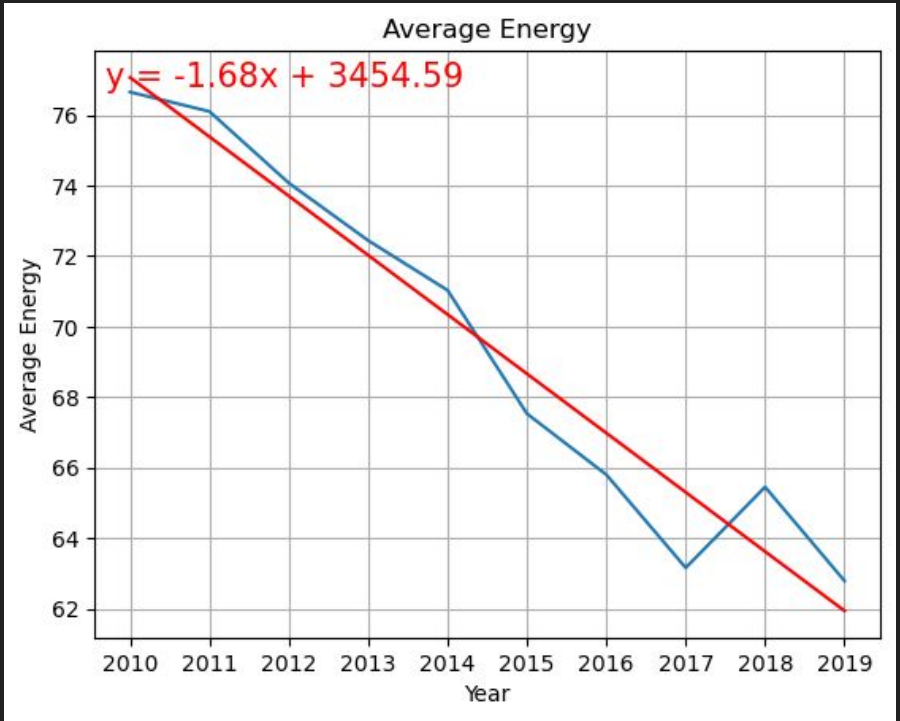
Null Hypothesis:
The Average Track Attributes do not follow any trends over the 2010s

Alternate Hypothesis:
The Average Track Attributes do follow a trend over the 2010s

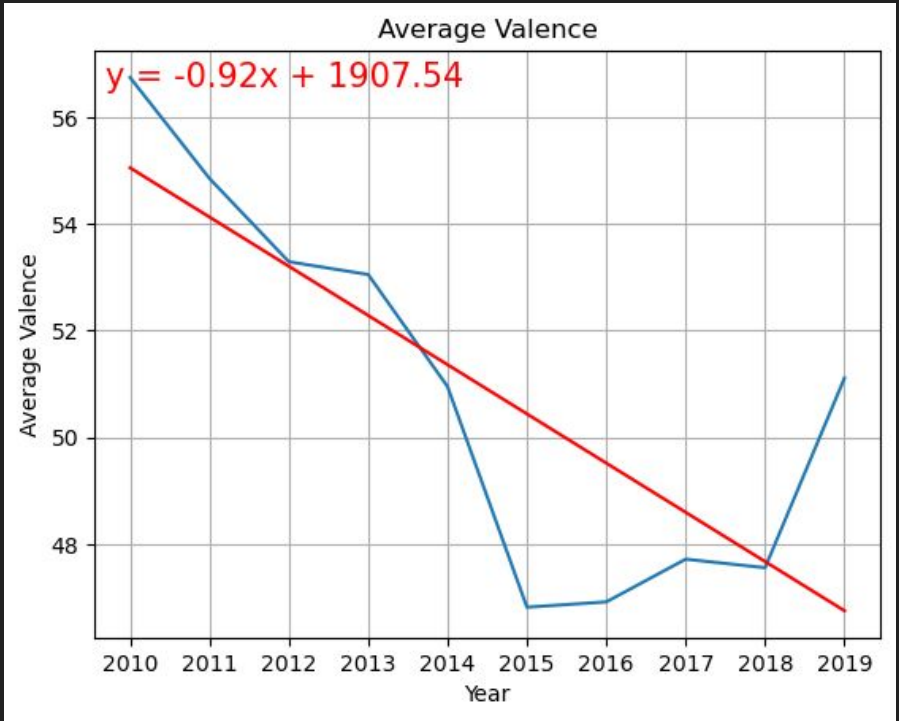


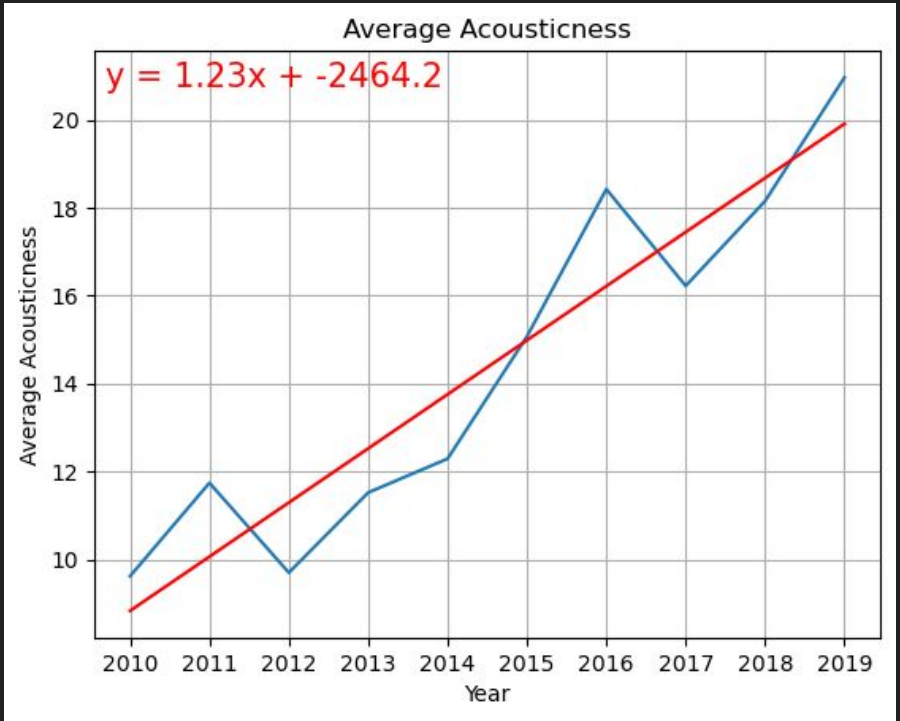
Q1. Average Song Attributes by year (2010s)



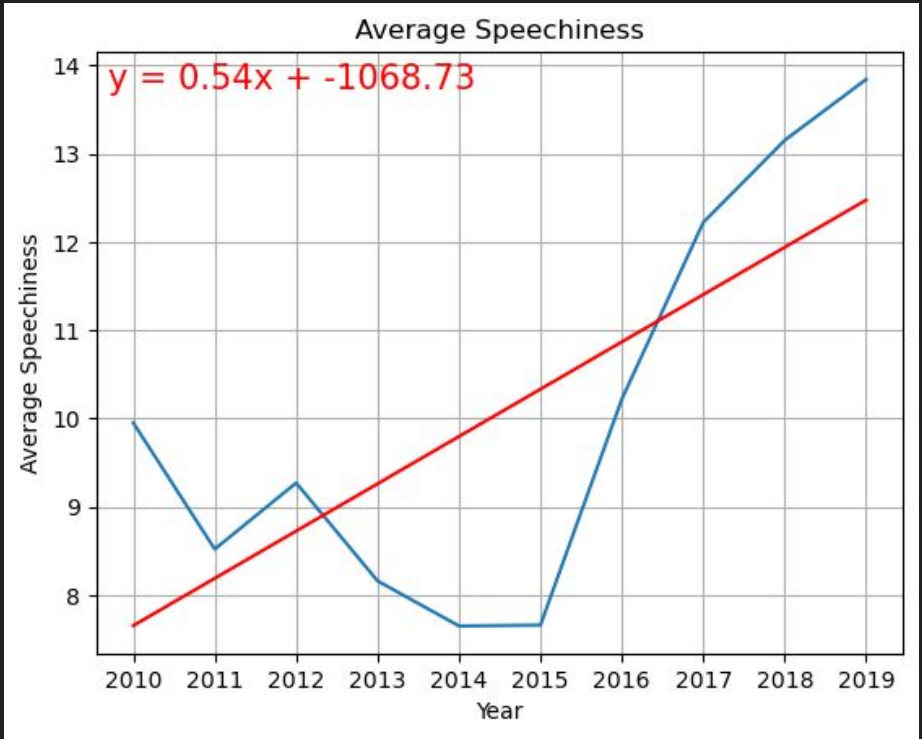


Q1. Average Song Attributes by year (2010s) cont.





Q1. Average Song Attributes by year (2010s) cont.



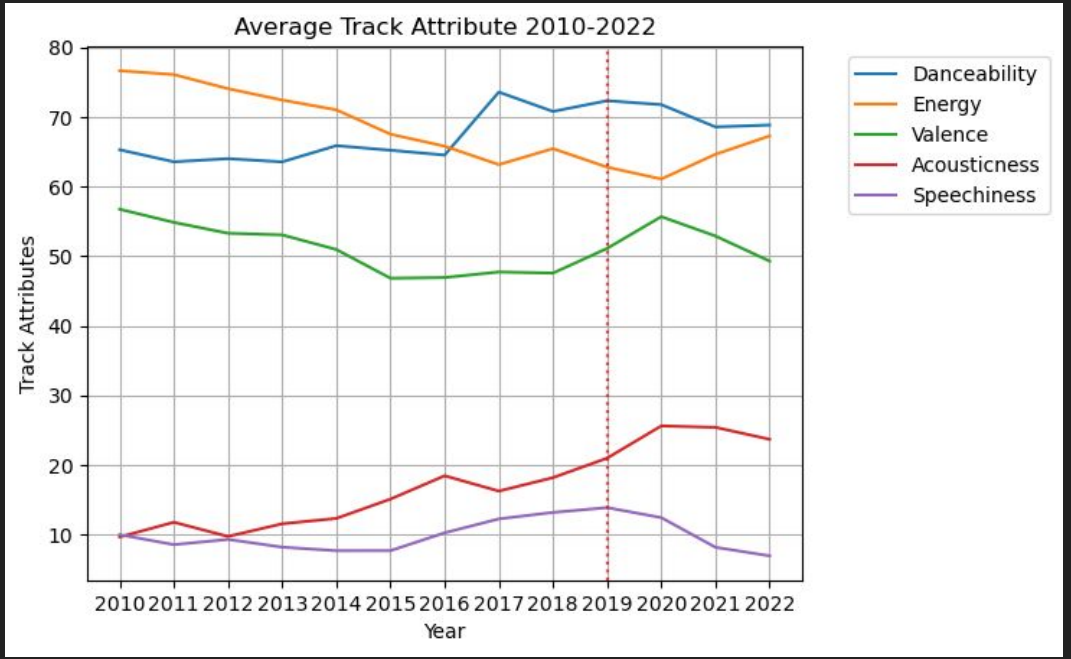
2010s P-Values	
Trait	P-Value
Danceability	2.06e-5
Energy	.19
Valence	.1.05e-17
Acousticness	1.77e-05
Speechiness	9.76e-09

In Summary, Danceability, Valence, Acousticness, and Speechiness, do seem to follow a trend in the 2010s

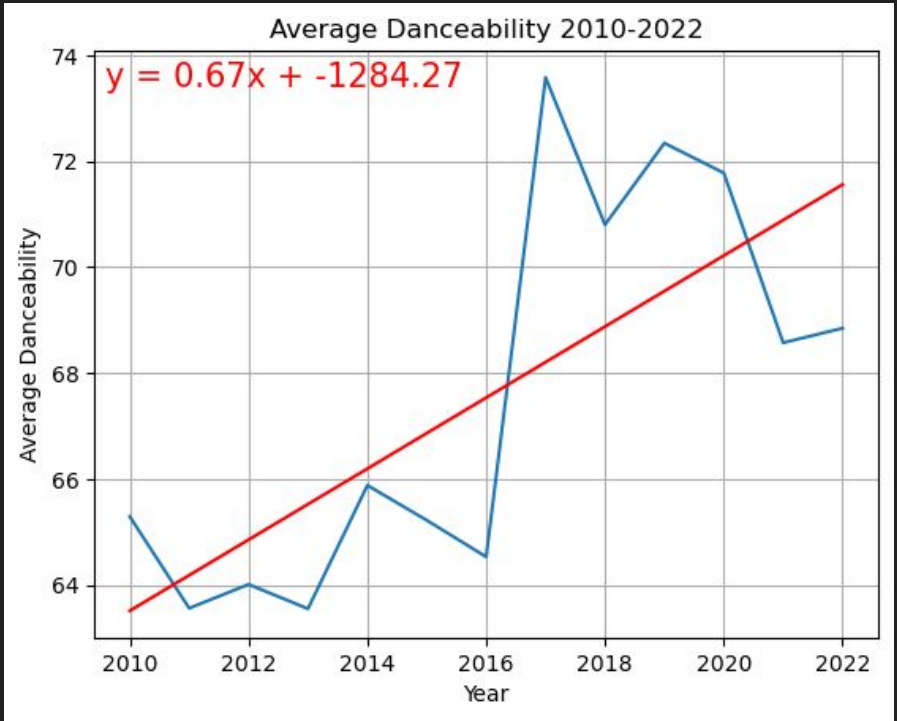
Question 2: Was the COVID 19 outbreak a cause for large shift in trends?

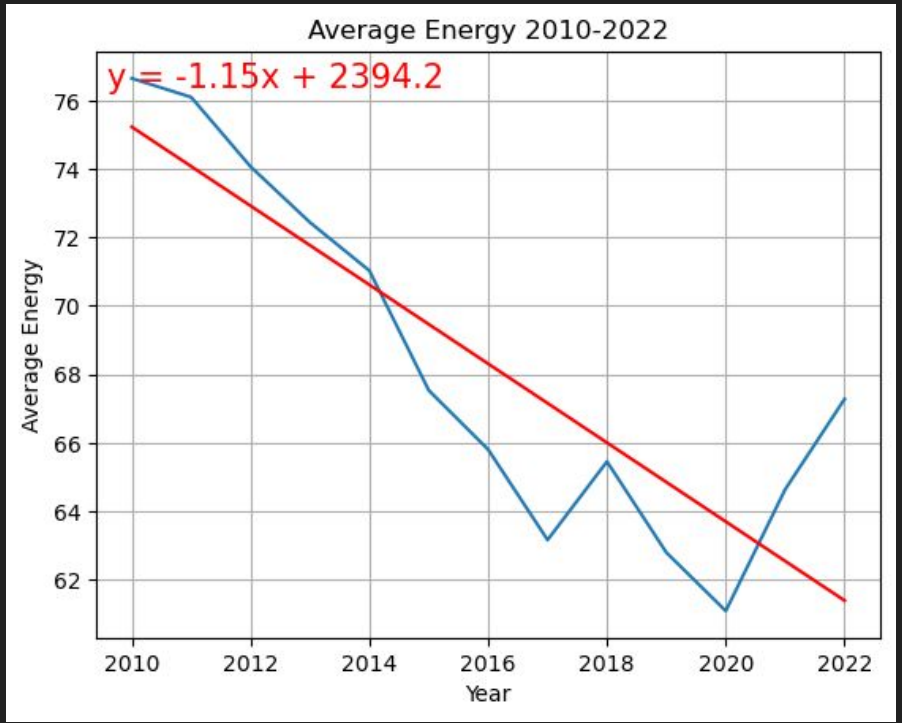
Null Hypothesis:
The Average Track Attribute Trends do not change from the 2010s to the 2020s

Alternate Hypothesis:
The Average Track Attribute trends do shift from the 2010s to the 2020s

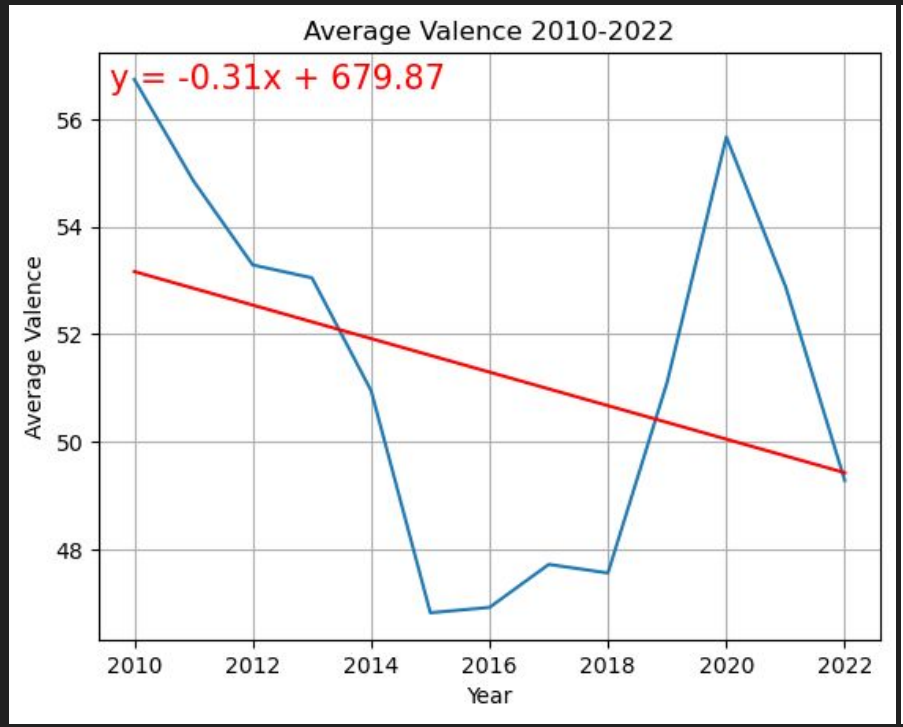


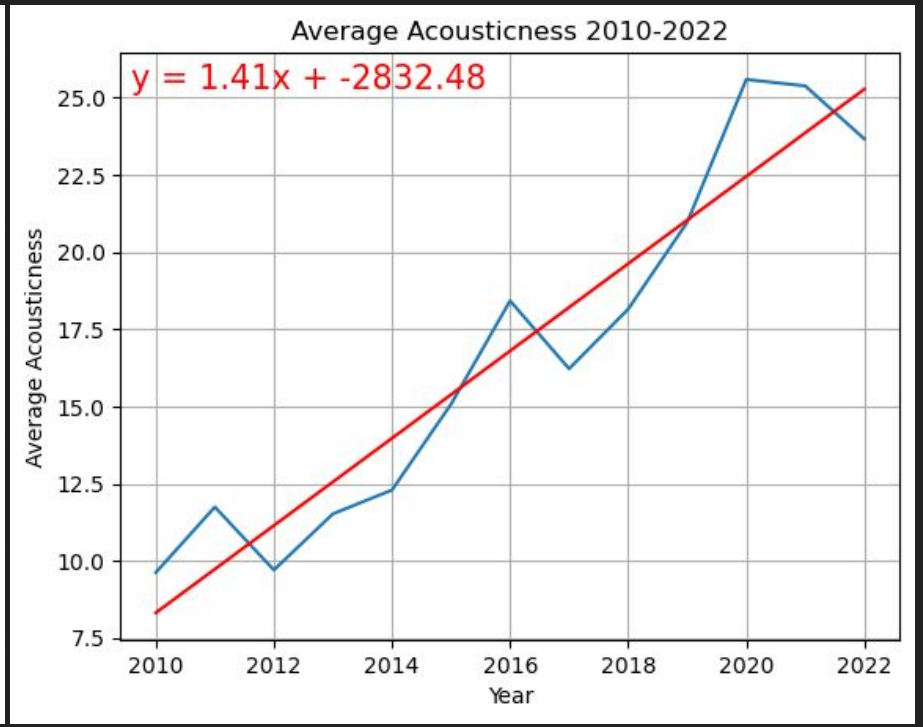
Q2. Average Song Attributes by year (2010-2022)



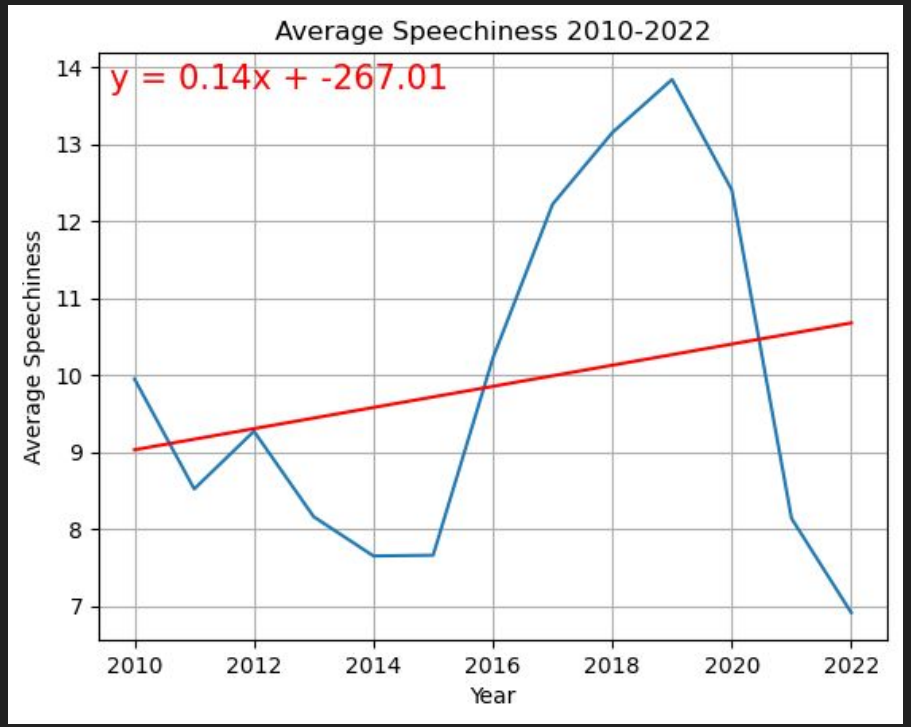


Q2. Average Song Attributes by year (2010-2022) cont.





Q2. Average Song Attributes by year (2010-2022) cont.



2010s vs 2020s T-Test	
Trait	P-Value
Danceability	.0121
Energy	7.68e-5
Valence	.3967
Acousticness	1.58e-6
Speechiness	.2218

In Summary, Danceability, Energy, and Acousticness appear to NOT follow the 2010's trend, indicating COVID-19 may be a cause for a shift in trait values.

Valence and Speechiness do appear to be in range of the 2010's trends.

Data Considerations/Limitations

- Attributes studied are defined and calculated by Spotify
  - Limited knowledge of how exact values are produced
- Starting in 2020 Spotify Top Tracks playlist lowered from 100 to 50
  - Less years and Less Data Per Year
- Spotify did not create an “Official” playlist for 2021, data gathered from Kaggle Dataset

Questions?

Credits/Sources

- Spotify Top 100 Songs of 2010-2019: By Michael Morris
  - Spotify Top 100 Songs of 2010-2019 | Kaggle
- Spotify top 50 songs in 2021: By Hanna Yukjymenko
  - Spotify top 50 songs in 2021 | Kaggle
- Spotify API Web API | Spotify for Developers
- Spotify API with Python Accessing the Spotify API with Python

Listen To The Hits

Top Hits of 2010 - playlist by Spotify

Top Hits of 2011 - playlist by Spotify

Top Hits of 2012 - playlist by Spotify

Top Hits of 2013 - playlist by Spotify

Top Hits of 2014 - playlist by Spotify

Top Hits of 2015 - playlist by Spotify

Top Hits of 2016 - playlist by Spotify

Top Hits of 2017 - playlist by Spotify

Top Hits of 2018 - playlist by Spotify

Top Hits of 2019 - playlist by Spotify

Top Tracks of 2020 - playlist by Spotify

Top Hits of 2021 - playlist by Cody Allison

Top Tracks of 2022 - playlist by Spotify