Movie Recommendation System

In [5]:
```python
# Importing required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
for dirname, _, filenames in os.walk('/dataset'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
# Exporting dataset
movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
movies.head()
```

Out[5]:

	budget	genres	homepage	id	keywords	original_lan
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	

In [6]:
```
movies.shape
```
Out[6]:
```
(4803, 20)
```

In [7]:
```
credits.head()
```
Out[7]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [8]:
```
credits.shape
```
Out[8]:
```
(4803, 4)
```

In [9]:
```
# Merging two tables (movies and credits tables)
movies = movies.merge(credits,on='title')
```

In [10]:
```
movies.head()
```

Out[10]:

	budget	genres	homepage	id	keywords	original_lan
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	

5 rows × 23 columns

In [11]:
```
movies.shape
```
Out[11]:
```
(4809, 23)
```

In [12]:
```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4809 entries, 0 to 4808
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   budget                4809 non-null   int64  
 1   genres                4809 non-null   object 
 2   homepage              1713 non-null   object 
 3   id                    4809 non-null   int64  
 4   keywords              4809 non-null   object 
 5   original_language     4809 non-null   object 
 6   original_title        4809 non-null   object 
 7   overview              4806 non-null   object 
 8   popularity            4809 non-null   float64
 9   production_companies  4809 non-null   object 
 10  production_countries  4809 non-null   object 
 11  release_date          4808 non-null   object 
 12  revenue               4809 non-null   int64  
 13  runtime               4807 non-null   float64
 14  spoken_languages      4809 non-null   object 
 15  status                4809 non-null   object 
 16  tagline               3965 non-null   object 
 17  title                 4809 non-null   object 
 18  vote_average          4809 non-null   float64
 19  vote_count            4809 non-null   int64  
 20  movie_id              4809 non-null   int64  
 21  cast                  4809 non-null   object 
 22  crew                  4809 non-null   object 
dtypes: float64(3), int64(5), object(15)
memory usage: 901.7+ KB
```

In [13]:
```
# Remove unnecessary columns "budget,homepage,id,original_language,original_title,popu
# Accesing required and important columns are('movie_id','title','overview','genres',
movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

In [14]:
```
movies.head()
```

Out[14]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 1463, "name": "culture clash"}, {"id"...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	[{"id": 818, "name": "based on novel"}, {"id"...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [15]:
```
# Finding NULL values
movies.isna().sum()
```

Out[15]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [16]:
```
# Removing NULL values
movies.dropna(inplace=True)
```

In [17]:
```
# This Library perform a collection of nodes which are linked together based on the gr
import ast
```

In [18]:
```
def convert(txt):
    L = []
    for i in ast.literal_eval(txt):
        L.append(i['name'])
    return L
```

In [19]:
```
# Accesing only genres 'name' from the genres table
movies['genres'] = movies['genres'].apply(convert)
movies.head()
```

Out[19]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id":...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id":...	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [20]:
```
# Accesing only keywords 'name' from the keywords table
movies['keywords'] = movies['keywords'].apply(convert)
movies.head()
```

Out[20]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[{"cast_id": 5, "character": "John Carter", "r...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [21]:
```
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"ic
```

Out[21]:
```
[{'id': 28, 'name': 'Action'},
 {'id': 12, 'name': 'Adventure'},
 {'id': 14, 'name': 'Fantasy'},
 {'id': 878, 'name': 'Science Fiction'}]
```

In [22]:
```
def convert3(text):
    L = []
    counter = 0
    for i in ast.literal_eval(text):
        if counter < 3:
            L.append(i['name'])
        counter+=1
    return L
```

In [23]:
```
movies['cast']
```

Out[23]:
```
0       [{"cast_id": 242, "character": "Jake Sully", "...
1       [{"cast_id": 4, "character": "Captain Jack Spa...
2       [{"cast_id": 1, "character": "James Bond", "cr...
3       [{"cast_id": 2, "character": "Bruce Wayne / Ba...
4       [{"cast_id": 5, "character": "John Carter", "c...
                              ...
4804    [{"cast_id": 1, "character": "El Mariachi", "c...
4805    [{"cast_id": 1, "character": "Buzzy", "credit_...
4806    [{"cast_id": 8, "character": "Oliver O\u2019To...
4807    [{"cast_id": 3, "character": "Sam", "credit_id...
4808    [{"cast_id": 3, "character": "Herself", "credi...
Name: cast, Length: 4806, dtype: object
```

In [24]:
```
# Converting  top 3 list of dictionary into just list
movies['cast'] = movies['cast'].apply(convert)
movies.head()
```

Out[24]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman, A...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [25]:
```
movies['cast']
```

Out[25]:
```
0       [Sam Worthington, Zoe Saldana, Sigourney Weave...
1       [Johnny Depp, Orlando Bloom, Keira Knightley, ...
2       [Daniel Craig, Christoph Waltz, Léa Seydoux, R...
3       [Christian Bale, Michael Caine, Gary Oldman, A...
4       [Taylor Kitsch, Lynn Collins, Samantha Morton,...
                              ...
4804    [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...
4805    [Edward Burns, Kerry Bishé, Marsha Dietlein, C...
4806    [Eric Mabius, Kristin Booth, Crystal Lowe, Geo...
4807    [Daniel Henney, Eliza Coupe, Bill Paxton, Alan...
4808    [Drew Barrymore, Brian Herzlinger, Corey Feldm...
Name: cast, Length: 4806, dtype: object
```

In [27]:
```
# Accesing  top 3 actor name from the cast table we dont require lots of actor name
movies['cast'] = movies['cast'].apply(lambda x:x[0:3])
movies['cast']
```

Out[27]:
```
0         [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1            [Johnny Depp, Orlando Bloom, Keira Knightley]
2             [Daniel Craig, Christoph Waltz, Léa Seydoux]
3             [Christian Bale, Michael Caine, Gary Oldman]
4           [Taylor Kitsch, Lynn Collins, Samantha Morton]
                              ...
4804    [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...
4805          [Edward Burns, Kerry Bishé, Marsha Dietlein]
4806            [Eric Mabius, Kristin Booth, Crystal Lowe]
4807             [Daniel Henney, Eliza Coupe, Bill Paxton]
4808    [Drew Barrymore, Brian Herzlinger, Corey Feldman]
Name: cast, Length: 4806, dtype: object
```

In [28]:
```
# This function fetch the director name from the crew feature
def fetch_director(text):
    L = []
    for i in ast.literal_eval(text):
        if i['job'] == 'Director':
            L.append(i['name'])
    return L
```

In [29]:
```
# Accessing director name from crew column
movies['crew'] = movies['crew'].apply(fetch_director)
movies.sample(5)
```

Out[29]:

	movie_id	title	overview	genres	keywords	cast	crew
[illegible]	11	Star Wars	Princess Leia is captured and held hostage by ...	[Adventure, Action, Science Fiction]	[android, galaxy, hermit, death star, lightsab...	[Mark Hamill, Harrison Ford, Carrie Fisher]	[George Lucas]
2766	264644	Room	Jack is a young boy of 5 years old who has liv...	[Drama, Thriller]	[based on novel, carpet, isolation, kidnapping...	[Brie Larson, Jacob Tremblay, Joan Allen]	[Lenny Abrahamson]
[illegible]	11835	Death Sentence	Nick Hume is a mild-mannered executive with a ...	[Action, Crime, Drama, Thriller]	[loss of son, repayment, revenge, murder, gang...	[Kevin Bacon, Garrett Hedlund, Kelly Preston]	[James Wan]
[illegible]	[illegible]	No Good Deed	Terri is a devoted wife and mother of two, liv...	[Crime, Thriller]	[sadistic, hostage, sociopath, serial killer, ...	[Idris Elba, Taraji P. Henson, Leslie Bibb]	[Sam Miller]
[illegible]	[illegible]	The Banger Sisters	In the late '60s, the self-proclaimed Ballas a...	[Comedy, Drama]	[tattoo, rock star, groupie, past, family]	[Goldie Hawn, Susan Sarandon, Geoffrey Rush]	[Bob Dolman]

In [30]:
```
# This functions removes the whitespace from the name because we want name unique some
# name sam biilings and sam johny but we focusing only first name but our ML Model wil
# removing white spaces from features
def collapse(L):
    L1 = []
    for i in L:
        L1.append(i.replace(" ",""))
    return L1
```

In [31]:
```
movies['cast'].head()
```

Out[31]:
```
0    [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1       [Johnny Depp, Orlando Bloom, Keira Knightley]
2        [Daniel Craig, Christoph Waltz, Léa Seydoux]
3        [Christian Bale, Michael Caine, Gary Oldman]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]
Name: cast, dtype: object
```

In [33]:
```
# Applying whitespace remove function to cast-->'Actor name'
movies['cast'] = movies['cast'].apply(collapse)
```

In [34]:
```
movies['cast']
```

Out[34]:
```
0         [SamWorthington, ZoeSaldana, SigourneyWeaver]
1            [JohnnyDepp, OrlandoBloom, KeiraKnightley]
2             [DanielCraig, ChristophWaltz, LéaSeydoux]
3             [ChristianBale, MichaelCaine, GaryOldman]
4           [TaylorKitsch, LynnCollins, SamanthaMorton]
                             ...
4804    [CarlosGallardo, JaimedeHoyos, PeterMarquardt]
4805         [EdwardBurns, KerryBishé, MarshaDietlein]
4806           [EricMabius, KristinBooth, CrystalLowe]
4807            [DanielHenney, ElizaCoupe, BillPaxton]
4808    [DrewBarrymore, BrianHerzlinger, CoreyFeldman]
Name: cast, Length: 4806, dtype: object
```

In [35]:
```
movies['crew'].head()
```

Out[35]:
```
0         [James Cameron]
1        [Gore Verbinski]
2            [Sam Mendes]
3     [Christopher Nolan]
4        [Andrew Stanton]
Name: crew, dtype: object
```

In [36]:
```
# Applying whitespace remove function to crew tabele-->'director name'
movies['crew'] = movies['crew'].apply(collapse)
```

In [37]:
```
movies['crew'].head()
```

Out[37]:
```
0         [JamesCameron]
1        [GoreVerbinski]
2            [SamMendes]
3     [ChristopherNolan]
4        [AndrewStanton]
Name: crew, dtype: object
```

In [38]:
```
# Applying whitespace remove function to genres ---> type of movies
movies['genres'] = movies['genres'].apply(collapse)

# Applying whitespace remove function to keywords
movies['keywords'] = movies['keywords'].apply(collapse)
```

In [39]:
```
movies.head()
```

Out[39]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]
2	206647	Spectre	A cryptic message from Bond’s past sends him o...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]

In [40]:
```python
# Splitting the each word in overview columns
movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

In [41]:
```python
# Adding overview,genres,keywords,cast and crew added into a one table --> tag table b
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['
```

In [42]:
```python
# Removing the columns overview,genres,keywords,cast, crew
new = movies.drop(columns=['overview','genres','keywords','cast','crew'])
```

In [43]:
```python
new['tags']
```

Out[43]:
```
0       [In, the, 22nd, century,, a, paraplegic, Marin...
1       [Captain, Barbossa,, long, believed, to, be, d...
2       [A, cryptic, message, from, Bond’s, past, send...
3       [Following, the, death, of, District, Attorney...
4       [John, Carter, is, a, war-weary,, former, mili...
                              ...                        
4804    [El, Mariachi, just, wants, to, play, his, gui...
4805    [A, newlywed, couple's, honeymoon, is, upended...
4806    ["Signed,, Sealed,, Delivered", introduces, a,...
4807    [When, ambitious, New, York, attorney, Sam, is...
4808    [Ever, since, the, second, grade, when, he, fi...
Name: tags, Length: 4806, dtype: object
```

In [44]:
```python
new['tags'] = new['tags'].apply(lambda x: " ".join(x))
new.head()
```

Out[44]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...
2	206647	Spectre	A cryptic message from Bond’s past sends him o...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...
4	49529	John Carter	John Carter is a war-weary, former military ca...

In [45]:
```python
new['tags']
```

Out[45]:
```
0       In the 22nd century, a paraplegic Marine is di...
1       Captain Barbossa, long believed to be dead, ha...
2       A cryptic message from Bond’s past sends him o...
3       Following the death of District Attorney Harve...
4       John Carter is a war-weary, former military ca...
                              ...                        
4804    El Mariachi just wants to play his guitar and ...
4805    A newlywed couple's honeymoon is upended by th...
4806    "Signed, Sealed, Delivered" introduces a dedic...
4807    When ambitious New York attorney Sam is sent t...
4808    Ever since the second grade when he first saw ...
Name: tags, Length: 4806, dtype: object
```

In [46]:
```python
## Vectorization
# Importing Vectorization library
from sklearn.feature_extraction.text import CountVectorizer

# Vectorizing data into 5000 features and applying stop words it removes the unccessar
cv = CountVectorizer(max_features=5000,stop_words='english')
```

In [47]:
```python
# Fit the tag table data
vector = cv.fit_transform(new['tags']).toarray()
```

In [48]:
```python
vector.shape
```

Out[48]:
```
(4806, 5000)
```

In [49]:
```
# Importing cosine similarity
from sklearn.metrics.pairwise import cosine_similarity
```

In [50]:
```
# cosine similarity ---> this perform like it get the words which is similar like (lov
similarity = cosine_similarity(vector)
```

In [51]:
```
# Accesing index value from movie title column
new[new['title'] == 'The Lego Movie'].index[0]
```

Out[51]:
```
744
```

In [52]:
```
# In this function we first add the movie title
def recommend(movie):
    index = new[new['title'] == movie].index[0]
    # Getting similarity movies in ascending order
    distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x:
    # Just picking top 5 movie which have more similarity to the movie by through inde
    for i in distances[1:6]:
        print(new.iloc[i[0]].title)
```

Recommendation Movies (Test)

In [53]:
```
# Recommending the movie 1
recommend('Gandhi')
```

```
Gandhi, My Father
The Wind That Shakes the Barley
A Passage to India
Guiana 1838
Ramanujan
```

In [54]:
```
# Recommending the movie 2
recommend('The Lego Movie')
```

```
The Adventures of Rocky & Bullwinkle
The Boxtrolls
Curious George
Percy Jackson: Sea of Monsters
Penguins of Madagascar
```

In [55]:
```
# Recommending the movie 3
recommend('Lucy')
```

```
The Helix... Loaded
Species
Jupiter Ascending
Phenomenon
Godzilla 2000
```

In [56]:
```
# Recommending the movie 4
recommend('Madagascar')
```

```
Roar
Madagascar: Escape 2 Africa
The Wild Thornberrys Movie
Rugrats Go Wild
Nicholas Nickleby
```

In [57]:
```
# Recommending the movie 5
recommend('Gossip')
```

```
The Lunchbox
Griff the Invisible
Rocket Singh: Salesman of the Year
Far from Heaven
The Bridges of Madison County
```

In [ ]: