In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import spotipy
import playlist
import config
import time
import create_dataframes
from sklearn.metrics import pairwise
from sklearn.preprocessing import MinMaxScaler, StandardScaler
from spotipy.oauth2 import SpotifyOAuth
```

In [2]:
```
#create spotipy object to interact with spotify web API
from spotipy.oauth2 import SpotifyClientCredentials
sp = spotipy.Spotify(
    auth_manager=SpotifyClientCredentials
        (client_id=config.SPOTIFY_CLIENT_ID,
        client_secret=config.SPOTIFY_SECRET
        ))
```

avg_audio_values = pd.DataFrame(playlist1.normalized_numeric_features.loc[:,0:].mean(axis=0)) was changed from using just the normalized audio features to using all the normalized numeric features

In [3]:
```
#playlist_link, final_playlist_len = create_dataframes.get_user_playlist()
playlist_link, final_playlist_len = 'https://open.spotify.com/playlist/37i9dQZE
playlist1 = playlist.Playlist(playlist_link, sp)
artist_counts, artist_id_counts = playlist1.get_artist_counts()
avg_audio_values = pd.DataFrame(playlist1.normalized_numeric_features.loc[:,0:]
artists_search = create_dataframes.artist_search_results(artist_counts.columns)
recs = playlist1.get_recommendations(artists_search, avg_audio_values, k=5)
#create_dataframes.create_new_playlist(recs, final_playlist_len)
```

In [4]:
```
avg_audio_values
```

Out[4]:

	0
0	0.601267
1	0.463976
2	0.457576
3	0.673927
4	0.566667
5	0.450619
6	0.257194
7	0.070887
8	0.274420
9	0.473408
10	0.484486
11	0.365575
12	0.412057

- standardize ratings + audio features
- compute PCA on the dataframe with the standardized numeric features
- check how many components explain 80% ish of the variance, & keep just those (set playlist1.normalized_audio_features = PCA reduced data)
- compute the average value of the selected components
- perform artists search
- NORMALIZE THE AUDIO FEATURES FROM THE SONGS IN THE SEARCH RESULT**** (add a PCA flag to playlist.get_recommendations which will perform PCA transformation on the data it gets back after line 192 in playlist.py)
- find KNN with PCA performed data

In [5]:
```
print(playlist1.audio_features.shape)
playlist1.audio_features
```

```
(30, 13)
```

Out[5]:

	track_id	danceability	energy	key	loudness	mode	speechiness	ac
0	2XHzzp1j4IfTNp1FTn7YFg	0.669	0.634	11	-6.476	1	0.0327	
1	01A7PEPSnmtixFPfB2UTal	0.793	0.631	11	-6.109	0	0.0998	
2	6AaOtHsKd195ec0Y4kC9ER	0.721	0.607	0	-7.078	0	0.2770	
3	63M8PK8yavNITSViKUB62p	0.752	0.766	1	-5.110	1	0.2510	
4	17Lmf8pUgrRrEKIKkuZirE	0.799	0.499	11	-7.785	1	0.1820	
5	2LvRR121MWFmmEGkuV2vQP	0.919	0.506	1	-7.140	1	0.3040	
6	1s9hBKw37JvYV9KYiR9oef	0.744	0.539	10	-6.757	1	0.0596	
7	1UYLxZIL4IicS6TaPUZFTH	0.893	0.635	2	-9.000	1	0.0514	
8	5BviN9XWqCYE88AOJg6nEK	0.801	0.521	4	-9.871	0	0.2800	
9	4fUh9VvKW4shhWN5LfN1Ba	0.762	0.611	9	-5.462	0	0.0876	
10	3m8CQnnfJJp4eQMWWI3zay	0.451	0.632	8	-7.980	0	0.2970	
11	5tFep7dXGd7vEJ668wTPux	0.850	0.405	7	-7.509	0	0.2250	
12	2JfOSsfYs8QNneR3oTpNv4	0.892	0.891	5	-4.428	0	0.3050	
13	29UgEHyxg63NDzckA3hLk5	0.615	0.648	5	-3.792	0	0.2200	
14	5xwpXWWkfJRqg1S27oVxh4	0.800	0.478	8	-7.874	1	0.1810	
15	5W7xC99N2Zzfh69r7I7zWK	0.570	0.580	1	-5.547	1	0.2100	
16	7MDKvOzNgAJ3KMCtaP2UOa	0.633	0.853	8	-6.491	1	0.3890	
17	5YCs9rP6ZcMJW6me3QhtWr	0.774	0.663	1	-7.283	1	0.1200	
18	5MGeW5mHnqeJiM0NPBbkEZ	0.611	0.543	1	-5.288	0	0.0672	
19	0w3Q3VFdrYzo24QUIGnBNy	0.545	0.763	11	-5.528	0	0.4070	
20	3r9n4SFmaqViN6G4Z6TbV7	0.705	0.646	4	-5.873	0	0.0690	
21	0FwezIaHwoxFy8IOLbF2UQ	0.854	0.401	9	-8.553	1	0.3950	
22	07J4gvixM0ksqSpgDw15J1	0.880	0.346	1	-9.145	1	0.3390	
23	1OBWQFd7hcTo18GWxS1bnZ	0.836	0.633	0	-5.919	0	0.1540	
24	5uNhj4KHA7nZO1CMwbnnqG	0.725	0.545	6	-8.485	1	0.0842	
25	1lqKgoo11HKUoMpbouC7I4	0.881	0.416	1	-13.488	1	0.1580	
26	6yxgLlHHp9EMklVUUdyjKh	0.951	0.483	6	-5.592	1	0.3000	
27	20X0AnKMbmD4PG6D5W1G3G	0.769	0.666	1	-7.543	1	0.0645	
28	4IeUAegXVbETb7FpgNDnSM	0.643	0.592	1	-6.684	0	0.2770	
29	0jkeRqlWciqKxU3iHQfdWj	0.711	0.833	7	-4.818	1	0.1540	

In [6]:
```
print(playlist1.normalized_audio_features.shape) #make sure its the same as the
playlist1.normalized_audio_features
```

```
(30, 13)
```

Out[6]:

	track_id	0	1	2	3	4	5	
0	2XHzzp1j4IfTNp1FTn7YFg	0.436	0.528440	1.000000	0.723185	1.0	0.000000	0.0183
1	01A7PEPSnmtixFPfB2UTal	0.684	0.522936	1.000000	0.761035	0.0	0.179268	0.0849
2	6AaOtHsKd195ec0Y4kC9ER	0.540	0.478899	0.000000	0.661097	0.0	0.652685	0.2156
3	63M8PK8yavNITSViKUB62p	0.602	0.770642	0.090909	0.864068	1.0	0.583222	0.7892
4	17Lmf8pUgrRrEKIKkuZirE	0.696	0.280734	1.000000	0.588181	1.0	0.398878	0.0131
5	2LvRR121MWFmmEGkuV2vQP	0.936	0.293578	0.090909	0.654703	1.0	0.724820	0.3013
6	1s9hBKw37JvYV9KYiR9oef	0.586	0.354128	0.909091	0.694204	1.0	0.071867	0.3368
7	1UYLxZIL4IicS6TaPUZFTH	0.884	0.530275	0.181818	0.462871	1.0	0.049960	0.0363
8	5BviN9XWqCYE88AOJg6nEK	0.700	0.321101	0.363636	0.373040	0.0	0.660700	0.0209
9	4fUh9VvKW4shhWN5LfN1Ba	0.622	0.486239	0.818182	0.827764	0.0	0.146674	0.2776
10	3m8CQnnfJJp4eQMWWI3zay	0.000	0.524771	0.727273	0.568069	0.0	0.706118	0.0157
11	5tFep7dXGd7vEJ668wTPux	0.798	0.108257	0.636364	0.616646	0.0	0.513759	0.2705
12	2JfOSsfYs8QNneR3oTpNv4	0.882	1.000000	0.454545	0.934406	0.0	0.727491	0.1161
13	29UgEHyxg63NDzckA3hLk5	0.328	0.554128	0.454545	1.000000	0.0	0.500401	0.0860
14	5xwpXWWkfJRqg1S27oVxh4	0.698	0.242202	0.727273	0.579002	1.0	0.396206	1.0000
15	5W7xC99N2Zzfh69r7I7zWK	0.238	0.429358	0.090909	0.818998	1.0	0.473684	0.0181
16	7MDKvOzNgAJ3KMCtaP2UOa	0.364	0.930275	0.727273	0.721638	1.0	0.951910	0.7039
17	5YCs9rP6ZcMJW6me3QhtWr	0.646	0.581651	0.090909	0.639955	1.0	0.233235	0.0410
18	5MGeW5mHnqeJiM0NPBbkEZ	0.320	0.361468	0.090909	0.845710	0.0	0.092172	0.7679
19	0w3Q3VFdrYzo24QUIGnBNy	0.188	0.765138	1.000000	0.820957	0.0	1.000000	0.0056
20	3r9n4SFmaqViN6G4Z6TbV7	0.508	0.550459	0.363636	0.785375	0.0	0.096981	0.2023
21	0FwezIaHwoxFy8IOLbF2UQ	0.806	0.100917	0.818182	0.508973	1.0	0.967940	0.2682
22	07J4gvixM0ksqSpgDw15J1	0.858	0.000000	0.090909	0.447917	1.0	0.818328	0.0555
23	1OBWQFd7hcTo18GWxS1bnZ	0.770	0.526606	0.000000	0.780631	0.0	0.324072	0.0714
24	5uNhj4KHA7nZO1CMwbnnqG	0.548	0.365138	0.545455	0.515986	1.0	0.137590	0.0254
25	1lqKgoo11HKUoMpbouC7I4	0.860	0.128440	0.090909	0.000000	1.0	0.334758	0.7513
26	6yxgLlHHp9EMklVUUdyjKh	1.000	0.251376	0.545455	0.814356	1.0	0.714133	0.1500
27	20X0AnKMbmD4PG6D5W1G3G	0.636	0.587156	0.090909	0.613139	1.0	0.084959	0.7631
28	4IeUAegXVbETb7FpgNDnSM	0.384	0.451376	0.090909	0.701733	0.0	0.652685	0.0000
29	0jkeRqlWciqKxU3iHQfdWj	0.520	0.893578	0.636364	0.894183	1.0	0.324072	0.3084

In [7]:
```
print(playlist1.numeric_features.shape)
playlist1.numeric_features
#includes popularity
```

```
(30, 14)
```

Out[7]:

	track_id	danceability	energy	key	loudness	mode	speechiness	ac
0	2XHzzp1j4IfTNp1FTn7YFg	0.669	0.634	11	-6.476	1	0.0327	
1	01A7PEPSnmtixFPfB2UTal	0.793	0.631	11	-6.109	0	0.0998	
2	6AaOtHsKd195ec0Y4kC9ER	0.721	0.607	0	-7.078	0	0.2770	
3	63M8PK8yavNITSViKUB62p	0.752	0.766	1	-5.110	1	0.2510	
4	17Lmf8pUgrRrEKIKkuZirE	0.799	0.499	11	-7.785	1	0.1820	
5	2LvRR121MWFmmEGkuV2vQP	0.919	0.506	1	-7.140	1	0.3040	
6	1s9hBKw37JvYV9KYiR9oef	0.744	0.539	10	-6.757	1	0.0596	
7	1UYLxZIL4IicS6TaPUZFTH	0.893	0.635	2	-9.000	1	0.0514	
8	5BviN9XWqCYE88AOJg6nEK	0.801	0.521	4	-9.871	0	0.2800	
9	4fUh9VvKW4shhWN5LfN1Ba	0.762	0.611	9	-5.462	0	0.0876	
10	3m8CQnnfJJp4eQMWWI3zay	0.451	0.632	8	-7.980	0	0.2970	
11	5tFep7dXGd7vEJ668wTPux	0.850	0.405	7	-7.509	0	0.2250	
12	2JfOSsfYs8QNneR3oTpNv4	0.892	0.891	5	-4.428	0	0.3050	
13	29UgEHyxg63NDzckA3hLk5	0.615	0.648	5	-3.792	0	0.2200	
14	5xwpXWWkfJRqg1S27oVxh4	0.800	0.478	8	-7.874	1	0.1810	
15	5W7xC99N2Zzfh69r7I7zWK	0.570	0.580	1	-5.547	1	0.2100	
16	7MDKvOzNgAJ3KMCtaP2UOa	0.633	0.853	8	-6.491	1	0.3890	
17	5YCs9rP6ZcMJW6me3QhtWr	0.774	0.663	1	-7.283	1	0.1200	
18	5MGeW5mHnqeJiM0NPBbkEZ	0.611	0.543	1	-5.288	0	0.0672	
19	0w3Q3VFdrYzo24QUIGnBNy	0.545	0.763	11	-5.528	0	0.4070	
20	3r9n4SFmaqViN6G4Z6TbV7	0.705	0.646	4	-5.873	0	0.0690	
21	0FwezIaHwoxFy8IOLbF2UQ	0.854	0.401	9	-8.553	1	0.3950	
22	07J4gvixM0ksqSpgDw15J1	0.880	0.346	1	-9.145	1	0.3390	
23	1OBWQFd7hcTo18GWxS1bnZ	0.836	0.633	0	-5.919	0	0.1540	
24	5uNhj4KHA7nZO1CMwbnnqG	0.725	0.545	6	-8.485	1	0.0842	
25	1lqKgoo11HKUoMpbouC7I4	0.881	0.416	1	-13.488	1	0.1580	
26	6yxgLlHHp9EMklVUUdyjKh	0.951	0.483	6	-5.592	1	0.3000	
27	20X0AnKMbmD4PG6D5W1G3G	0.769	0.666	1	-7.543	1	0.0645	
28	4IeUAegXVbETb7FpgNDnSM	0.643	0.592	1	-6.684	0	0.2770	
29	0jkeRqlWciqKxU3iHQfdWj	0.711	0.833	7	-4.818	1	0.1540	

In [8]:
```
print(playlist1.normalized_numeric_features.shape)
playlist1.normalized_numeric_features
```

```
(30, 14)
```

Out[8]:

	track_id	0	1	2	3	4	5	
0	2XHzzp1j4IfTNp1FTn7YFg	0.436	0.528440	1.000000	0.723185	1.0	0.000000	0.0183
1	01A7PEPSnmtixFPfB2UTal	0.684	0.522936	1.000000	0.761035	0.0	0.179268	0.0849
2	6AaOtHsKd195ec0Y4kC9ER	0.540	0.478899	0.000000	0.661097	0.0	0.652685	0.2156
3	63M8PK8yavNITSViKUB62p	0.602	0.770642	0.090909	0.864068	1.0	0.583222	0.7892
4	17Lmf8pUgrRrEKIKkuZirE	0.696	0.280734	1.000000	0.588181	1.0	0.398878	0.0131
5	2LvRR121MWFmmEGkuV2vQP	0.936	0.293578	0.090909	0.654703	1.0	0.724820	0.3013
6	1s9hBKw37JvYV9KYiR9oef	0.586	0.354128	0.909091	0.694204	1.0	0.071867	0.3368
7	1UYLxZIL4IicS6TaPUZFTH	0.884	0.530275	0.181818	0.462871	1.0	0.049960	0.0363
8	5BviN9XWqCYE88AOJg6nEK	0.700	0.321101	0.363636	0.373040	0.0	0.660700	0.0209
9	4fUh9VvKW4shhWN5LfN1Ba	0.622	0.486239	0.818182	0.827764	0.0	0.146674	0.2776
10	3m8CQnnfJJp4eQMWWl3zay	0.000	0.524771	0.727273	0.568069	0.0	0.706118	0.0157
11	5tFep7dXGd7vEJ668wTPux	0.798	0.108257	0.636364	0.616646	0.0	0.513759	0.2705
12	2JfOSsfYs8QNneR3oTpNv4	0.882	1.000000	0.454545	0.934406	0.0	0.727491	0.1161
13	29UgEHyxg63NDzckA3hLk5	0.328	0.554128	0.454545	1.000000	0.0	0.500401	0.0860
14	5xwpXWWkfJRqg1S27oVxh4	0.698	0.242202	0.727273	0.579002	1.0	0.396206	1.0000
15	5W7xC99N2Zzfh69r7I7zWK	0.238	0.429358	0.090909	0.818998	1.0	0.473684	0.0181
16	7MDKvOzNgAJ3KMCtaP2UOa	0.364	0.930275	0.727273	0.721638	1.0	0.951910	0.7039
17	5YCs9rP6ZcMJW6me3QhtWr	0.646	0.581651	0.090909	0.639955	1.0	0.233235	0.0410
18	5MGeW5mHnqeJiM0NPBbkEZ	0.320	0.361468	0.090909	0.845710	0.0	0.092172	0.7679
19	0w3Q3VFdrYzo24QUIGnBNy	0.188	0.765138	1.000000	0.820957	0.0	1.000000	0.0056
20	3r9n4SFmaqViN6G4Z6TbV7	0.508	0.550459	0.363636	0.785375	0.0	0.096981	0.2023
21	0FwezIaHwoxFy8IOLbF2UQ	0.806	0.100917	0.818182	0.508973	1.0	0.967940	0.2682
22	07J4gvixM0ksqSpgDw15J1	0.858	0.000000	0.090909	0.447917	1.0	0.818328	0.0555
23	1OBWQFd7hcTo18GWxS1bnZ	0.770	0.526606	0.000000	0.780631	0.0	0.324072	0.0714
24	5uNhj4KHA7nZO1CMwbnnqG	0.548	0.365138	0.545455	0.515986	1.0	0.137590	0.0254
25	1lqKgoo11HKUoMpbouC7I4	0.860	0.128440	0.090909	0.000000	1.0	0.334758	0.7513
26	6yxgLlHHp9EMklVUUdyjKh	1.000	0.251376	0.545455	0.814356	1.0	0.714133	0.1500
27	20X0AnKMbmD4PG6D5W1G3G	0.636	0.587156	0.090909	0.613139	1.0	0.084959	0.7631
28	4IeUAegXVbETb7FpgNDnSM	0.384	0.451376	0.090909	0.701733	0.0	0.652685	0.0000
29	0jkeRqlWciqKxU3iHQfdWj	0.520	0.893578	0.636364	0.894183	1.0	0.324072	0.3084

In [9]:
```
playlist1.df_playlist
```

Out[9]:

	index	song_title	artist	artist_id	album	
0	1	Love Me	[Lil Wayne, Drake, Future]	[55Aa2cqylxrFIXC767Z865, 3TVXtAsR1Inumwj472S9r...	I Am Not A Human Being II (Deluxe)	2XH
1	2	Buy The World	[Mike WiLL Made-It, Lil Wayne, Kendrick Lamar,...	[0NWbwDZY1VkRqFafuQm6wk, 55Aa2cqylxrFIXC767Z86...	Buy The World	01A
2	3	DnF (feat. Drake & Future)	[Preme, Drake, Future]	[0bdJZl7TDeiymDYzMJnVh2, 3TVXtAsR1Inumwj472S9r...	DnF (feat. Drake & Future)	6AaOt
3	4	Schemin Up (feat. Drake and P. Reign)	[OB OBrien, Drake, Preme]	[4thGcsymOK5oc43gJVtAqZ, 3TVXtAsR1Inumwj472S9r...	Schemin Up (feat. Drake and P. Reign)	63M8
4	5	I Just Wanna Party	[YG, ScHoolboy Q, Jay Rock]	[0A0FS04o6zMoto8OKPsDwY, 5IcR3N7QB1j6KBL8eImZ8...	My Krazy Life	17L
5	6	HeadBand (feat. 2 Chainz)	[B.o.B, 2 Chainz]	[5ndkK3dpZLKtBklKjxNQwT, 17lzZA2AlOHwCwFALHttmp]	Underground Luxury	2LvRR12
6	7	Fefe On The Block	[Stunt Taylor]	[6OlcRxdAiJfOY2SiexBMej]	Stunt'n On Turbo	1s9h
7	8	Blasé (feat. Future & Rae Sremmurd)	[Ty Dolla $ign, Future, Rae Sremmurd]	[7c0XG5clJTrrAgEC3ULPiq, 1RyvyyTE3xzB2ZywiAwp0...	Free TC	1UY
8	9	100	[The Game, Drake]	[0NbfKEOTQCcwd6o7wSDOHI, 3TVXtAsR1Inumwj472S9r4]	The Documentary 2	5BviN9
9	10	Be Real (feat. DeJ Loaf)	[Kid Ink, DeJ Loaf]	[6KZDXtSj0SzGOV705nNeh3, 7kFfY4UjNdNyaeUgLIEbIF]	Full Speed	4fUh9V
10	11	Drank in My Cup	[Kirko Bangz]	[2r8r62VGJKGi463aH1HJUZ]	Drank in My Cup	3m8CQ
11	12	Both (feat. Drake)	[Gucci Mane, Drake]	[13y7CgLHjMVRMDqxdx0Xdo, 3TVXtAsR1Inumwj472S9r4]	The Return of East Atlanta Santa	5tFep
12	13	OG Bobby Johnson (feat. Snoop Dogg, A$AP Ferg,...	[QUE., A$AP Ferg, Pusha T, Snoop Dogg]	[4Mop3hBjHoSoKwwkzKD3Gi, 5dHt1vcEm9qb8fCyLcB3H...	Who Is QUE. EP	2JfOS

	index	song_title	artist	artist_id	album	
13	14	Lockjaw (feat. Kodak Black)	[French Montana, Kodak Black]	[6vXTefBL93Dj5IqAWq6OTv, 46SHBwWsqBkxI7EeeBEQG7]	MONTANA	29UgE
14	15	All I Need (One Mo Drank) (feat. K Camp)	[Juicy J, K CAMP]	[5gCRApTajqwbnHHPbr2Fpi, 5bgfj5zUoWpyeVatGDjn6H]	All I Need (One Mo Drank) (feat. K Camp)	5xwpX
15	16	Aston Martin Music	[Rick Ross, Drake, Chrisette Michele]	[1sBkRIssrMs1AbVkOJbc7a, 3TVXtAsR1Inumwj472S9r...	Teflon Don	5W7x(
16	17	Hella Hoes (feat. A $APRocky$, A AP Ferg, A$AP ...	[A$APMob$, AAP Rocky, A $APFerg$, AAP NAST, A...	[7yO4IdJjCEPz7YgZMe25iS, 13ubrt8QOOCPljQ2FL1Kc...	Hella Hoes (feat. A $APRocky$, A AP Ferg, A$AP ...	7MDKvC
17	18	Uber Everywhere (feat. Travis Scott)	[MadeinTYO, Travis Scott]	[5SyGEPymt1G2uto47tVWvZ, 0Y5tJX1MQlPlqiwlOH1tJY]	You Are Forgiven (Deluxe Edition)	5YCs9r
18	19	I Like Tuh	[Carnage, ILOVEMAKONNEN]	[7CCjtD0hCK005Bvg2WG1a7, 3aGFCoR8xGN6DKwvdzeSja]	Papi Gordo	5MGeW5
19	20	679 (feat. Monty)	[Fetty Wap, Monty]	[6PXS4YHDkKvl1wkIl4V8DL, 1Wnfj5qZsp8nPsGBBRRa4W]	Fetty Wap (Deluxe)	0w3Q3
20	21	Valley	[Young Chop, Chief Keef]	[5L0n62BVUljSzKMLyMaFHR, 15iVAtD3s3FsQR4w1v6M0P]	Still	3r9n4S
21	22	Imma Ride	[Rich Homie Quan]	[5lHRUCqkQZCIWeX7xG4sYT]	Rich Homie Cartel Vol 1	0Fwe
22	23	U Guessed It	[OG Maco, 2 Chainz]	[0nvvVQhPQGkYTJlThLbJeu, 17lzZA2AlOHwCwFALHttmp]	U Guessed It (feat. 2 Chainz) - Single	07J4
23	24	Awwsome	[Shy Glizzy]	[1DvtabXAjfrMihPP6JQdHs]	Awwsome	1OBWC
24	25	I Love My Squad	[Iamsu!]	[2ZavqCJe7uqkRpISes0NFi]	Sincerely Yours, IAMSU!	5uNhj4

	index	song_title	artist	artist_id	album	
25	26	UP! (Beat The P*ssy UP) [Street]	[LoveRance, Iamsu!, Skipper]	[1rBxtaN521NYi8Z35G7fUn, 2ZavqCJe7uqkRpISes0NF...	UP! (Beat The P*ssy Up)	1lqKg
26	27	Flicka Da Wrist	[Chedda Da Connect]	[0KF35OGFXQttk0yWReabtG]	Chedda World "The Album"	6yxgI
27	28	Like Whaaat (feat. Bad Lucc)	[Problem, Bad Lucc]	[0399oiMcmbOzzsYQDNYqxn, 77kP3V3j9SPpyF5wem7tck]	Welcome to Mollywood, Pt. 2	20X0AnKI
28	29	I Luv This Shit	[August Alsina, Trinidad James]	[19Fi1Rj7kk8kyiwxpXy3yM, 0I5HubncQ8E1MFZOlPDY4J]	Downtown: Life Under The Gun	4IeUAe
29	30	Seen It All	[Jeezy, JAY-Z]	[4yBK75WVCQXej1p04GWqxH, 3nFkdlSjzX9mRTtwJOzDYB]	Seen It All: The Autobiography	0jke

30 rows × 22 columns

In [10]:
```
pca_obj, playlist1.pca_transformed_features = playlist1.pca_transform(playlist1
```

In [11]:
```
pca_obj.explained_variance_
```

Out[11]:
```
array([0.30594016, 0.15391683, 0.13724652, 0.10479275, 0.09413014,
       0.07207239, 0.06556575, 0.05663633, 0.03776728, 0.03538319,
       0.01973168, 0.01348896, 0.00786931])
```

In [12]:
```
len(pca_obj.explained_variance_)
```

Out[12]:
```
13
```

In [13]:
```
sum(pca_obj.explained_variance_[:6]) #reduce playlist1.pca_transformed_features
```

Out[13]:
```
0.8680987951170095
```

In [14]:
```
#automate the process to find # of components that explain 75% of the variance
# should be 'automated' because this may vary per playlist
def choose_components(pca, pct):
    '''Will find the number of components in PCA necessary to explain pct of th
       ----------------------------------------
       parameters:
       - pca: pandas.decomposition.PCA object
       - pct: % variance youd like explained by the components
       ----------------------------------------
       returns:
       - num_components: number of components to explain desired % variance
    '''

    for num_components in range(0, len(pca.explained_variance_)):
        var_explained = sum(pca.explained_variance_[:num_components])
```

```
        if var_explained >= pct:
            return num_components
    return -1
```

In [15]:
```
components = choose_components(pca_obj, .85)
components
```

Out[15]: 6

In [16]:
```
#reduce the pca data
playlist1.pca_transformed_features = pd.DataFrame(playlist1.pca_transformed_fea
playlist1.pca_transformed_features
```

Out[16]:

	0	1	2	3	4	5
0	-0.310691	0.783396	0.100749	0.048871	0.297630	0.364042
1	0.638580	0.712854	-0.500507	0.056164	-0.392342	0.385841
2	0.501253	-0.499147	-0.116262	-0.110061	-0.097786	-0.004071
3	-0.486820	-0.505744	0.452408	0.193701	0.285574	0.503326
4	-0.420821	0.666693	-0.183224	-0.345210	-0.130369	-0.154391
5	-0.609557	-0.158517	-0.236944	-0.356428	0.144582	0.069139
6	-0.334146	0.450530	0.191466	0.437847	-0.251508	-0.402911
7	-0.540953	-0.115519	-0.295052	-0.057181	0.116795	0.172252
8	0.409572	-0.428562	-0.157281	-0.559135	-0.190773	0.279345
9	0.544823	0.441660	-0.384768	0.223262	-0.397805	0.348767
10	0.653613	0.311294	-0.110727	-0.250039	0.443281	-0.005960
11	0.373689	0.211862	-0.441480	-0.093983	-0.094275	0.066486
12	0.735476	-0.524488	0.610000	-0.247955	-0.284151	0.327193
13	0.669450	0.024402	-0.114444	0.122388	0.428366	-0.200116
14	-0.556330	-0.033684	0.521065	0.349182	-0.370110	0.203645
15	-0.341449	-0.041329	-0.010624	-0.024681	0.844146	0.073123
16	-0.280303	0.138056	0.752552	0.116114	0.146298	0.049104
17	-0.405919	0.061229	-0.374312	0.106415	0.413192	-0.358666
18	0.471242	-0.494943	-0.052388	0.900358	-0.001743	-0.279588
19	1.041642	0.399960	0.782979	-0.214625	0.109845	-0.299324
20	0.677814	-0.230699	-0.051709	0.326347	-0.203415	-0.338365
21	-0.286796	0.037791	0.544484	-0.543915	-0.330951	-0.379722
22	-0.609065	-0.336896	-0.358573	-0.459379	-0.124919	-0.499380
23	0.463845	-0.483526	-0.305983	0.183427	-0.038740	-0.141745
24	-0.413191	0.270706	-0.216801	0.051652	-0.225938	-0.140255
25	-0.810139	-0.448412	-0.175074	-0.004594	-0.374182	0.185311
26	-0.445176	-0.022255	0.217349	-0.246685	-0.223581	-0.142255
27	-0.652644	-0.175106	-0.285816	0.437017	-0.001165	0.114949
28	0.547888	-0.334853	-0.201504	-0.270412	0.368690	0.128456
29	-0.224887	0.323247	0.400420	0.231539	0.135352	0.075772

In [17]:
```
recs
```

Out[17]:

	artist	track	track_id	similarity
0	Lil Wayne	Hittas	5s6pKnF3ZPtuqvEx1i1V7U	0.939879
1	Lil Wayne	Mona Lisa (feat. Kendrick Lamar)	0dbTQYW3Ad1FTzIA9t90E8	0.929644
2	Lil Wayne	No Worries	6bdFbw9THAMceZtkPWXu8e	0.921159
3	Lil Wayne	Drop The World	3e21cX0CVwzkQXiHz7WUQZ	0.916111
4	Lil Wayne	Mrs. Officer	0EHR9OObFtjlhQB8wSt1m7	0.911495
...	...	...	...	...
270	JAY-Z	Empire State Of Mind	2igwFfvr1OAGX9SKDCPBwO	0.937264
271	JAY-Z	Neck & Wrist (feat. JAY-Z & Pharrell Williams)	7kQJCw0ZkvHgfJqRwPblmG	0.922061
272	JAY-Z	Ni**as In Paris	4Li2WHPkuyCdtmokzW2007	0.919346
273	JAY-Z	Heartbreaker (feat. Jay-Z)	0jsANwwkkHyyeNyuTFq2XO	0.918459
274	JAY-Z	JAY-Z	1kMzZkdIv1KYLZFPsXeUQb	0.916069

275 rows × 4 columns

In [18]:
```
playlist1.pca_transformed_features.iloc[:,5]
```

Out[18]:
```
0     0.364042
1     0.385841
2    -0.004071
3     0.503326
4    -0.154391
5     0.069139
6    -0.402911
7     0.172252
8     0.279345
9     0.348767
10   -0.005960
11    0.066486
12    0.327193
13   -0.200116
14    0.203645
15    0.073123
16    0.049104
17   -0.358666
18   -0.279588
19   -0.299324
20   -0.338365
21   -0.379722
22   -0.499380
23   -0.141745
24   -0.140255
25    0.185311
26   -0.142255
27    0.114949
28    0.128456
29    0.075772
Name: 5, dtype: float64
```

In [19]:
```
#recalculate average audio values on the pca reduced data
avg_audio_values = pd.DataFrame(playlist1.pca_transformed_features.mean(axis=0)
```

In [20]:
```
avg_audio_values
```

Out[20]:

	0
0	2.220446e-17
1	2.405483e-17
2	1.850372e-18
3	4.348374e-17
4	8.511710e-17
5	-2.081668e-17

In [21]:
```
artists_search = create_dataframes.artist_search_results(artist_counts.columns)
recs = playlist1.get_recommendations(artists_search, avg_audio_values, pca_comp
```

In [22]:
```
recs
```

Out[22]:

	artist	track	track_id	similarity
0	Lil Wayne	SMUCKERS (feat. Lil Wayne & Kanye West)	078C2jXg7XsMgW78Gfx1JA	0.0
1	Lil Wayne	Down	6cmm1LMvZdB5zsCwX5BjqE	0.0
2	Lil Wayne	Beware	6bxUnsSGZCmoHHU5auwtps	0.0
3	Lil Wayne	6 Foot 7 Foot	5Qy6a5KzM4XlRxsNcGYhgH	0.0
4	Lil Wayne	Lollipop	4P7VFiaZb3xrXoqGwZXC3J	0.0
...	...	...	...	...
270	JAY-Z	4:44	1gT5TGwbkkkUliNzHRIGi1	0.0
271	JAY-Z	JAY-Z	684MKCXtMZF58P24BkGFkJ	0.0
272	JAY-Z	GOD DID (feat. Rick Ross, Lil Wayne, Jay-Z, Jo...	2sOj9vyd6yiss9W1IK6chU	0.0
273	JAY-Z	Neck & Wrist (feat. JAY-Z & Pharrell Williams)	7kQJCw0ZkvHgfJqRwPblmG	0.0
274	JAY-Z	Empire State Of Mind	2igwFfvr1OAGX9SKDCPBwO	0.0

275 rows × 4 columns

why are similarities = 0...

In [23]:
```
l = [-4.870593456842883e-17, -2.1693136948257682e-17, 1.92738037764736e-17, -5.
max(l)
```

Out[23]:
```
7.60854490150577e-17
```

because values in similarities are tiny. how? if we used cosine distance

try using standard scalers

- changed line 206 in create_dataframes.py to use standardscaler instead of maxscaler.
- should try using another distance measure too