WHAT MAKES A MUSICAL HIT?

NANDO N' BASS

The top ways to break the Topcharts…

Using Data Analysis to bring insights to a music production Studio

João Fernando
07/2024

Exploring the dataset

NANDO N' BASS

Spotify Tracks Dataset (extracted October 2022):

- 114K entries
- 89740 unique songs
- 114 musical genres

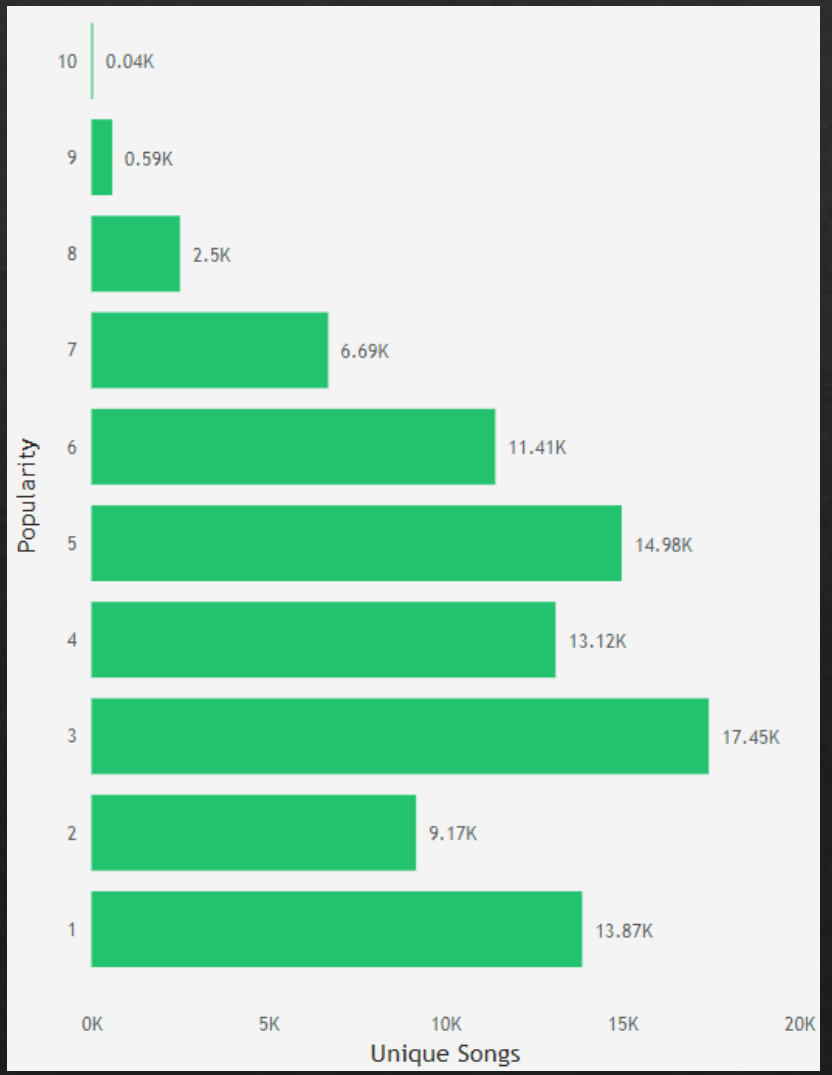
20 most popular genres

Track genre	AVGpopularity
pop-film	59.2830
k-pop	56.9530
chill	53.6510
sad	52.3790
grunge	49.5940
indian	49.5390
anime	48.7720
emo	48.1280
sertanejo	47.8660
pop	47.5760
progressive-house	46.6150
piano	45.2730
mandopop	45.0250
deep-house	44.8080
brazil	44.6700
electronic	44.3250
pagode	44.2980
ambient	44.1910
british	43.8020
metal	43.7050

20 least popular genres

Track genre	AVGpopularity
iranian	2.2100
romance	3.2450
latin	8.2970
detroit-techno	11.1740
chicago-house	12.3390
classical	13.0550
jazz	13.6280
grindcore	14.6250
kids	14.8880
idm	15.7660
honky-tonk	16.3550
country	17.0280
rock	19.0010
soul	19.7950
tango	19.8710
breakbeat	20.1230
reggae	20.6300
party	20.9820
happy	21.7160
black-metal	22.4490

Popularity Distribution



Non numeric variables

NANDO N' BASS

Explicit:

- Measures if a track has explicit lyrics (True or False)

Easy to understand

Key:

- The key/note for this track. Uses standard Pitch Class notation

Mode:

- Measures if a track is on Major or Minor scale (1 or 0)

Time Signature:

- Measures the beats per measure (4/4, 3/4…)

Technically complex

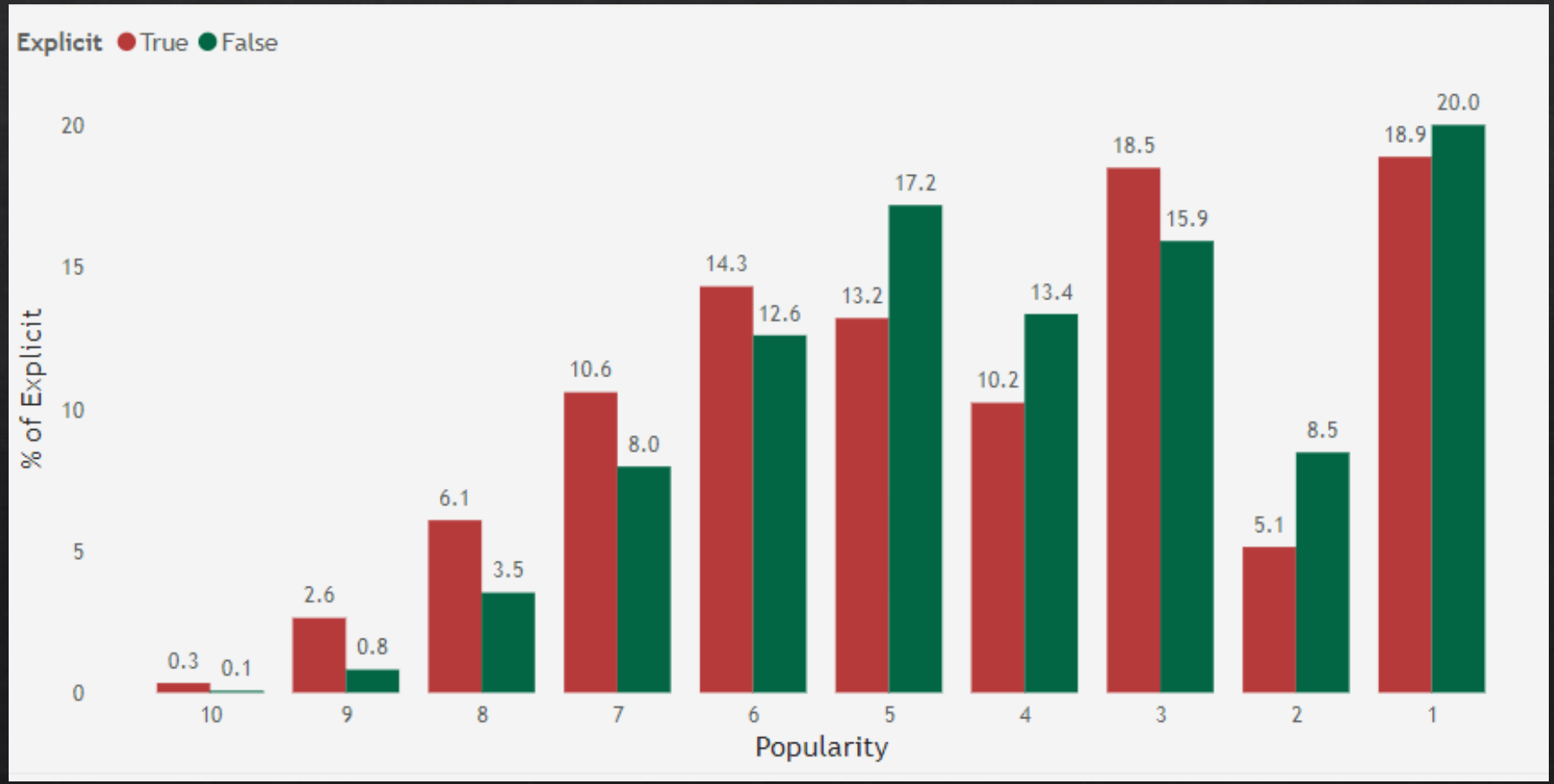
Explicit

NANDO N' BASS

explicit	Tracks	ID%	AVGpopularity
True	7704	8.5848	36.4542
False	82036	91.4152	32.9382

Popularity Rank	Tracks	% of Explicit	Cumulative %
10	**43**	**0.09**	**0**
True	14	0.34	0
False	29	0.06	0
9	**585**	**0.97**	**1**
True	158	2.64	3
False	427	0.81	1
8	**2501**	**3.75**	**5**
True	387	6.08	9
False	2114	3.53	4
7	**6689**	**8.20**	**13**
True	790	10.60	20
False	5899	7.98	12
6	**11413**	**12.76**	**26**
True	1139	14.33	34
False	10274	12.61	25
5	**14984**	**16.86**	**43**
True	1104	13.21	47
False	13880	17.20	42
4	**13115**	**13.09**	**56**
True	879	10.23	57
False	12236	13.36	56
3	**17445**	**16.16**	**72**
True	1698	18.52	76
False	15747	15.94	71
2	**9174**	**8.19**	**80**
True	491	5.14	81
False	8683	8.48	80
1	**13866**	**19.94**	**100**
True	1047	18.91	100
False	12819	20.03	100
Total	**89741**	**100.00**	**100**

- Being explicit has a **SLIGHT** positive impact on popularity



Explicit

NANDO N' BASS

Most explicit genres	Count	AVGpopularity
comedy	656	24.6280
emo	465	48.1280
sad	450	52.3790
j-dance	391	26.6560
hardcore	325	36.1050
hip-hop	319	37.7590
funk	304	32.3220
dancehall	302	33.4670
metalcore	291	43.4770
death-metal	251	32.1690
latino	249	25.6560
industrial	236	31.0410
french	219	41.0670
turkish	218	40.6980
reggaeton	212	23.8580
dance	174	22.6900
chill	171	53.6510
reggae	167	20.6300
alternative	164	24.3370
grindcore	153	14.6250

It's hard to quantify the impact of a song being explicit:

- Instrumental genres don't have lyrics, and as such are not explicit. These genres tend to be less popular.
- There can external factors limiting access to explicit songs, like radio play or age restrictions.

Recommendation:

No need to shy away from explicit songs. But no need to push for them.

Key, mode and time signature

NANDO N' BASS

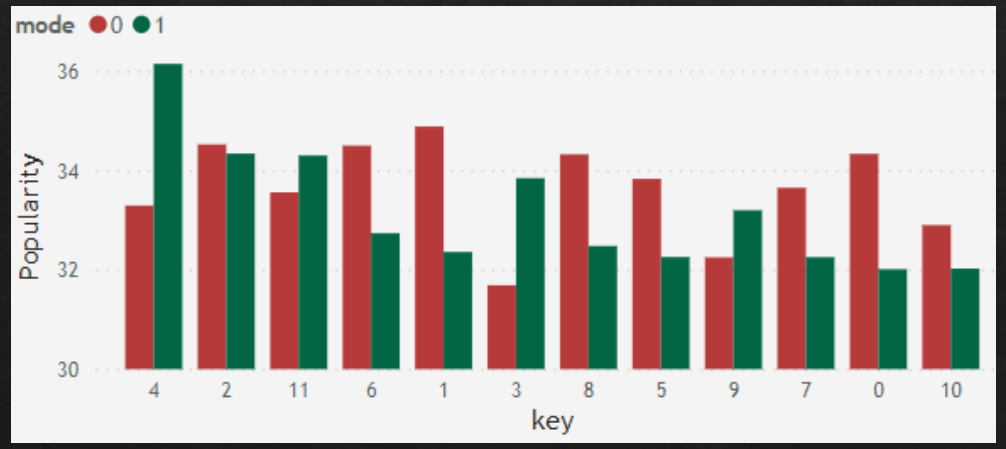
key	0	1	Total
7	2448	8103	**10551**
0	2307	8045	**10352**
2	2120	7207	**9327**
9	3583	5415	**8998**
1	2766	5810	**8576**
5	3176	4132	**7308**
4	3721	3412	**7133**
11	4014	3115	**7129**
6	3002	3137	**6139**
10	3001	2888	**5889**
8	1462	4108	**5570**
3	979	1790	**2769**
Total	**32579**	**57162**	**89741**

Translating key and mode analysis for non-musicians:

- Keys 7, 0, 2 are Sol, Re, Do.
  These are the most common music keys in the Western World.
- Major mode is 1, Minor mode is 0.
  Major scales sound “happier” than their “gloomier” Minor counterparts.

25% of all songs are in Major Sol, Re or Do

Average Popularity by key and mode



Hard to quantify the impact of key or mode on popularity:

- Certain genres are much more biased to certain key/mode combinations, and just looking at averages and distributions can lead to wrong conclusions.

Key, mode and time signature

NANDO N' BASS

time_signature	ID%
4	89.3359
3	8.0659
5	1.6018
1	0.8535
0	0.1430

Time signature is extremely homogeneous:

- 89% of all tracks are in 4/4 time signature

These findings reflect common musical knowledge. Times outside of 4/4 and 3/4 tend to have an arrhythmic feeling to them.

The popular keys, modes and time signatures tend to have a more familiar sound to them.
A slight change can make a song sound unnatural or exotic.

Genres with the most songs not in 4/4

track_genre	Tracks	AVGpopularity
sleep	568	35.0710
comedy	461	24.6280
romance	371	3.2450
ambient	352	44.1910
opera	341	24.6210
iranian	332	2.2100
classical	309	13.0550
new-age	305	29.0650
disney	245	27.5360
piano	240	45.2730
j-dance	223	26.6560
show-tunes	222	31.2460
guitar	222	29.5260
tango	216	19.8710
folk	198	38.0060
idm	198	15.7660
grindcore	196	14.6250
honky-tonk	193	16.3550
jazz	193	13.6280
songwriter	189	37.8130

<u>Recommendation:</u>

There is little incentive to deviate from the norm, but success can also be found in niche communities.

Numeric variables - Popularity

NANDO N' BASS

poprank	Tracks	id%	duration(ms)	minutes	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
10	43	0.0479	207237.57	3.45	0.7	0.68	-6.08	0.1	0.22	0.002	0.18	0.44	120.03
9	585	0.6519	213182.21	3.55	0.64	0.68	-6.15	0.08	0.19	0.026	0.17	0.52	119.37
8	2501	2.7869	220066.96	3.67	0.61	0.67	-6.81	0.07	0.23	0.04	0.17	0.5	120.33
7	6689	7.4538	219003.87	3.65	0.58	0.63	-7.99	0.08	0.3	0.094	0.18	0.47	121.05
6	11413	12.7179	222720.90	3.71	0.57	0.62	-8.46	0.07	0.32	0.157	0.19	0.42	122.47
5	14984	16.6971	237204.34	3.95	0.58	0.63	-8.05	0.07	0.34	0.138	0.24	0.46	123.42
4	13115	14.6144	230996.64	3.85	0.56	0.66	-8.18	0.08	0.31	0.132	0.23	0.51	123.24
3	17445	19.4395	227849.72	3.80	0.52	0.65	-8.74	0.12	0.35	0.178	0.25	0.47	122.3
2	9174	10.2229	240743.95	4.01	0.55	0.68	-8.64	0.09	0.27	0.354	0.21	0.44	125.16
1	13865	15.4502	222669.14	3.71	0.59	0.62	-8.31	0.08	0.32	0.142	0.18	0.51	119.72

Characteristics of more popular songs:

- Slightly shorter.
- More danceable and energetic.
- Louder.
- Less acoustic.
- Not instrumental.
- Average valence (positivity)

Are these reflected in the most popular genres?

Numeric variables - Genres

Track genre	AVGpopularity	duration(ms)	minutes	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
pop-film	59.2830	279657.08	4.66	0.6	0.6	-7.87	0.06	0.44	0.008	0.17	0.53	117.26
k-pop	56.9530	251528.70	4.19	0.65	0.68	-6.47	0.09	0.29	0.01	0.18	0.56	119.22
chill	53.6510	169009.97	2.82	0.66	0.43	-10.46	0.1	0.53	0.178	0.17	0.4	115.48
sad	52.3790	153800.88	2.56	0.69	0.46	-10.28	0.13	0.47	0.106	0.17	0.42	119.06
grunge	49.5940	235579.06	3.93	0.46	0.8	-5.66	0.06	0.05	0.037	0.22	0.4	129.35
indian	49.5390	245473.10	4.09	0.59	0.57	-8.85	0.07	0.48	0.039	0.17	0.46	116.14
anime	48.7720	210204.08	3.50	0.54	0.67	-7.93	0.09	0.27	0.263	0.2	0.43	123.53
emo	48.1280	189690.33	3.16	0.6	0.67	-6.72	0.11	0.2	0.03	0.18	0.44	126.99
sertanejo	47.8660	204583.55	3.41	0.59	0.71	-5.49	0.06	0.44	0	0.47	0.62	127.05
pop	47.5760	220672.78	3.68	0.63	0.61	-6.95	0.08	0.34	0.009	0.16	0.51	120.93
progressive-house	46.6150	206842.19	3.45	0.62	0.81	-5.42	0.07	0.06	0.2	0.21	0.37	125.09
piano	45.2730	203966.54	3.40	0.46	0.32	-17.01	0.05	0.7	0.516	0.14	0.31	118.08
mandopop	45.0250	251550.05	4.19	0.55	0.5	-8.18	0.04	0.48	0.004	0.16	0.35	123.48
deep-house	44.8080	219344.58	3.66	0.71	0.74	-6.97	0.06	0.1	0.238	0.17	0.45	120.87
brazil	44.6700	274230.48	4.57	0.56	0.62	-7.6	0.1	0.33	0.013	0.3	0.47	121.94
electronic	44.3250	233013.33	3.88	0.65	0.69	-7.14	0.08	0.18	0.249	0.19	0.39	122.95
pagode	44.2980	228321.44	3.81	0.58	0.71	-7.27	0.06	0.54	0.001	0.52	0.69	126.86
ambient	44.1910	237059.04	3.95	0.37	0.24	-18.6	0.04	0.78	0.676	0.13	0.17	111.11
british	43.8020	221502.45	3.69	0.5	0.51	-10.52	0.05	0.43	0.194	0.17	0.44	118.83
metal	43.7050	256462.96	4.27	0.46	0.84	-5.43	0.07	0.04	0.065	0.21	0.42	127.91

Track genre	AVGpopularity	duration(ms)	minutes	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
iranian	2.2100	319709.54	5.33	0.3	0.55	-13.08	0.09	0.41	0.59	0.19	0.15	114.77
romance	3.2450	200340.67	3.34	0.43	0.29	-13.21	0.06	0.87	0.054	0.21	0.39	109.69
latin	8.2970	213894.43	3.56	0.72	0.73	-5.26	0.09	0.18	0.009	0.17	0.63	119.81
detroit-techno	11.1740	372012.40	6.20	0.72	0.71	-11.09	0.07	0.07	0.7	0.15	0.47	126.59
chicago-house	12.3390	366853.87	6.11	0.77	0.73	-9.15	0.07	0.04	0.483	0.14	0.59	123.92
classical	13.0550	233075.00	3.88	0.38	0.19	-20.15	0.05	0.92	0.619	0.17	0.38	107.95
jazz	13.6280	190342.20	3.17	0.51	0.35	-11.58	0.06	0.72	0.08	0.17	0.49	112.64
grindcore	14.6250	133753.51	2.23	0.27	0.92	-6.24	0.14	0.01	0.538	0.28	0.22	119.23
kids	14.8880	156693.05	2.61	0.78	0.61	-8.12	0.15	0.24	0.02	0.17	0.68	121.75
idm	15.7660	267386.54	4.46	0.53	0.56	-12.41	0.08	0.38	0.675	0.17	0.3	123.3
honky-tonk	16.3550	162327.93	2.71	0.57	0.37	-11.76	0.04	0.71	0.057	0.19	0.63	116.56
country	17.0280	205999.03	3.43	0.56	0.6	-7.28	0.04	0.32	0.006	0.18	0.52	123.96
rock	19.0010	219674.52	3.66	0.54	0.68	-6.66	0.05	0.21	0.021	0.17	0.54	126.32
soul	19.7950	199718.00	3.33	0.6	0.53	-8.28	0.06	0.4	0.014	0.18	0.52	114.63
tango	19.8710	182868.04	3.05	0.54	0.37	-10.06	0.08	0.85	0.14	0.23	0.58	114.55
breakbeat	20.1230	321762.22	5.36	0.65	0.85	-6.03	0.09	0.03	0.396	0.22	0.47	133.03
reggae	20.6300	214684.05	3.58	0.75	0.73	-5.29	0.12	0.18	0.003	0.19	0.65	119.87
party	20.9820	201067.56	3.35	0.67	0.87	-4.41	0.09	0.09	0.002	0.29	0.68	131.3
happy	21.7160	241986.19	4.03	0.55	0.91	-5.42	0.12	0.06	0.264	0.28	0.33	152.91
black-metal	22.4490	311103.12	5.19	0.3	0.87	-6.53	0.09	0.03	0.441	0.24	0.19	128.6

Track genre	AVGpopularity	duration(ms)	minutes	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
emo	48.1280	189690.33	3.16	0.6	0.67	-6.72	0.11	0.2	0.03	0.18	0.44	126.99
latin	8.2970	213894.43	3.56	0.72	0.73	-5.26	0.09	0.18	0.009	0.17	0.63	119.81

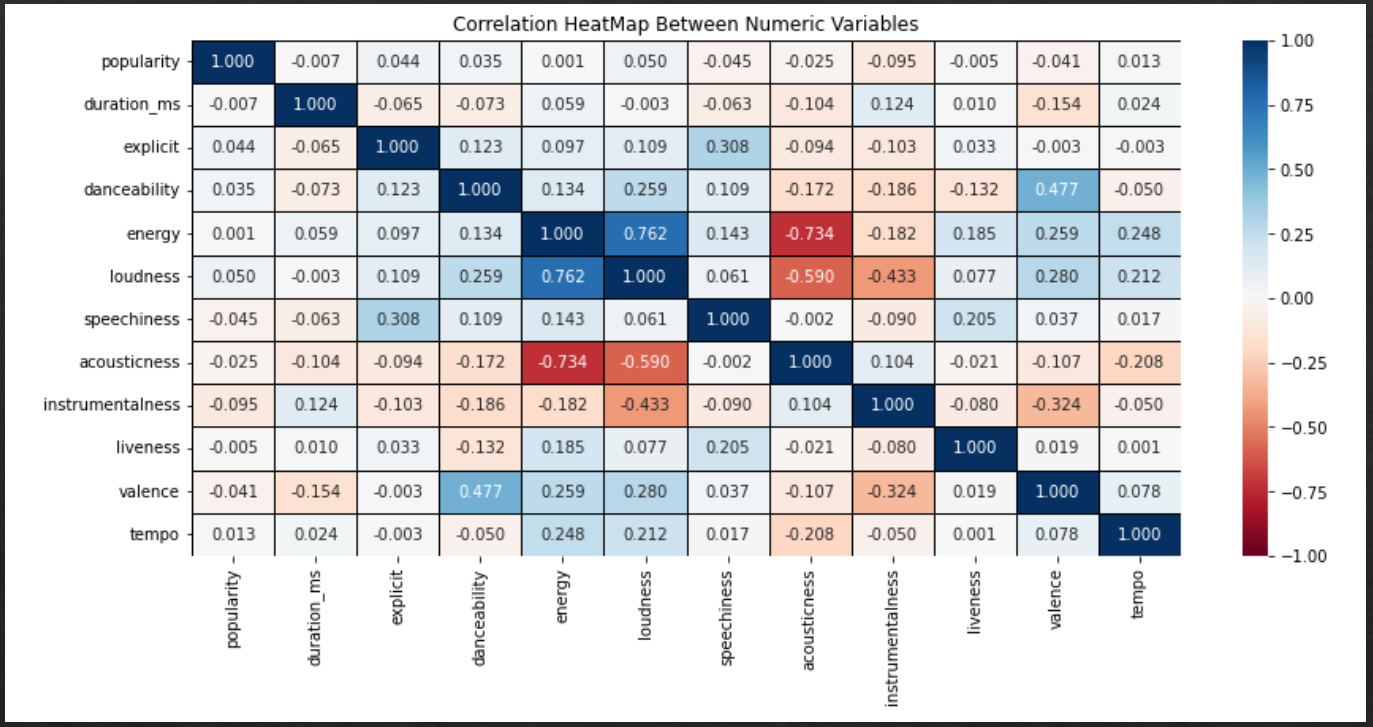
Characteristics of more popular genres:

- Slightly shorter ?
  Not too long, not too short
- More danceable and energetic ?
  High but moderate values
- Louder ?
  No easily visible impact
- Less acoustic ✓
  Popular genres tend to be slightly less acoustic
- Not instrumental ✓
  Popular genres tend to be less instrumental

Recommendation:

Dataset not adequate to analyze genre popularity.
Focus instead on song popularity.

Numeric variables - Correlation



Overall low Popularity correlations

Positive:

- Explicit
- Danceability
- Loudness
- Tempo

Negative:

- Speechiness
- Acousticness
- Instrumentalness
- Valence

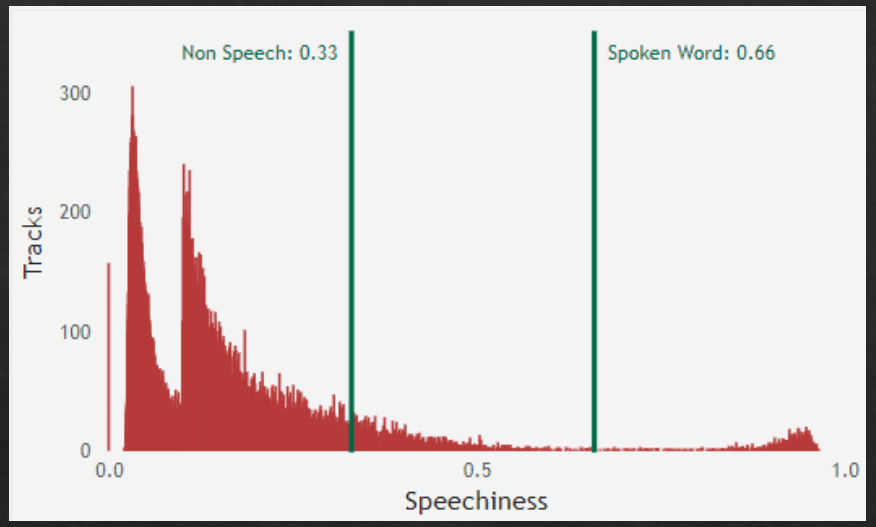
Non Popularity correlations

- Explicit songs have more speechiness
- Danceabilty and Energy are both highly positively correlated with loudness and valence.
- Danceability and Energy are not highly correlated. Being energetic does not make a song danceable.
- Acousticness has a high negative impact on both energy and loudness.
- Live songs have a positive correlation with energy and speechiness.

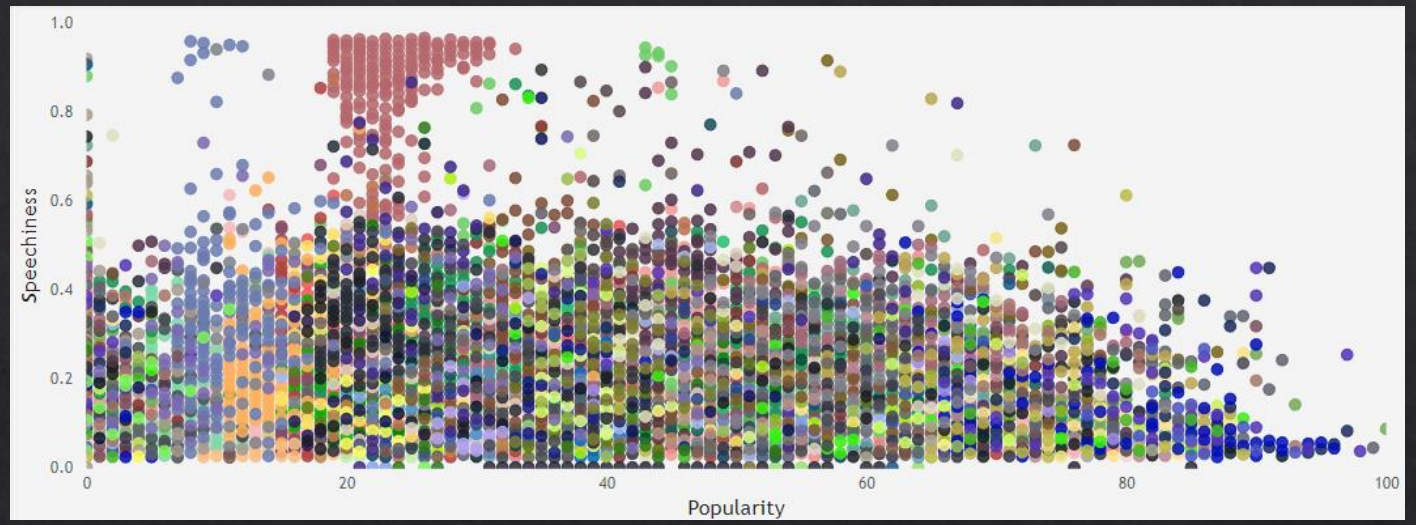
Numeric variables - Speechiness

NANDO N' BASS

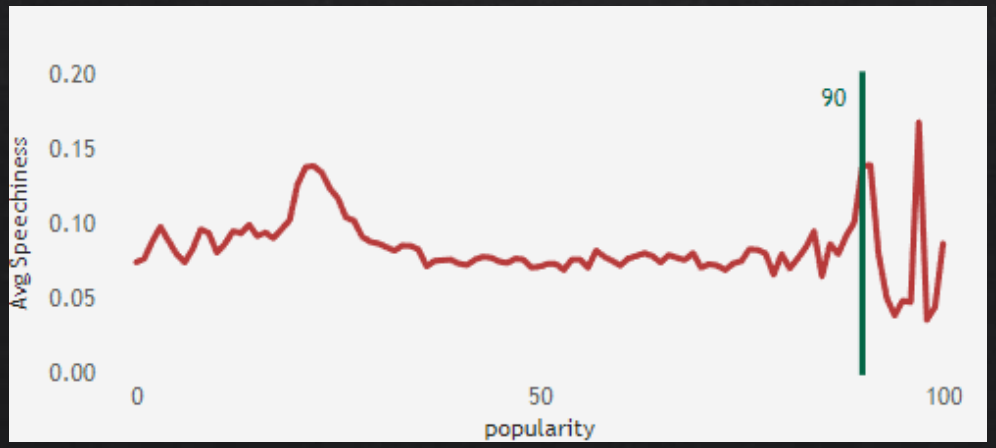
Speechiness Distribution



Speechiness by gender across Popularity



Avg Speechiness across Popularity



- Relevant negative correlation.

  BUT

- Homogeneous across popularity ranks.

Recommendations

NANDO N' BASS

- No need to shy away from explicit songs. But no need to push for them.
- Regarding musical structure, there is little incentive to deviate from the norm, but success can also be found in niche communities.
- Energy has a positive impact but is not equal to danceability (metal VS house).
- Higher danceability, and loudness should lead to a higher popularity.
- Higher acousticness and instrumental should lead to a lower popularity.

Next Steps

NANDO N' BASS

Expand the dataset:

- None of the variables has a very high correlation value with popularity, we can expand the amount of tracks we have, or even our timeframe.
- We may be missing relevant variables, like release year, which could add to this analysis.

Delve deeper into the different variables:

- Valence having a negative correlation is curious. After looking at the genres, “Party” and “Kids” show very high valence values, but very low popularity, skewing this results.
  In this case, Natural Language Processing and Sentiment Analysis could be good approaches.

Advanced predictive models:

- By doing a more thorough statistical approach, like looking at the distributions of the different variables or dimensionality reduction, we could get some deeper insights with the data own.

Redefining musical genres:

- At a superficial level, the current data set does not allow for a deep analysis focused on musical genres. By clustering genres we could shine a new light on this topic.