In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
def normalize(df):
    df['duration_ms_stand'] = np.where(df['duration_ms'] >= 10 * 60 * 1000, 10 * 60 * 1
000, df['duration_ms'])
#     df = df[df['duration_ms'] <= 10 * 60 * 1000]
    df['duration_norm'] = (df['duration_ms']-df['duration_ms'].min())/(df['duration_ms'
].max()-df['duration_ms'].min())

    df['loudness_stand'] = np.where(df['loudness'] <= -40, -40 , df['loudness'])
    df['loudness_norm'] = (df['loudness_stand']-df['loudness_stand'].min())/(df['loudne
ss_stand'].max()-df['loudness_stand'].min())

    df['tempo_norm'] = (df['tempo']-df['tempo'].min())/(df['tempo'].max()-df['tempo'].m
in())
    return df
```

In [3]:

```
random_songs = normalize(pd.read_csv('SpotifyAudioFeaturesNov2018.csv',encoding = 'utf-
8'))
```

let's take the most popular songs - we arbitrary chose 90 and up:

In [4]:

```
popularity_90_or_more= random_songs[random_songs.popularity>=90]
popularity_90_or_more.sort_values(by ='popularity',ascending=False).head()
```

Out[4]:

	artist_name	track_id	track_name	acousticness	danceability	d
109531	Ariana Grande	2rPE9A1vEgShuZxxzR2tZH	thank u, next	0.2800	0.724	
109546	DJ Snake	4w8niZpiMy6qz1mntFA5uM	Taki Taki (with Selena Gomez, Ozuna & Cardi B)	0.1530	0.841	
109544	Bad Bunny	116H0KvKr2ZI4RPuVBruDO	MIA (feat. Drake)	0.0141	0.817	
109540	Marshmello	2dpaYNEQHiRxtZbfNsse99	Happier	0.1910	0.687	
109535	Kodak Black	7l3E7IcozEodtVsSTCkcaA	ZEZE (feat. Travis Scott & Offset)	0.0710	0.826	

5 rows × 22 columns

In [5]:

```
features_list = ['acousticness','danceability','energy','instrumentalness','liveness',
'mode','speechiness','valence',
                 'duration_norm','loudness_norm','tempo_norm']
```

In [6]:

```

```

Out[6]:

```
'thank u, next'
```

using Euclidean distance

In [7]:

```
for i in range(5):
    random_songs[popularity_90_or_more.iloc[i]['track_name']] = random_songs[features_l
ist].apply(lambda x: np.linalg.norm(x-np.array(popularity_90_or_more.iloc[i][features_l
ist])),axis =1)
```

In [10]:

```
random_songs.sort_values(popularity_90_or_more.iloc[0]['track_name']).head(2)[['artist_
name','track_name','popularity'] + features_list]
```

Out[10]:

	artist_name	track_name	popularity	acousticness	danceability	energy	instrum
109531	Ariana Grande	thank u, next	100	0.28	0.724	0.647	
212	Cody Johns	Smile	34	0.28	0.747	0.634	

In [11]:

```
random_songs.sort_values(popularity_90_or_more.iloc[1]['track_name']).head(2)[['artist_
name','track_name','popularity'] + features_list]
```

Out[11]:

	artist_name	track_name	popularity	acousticness	danceability	energy	instrum
109533	Travis Scott	SICKO MODE	95	0.00513	0.834	0.730	
85256	Minty Burns	Green Man	23	0.03040	0.800	0.734	

In [12]:

```
random_songs.sort_values(popularity_90_or_more.iloc[2]['track_name']).head(2)[['artist_name','track_name','popularity'] + features_list]
```

Out[12]:

	artist_name	track_name	popularity	acousticness	danceability	energy	instrumenta
109534	Sheck Wes	Mo Bamba	93	0.194	0.729	0.625	0.(
106666	Poo Bear	Hard 2 Face Reality	68	0.220	0.625	0.626	0.(

In [13]:

```
random_songs.sort_values(popularity_90_or_more.iloc[3]['track_name']).head(2)[['artist_name','track_name','popularity'] + features_list]
```

Out[13]:

	artist_name	track_name	popularity	acousticness	danceability	energy	instrum
109535	Kodak Black	ZEZE (feat. Travis Scott & Offset)	95	0.0710	0.826	0.615	
96911	Kap G	Lenox Square (feat. Chef Cook It Up)	39	0.0594	0.822	0.662	

In [14]:

```
random_songs.sort_values(popularity_90_or_more.iloc[4]['track_name']).head(2)[['artist_name','track_name','popularity'] + features_list]
```

Out[14]:

	artist_name	track_name	popularity	acousticness	danceability	energy	instrum
109536	Halsey	Without Me	95	0.297	0.752	0.488	
20671	Landon Austin	Psycho, Wait (Acoustic Mashup)	26	0.287	0.748	0.500	