Music Prediction Model

Teng Liang, Richard Yang, Elaine He

Context and approach

- Music is one of the most important aspects of society and culture.
- Music has the potential to culturally, morally, and socially influence our society.
- We aim to see what influences the popularity of music throughout the years.
- We poise the question whether the mood of the music affects the mood of the society, comparing trends throughout the years.
- We also want to explore the influence of musician as influencer with thought of social media.

Dataset

	year	danceability	energy	valence	tempo	loudness	mode	key	acousticness	instrumentalness	liveness	speechiness	duration_ms	popularity
0	1921	0.425661	0.236784	0.425495	100.397758	-17.095437	1	7	0.895823	0.322330	0.215814	0.077258	229911.9141	0.351562
1	1922	0.480000	0.237026	0.534056	101.376139	-19.179958	1	10	0.939236	0.440470	0.238647	0.115419	167904.5417	0.138889
2	1923	0.568462	0.246936	0.624788	112.456598	-14.373882	1	0	0.976329	0.401932	0.236656	0.098619	178356.3018	5.727811
3	1924	0.548654	0.347033	0.668574	120.653359	-14.202304	1	10	0.935575	0.583955	0.237875	0.090210	188461.6498	0.603376
4	1925	0.571890	0.264373	0.616430	115.671715	-14.516707	1	5	0.965422	0.408893	0.243094	0.115457	184130.6996	2.707224
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	2016	0.599976	0.592877	0.430769	119.070344	-7.949913	1	0	0.280290	0.074646	0.180198	0.107298	219400.7638	61.371254
96	2017	0.612286	0.586739	0.414465	116.840277	-8.422697	1	1	0.289916	0.098209	0.194218	0.111752	209343.6130	64.861500
97	2018	0.664930	0.590591	0.447141	122.004325	-7.253666	1	1	0.271941	0.035948	0.171781	0.128140	200919.1190	67.276000
98	2019	0.644215	0.578796	0.465856	118.868163	-8.041738	1	1	0.289298	0.076518	0.167161	0.124799	197733.1330	69.655500
99	2020	0.673077	0.611914	0.482755	121.228704	-7.204024	1	1	0.247374	0.039052	0.177048	0.143505	197114.6623	63.111048

100 rows × 14 columns

Data by year

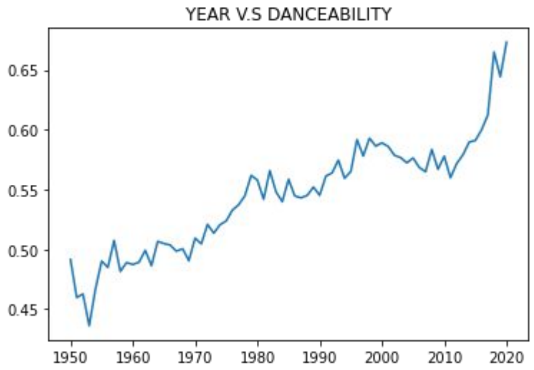
Analysis of dataset

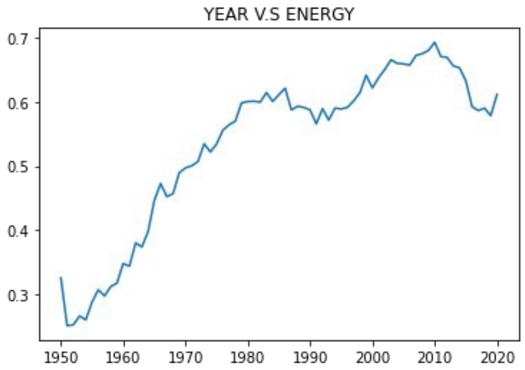
- The dataset includes data such as danceability, valence, tempo, and other technical aspects of music sorted by artist and year.
- With this dataset, we can visualize the change in trend and influence of music throughout the years.
- We are planning to exclude data from the time period 1920 - 1950 because messiness.
- We hypothesize that this is due to the disturbance of the two World Wars.

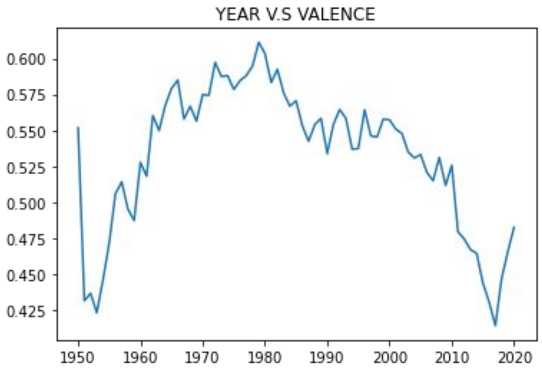
Some Interesting Points

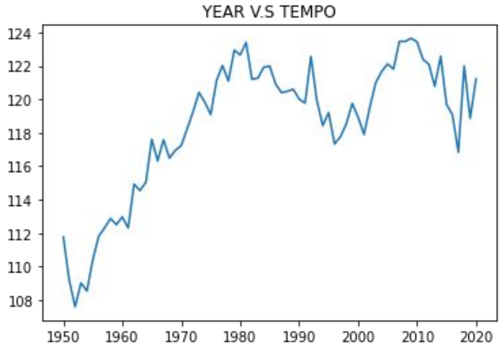
1. The recent popular music artworks appeared on the market seem to be labeled with “negative”, “sad”, “dark” labels more in comparison to the old musical works that usually would be considered as more positive. Is there a trend that can be reflected on the data that people are more inclined to listen more melancholy music?
2. There is a characteristic in the data called “danceability”. We expect it to keep growing as social media like tik tok or Instagram have been viral.

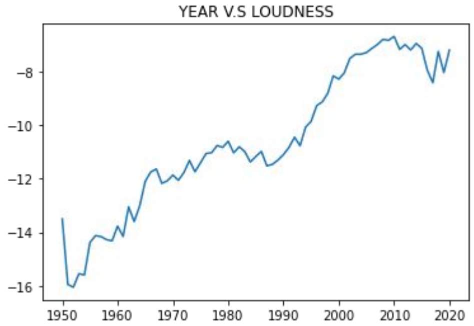
Graphs

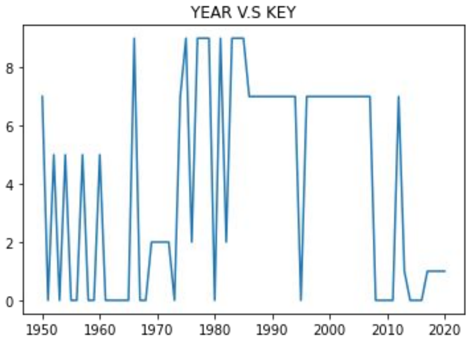

YEAR V.S DANCEABILITY
0.65
0.60
0.55
0.50
0.45
1950 1960 1970 1980 1990 2000 2010 2020


YEAR V.S ENERGY
0.7
0.6
0.5
0.4
0.3
1950 1960 1970 1980 1990 2000 2010 2020

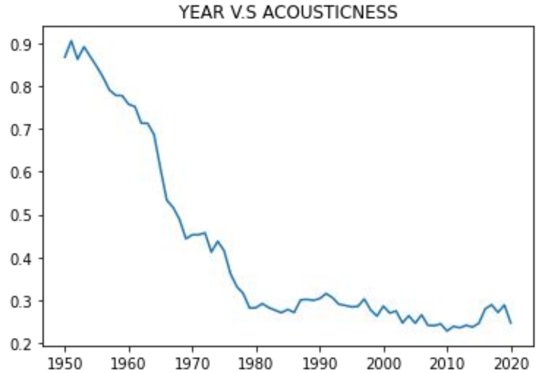

YEAR V.S VALENCE
0.600
0.575
0.550
0.525
0.500
0.475
0.450
0.425
1950 1960 1970 1980 1990 2000 2010 2020

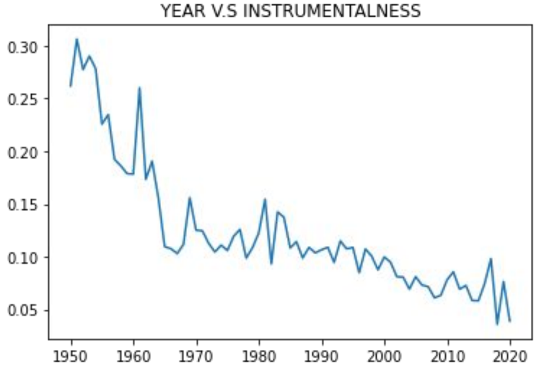

YEAR V.S TEMPO
124
122
120
118
116
114
112
110
108
1950 1960 1970 1980 1990 2000 2010 2020

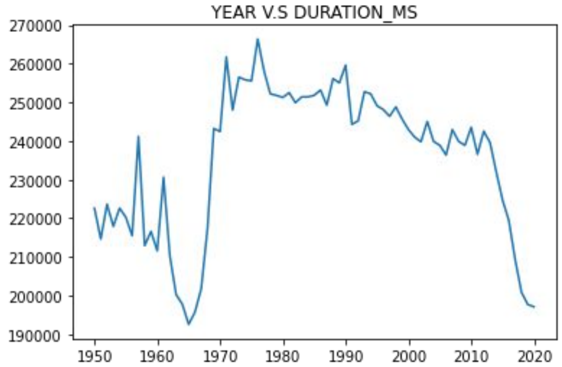

YEAR V.S LOUDNESS
−8
−10
−12
−14
−16
1950 1960 1970 1980 1990 2000 2010 2020

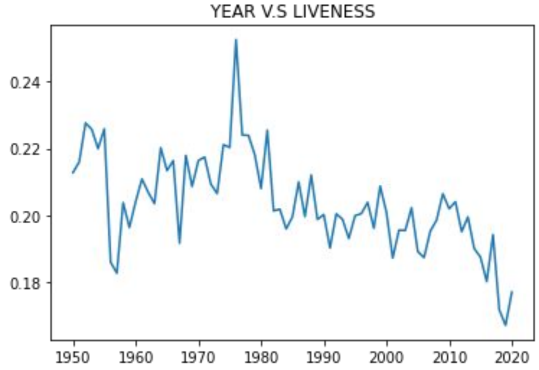

YEAR V.S KEY
8
6
4
2
0
1950 1960 1970 1980 1990 2000 2010 2020

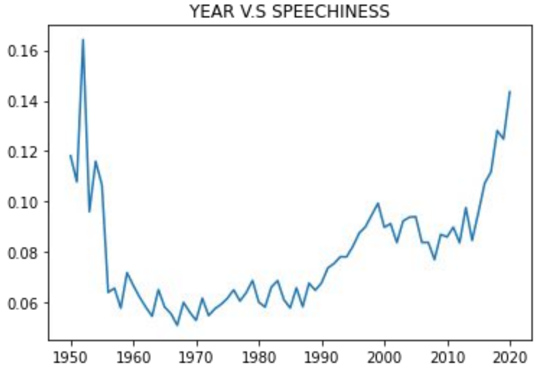
Graphs

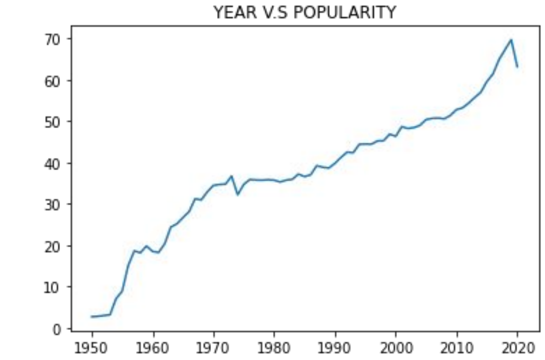

YEAR V.S ACOUSTICNESS
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
1950 1960 1970 1980 1990 2000 2010 2020


YEAR V.S INSTRUMENTALNESS
0.30
0.25
0.20
0.15
0.10
0.05
1950 1960 1970 1980 1990 2000 2010 2020

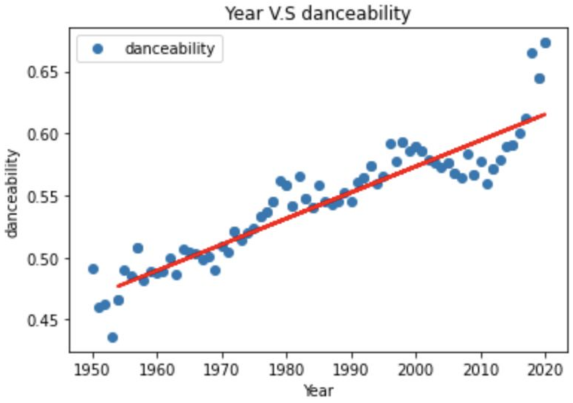

YEAR V.S DURATION_MS
270000
260000
250000
240000
230000
220000
210000
200000
190000
1950 1960 1970 1980 1990 2000 2010 2020


YEAR V.S LIVENESS
0.24
0.22
0.20
0.18
1950 1960 1970 1980 1990 2000 2010 2020

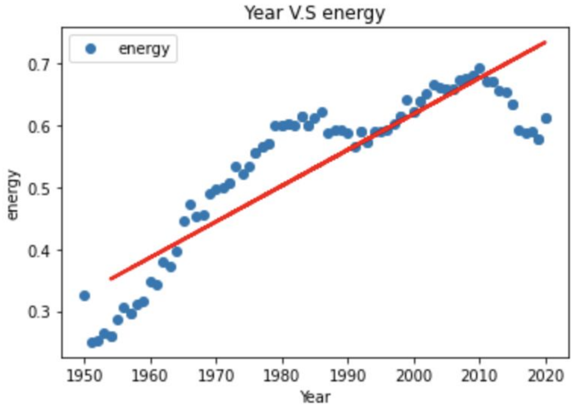

YEAR V.S SPEECHINESS
0.16
0.14
0.12
0.10
0.08
0.06
1950 1960 1970 1980 1990 2000 2010 2020


YEAR V.S POPULARITY
70
60
50
40
30
20
10
0
1950 1960 1970 1980 1990 2000 2010 2020

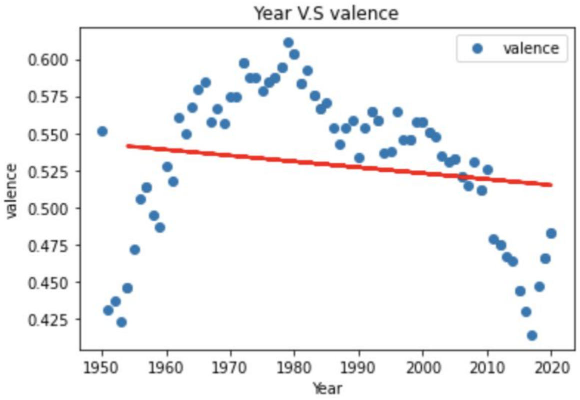
Graphs



0.7549867651410209

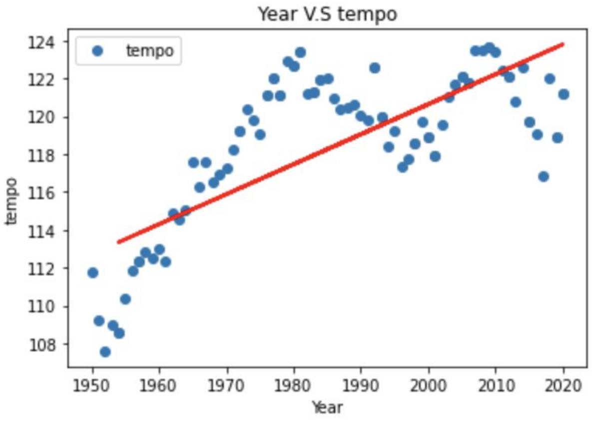


0.5201384978977612

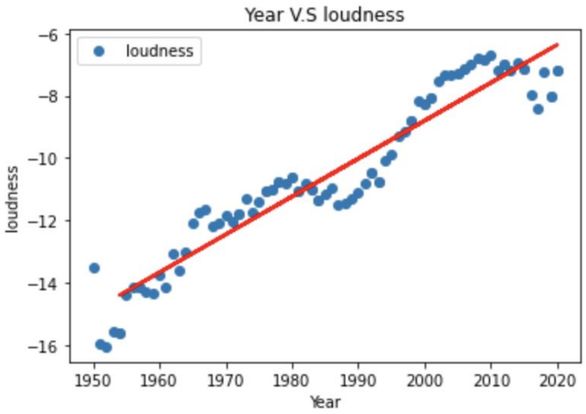


0.03494069110710540 4

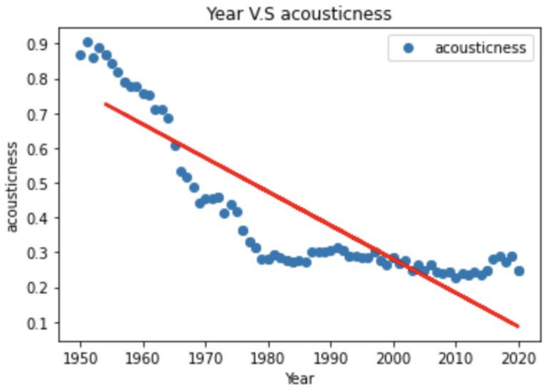
Graphs



0.05735336952947689 4

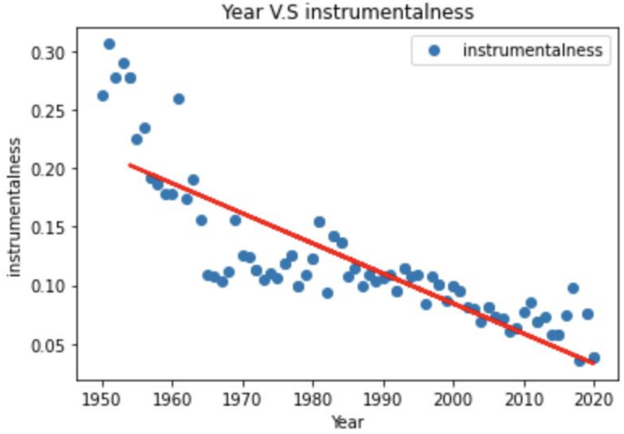


0.9025124054971821

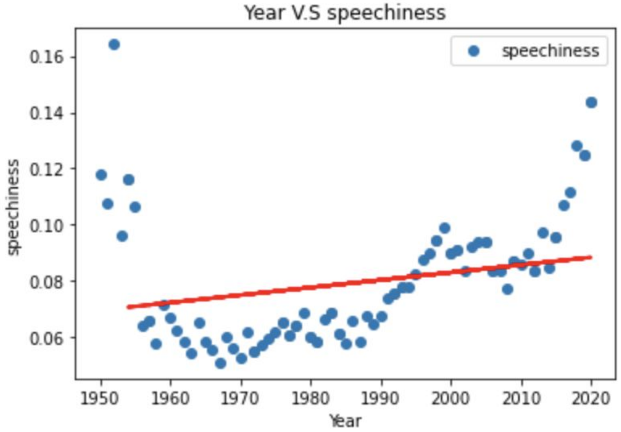


0.42162509549220484

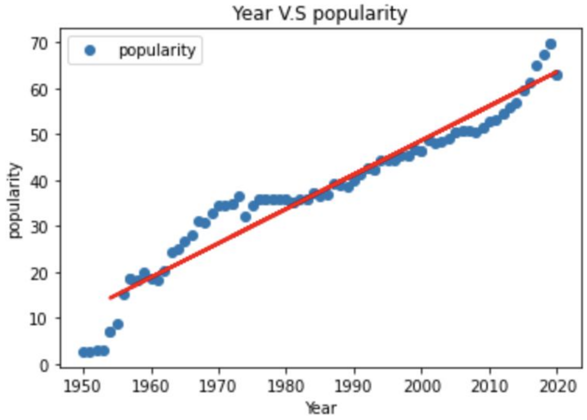
Graphs



0.7387038991509387

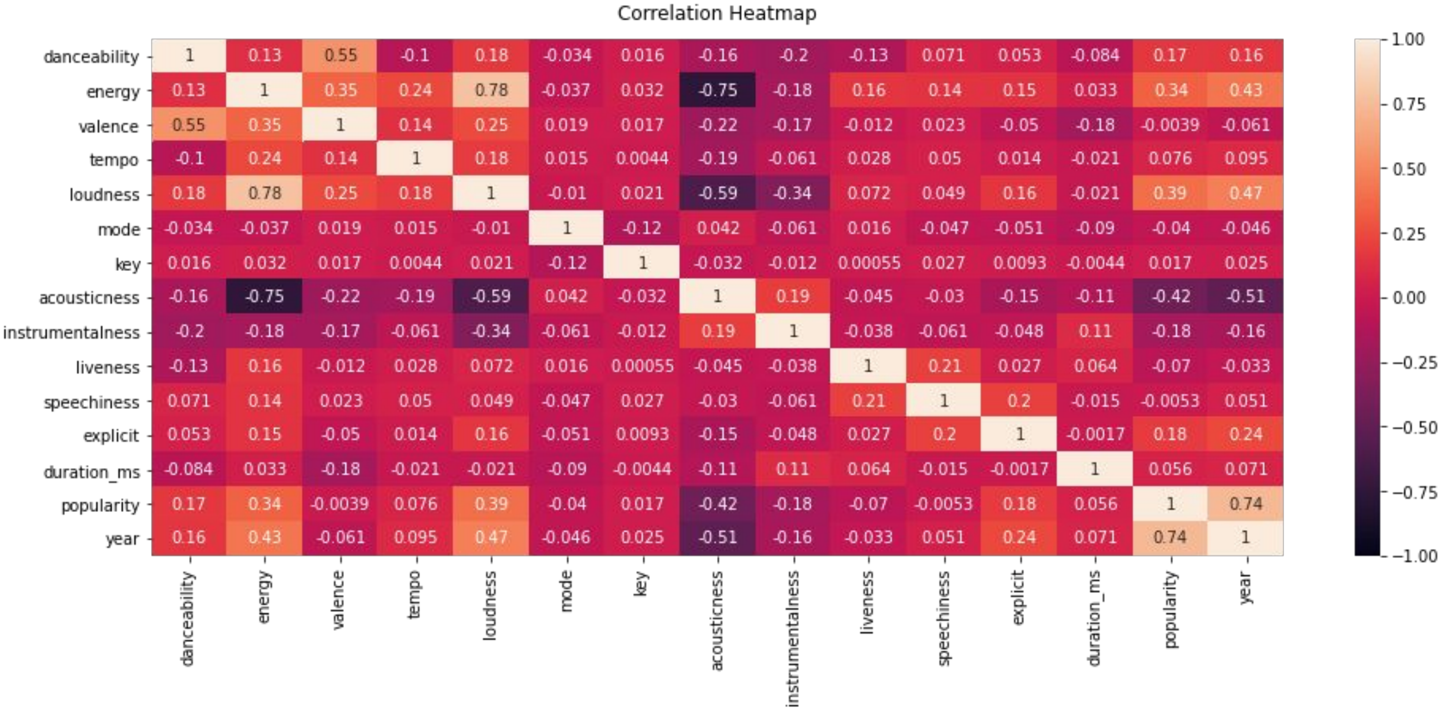


0.19407180529255663



0.9353958844494671

Correlation graph



Hypothesis

- Note: We assume all the data in the dataset is correct.
- Hypothesis: We predict that loudness and energy impacts popularity positively while accousticness and instrumental impacts popularity negatively.

OLS regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             popularity   R-squared:                       0.562
Model:                            OLS   Adj. R-squared:                  0.562
Method:                 Least Squares   F-statistic:                 1.009e+04
Date:                Thu, 05 Aug 2021   Prob (F-statistic):               0.00
Time:                        19:22:16   Log-Likelihood:            -3.5841e+05
No. Observations:               94364   AIC:                         7.168e+05
Df Residuals:                   94351   BIC:                         7.170e+05
Df Model:                          12
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const            -1203.4698      4.978   -241.749      0.000   -1213.227   -1193.713
danceability         3.4256      0.288     11.877      0.000       2.860       3.991
energy              -2.3609      0.300     -7.861      0.000      -2.950      -1.772
key                 -0.0076      0.010     -0.753      0.451      -0.027       0.012
valence              0.1639      0.196      0.837      0.403      -0.220       0.548
tempo                0.0014      0.001      1.106      0.269      -0.001       0.004
loudness             0.1592      0.013     12.543      0.000       0.134       0.184
acousticness        -2.6475      0.166    -15.946      0.000      -2.973      -2.322
instrumentalness    -2.9580      0.148    -20.027      0.000      -3.248      -2.669
liveness            -3.3042      0.199    -16.580      0.000      -3.695      -2.914
speechiness         -7.8389      0.487    -16.096      0.000      -8.793      -6.884
duration_ms       2.021e-06   3.39e-07      5.968      0.000    1.36e-06    2.68e-06
year                 0.6272      0.002    251.911      0.000       0.622       0.632
==============================================================================
Omnibus:                    11872.734   Durbin-Watson:                   0.577
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            22632.034
Skew:                           0.813   Prob(JB):                         0.00
Kurtosis:                       4.763   Cond. No.                     3.72e+07
==============================================================================
```

Musician Social Network

- How do we develop a model to measure influence?
- Visualizing and Exploring a Subnetwork

We do the social network with a simple dictionary, where each key represents an influencer, and each value represents that influencer’s followers.

Using that, we can calculate influence by implementing a breadth-first search algorithm (BFS) .

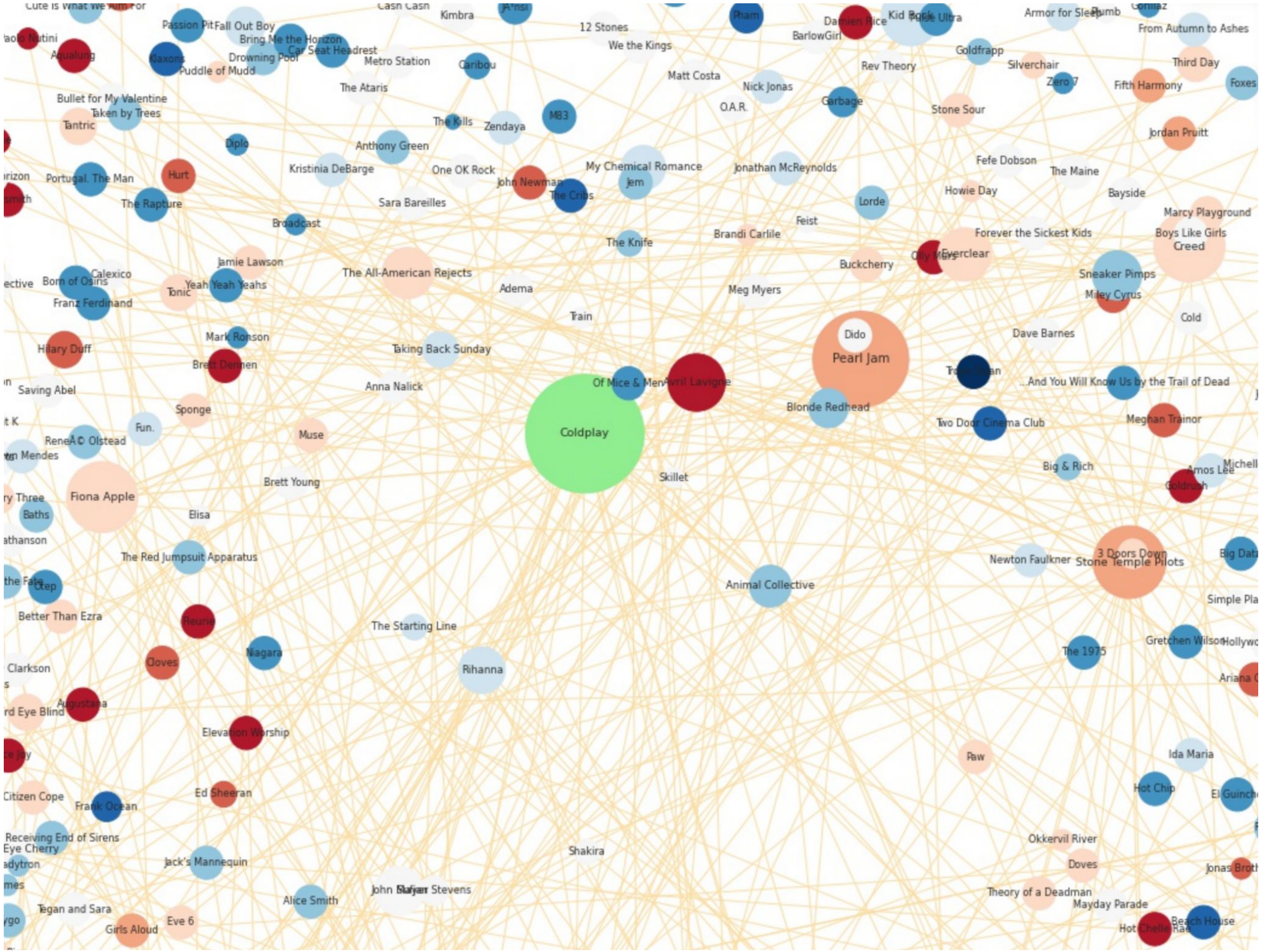
We Implement breadth-first search algorithm to extract the number of followers for each influencer.

Social Network

	Influencer	Depth 0	Depth 1	Depth 2	Depth 3	Depth 4	Depth 5	Depth 6	Depth 7	Depth 8	Depth 9	Depth 10	Total
0	The Beatles	1	615	2802	3744	4062	4258	4438	4531	4567	4599	4623	2002.53125
1	Bob Dylan	1	389	2325	3818	4291	4474	4577	4613	4630	4631	4631	1803.73046
2	Chuck Berry	1	160	1985	3910	4405	4557	4634	4660	4668	4669	4669	1627.00390
3	James Brown	1	155	1808	3937	4402	4535	4593	4622	4630	4631	4631	1581.519531
4	Elvis Presley	1	167	1975	3632	4162	4415	4533	4598	4619	4630	4631	1571.22460

Visualize Social Network

Although we can visualize a small portion of the network, it is hard to visualize the entire network in a meaningful way due to the sheer number of connections present. We do a a medium-sized subnetwork to Depth 10.



Coldplay
Pearl Jam
Dido
Avril Lavigne
Of Mice & Men
Fiona Apple
Stone Temple Pilots
3 Doors Down
Animal Collective
Blonde Redhead
Skillet
Rihanna
Shakira
Fall Out Boy
Bring Me the Horizon
Car Seat Headrest
Drowning Pool
Puddle of Mudd
Metro Station
Caribou
The Ataris
Kimbra
12 Stones
We the Kings
Matt Costa
Nick Jonas
O.A.R.
Garbage
BarlowGirl
Rev Theory
Stone Sour
Goldfrapp
Silverchair
Zero 7
Fifth Harmony
Third Day
From Autumn to Ashes
Armor for Sleep
Foxes
Jordan Pruitt
Passion Pit
Aqualung
Bullet for My Valentine
Taken by Trees
Tantric
Diplo
Anthony Green
The Kills
Zendaya
M83
Kristinia DeBarge
One OK Rock
John Newman
The Cribs
Jem
My Chemical Romance
Jonathan McReynolds
Fefe Dobson
The Maine
Bayside
Marcy Playground
Boys Like Girls
Creed
Portugal. The Man
Hurt
The Rapture
Sara Bareilles
Lorde
Howie Day
Broadcast
The Knife
Brandi Carlile
Feist
Forever the Sickest Kids
Everclear
Buckcherry
Sneaker Pimps
Miley Cyrus
Jamie Lawson
The All-American Rejects
Adema
Meg Myers
Calexico
Born of Osiris
Tonic
Yeah Yeah Yeahs
Franz Ferdinand
Train
Cold
Dave Barnes
Mark Ronson
Taking Back Sunday
Hilary Duff
Brett Dennen
...And You Will Know Us by the Trail of Dead
Saving Abel
Anna Nalick
Sponge
Two Door Cinema Club
Meghan Trainor
Fun.
Muse
Big & Rich
Amos Lee
Brett Young
Baths
Elisa
The Red Jumpsuit Apparatus
Newton Faulkner
Better Than Ezra
The Starting Line
Simple Plan
Gretchen Wilson
The 1975
Clarkson
Cloves
Niagara
Augustana
Elevation Worship
Paw
Ida Maria
Hot Chip
Citizen Cope
Frank Ocean
Ed Sheeran
Receiving End of Sirens
Eye Cherry
Jack's Mannequin
Okkervil River
Doves
Theory of a Deadman
Mayday Parade
Alice Smith
Tegan and Sara
Girls Aloud
Eve 6
Beach House

Other possible datasets

	artist_name	artist_id	danceability	energy	valence	tempo	loudness	mode	key	acousticness	instrumentalness	liveness	speechiness	duration_ms	popularity	count
0	Frank Sinatra	792507	0.384478	0.238017	0.364288	110.181698	-14.271141	1	5	0.735648	0.020855	0.232106	0.049614	189179.9255	26.004383	1369
1	Vladimir Horowitz	119107	0.343210	0.118844	0.225951	94.900679	-23.193418	1	1	0.990070	0.879508	0.183812	0.043360	266541.1251	3.592378	1207
2	Johnny Cash	816890	0.619803	0.449381	0.680662	115.037747	-11.593104	1	10	0.685637	0.022647	0.242243	0.098216	162279.2672	26.614130	1104
3	Billie Holiday	79016	0.572637	0.201368	0.498934	109.912172	-13.225966	1	5	0.908499	0.013064	0.217727	0.062432	185131.4530	15.621005	1095
4	Bob Dylan	66915	0.512598	0.477932	0.551934	126.160149	-11.184330	1	7	0.562567	0.034211	0.308978	0.064535	256713.4203	30.860806	1092
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5849	Natalie La Rose	3359519	0.830000	0.520000	0.735000	104.990000	-8.714000	1	0	0.000792	0.000013	0.065600	0.037600	189907.0000	64.000000	1
5850	Sarah Ross	3381566	0.721000	0.944000	0.626000	85.002000	-5.982000	1	8	0.013000	0.000000	0.320000	0.159000	262760.0000	52.000000	1
5851	Rotimi	3410250	0.637000	0.501000	0.431000	103.993000	-6.148000	0	0	0.229000	0.000059	0.099000	0.187000	185461.0000	71.000000	1
5852	Jillian Jacqueline	3455945	0.547000	0.672000	0.283000	155.791000	-5.023000	1	11	0.304000	0.000000	0.099600	0.049600	213133.0000	58.000000	1
5853	Jaira Burns	3639618	0.566000	0.769000	0.385000	170.036000	-4.342000	1	7	0.018300	0.000000	0.108000	0.087200	191100.0000	74.000000	1

5854 rows × 16 columns

Data by artist

Other possible datasets

	influencer_id	influencer_name	influencer_main_genre	influencer_active_start	follower_id	follower_name	follower_main_genre	follower_active_start
0	759491	The Exploited	Pop/Rock	1980	74	Special Duties	Pop/Rock	1980
1	25462	Tricky	Electronic	1990	335	PJ Harvey	Pop/Rock	1990
2	66915	Bob Dylan	Pop/Rock	1960	335	PJ Harvey	Pop/Rock	1990
3	71209	Leonard Cohen	Pop/Rock	1950	335	PJ Harvey	Pop/Rock	1990
4	91438	The Gun Club	Pop/Rock	1980	335	PJ Harvey	Pop/Rock	1990
...	...	...	...	...	...	...	...	...
42765	580300	Sufjan Stevens	Pop/Rock	1990	3661738	Rosemary & Garlic	Pop/Rock	2010
42766	261309	Vybz Kartel	Reggae	2000	3670556	Trinidad Cardona	R&B;	2010
42767	467203	Michael Jackson	R&B;	1960	3670556	Trinidad Cardona	R&B;	2010
42768	2518003	Popcaan	Reggae	2000	3670556	Trinidad Cardona	R&B;	2010
42769	2896351	Tommy Lee	Reggae	2000	3670556	Trinidad Cardona	R&B;	2010

42770 rows × 8 columns

Influencer data

What’s next

- Use neural networks to run multiple regressions with popularity being the dependent variable and year and other variables being the independent variables.
- Combine social media data of these musicians with their own characteristics and analyze their influence.
- Using the data by artist and influencer data to build a model to analyze the following trend of different musical genres.
- Build a model categorizing the danceability, valence, tempo, and other technical aspects of music into different musical genres.