

History of Rock in Spotify

The following analysis gets the data from Kaggle, free license

What is Kaggle? According to online definitions, Kaggle, a subsidiary of Google LLC, is an online community of data scientists and machine learning practitioners. Inside the website can be found courses, datasets, contest/challenges including money.

Dataset can be uploaded by single usernames or by companies during a competition.

Scope of the Study

A lot of considerations can be made from the history of rock music, but the scope of this study is to support the changes that music rock did during the years.

Rock music, as an alternative of pop music (intended as common or soft) in the beginning was an underground music that gained fame during the years, with a constant increase. Some people or critics claim that rock is dead, but we will seek if there is a truth on this sentence.

Data

Dataset is from 2020 retrieved from Spotify covering rock songs from 1950 to 2020 with 5484 songs and 17 tags/label to identify and classify a song. From the tag list, only popularity is an index from the audience feedback while the remaining tags describe the song characteristics.

1. **Index**
2. **Name:** Song's name
3. **Artist**
4. **Release date**
5. **Length:** in minutes
6. **Popularity:** A value from 0 to 100
7. **Danceability:** Describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
8. **Acousticness:** A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
9. **Energy:** Represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.
10. **Instrumentalness:** Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal".
11. **Key:** The estimated overall key of the track. Integers map to pitches using standard [Pitch Class notation](#). E.g. 0 = C, 1 = C#/D \flat , 2 = D, and so on.

12. **Liveness:** Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
13. **Loudness:** The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks.
14. **Speechiness:** This 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.
15. **Tempo:** The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
16. **Time Signature:** An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
17. **Valence:** Describes 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).

Popularity requires some clarification from analytical point of view and need some assumptions. We don't know when the popularity was measured, monthly or yearly, and also in which year. Considering this lack of information, we will assume likelihood that popularity was calculated in 2020 when considering songs from 1950 to 2019.

Loading data

Data set of rock music is made of 5848 rows, one row for one song.

Data Pre-processing & Feature Engineering

After loading the data, we need to manipulate it according to our scope of the study, more specifically we will count the letters both in the artist's name and song's name.

The name of the song contains some noise created by the versions mastered or remastered. this creates a distortion in the real name of the song. Most of time, remastering a song has the only effect to clean using new technologies and also to refresh the mind of people.

Since there are 5848 rows in the data, this creates a lot of noise, so the best way for filtering data, is to preprocess in aggregated way following statistical parameters, *mean, max & min of the values for each year from 1956 to 2020*. This leads to a new data set of 65 rows where every row is one year.

As an example, find below the first 100 songs listed.

	index	name	artist
1	0	'Smells Like Teen Spirit'	'Nirvana'
2	1	'Stairway to Heaven - Remaster'	'Led Zeppelin'
3	2	'Bohemian Rhapsody - Remastered 2011'	'Queen'
4	3	'Imagine - Remastered 2010'	'John Lennon'
5	4	'(I Can't Get No) Satisfaction - Mono Version'	'The Rolling Stones'

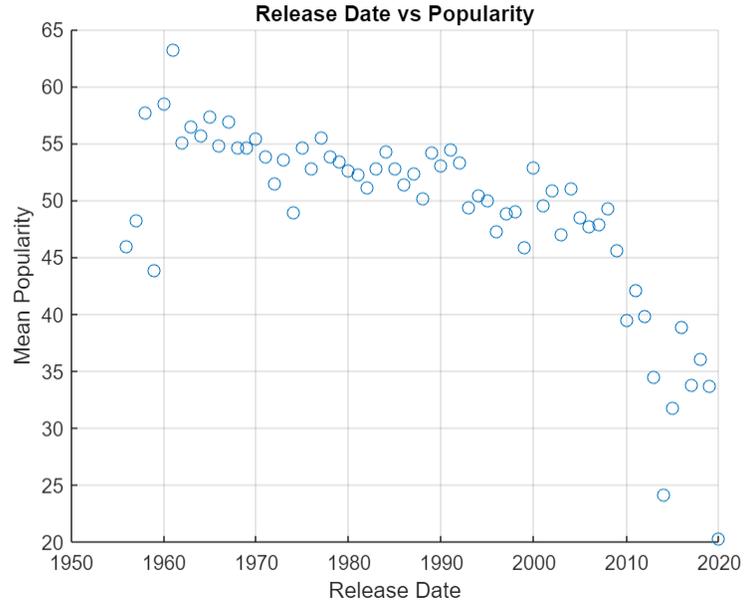
	index	name	artist
6	5	'Hotel California - 2013 Remaster'	'Eagles'
7	6	'Enter Sandman'	'Metallica'
8	7	'Whole Lotta Love - 1990 Remaster'	'Led Zeppelin'
9	8	'Comfortably Numb'	'Pink Floyd'
10	9	'One'	'U2'
11	10	'Born to Run'	'Bruce Springsteen'
12	11	'Smoke on the Water'	'Deep Purple'
13	12	'Hey Jude - Remastered 2015'	'The Beatles'
14	13	'Paranoid - 2012 - Remaster'	'Black Sabbath'
15	14	'Seven Nation Army'	'The White Stripes'
16	15	'Free Bird'	'Lynyrd Skynyrd'
17	16	'Back In Black'	'AC/DC'
18	17	'Sympathy For The Devil - 50th Anniversary Edition'	'The Rolling Stones'
19	18	'Wish You Were Here'	'Pink Floyd'
20	19	'With Or Without You - Remastered'	'U2'
21	20	'Sweet Child O' Mine'	'Guns N' Roses'
22	21	'I Still Haven't Found What I'm Looking For'	'U2'
23	22	'Like a Rolling Stone'	'Bob Dylan'
24	23	'All Along the Watchtower'	'Jimi Hendrix'
25	24	'Under the Bridge'	'Red Hot Chili Peppers'
26	25	'Creep'	'Radiohead'
27	26	'Kashmir - 1990 Remaster'	'Led Zeppelin'
28	27	'Light My Fire'	'The Doors'
29	28	'My Generation - Stereo Version'	'The Who'
30	29	'More Than a Feeling'	'Boston'
31	30	'Another Brick in the Wall, Pt. 2'	'Pink Floyd'
32	31	'Layla'	'Derek & The Dominos'
33	32	'Gimme Shelter'	'The Rolling Stones'
34	33	'You Shook Me All Night Long'	'AC/DC'
35	34	'Alive'	'Pearl Jam'
36	35	'Won't Get Fooled Again - Original Album Version'	'The Who'
37	36	'London Calling - Remastered'	'The Clash'
38	37	'Come As You Are'	'Nirvana'
39	38	'Purple Haze'	'Jimi Hendrix'

	index	name	artist
40	39	'Pride (In The Name Of Love) - Remastered 2009'	'U2'
41	40	'Dream On'	'Aerosmith'
42	41	'Sunday Bloody Sunday - Remastered 2008'	'U2'
43	42	'Rock and Roll - Remaster'	'Led Zeppelin'
44	43	'In My Life - Remastered 2009'	'The Beatles'
45	44	'Let It Be - Remastered 2009'	'The Beatles'
46	45	'Baba O'Riley'	'The Who'
47	46	'Jeremy'	'Pearl Jam'
48	47	'Welcome To The Jungle'	'Guns N' Roses'
49	48	'Bitter Sweet Symphony'	'The Verve'
50	49	'Roxanne'	'The Police'
51	50	'Money'	'Pink Floyd'
52	51	'Every Breath You Take - Remastered 2003'	'The Police'
53	52	'A Day In The Life - Remastered 2009'	'The Beatles'
54	53	'Sweet Emotion'	'Aerosmith'
55	54	'Born To Be Wild - Single Version'	'Steppenwolf'
56	55	'New Year's Day - Single Edit - Remastered'	'U2'
57	56	'Paint It, Black'	'The Rolling Stones'
58	57	'Black Dog - Remaster'	'Led Zeppelin'
59	58	'Basket Case'	'Green Day'
60	59	'November Rain'	'Guns N' Roses'
61	60	'Even Flow'	'Pearl Jam'
62	61	'You Can't Always Get What You Want'	'The Rolling Stones'
63	62	'Money for Nothing'	'Dire Straits'
64	63	'Nothing Else Matters'	'Metallica'
65	64	'Everlong'	'Foo Fighters'
66	65	'Angie'	'The Rolling Stones'
67	66	'Jump - 2015 Remaster'	'Van Halen'
68	67	'(Don't Fear) The Reaper'	'Blue Öyster Cult'
69	68	'Come Together - Remastered 2009'	'The Beatles'
70	69	'Jumpin' Jack Flash - Live'	'The Rolling Stones'
71	70	'Time'	'Pink Floyd'
72	71	'Should I Stay or Should I Go - Remastered'	'The Clash'
73	72	'Mr. Brightside'	'The Killers'

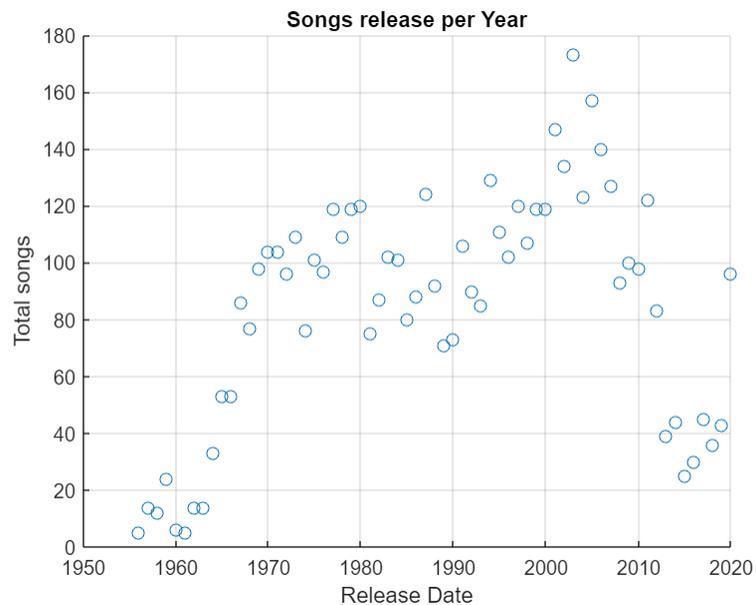
	index	name	artist
74	73	'Beautiful Day'	'U2'
75	74	'House Of The Rising Sun'	'The Animals'
76	75	'Lola - Coca Cola Version'	'The Kinks'
77	76	'Loser'	'Beck'
78	77	'Killing In The Name'	'Rage Against The Machine'
79	78	'All My Life'	'Foo Fighters'
80	79	'Walk This Way'	'Aerosmith'
81	80	'Sultans of Swing'	'Dire Straits'
82	81	'We Will Rock You - Remastered'	'Queen'
83	82	'Song 2 - 2012 Remaster'	'Blur'
84	83	'Love Will Tear Us Apart - 2010 Remaster'	'Joy Division'
85	84	'Where The Streets Have No Name - Remastered'	'U2'
86	85	'In the End'	'Linkin Park'
87	86	'Good Vibrations - Remastered'	'The Beach Boys'
88	87	'All Right Now'	'Free'
89	88	'Changes - 2015 Remaster'	'David Bowie'
90	89	'Paradise City'	'Guns N' Roses'
91	90	'One (Remastered)'	'Metallica'
92	91	'Hey Joe'	'Jimi Hendrix'
93	92	'Sunshine Of Your Love'	'Cream'
94	93	'Give It Away'	'Red Hot Chili Peppers'
95	94	'Take Me Out'	'Franz Ferdinand'
96	95	'Sweet Home Alabama'	'Lynyrd Skynyrd'
97	96	'Highway to Hell'	'AC/DC'
98	97	'L.A. Woman'	'The Doors'
99	98	'American Idiot'	'Green Day'
100	99	'Black Hole Sun'	'Soundgarden'

Data Visualization

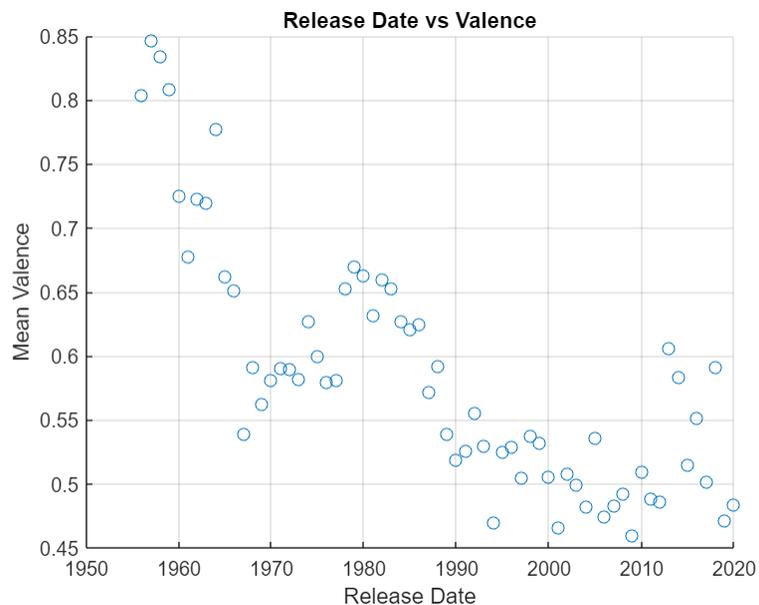
After processing data, it is time to proceed to visualization and get some deep insight about rock music in the last 65 years. There was an evolution over the years, from the technology used in the music industry as well as audience. The graph below shows the trend of song's popularity from 50s to 2020. It is very important to recall the fact that we assume that these values are calculated in 2020 considering songs from 1950 to 2020.



The total song released on yearly basis



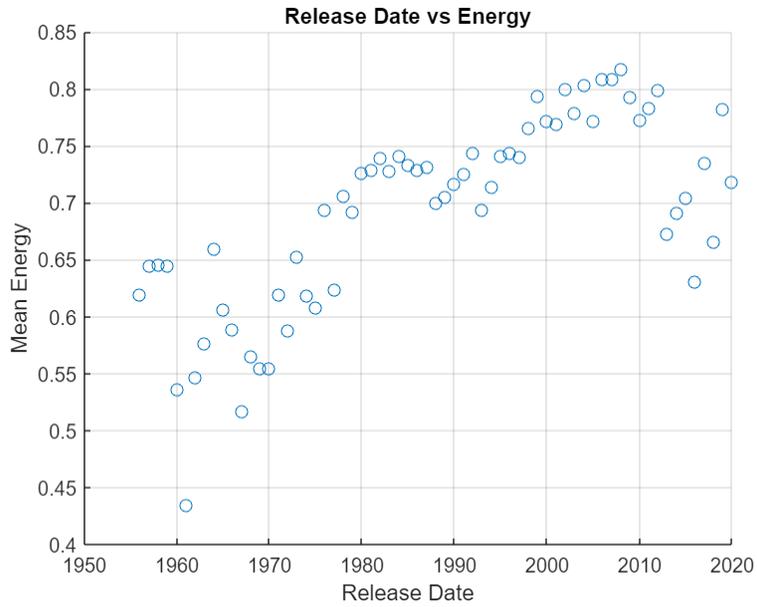
As recalled in the beginning, Valence describes the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g., happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g., sad, depressed, angry).



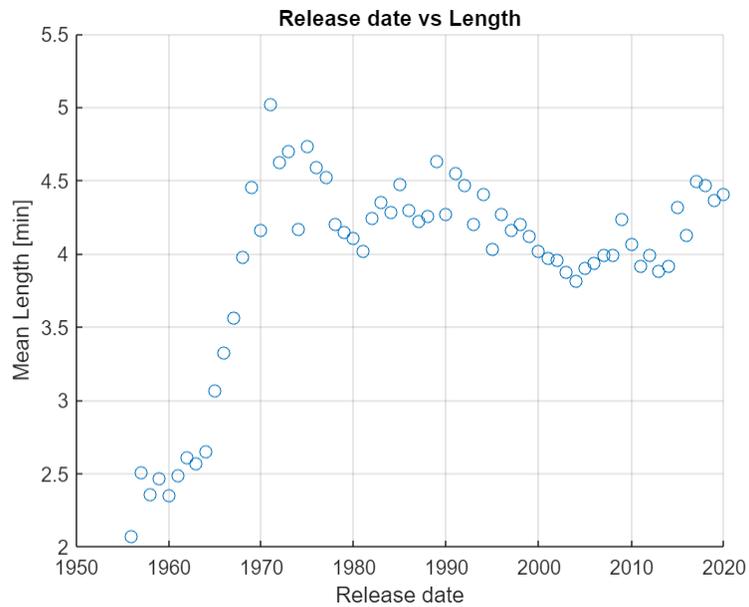
From above it is possible to notice the descending trend in the Valence value through the years.

Despite the descending trend in the Valence, Energy of the song is constantly increasing. Technology and audience evolution plays a role in this trend that can confirm that modern rock songs have more energy but they still have a descending valence value.

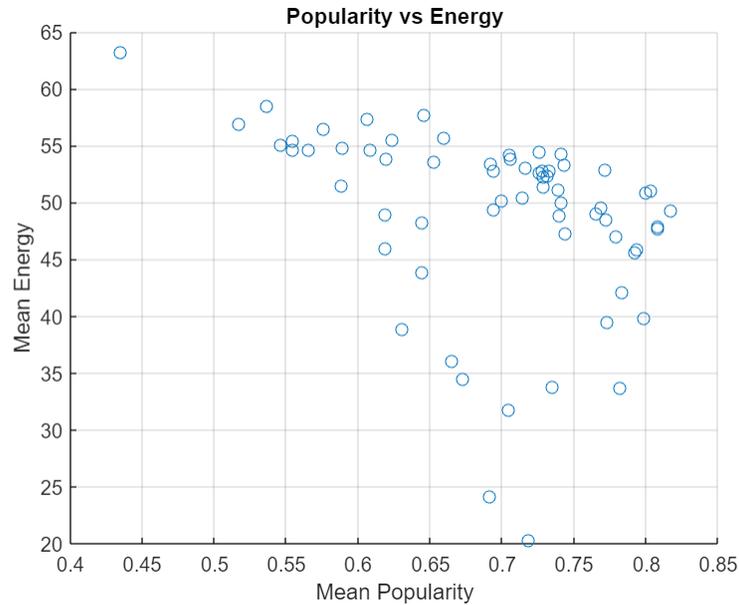
Energy represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.



Another important parameter to consider when analysing data is the length of the song in minutes. We see that the first rock songs were very short, mainly due to recording technology. From 1970 onwards, the length of the song remains mostly the same.



From below, a good insight shows that a song with an average Energy between 50 & 55, the expected Popularity can be between 0.7 and 0.75.



When talking about songs, it is nice to see if there was an evolution of the key note during the years, from the beginning of rock music. pitch class is calculated through sound processing and is an automatic output.

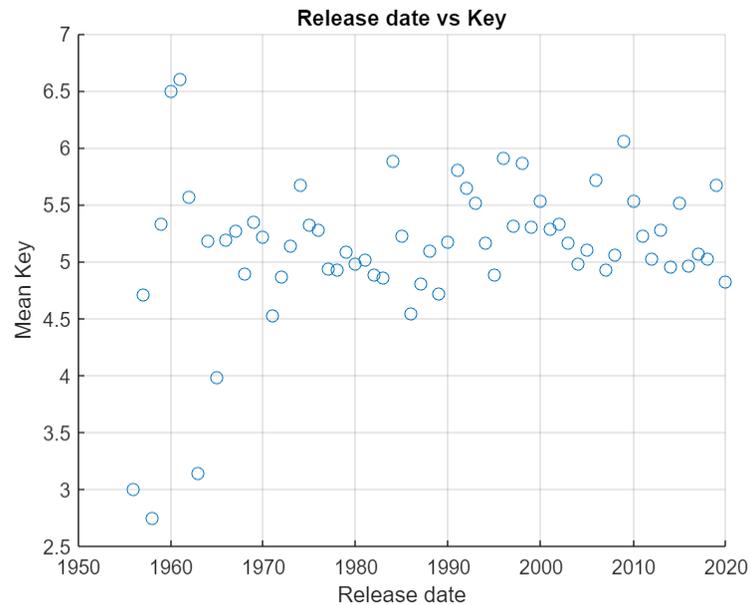
Pitch class	Tonal counterparts	Solfege
0	C (also B \sharp , Ddouble flat)	do
1	C \sharp , D \flat (also Bdouble sharp)	
2	D (also Cdouble sharp, Edouble flat)	re
3	D \sharp , E \flat (also Fdouble flat)	
4	E (also Ddouble sharp, F \flat)	mi
5	F (also E \sharp , Gdouble flat)	fa
6	F \sharp , G \flat (also Edouble sharp)	
7	G (also Fdouble sharp, Adouble flat)	sol
8	G \sharp , A \flat	
9	A (also Gdouble sharp, Bdouble flat)	la
10, t or A	A \sharp , B \flat (also Cdouble flat)	

11, e or B

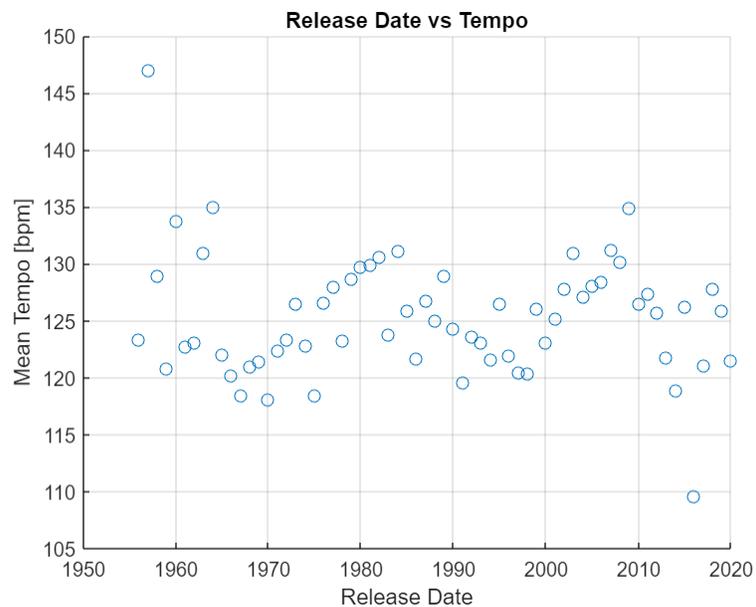
B (also Adouble sharp, Cb)

si

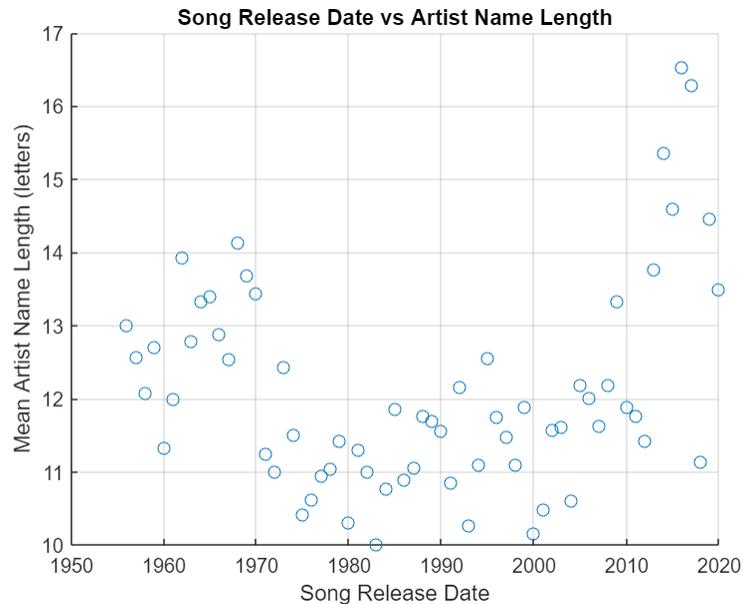
The list from Spotify shows that 5 is the common keynote from rock songs from almost 60 years. There is some noise in the data, but since decimals are not allowed, we can confirm a flat 5 or F as keynote.



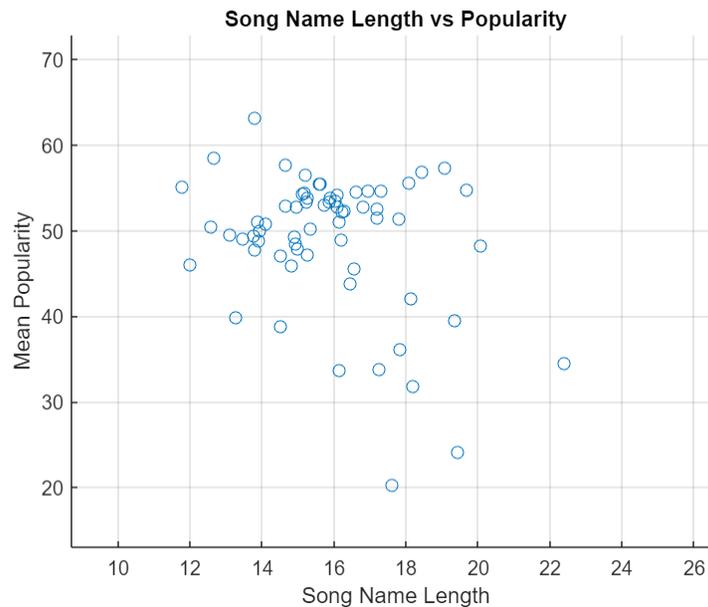
As done previously for keynote, tempo can be an indicator of the music trend within musicians and of course companies. It is worth to recall that the values from the tables refer to the mean value of all songs release in the reference year, so from the graph below we can see that the mean value of the tempo was in between 120 & 130 bpm.



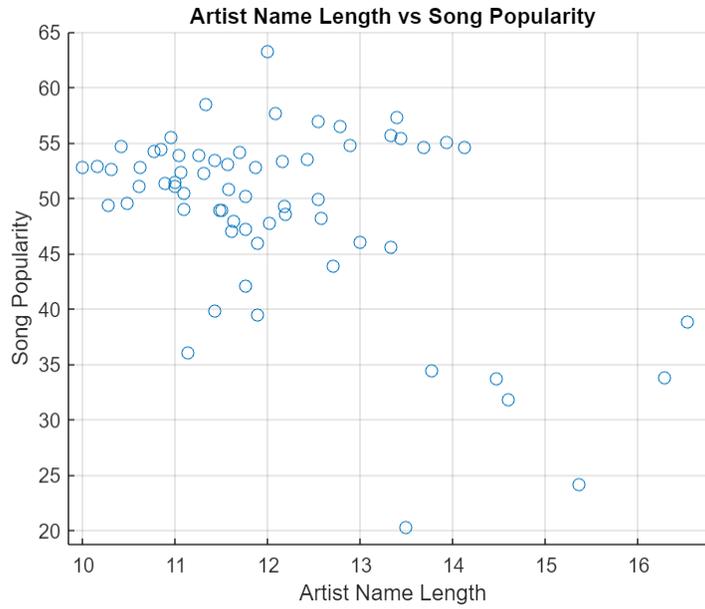
The second graph is about how the artist's name has changed through the years. After some stability from the 70s to 2010, the length increases up to 16 letters in the name and return to "normal" values in the 2020.



After some presentation, one the important information extracted is the popularity of the songs categorized by name length. It is very evident that popularity stays in the range of 16 letters length of the songs.

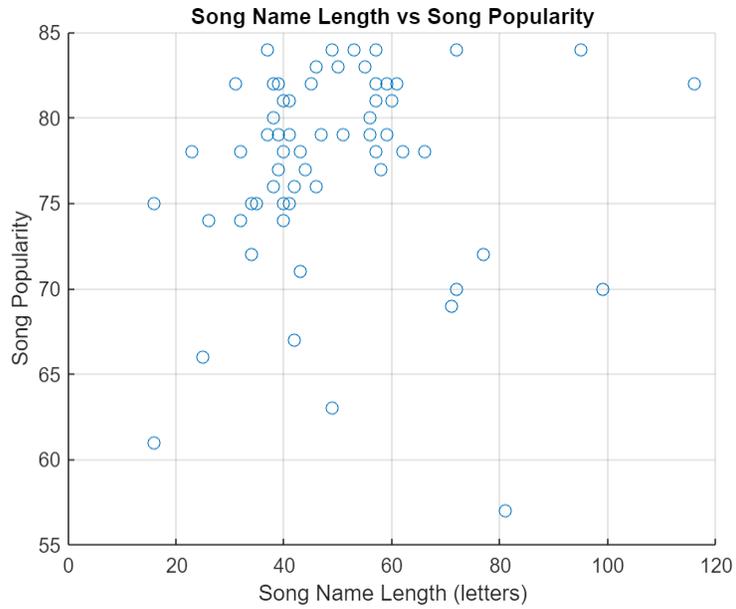


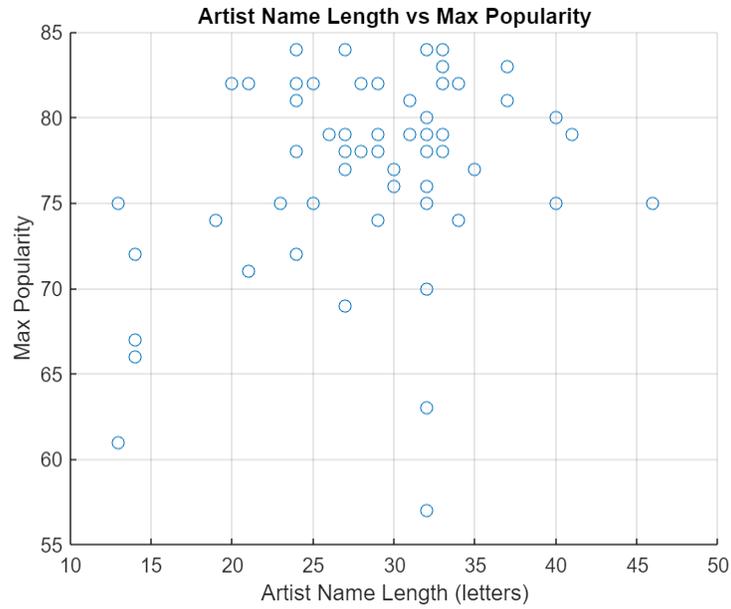
Another curious information is to graph the artist's name length and his popularity. As confirmation of a non-written rule, shorter artist name, highest popularity, maybe related to mnemonic and or sympathy.



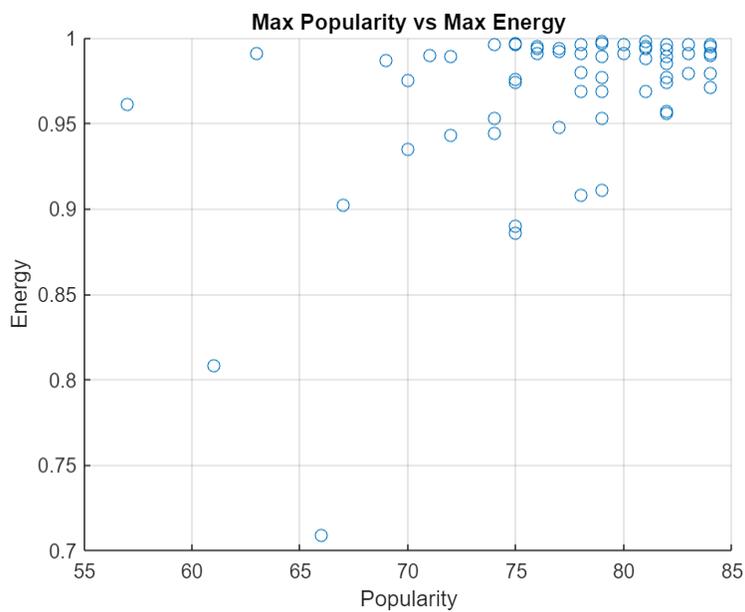
Analysis using Max values over the years

The graph below are done using only maximum values over the years.

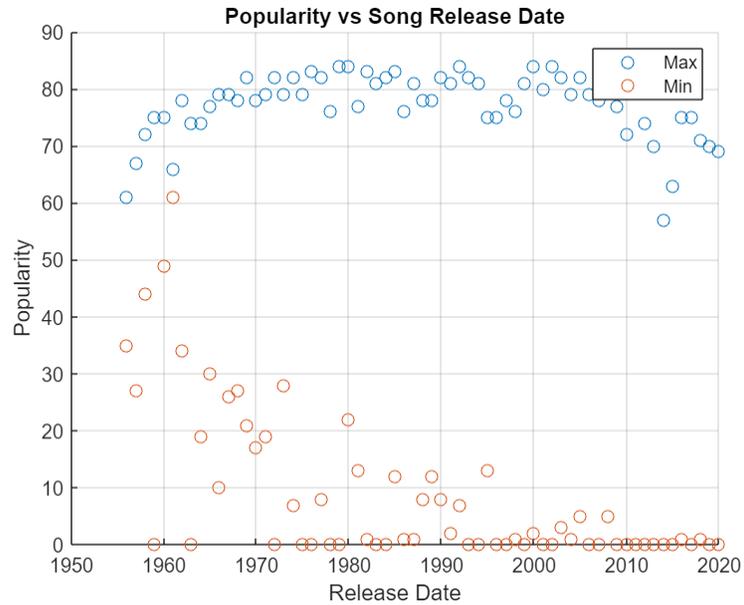




The max graph below, show that the maximum Energy correspond a Popularity between 75 & 85.



The last graph from the analysis recalls the Maximum and Minimum Popularity over the years showing an increasing trend for the songs from 1956 until 1980 and a descending trend from then until 2020. Even if we consider zero values as outliers, the descending trend of the minimum Popularity over the years is evident.



Conclusion

The scope of this study to identify a trend that show a decadence in the rock. If we consider popularity alone, the result is misleading because, as mentioned previously, the popularity of a song of the 50s was not calculated in the fifties but we assumed that was in done in 2022. From this point, older songs seem to be more popular than newest, reaching the maximum Popularity for songs released in between 80s and 90s.

From the graph "Release Date vs Energy" we can notice an increasing trend from the starting year having a little effect in terms of Popularity of the song released after the 90s.

The data presented show a descending interest in the rock music from 90's onwards, which reflect the audience sensation about music nowadays. From the other side, 80s & 90s still are the most popular year for rock music. One of the factors that can influence this result is that the age of the people who listen to rock music now. If we consider that teenage people in the 80s & 90s, now are around 30 to 40 years old, it is easy to assume the audience target.

If a new song will be released tomorrow following having tag values given, for sure will be more appreciated by the audience in that range of age.

Thank you for reading