

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 [ ]: import pandas as pd
import numpy as np
import Recommenders as Recommenders
```

Loading the dataset

```
In [ ]: song_df_1 = pd.read_csv('triplets_file.csv')
song_df_1.head()
```

```
Out[ ]:
           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 [ ]: song_df_2 = pd.read_csv('song_data.csv')
song_df_2.head()
```

```
Out[ ]:
           song_id  title  release  artist_name  year
0  SOQMMHC12AB0180CB8  Silent Night  Monster Ballads X-Mas  Faster Pussay cat  2003
1  SOVFAK12A8C1350D9  Tanssi vaan  Karkuteillä  Karkkiautomaatti  1995
2  SOGTUKY12AB017F4F1  No One Could Ever  Butler  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 [ ]: # combine both data
song_df = pd.merge(song_df_1, song_df_2.drop_duplicates(['song_id']), on='song_id', how='left')
song_df.head()
```

```
Out[ ]:
           user_id  song_id  listen_count  title  release  artist_name  year
0  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOAKIMP12A8C130995      1  The Cove  Thicker Than Water  Jack Johnson  0
1  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBBMDR12A8C13253B      2  Entre Dos Aguas  Flamenco Para Niños  Paco De Lucia  1976
2  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBXHDL12A81C204C0      1  Stronger  Graduation  Kanye West  2007
3  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBYHAJ12A6701BF1D      1  Constellations  In Between Dreams  Jack Johnson  2005
4  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SODACBL12A8C13C273      1  Learn To Fly  There Is Nothing Left To Lose  Foo Fighters  1999
```

```
In [ ]: print(len(song_df_1), len(song_df_2))
2000000 1000000
```

```
In [ ]: len(song_df)
```

```
Out[ ]: 2000000
```

Data Preprocessing

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

```
Out[ ]:
           user_id  song_id  listen_count  title  release  artist_name  year  song
0  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOAKIMP12A8C130995      1  The Cove  Thicker Than Water  Jack Johnson  0  The Cove - Jack Johnson
1  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBBMDR12A8C13253B      2  Entre Dos Aguas  Flamenco Para Niños  Paco De Lucia  1976  Entre Dos Aguas - Paco De Lucia
2  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBXHDL12A81C204C0      1  Stronger  Graduation  Kanye West  2007  Stronger - Kanye West
3  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SOBYHAJ12A6701BF1D      1  Constellations  In Between Dreams  Jack Johnson  2005  Constellations - Jack Johnson
4  b80344d063b5ccb3212f76538f3d9e43d87dca9e  SODACBL12A8C13C273      1  Learn To Fly  There Is Nothing Left To Lose  Foo Fighters  1999  Learn To Fly - Foo Fighters
```

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

```
In [ ]: # cumulative sum of listen count of the songs
song_grouped = song_df.groupby(['song']).agg(['listen_count': 'count']).reset_index()
song_grouped.head()
```

```
Out[ ]:
           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 [ ]: 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[ ]:
           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 + Th...      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? [Plano Y Voz] - Alejandro Sanz      1      0.01
5150  Época - Gotan Project      1      0.01
```

5151 rows × 3 columns

Popularity Recommendation Engine

```
In [ ]: pr = Recommenders.popularity_recommender_py()
```

```
In [ ]: pr.create(song_df, 'user_id', 'song')
```

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

```
Out[ ]:
           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 + Th...      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 - Chattraxx Karaoke      24      9.0
1862  b80344d063b5ccb3212f76538f3d9e43d87dca9e  Horn Concerto No. 4 in E flat K495: II. Romanc...      23      10.0
```

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

```
Out[ ]:
           user_id  song  score  Rank
3660  e006b1a48466bf9feefed32bec6494495a4436  Sehr kosmisch - Harmonia      45      1.0
4678  e006b1a48466bf9feefed32bec6494495a4436  Undo - Björk      32      2.0
5105  e006b1a48466bf9feefed32bec6494495a4436  You're The One - Dwight Yoakam      32      3.0
1071  e006b1a48466bf9feefed32bec6494495a4436  Dog Days Are Over (Radio Edit) - Florence + Th...      28      4.0
3655  e006b1a48466bf9feefed32bec6494495a4436  Secrets - OneRepublic      28      5.0
4378  e006b1a48466bf9feefed32bec6494495a4436  The Scientist - Coldplay      27      6.0
4712  e006b1a48466bf9feefed32bec6494495a4436  Use Somebody - Kings Of Leon      27      7.0
3476  e006b1a48466bf9feefed32bec6494495a4436  Revelry - Kings Of Leon      26      8.0
1387  e006b1a48466bf9feefed32bec6494495a4436  Fireflies - Chattraxx Karaoke      24      9.0
1862  e006b1a48466bf9feefed32bec6494495a4436  Horn Concerto No. 4 in E flat K495: II. Romanc...      23      10.0
```

Item Similarity Recommendation

```
In [ ]: ir = Recommenders.item_similarity_recommender_py()
ir.create(song_df, 'user_id', 'song')
```

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

```
In [ ]: # 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
Apesta 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 Adams
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 [ ]: # 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 cooccurrence_matrix : 6844

```
Out[ ]:
           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 [ ]: # 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 cooccurrence_matrix : 75

```
Out[ ]:
           user_id  song  score  rank
0  Quiet Houses - Fleet Foxes  0.75      1
1  St. Elsewhere - Dave Grusin  0.75      2
2  Misled - Céline Dion  0.75      3
3  Your Protector - Fleet Foxes  0.75      4
4  Oil And Water - Incubus  0.75      5
5  Tiger Mountain Peasant Song - Fleet Foxes  0.75      6
6  Meadowlarks - Fleet Foxes  0.75      7
7  Sun It Rises - Fleet Foxes  0.75      8
8  I'd Die Without You - P.M. Dawn  0.75      9
```

