Dataset Information

Million Songs Dataset contains of two files: triplet_file and metadata_file. The triplet_file contains user_id, song_id and listen time. The metadata_file contains song_id, title, release, year and artist_name. Million Songs Dataset is a mixture of song from various website with the rating that users gave after listening to the song.

There are 3 types of recommendation system: content-based, collaborative and popularity.

Import modules

In [3]:

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

Loading the dataset

In [4]:

```
song_df_1 = pd.read_csv('triplets_file.csv')
song_df_1.head()
```

Out[4]:

	user_id	song_id	listen_count
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBYHAJ12A6701BF1D	1
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1

In [5]:

```
song_df_2 = pd.read_csv('song_data.csv')
song_df_2.head()
```

Out[5]:

	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

In [9]:

```
# combine both data
song_df = pd.merge(song_df_1, song_df_2.drop_duplicates(['song_id']), on='song_
song_df.head()
```

Out[9]:

	user_id	song_id	listen_count	title
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2	Entre Dos Aguas
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1	Stronger
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBYHAJ12A6701BF1D	1	Constellations
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1	Learn To Fly

In [11]:

```
print(len(song_df_1), len(song_df_2))
```

```
2000000 1000000
```

In [12]:

```
len(song_df)
```

Out[12]:

```
2000000
```

Data Preprocessing

In [13]:

```
# creating new feature combining title and artist name
song_df['song'] = song_df['title']+'-'+song_df['artist_name']
song_df.head()
```

Out[13]:

	user_id	song_id	listen_count	title
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOAKIMP12A8C130995	1	The Cove
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBBMDR12A8C13253B	2	Entre Dos Aguas
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBXHDL12A81C204C0	1	Stronger
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SOBYHAJ12A6701BF1D	1	Constellations
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	SODACBL12A8C13C273	1	Learn To Fly

In [14]:

```
# taking top 10k samples for quick results
song_df = song_df.head(10000)
```

In [16]:

```
# cummulative sum of listen count of the songs
song_grouped = song_df.groupby(['song']).agg({'listen_count' : 'count'}).reset_
song_grouped.head()
```

Out[16]:

	song	listen_count
0	#40-DAVE MATTHEWS BAND	1
1	& Down-Boys Noize	4
2	'97 Bonnie & Clyde-Eminem	2
3	'Round Midnight-Miles Davis	3
4	'Till I Collapse-Eminem / Nate Dogg	6

In [17]:

```
grouped_sum = song_grouped['listen_count'].sum()
song_grouped['percentage'] = (song_grouped['listen_count'] / grouped_sum) * 100
song_grouped.sort_values(['listen_count','song'],ascending=[0,1])
```

Out[17]:

	song	listen_count	percentage
3660	Sehr kosmisch-Harmonia	45	0.45
4678	Undo-Björk	32	0.32
5105	You're The One-Dwight Yoakam	32	0.32
1071	Dog Days Are Over (Radio Edit)-Florence + The ...	28	0.28
3655	Secrets-OneRepublic	28	0.28
...	...	...	...
5139	high fives-Four Tet	1	0.01
5140	in white rooms-Booka Shade	1	0.01
5143	paranoid android-Christopher O'Riley	1	0.01
5149	¿Lo Ves? [Piano Y Voz]-Alejandro Sanz	1	0.01
5150	Época-Gotan Project	1	0.01

5151 rows × 3 columns

Popularity Recommendation Engine

In [18]:

```
pr = Recommenders.popularity_recommender_py()
```

In [19]:

```
pr.create(song_df, 'user_id','song')
```

In [20]:

```
# display the top 10 popular songs
pr.recommend(song_df['user_id'][5])
```

Out[20]:

	user_id	song	score	Rank
3660	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Sehr kosmisch-Harmonia	45	1.0
4678	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Undo-Björk	32	2.0
5105	b80344d063b5ccb3212f76538f3d9e43d87dca9e	You're The One-Dwight Yoakam	32	3.0
1071	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Dog Days Are Over (Radio Edit)-Florence + The ...	28	4.0
3655	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Secrets-OneRepublic	28	5.0
4378	b80344d063b5ccb3212f76538f3d9e43d87dca9e	The Scientist-Coldplay	27	6.0
4712	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Use Somebody-Kings Of Leon	27	7.0
3476	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Revelry-Kings Of Leon	26	8.0
1387	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Fireflies-Charttraxx Karaoke	24	9.0
1862	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Horn Concerto No. 4 in E flat K495: II. Romanc...	23	10.0

In [21]:

```
pr.recommend(song_df['user_id'][100])
```

Out[21]:

	user_id	song	score	Rank
3660	e006b1a48f466bf59feefed32bec6494495a4436	Sehr kosmisch-Harmonia	45	1.0
4678	e006b1a48f466bf59feefed32bec6494495a4436	Undo-Björk	32	2.0
5105	e006b1a48f466bf59feefed32bec6494495a4436	You're The One-Dwight Yoakam	32	3.0
1071	e006b1a48f466bf59feefed32bec6494495a4436	Dog Days Are Over (Radio Edit)-Florence + The ...	28	4.0
3655	e006b1a48f466bf59feefed32bec6494495a4436	Secrets-OneRepublic	28	5.0
4378	e006b1a48f466bf59feefed32bec6494495a4436	The Scientist-Coldplay	27	6.0
4712	e006b1a48f466bf59feefed32bec6494495a4436	Use Somebody-Kings Of Leon	27	7.0
3476	e006b1a48f466bf59feefed32bec6494495a4436	Revelry-Kings Of Leon	26	8.0
1387	e006b1a48f466bf59feefed32bec6494495a4436	Fireflies-Charttraxx Karaoke	24	9.0
1862	e006b1a48f466bf59feefed32bec6494495a4436	Horn Concerto No. 4 in E flat K495: II. Romanc...	23	10.0

Item Similarity Recommendation

In [22]:

```
ir = Recommenders.item_similarity_recommender_py()
ir.create(song_df,'user_id','song')
```

In [23]:
```
user_items = ir.get_user_items(song_df['user_id'][5])
```

In [24]:
```
# display user songs history
for user_item in user_items:
    print(user_item)
```

```
The Cove-Jack Johnson
Entre Dos Aguas-Paco De Lucia
Stronger-Kanye West
Constellations-Jack Johnson
Learn To Fly-Foo Fighters
Apuesta Por El Rock 'N' Roll-Héroes del Silencio
Paper Gangsta-Lady GaGa
Stacked Actors-Foo Fighters
Sehr kosmisch-Harmonia
Heaven's gonna burn your eyes-Thievery Corporation feat. Emiliana Torrini
Let It Be Sung-Jack Johnson / Matt Costa / Zach Gill / Dan Lebowitz / Steve A
dams
I'll Be Missing You (Featuring Faith Evans & 112)(Album Version)-Puff Daddy
Love Shack-The B-52's
Clarity-John Mayer
I?'m A Steady Rollin? Man-Robert Johnson
The Old Saloon-The Lonely Island
Behind The Sea [Live In Chicago]-Panic At The Disco
Champion-Kanye West
Breakout-Foo Fighters
Ragged Wood-Fleet Foxes
Mykonos-Fleet Foxes
Country Road-Jack Johnson / Paula Fuga
Oh No-Andrew Bird
Love Song For No One-John Mayer
Jewels And Gold-Angus & Julia Stone
Warning-Incubus
83-John Mayer
Neon-John Mayer
The Middle-Jimmy Eat World
High and dry-Jorge Drexler
All That We Perceive-Thievery Corporation
The Christmas Song  (LP Version)-King Curtis
Our Swords (Soundtrack Version)-Band Of Horses
Are You In?-Incubus
Drive-Incubus
Generator-Foo Fighters
Come Back To Bed-John Mayer
He Doesn't Know Why-Fleet Foxes
Trani-Kings Of Leon
Bigger Isn't Better-The String Cheese Incident
Sun Giant-Fleet Foxes
City Love-John Mayer
Right Back-Sublime
Moonshine-Jack Johnson
Holes To Heaven-Jack Johnson
```

In [25]:

```
# give song recommendation for that user
ir.recommend(song_df['user_id'][5])
```

```
No. of unique songs for the user: 45
no. of unique songs in the training set: 5151
Non zero values in cooccurence_matrix :6844
```

Out[25]:

	user_id	song	score	rank
0	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Oliver James-Fleet Foxes	0.043076	1
1	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Quiet Houses-Fleet Foxes	0.043076	2
2	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Your Protector-Fleet Foxes	0.043076	3
3	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Tiger Mountain Peasant Song-Fleet Foxes	0.043076	4
4	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Sun It Rises-Fleet Foxes	0.043076	5
5	b80344d063b5ccb3212f76538f3d9e43d87dca9e	The End-Pearl Jam	0.037531	6
6	b80344d063b5ccb3212f76538f3d9e43d87dca9e	St. Elsewhere-Dave Grusin	0.037531	7
7	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Misled-Céline Dion	0.037531	8
8	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Oil And Water-Incubus	0.037531	9
9	b80344d063b5ccb3212f76538f3d9e43d87dca9e	Meadowlarks-Fleet Foxes	0.037531	10

In [28]:

```
# give related songs based on the words
ir.get_similar_items(['Oliver James - Fleet Foxes', 'The End - Pearl Jam'])
```

```
no. of unique songs in the training set: 5151
Non zero values in cooccurence_matrix :0
```

Out[28]:

	user_id	song	score	rank
0		Nice Weather For Ducks-Lemon Jelly	0.0	1
1		The Irony Of It All (Album Version)-The Streets	0.0	2
2		Officially Missing You (Radio Version)-Tamia	0.0	3
3		On The Road Again (Pigna People Remix)-Telex	0.0	4
4		What Can Be Safely Written-Nile	0.0	5
5		Lord I Guess I'll Never Know-The Verve	0.0	6
6		Take Em To Church-Cam'Ron / Juelz Santana / Un...	0.0	7
7		This Ain't A Scene_ It's An Arms Race-Fall Out...	0.0	8
8		Praise You-Fatboy Slim	0.0	9
9		Yeah Yeah-Bodyrox	0.0	10

In [ ]: