

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

```
In [27]: df = pd.read_csv("songdata.csv")
```

```
In [28]: df.head(5)
```

```
Out[28]:
```

	artist	song	link	text
0	ABBA	Ahe's My Kind Of Girl	/a/abba/ahes+my+kind+of+girl_20598417.html	Look at her face, it's a wonderful face \nAnd...
1	ABBA	Andante, Andante	/a/abba/andante+andante_20002708.html	Take it easy with me, please \nTouch me gentl...
2	ABBA	As Good As New	/a/abba/as+good+as+new_20003033.html	I'll never know why I had to go \nWhy I had t...
3	ABBA	Bang	/a/abba/bang_20598415.html	Making somebody happy is a question of give an...
4	ABBA	Bang-A-Boomerang	/a/abba/bang+a+boomerang_20002668.html	Making somebody happy is a question of give an...

```
In [29]: df.tail(5)
```

```
Out[29]:
```

	artist	song	link	text
57645	Ziggy Marley	Good Old Days	/z/ziggy+marley/good+old+days_10198588.html	Irie days come on play \nLet the angels fly l...
57646	Ziggy Marley	Hand To Mouth	/z/ziggy+marley/hand+to+mouth_20531167.html	Power to the workers \nMore power \nPower to...
57647	Zwan	Come With Me	/z/zwan/come+with+me_20148981.html	all you need \nis something i'll believe \nf...
57648	Zwan	Desire	/z/zwan/desire_20148986.html	northern star \nam i frightened \nwhere can ...
57649	Zwan	Heartsong	/z/zwan/heartsong_20148991.html	come in \nmake yourself at home \ni'm a bit ...

```
In [31]: df.shape
```

```
Out[31]: (57650, 4)
```

```
In [32]: df.isnull().sum()
```

```
Out[32]: artist    0
         song      0
         link      0
         text      0
         dtype: int64
```

```
In [33]: df =df.sample(5000).drop('link', axis=1).reset_index(drop=True)
```

```
In [34]: df
```

```
Out[34]:
```

	artist	song	text
0	Leo Sayer	Don't Say It's Over	They say that we could be in love \nThey say ...
1	Elton John	Boogie Pilgrim	Crime in the streets \nl read about it everyd...
2	Norah Jones	4 Broken Hearts	Ooh, ooh ooh \n \nPeople can't be hand-made ...
3	Chris Brown	From My Heart	Verse 1 \n \nLove jus hear me out for a seco...
4	Vince Gill	What The Cowgirls Do	Have you ever been down to Texas \nDown aroun...
...	...	...	...
4995	Lata Mangeshkar	Dil Ne Phir Yaad Kiya Bewafa Laut Bhi Aa	Dil ne phir yaad kiya bewafa laut bhi aa - 2 ...
4996	Dusty Springfield	Love Me By Name	I tried so hard to be carefree \nRolling out ...
4997	Glee	Dream On	Every time that I look in the mirror \nAll th...
4998	Emmylou Harris	Heartbreak Hill	Long long time ago when we were young \nYou s...
4999	Backstreet Boys	Yes, I Will	I open my eyes \nl see your face \nl cannot ...

5000 rows × 3 columns

```
In [35]: df.tail(10)
```

Out[35]:

	artist	song	text
4990	Iggy Pop	Sickness	Hey little girl \nYou're black and blue \nHe...
4991	Bob Dylan	Beyond The Horizon	Beyond the horizon, behind the sun \nAt the e...
4992	Nicki Minaj	Get On Your Knees	Get on your knees, get on your knees, get on y...
4993	Kiss	I	I was so frightened \nl almost ran away \nl ...
4994	Sting	Dienda	How like the fall \nTo be gone in a day \nJu...
4995	Lata Mangeshkar	Dil Ne Phir Yaad Kiya Bewafa Laut Bhi Aa	Dil ne phir yaad kiya bewafa laut bhi aa - 2 ...
4996	Dusty Springfield	Love Me By Name	I tried so hard to be carefree \nRolling out ...
4997	Glee	Dream On	Every time that I look in the mirror \nAll th...
4998	Emmylou Harris	Heartbreak Hill	Long long time ago when we were young \nYou s...
4999	Backstreet Boys	Yes, I Will	I open my eyes \nl see your face \nl cannot ...

In [36]: `df['text'][0]`

Out[36]: "They say that we could be in love \nThey say that we could be \nThey say we'll fall \nLike others will fall \nDo n't tell me that it's over \nYou know you'll break my heart \nDon't tell me that it's over \nYou'll see me fall apar t \nDon't tell that it's over \nOr else I, well I might fall apart \nNow, when we first met \nI was just a poor boy on the street \nAnd you looked at me and took me in \nYeah, and knocked me off my feet \nNow you don't really want t o break things up \nThe things we've been through, boy they're enough \nDon't tell me that it's over \nDon't say tha t it's over \nYou know you'll break my heart \nDon't tell me that it's over \nYou'll see me fall apart \nDon't tell that it's over \nOr else I, well I might fall apart \nThey say that we could be in love \nThey say that we could be \n\n"

In [38]: `#df = df.sample(5000)`In [39]: `df.shape`Out[39]: `(5000, 3)`In [43]: `df['text'] = df['text'].str.lower().replace(r'^\w\s', ' ').replace(r'\n', ' ', regex = True)`

In [44]: `import nltk`  
`from nltk.stem.porter import PorterStemmer`

```
In [45]: stemmer = PorterStemmer()
```

```
In [47]: def token(txt):
          token = nltk.word_tokenize(txt)
          a = [stemmer.stem(w) for w in token]
          return " ".join(a)
```

```
In [48]: df['text'].apply(lambda x: token(x))
```

```
Out[48]: 0      they say that we could be in love they say tha...
          1      crime in the street i read about it everyday i...
          2      ooh , ooh ooh peopl ca n't be hand-mad but he ...
          3      vers 1 love ju hear me out for a second mind i...
          4      have you ever been down to texa down around sa...

          ...
          4995     dil ne phir yaad kiya bewafa laut bhi aa - 2 b...
          4996     i tri so hard to be carefre roll out of someon...
          4997     everi time that i look in the mirror all these...
          4998     long long time ago when we were young you swor...
          4999     i open my eye i see your face i can not hide ,...
          Name: text, Length: 5000, dtype: object
```

```
In [49]: from sklearn.feature_extraction.text import TfidfVectorizer
          from sklearn.metrics.pairwise import cosine_similarity
```

```
In [51]: tfidf = TfidfVectorizer(analyzer='word',stop_words='english')
```

```
In [53]: matrix = tfidf.fit_transform(df['text'])
```

```
In [55]: similarity = cosine_similarity(matrix)
```

```
In [56]: similarity[0]
```

```
Out[56]: array([1.          , 0.01725179, 0.07663282, ..., 0.01647249, 0.00969559,
          0.03922488])
```

```
In [59]: df[df['song'] == 'Heartbreak Hill'].index[0]
```

```
Out[59]: 4998
```

```
In [60]: def recommender(song_df):
          idx = df[df['song'] == song_df].index[0]
```

```
distances = sorted(list(enumerate(similarity[idx])),reverse=True,key=lambda x:x[1])

songs = []
for m_id in distances[1:20]:
    songs.append(df.iloc[m_id[0]].song)

return songs
```

```
In [61]: recommender('Heartbreak Hill')
```

```
Out[61]: ['A Room At The Heartbreak Hotel',
          'Red Hill Mining Town',
          'Heartbreak Express',
          'Heartbreak Hotel',
          'Heartbreak Hotel',
          'Blue Prelude',
          'Lonely Is The Night',
          'Running Up That Hill',
          'In The Still Of The Night',
          'Night Lights',
          'A Mansion On The Hill',
          'Look At The Sky',
          'All Night Long',
          'The Side Of A Hill',
          'Dream The Night Away',
          'Fools',
          'Lonely Love',
          'Lonely Days',
          'Oh, What A Night']
```

```
In [62]: import pickle
         pickle.dump(similarity,open('similarity.pkl','wb'))
         pickle.dump(df,open('df.pkl','wb'))
```

```
In [ ]:
```