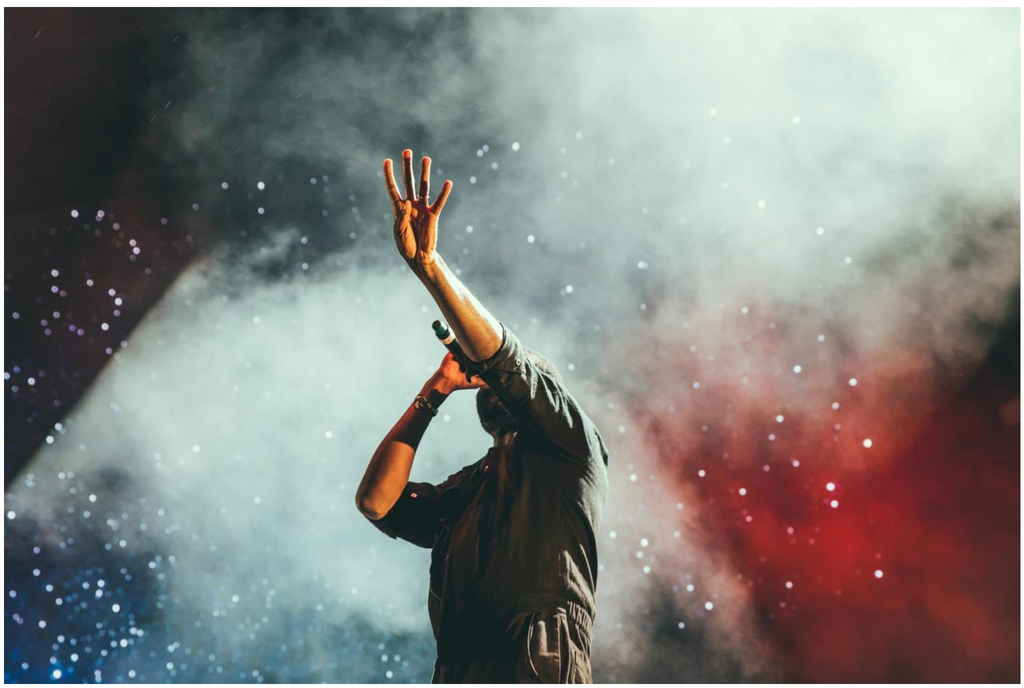
SPOTIFY RECOMMENDER SYSTEM



1/27/2021

Using continuous DATA to recommend products

Here I attempt to create a recommendation system utilizing objective measures that define a product, rather than via collaborative filtering using user ratings.

Spotify Recommender System

USING CONTINUOUS DATA TO RECOMMEND PRODUCTS

WHY?

If you are familiar with recommendation systems, then you are aware that most companies that implement them utilize some method of collaborative filtering, namely, user-based, or item-based. This is ultimately the ideal way to do this, as it makes it so that products do not have to have some abstract measure of being to evaluate where they stand in relation to others. However, I wanted to see if utilizing objective numerical measures assigned to different ‘products’ could produce a recommendation system in some what that mirrored the effectiveness of collaborative filtering systems, which rely on users’ ratings of products to make recommendations.

THE DATA

My interest was piqued when I came across a dataset containing 170,000 songs from Spotify, ranging from 1920 to 2020. What interested me is that the songs all had associated numerical values assigned to them that described their being. An example of some of these measures included energy, loudness, acoustic-ness, dance-ability, and other similar continuous measures mostly lying between zero and one. The reason this is interesting is because it provides a unique opportunity to attempt to find similarities in numerical measures, that is, relationships between the products themselves rather than using user ratings to find them.

INITIAL FINDINGS | EDA | CORRELATIONS IN THE DATA

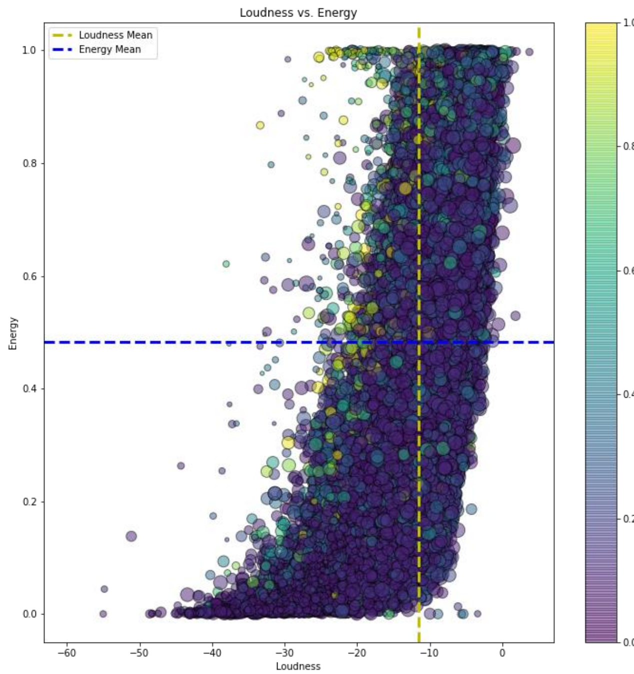
The data was simple upon import. There were no missing values, and nothing to take care of regarding incorrect or clearly misplaced values. Some valuable findings may have arisen from some EDA and visualization.

A PEEK INTO THE DATA:

	artists	name	year	acousticness	danceability	energy	explicit	instrumentalness	key	liveness	loudness	mode	popularity
0	['Sergei Rachmaninoff', 'James Levine', 'Berli...	Piano Concerto No. 3 in D Minor, Op. 30: III. ...	1921	0.982	0.279	0.211	0	0.878000	10	0.665	-20.096	1	4
1	['Dennis Day']	Clancy Lowered the Boom	1921	0.732	0.819	0.341	0	0.000000	7	0.160	-12.441	1	5
2	['KHP Kridhamardawa Karaton Ngayogyakarta Hadi...	Gati Bali	1921	0.961	0.328	0.166	0	0.913000	3	0.101	-14.850	1	5
3	['Frank Parker']	Danny Boy	1921	0.967	0.275	0.309	0	0.000028	5	0.381	-9.316	1	3
4	['Phil Regan']	When Irish Eyes Are Smiling	1921	0.957	0.418	0.193	0	0.000002	3	0.229	-10.096	1	2

The inspiration for this project came from the data layout. It interests me to think that you might be able to objectively describe songs with numbers, as opposed to user ratings, and as such group them as a result.

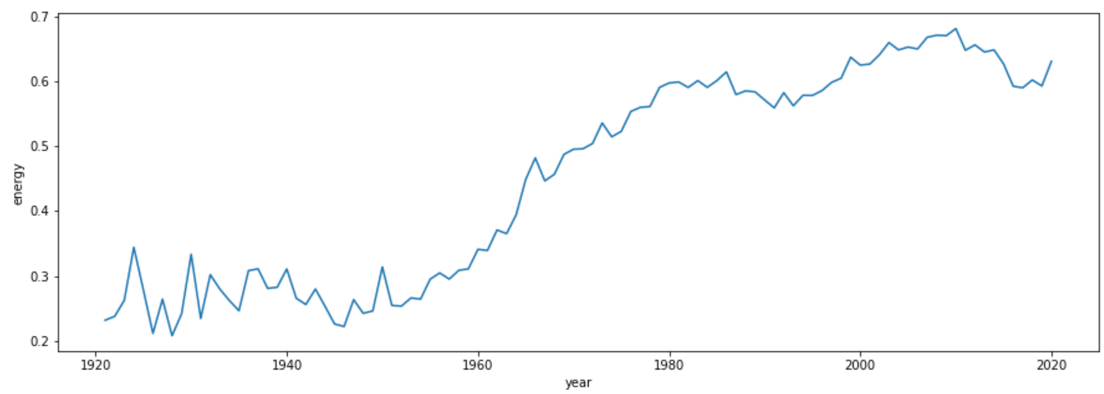
LOUDNESS VS. ENERGY



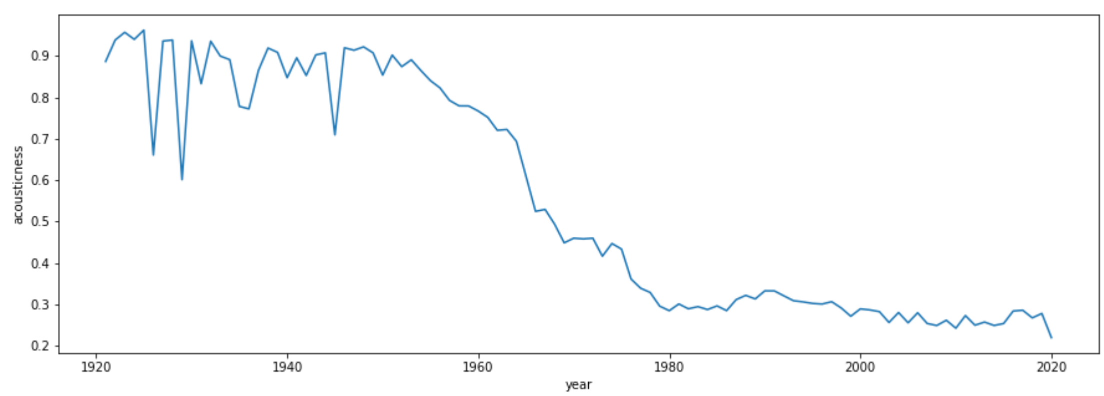
An interesting correlation found between loudness and energy. Louder songs have more energy than music that is quiet within the dataset. This becomes important later in discussing how music has gotten louder over the years.

MUSIC OVER THE YEARS

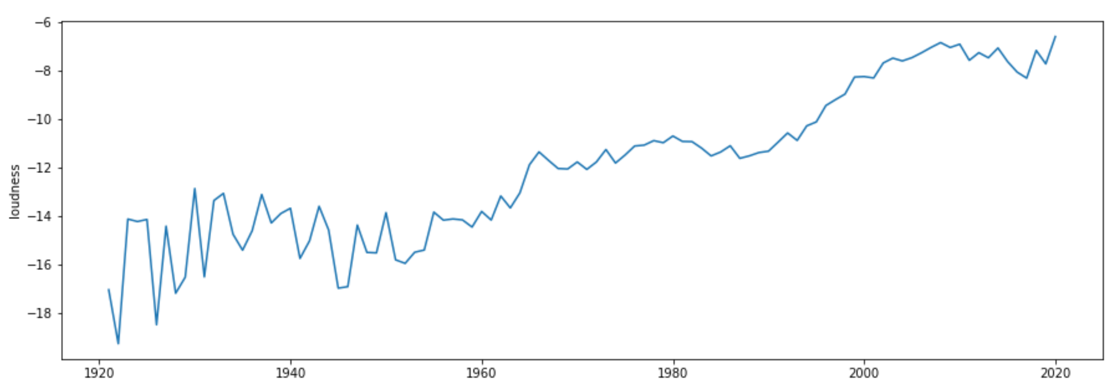
A few interesting changes in music over the years was found through simple time series analysis. Graphs of these findings follow:



Since the 1920's, music has become more 'energetic' in recent years.



'Acoustic-ness' likely references the nature of the music, that is, does it contain instrumentation that utilizes more acoustic elements rather than electronic ones. As you can see, acoustic-ness has decreased heavily on average since 1920.



Finally, music has gotten louder over the years.

THE LOUDNESS WAR

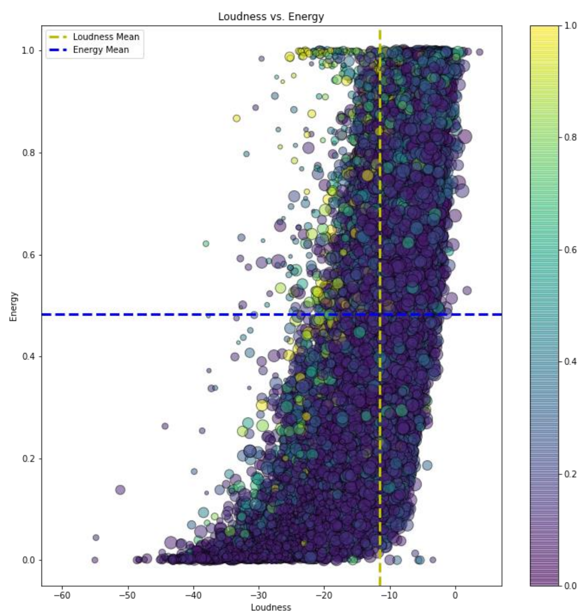
There is a correlation between two measures that weave a significant story.

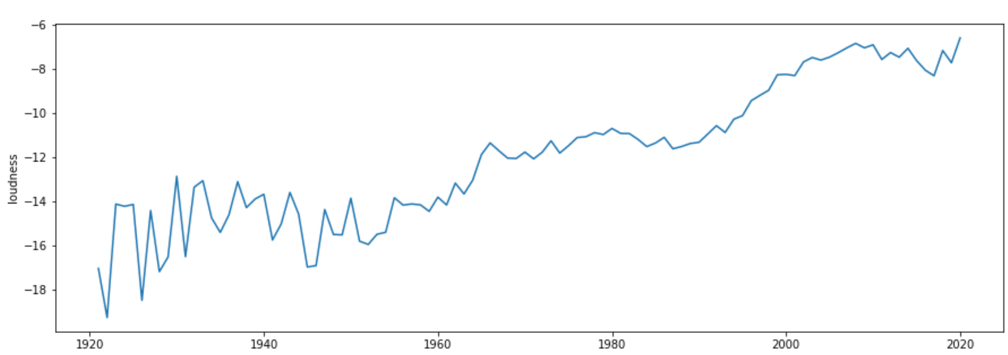
1. Energy
2. Loudness

This is rather unsurprising when you investigate the history of music. Since the 1950's especially, audio engineers began to intuitively take note of this correlation – Louder music has more energy. Since this correlation was discovered in recorded music, the goal of audio engineers has been to bring music to a threshold below 0 decibels that is as "loud" as possible. Zero decibels represent a threshold in recorded music, as anything louder than that will 'peak' and produce unpleasant, scratchy sounds.

Because of this correlation, audio engineers have pushed music as close as possible to this threshold, sometimes beyond the capability of hardware and software at the time. This reduces the dynamic range in music, but also increases the perceived energy as well.

RECAP: THE LOUDNESS WAR IN TWO GRAPHS





MAKING RECOMMENDATIONS

Cosine Similarities by Hand:

1. Dimensionality reduction via PCA
    a. Reduced dimensions to a desired number of components
2. Normalize Data
    a. Normalize the data as part of the cosine similarity computation.
3. Dot Product of a chosen Artist
    a. Take the dot product of the original dataset in relation to a chosen artist catalogue to complete the cosine similarity computation.
    b. This returns a dataset of cosine similarities, for which we then sort from largest to smallest.

RESULTS

Noticing a Blend in the Data:

1. There is something about rap music that makes it easily identifiable.
2. The lines start to get blurry when the genre starts to lean more towards softer genres.
    a. Many mainstream artists have a mix of songs that are both low and high energy, ranging in acoustic-ness, danceability, and so on. It is not hard to picture one of a few Taylor Swift songs, for example – She has songs that range from country with high levels of 'acoustic-ness' to more pop oriented music, and maybe a few slow, sad songs in there as well. All of these could have a range of values that blend in certain areas yet act as opposites in others.

ARTIST RECOMMENDATION ONLY – THE GOOD RESULTS

The following are examples of artists return that make a lot of sense. A quick YouTube search corroborates these results. Bear in mind these results are based on the song chosen as well:

```
1  recommend_artist('Bring Me The Horizon', 2)
```

```
Bring Me The Horizon        1.000000
A Day To Remember           0.987807
Trivium                     0.983088
Bullet For My Valentine     0.982074
Bad Wolves                  0.979866
Andy Mineo                  0.977085
Five Finger Death Punch     0.975015
Alison Wonderland           0.973705
Of Mice & Men               0.972341
Evans Blue                  0.972041
Foo Fighters                0.971632
Attack Attack!              0.970140
Linkin Park                 0.969872
Killswitch Engage           0.969451
Atreyu                      0.969222
dtype: float64


Song Referenced:
Sleepwalking
```

```
1  recommend_artist('Gunna', 1)
```

```
Gunna                          1.000000
Sheff G                        0.980424
Lil Baby, Moneybagg Yo         0.979180
Lil Skies                      0.975615
Yo Gotti                       0.974938
Duckwrth                       0.974671
Bhad Bhabie, Kodak Black       0.973527
Young Dolph                    0.972903
Flo Milli                      0.970787
Jack Harlow, jetsonmade        0.970739
Josh A, Iamjakehill            0.970136
Bad Bunny                      0.969864
Lizzo, Missy Elliott           0.969853
Five Finger Death Punch        0.969832
Ty Dolla $ign, E-40            0.969667
dtype: float64


Song Referenced:
Who You Foolin
```

ARTIST RECOMMENDATION – THE BAD RESULTS

The following are examples that do not make any sense at all. This is where I discovered that there are underlying connections in the data in which certain features blend into each other and thus make sense on an objective level of measurement, but do not make sense from a practicality standpoint.

```
1 recommend_artist('Bring Me The Horizon', 1)
```

```
Bring Me The Horizon                                                          1.000000
Calvin Harris, John Newman                                                    0.987953
Alison Wonderland                                                             0.981160
Chevelle                                                                      0.976773
One Direction                                                                 0.971795
Rihanna                                                                       0.971090
Jory Boy, Bad Bunny                                                           0.971080
Stereophonics                                                                 0.970913
Rihanna                                                                       0.970903
K/DA, Madison Beer, (G)I-DLE, Lexie Liu, Jaira Burns, Seraphine, League of Legends    0.970775
EXO                                                                           0.970318
Bad Wolves                                                                    0.969700
Jason Aldean                                                                  0.969397
Justin Bieber, Big Sean                                                       0.968851
Bring Me The Horizon                                                          0.968655
dtype: float64


Song Referenced:
Throne
```

```
1 recommend_artist('Mumford & Sons', 2)
```

```
Mumford & Sons        1.000000
Pierce The Veil       0.981555
Moderatto             0.975989
ZAYN, Zhavia Ward     0.973609
Hillsong Worship      0.970897
Julie Roberts         0.969972
LeAnn Rimes           0.969792
Lorde                 0.968064
Vince Staples         0.967429
Lorde                 0.967256
Taylor Swift          0.965110
ZAYN, Taylor Swift    0.965075
Taylor Swift          0.962409
Faith Hill            0.961677
Pentatonix            0.961104
dtype: float64


Song Referenced:
Awake My Soul
```

SONG RECOMMENDATION – THE GOOD RESULTS

The same process, but filtered by song:

```
Song Chosen: Picture To Burn
By: Taylor Swift


Recommended:


16846                                        Our Song
16874     Teardrops On My Guitar - Radio Single Remix
16885                                  Picture To Burn
16904                                       Tim McGraw
16932                                  Should've Said No
                               ...
138465                                    Silent Night
138467                                 White Christmas
140553                         It's Nice To Have A Friend
153303                                You Belong With Me
168573                      Drops Of Jupiter - Live/2011
Name: name, Length: 207, dtype: object


By: Taylor Swift


16582                               Take Your Mama
16882                               I Can't Decide
54484     I Don't Feel Like Dancin' - Radio Edit
72292                  I Don't Feel Like Dancin'
139341                            Let's Have A Kiki
152518                              Filthy/Gorgeous
Name: name, dtype: object


By: Scissor Sisters


18269           How Country Feels
18299     Runnin' Outta Moonlight
36194                    Boots On
73128             Whistlin' Dixie
73720              Like a Cowboy
90821              Goodnight Kiss
91455                     We Went
168204           My Kinda Country
Name: name, dtype: object


By: Randy Houser
```

```
Song Chosen: Who You Foolin
By: Gunna


Recommended:


57143                    Isis
75110    Isis (feat. Logic)
Name: name, dtype: object


By: Joyner Lucas, Logic


19424                            BOP
19459                           Suge
19496                          VIBEZ
38375                      Goin Baby
38447               Brother's Keeper
38460                  Gucci Peacoat
38494                        Shanyah
38499     More Money More Problems
74693                             21
74850                          INTRO
108299                     Next Song
108324           Walker Texas Ranger
124647                   OFF THE RIP
155362                           XXL
170105                Today - Intro
170425                  Taking It Out
Name: name, dtype: object


By: DaBaby


74532                    All Of A Sudden
170461    No Sucker (feat. Moneybagg Yo)
Name: name, dtype: object


By: Lil Baby, Moneybagg Yo
```

SONG RECOMMENDATION – THE BAD RESULTS

This also yielded some poor recommendations due to the aforementioned blending of measures

```
Song Chosen: Throne
By: Bring Me The Horizon


Recommended:


18570    Blame (feat. John Newman)
Name: name, dtype: object


By: Calvin Harris, John Newman


17422                                    Little Lion Man
17435                                           The Cave
17570                                      Awake My Soul
18042                                        I Will Wait
36366                                   White Blank Page
36440                                       Winter Winds
```

The first recommendations for this metal band are Calvin Harris, an EDM artist, and Mumford & Sons, a folk band. Very impractical and unlikely.

```
Song Chosen: El Farsante
By: Ozuna


Recommended:


74664     El Farsante - Remix
108316                  Ibiza
Name: name, dtype: object


By: Ozuna, Romeo Santos


18497    Break Free
Name: name, dtype: object


By: Ariana Grande, Zedd


92235    Tinted Eyes (feat. blackbear & 24kGoldn)
Name: name, dtype: object


By: DVBBS, 24kGoldn, blackbear
```

The model has no way of filtering by language, as no language was included. This would be an ideal way to improve results, as Ozuna is a Spanish artist, and we receive English artists in return.

FUTURE IMPROVEMENTS

1. Overall, the model performs great on some music, but poorly on others.
2. What this project revealed overall is that collaborative filtering techniques are easily the most valuable and reliable way to make reliable recommendations. It is likely that it would be impossible to maintain these kinds of measures across a wide variety of products, and subsequently maintain them accurately without subjective bias.
3. Further exploration of the data may lead to discovering trends in the dataset that would assist in figuring out how to make the recommendations more realistic and applicable.