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

In [2]:
```
df = pd.read_csv("spotify_millsongdata.csv")
```

In [3]:
```
df.head(5)
```

Out[3]:

	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 \r\nA...
1	ABBA	Andante, Andante	/a/abba/andante+andante_20002708.html	Take it easy with me, please \r\nTouch me gen...
2	ABBA	As Good As New	/a/abba/as+good+as+new_20003033.html	I'll never know why I had to go \r\nWhy I had...
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 [4]:
```
df.tail(5)
```

Out[4]:

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

In [5]:
```
df.shape
```

Out[5]:
```
(57650, 4)
```

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
artist    0
song      0
link      0
text      0
dtype: int64
```

In [7]:
```
df =df.sample(5000).drop('link', axis=1).reset_index(drop=True)
```

In [8]:
```
df.head(10)
```

Out[8]:

	artist	song	text
0	Iron Maiden	Rainmaker	When I was wandering in the desert \r\nAnd wa...
1	Dave Matthews Band	Belly Belly	Think about what you want \r\nThink about wha...
2	Everclear	Heroin Girl	I used to know a girl \r\nShe had two pierced...
3	Robbie Williams	Blue Swde Shoes	You're the Mona Lisa, \r\nYou're The ravin Pa...
4	Billie Holiday	I'm Yours	Ask the sky above \r\nAnd ask the earth below...
5	Snoop Dogg	Give It 2 'em Dogg	Give it up, give it up nigga, you know what's ...
6	Dewa 19	Arjuna	[Verse 1] \r\nSudah kudaki gunung tertinggi ...
7	Counting Crows	Perfect Blue Buildings	Just down the street from your hotel, baby \r...
8	Nat King Cole	If You Said No	The stars would cry in the blue if you said no...
9	Utopia	The Seven Rays	My great grandfather was a satisfied man \r\n...

In [9]:
```
df['text'][0]
```

Out[9]:
```
'When I was wandering in the desert  \r\nAnd was searching for the truth  \r\nI he
ard a choir of angels calling out my name  \r\nI had the feeling that my life woul
d never be the same again  \r\nI turned my face towards the barren sun  \r\n  \r\n
And I know of the pain that you feel the same as me  \r\nAnd I dream of the rain a
s it falls upon the leaves  \r\nAnd the cracks in our lives like the cracks upon t
he ground  \r\nThey are sealed and are now washed away  \r\n  \r\n[Chorus:]  \r\nY
ou tell me we can start the rain  \r\nYou tell me that we all can change  \r\nYou
tell me we can find something to wash the tears away  \r\nYou tell me we can start
the rain  \r\nYou tell me that we all can change  \r\nYou tell me we can find some
thing to wash the tears  \r\n  \r\nAnd I know of the pain that you feel the same a
s me  \r\nAnd I dream of the rain as it falls upon the leaves  \r\nAnd the cracks
in the ground like the cracks are in our lives  \r\nThey are sealed and now far aw
ay  \r\n  \r\n[Chorus]  \r\n  \r\nAnd I know of the pain that you feel the same as
me  \r\nAnd I dream of the rain as it falls upon the leaves  \r\nAnd the cracks in
the ground like the cracks are in our lives  \r\nThey are sealed and now far away
\r\n\r\n'
```

In [10]:
```
# df = df.sample(5000)
```

In [11]:
```
df.shape
```

Out[11]:
```
(5000, 3)
```

Text Cleaning/ Text Preprocessing

In [12]:
```
df['text'] = df['text'].str.lower().replace(r'^\w\s', ' ').replace(r'\n', ' ', rege
```

In [13]:
```
import nltk
from nltk.stem.porter import PorterStemmer
```

```
stemmer = PorterStemmer()

def tokenization(txt):
    tokens = nltk.word_tokenize(txt)
    stemming = [stemmer.stem(w) for w in tokens]
    return " ".join(stemming)
```

In [14]:
```
df['text'] = df['text'].apply(lambda x: tokenization(x))
```

In [15]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

In [16]:
```
tfidvector = TfidfVectorizer(analyzer='word',stop_words='english')
matrix = tfidvector.fit_transform(df['text'])
similarity = cosine_similarity(matrix)
```

In [17]:
```
similarity[0]
```

Out[17]:
```
array([1.        , 0.02415267, 0.03890341, ..., 0.03560189, 0.01148531,
       0.08239572])
```

In [18]:
```
df[df['song'] == 'Crying Over You']
```

Out[18]:

	artist	song	text

In [19]:
```
def recommendation(song_df):
    idx = df[df['song'] == song_df].index[0]
    distances = sorted(list(enumerate(similarity[idx])),reverse=True,key=lambda x:x

    songs = []
    for m_id in distances[1:21]:
        songs.append(df.iloc[m_id[0]].song)

    return songs
```

In [21]:
```
recommendation("I'm Yours")
```

Out[21]:
```
["I'm Yours",
 'Love To Love',
 "High Steppin' Proud",
 'Tomorrow',
 'Come Tomorrow',
 'Only Love',
 'Eternally',
 'Tomorrow',
 'Feet On The Ground',
 'Tomorrow',
 'For The Love Of A Woman',
 'Alone Tonight',
 'Borrowed Time',
 'I Have Dreamed',
 'More Than I Can Say',
 'Make Tomorrow',
 'Help Me Make It Through The Night',
 'Feels Like Heaven',
 'Why Do I Love You?',
 'My Kind Of Lady']
```

In [22]:
```
import pickle
pickle.dump(similarity,open('similarity.pkl','wb'))
pickle.dump(df,open('df.pkl','wb'))
```

In [ ]:
```

```