In [1]:
```
import pandas as pd
```

In [2]:
```
triplets_file = 'https://static.turi.com/datasets/millionsong/10000.txt'
songs_metadata_file = 'https://static.turi.com/datasets/millionsong/song_data.csv'
```

In [3]:
```
data=pd.read_table(triplets_file,header=None)  # TAKING 1st set of data containing user coun
t.....
```

In [4]:
```
data
```

Out[4]:

	0	1	2
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBYHAJ12A6701BF1D	1
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1
...	...	...	...
1999995	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJEYPO12AAA8C6B0E	2
1999996	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJJYDE12AF729FC16	4
1999997	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJKQSF12A6D4F5EE9	3
1999998	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJUXGA12AC961885C	1
1999999	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJYOLS12A8C13C06F	1

2000000 rows × 3 columns

In [5]:
```
data.columns=('user_id','song_id','count')
```

In [6]:
```
data
```

Out[6]:

	user_id	song_id	count
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBYHAJ12A6701BF1D	1
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1
...	...	...	...
1999995	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJEYPO12AAA8C6B0E	2
1999996	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJJYDE12AF729FC16	4
1999997	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJKQSF12A6D4F5EE9	3
1999998	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJUXGA12AC961885C	1
1999999	d8bfd4ec88f0f3773a9e022e3c1a0f1d3b7b6a92	SOJYOLS12A8C13C06F	1

2000000 rows × 3 columns

In [7]:
```
data2=pd.read_csv(songs_metadata_file) # Taking second set of data having information of the songs.....
```

In [8]:
```
data2
```

Out[8]:

	song_id	title	release	artist_name	year
0	SOQMMHC12AB0180CB8	Silent Night	Monster Ballads X-Mas	Faster Pussy cat	2003
1	SOVFVAK12A8C1350D9	Tanssi vaan	Karkuteillä	Karkkiautomaatti	1995
2	SOGTUKN12AB017F4F1	No One Could Ever	Butter	Hudson Mohawke	2006
3	SOBNYVR12A8C13558C	Si Vos Querés	De Culo	Yerba Brava	2003
4	SOHSBXH12A8C13B0DF	Tangle Of Aspens	Rene Ablaze Presents Winter Sessions	Der Mystic	0
...	...	...	...	...	...
999995	SOTXAME12AB018F136	O Samba Da Vida	Pacha V.I.P.	Kiko Navarro	0
999996	SOXQYIQ12A8C137FBB	Jago Chhadeo	Naale Baba Lassi Pee Gya	Kuldeep Manak	0
999997	SOHODZI12A8C137BB3	Novemba	Dub_Connected: electronic music	Gabriel Le Mar	0
999998	SOLXGOR12A81C21EB7	Faraday	The Trance Collection Vol. 2	Elude	0
999999	SOWXJXQ12AB0189F43	Fernweh feat. Sektion Kuchikäschtli	So Oder So	Texta	2004

1000000 rows × 5 columns

In [9]:
```
overall_data=pd.merge(data,data2,on="song_id")    # combining both data on basis of song id
.....
```

In [10]:
```
overall_data
```

Out[10]:

	user_id	song_id	count	title	rele
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove	Thic T W
1	7c86176941718984fed11b7c0674ff04c029b480	SOAKIMP12A8C130995	1	The Cove	Thic T W
2	76235885b32c4e8c82760c340dc54f9b608d7d7e	SOAKIMP12A8C130995	3	The Cove	Thic T W
3	250c0fa2a77bc6695046e7c47882ecd85c42d748	SOAKIMP12A8C130995	1	The Cove	Thic T W
4	3f73f44560e822344b0fb7c6b463869743eb9860	SOAKIMP12A8C130995	6	The Cove	Thic T W
...	...	...	...	...	
2086941	8d5be34165a0d2d20878abd6a48bb87af29b9f7a	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086942	23f8ab814cd41e4a3394e762cc7360eb6c04cbd7	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086943	4cc239fd4ab90eb599b2263e21dceebb252cf340	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086944	e0039fa2e1d0c51c729d2521a48eccc52a375cc2	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086945	6a5f74c28e6d091b31027965d402c93a6c7667e2	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L

2086946 rows × 7 columns

In [11]:
```
modified_data=overall_data.dropna(axis=0,subset=["title"]) #  Removing Nan values from the title column
```

In [12]:
```
modified_data=modified_data.dropna(axis=0,subset=["release"]) # Removing Nan "    " "    "   release column
```

In [13]:
```
modified_data=modified_data.dropna(axis=0,subset=["artist_name"]) # Removing Nan " "  " "   artist name column
```

In [14]:
```
modified_data
```

Out[14]:

	user_id	song_id	count	title	rele
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove	Thic T W
1	7c86176941718984fed11b7c0674ff04c029b480	SOAKIMP12A8C130995	1	The Cove	Thic T W
2	76235885b32c4e8c82760c340dc54f9b608d7d7e	SOAKIMP12A8C130995	3	The Cove	Thic T W
3	250c0fa2a77bc6695046e7c47882ecd85c42d748	SOAKIMP12A8C130995	1	The Cove	Thic T W
4	3f73f44560e822344b0fb7c6b463869743eb9860	SOAKIMP12A8C130995	6	The Cove	Thic T W
...	...	...	...	...	
2086941	8d5be34165a0d2d20878abd6a48bb87af29b9f7a	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086942	23f8ab814cd41e4a3394e762cc7360eb6c04cbd7	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086943	4cc239fd4ab90eb599b2263e21dceebb252cf340	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086944	e0039fa2e1d0c51c729d2521a48eccc52a375cc2	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086945	6a5f74c28e6d091b31027965d402c93a6c7667e2	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L

2086946 rows × 7 columns

In [15]:
```
total_count = (modified_data.groupby(by=['song_id'])['count'].count().reset_index().rename(colu
mns={'count':'total_count'})[['song_id','total_count']])
```

In [16]:
```
total_count  # storing total counts of the data....
print(total_count['total_count'].describe()) # to see threshold for data....
```

```
count    10000.000000
mean       208.694600
std        351.823049
min         48.000000
25%         89.000000
50%        126.000000
75%        205.000000
max       8277.000000
Name: total_count, dtype: float64
```

In [17]:
```
data_with_rating=overall_data.merge(total_count,left_on='song_id',right_on='song_id',how="left"
) ## Adding the total count to past data
```

In [18]:
```
total_count
```

Out[18]:

	song_id	total_count
0	SOAAAGQ12A8C1420C8	66
1	SOAACPJ12A81C21360	147
2	SOAACSG12AB018DC80	67
3	SOAAEJI12AB0188AB5	62
4	SOAAFAC12A67ADF7EB	190
...	...	...
9995	SOZZTNF12A8C139916	270
9996	SOZZVWB12AB0189C30	138
9997	SOZZWZV12A67AE140F	119
9998	SOZZYAO12A6701FF36	76
9999	SOZZZPV12A8C1444B5	71

10000 rows × 2 columns

In [19]:
```
data_with_rating # This data contains the total count additional column
```

Out[19]:

	user_id	song_id	count	title	rele
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove	Thic T W
1	7c86176941718984fed11b7c0674ff04c029b480	SOAKIMP12A8C130995	1	The Cove	Thic T W
2	76235885b32c4e8c82760c340dc54f9b608d7d7e	SOAKIMP12A8C130995	3	The Cove	Thic T W
3	250c0fa2a77bc6695046e7c47882ecd85c42d748	SOAKIMP12A8C130995	1	The Cove	Thic T W
4	3f73f44560e822344b0fb7c6b463869743eb9860	SOAKIMP12A8C130995	6	The Cove	Thic T W
...	...	...	...	...	
2086941	8d5be34165a0d2d20878abd6a48bb87af29b9f7a	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086942	23f8ab814cd41e4a3394e762cc7360eb6c04cbd7	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086943	4cc239fd4ab90eb599b2263e21dceebb252cf340	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086944	e0039fa2e1d0c51c729d2521a48eccc52a375cc2	SOPBPHJ12AAF3B59B6	2	Baby [You've Got What It Takes] [with Sharon J...	Cr L
2086945	6a5f74c28e6d091b31027965d402c93a6c7667e2	SOPBPHJ12AAF3B59B6	1	Baby [You've Got What It Takes] [with Sharon J...	Cr L

2086946 rows × 8 columns

In [28]:
```
data_table=data_with_rating.pivot_table(index='title',columns='user_id',values='count').fillna(0) # To use cosine transform , converted to pivot table
```

In [29]:
```
data_table
```

Out[29]:

user_id	00003a4459f33b92906be11abe0e93efc423c0ff	00005c6177188f12fb5e2e82cdbd93e8a3f35e64
title		
#!*@ You Tonight [Featuring R. Kelly] (Explicit Album Version)	0.0	0.0
#40	0.0	0.0
& Down	0.0	0.0
' Cello Song	0.0	0.0
'97 Bonnie & Clyde	0.0	0.0
...	...	...
the Love Song	0.0	0.0
you were there with me	0.0	0.0
¡Viva La Gloria! (Album Version)	0.0	0.0
¿Lo Ves? [Piano Y Voz]	0.0	0.0
Época	0.0	0.0

9593 rows × 76353 columns

In [30]:
```
from scipy.sparse import csr_matrix  # using cosine transforms

song_features_matrix= csr_matrix(data_table.values)

from sklearn.neighbors import NearestNeighbors # using neighrest neighbors

model=NearestNeighbors(metric='cosine',algorithm='brute')
model.fit(song_features_matrix)
```

Out[30]:
```
NearestNeighbors(algorithm='brute', metric='cosine')
```

In [31]:
```
import numpy as np
```

In [32]:
```
any_user=np.random.choice(data_table.shape[0]) # taking any random song
print(any_user)
distances,indices =model.kneighbors(data_table.iloc[any_user,:].values.reshape(1,-1),n_neighb
ors=6) # use 6 neighrest neigbours of paricular song
```

```
9426
```

In [34]:
```
 for i in range(0,len(distances.flatten())):
        if i==0:
            print('Recommendation for {0}:\n'.format(data_table.index[any_user]))
        else:
            print('{0}:{1},with distances of {2}:'.format(i,data_table.index[indices.flatten()[i]],distances.
flatten()[i])) # use 6 neighrest neigbours including the same movie
```

```
Recommendation for You And Me Jesus:

1:The Crow & The Butterfly (Album Version),with distances of 0.8064025040816024:
2:How's It Going To Be (LP Version),with distances of 0.8415733847286824:
3:For The Dishwasher,with distances of 0.8446658103761898:
4:What If?,with distances of 0.845193975332189:
5:Bedroom Suite,with distances of 0.848600647760634:
```

In [ ]:
```

```