

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

```
In [2]: movies = pd.read_csv('tmdb_5000_movies.csv')
credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [3]: # view top 5 data in movies and credits
movies.head()
```

Out[3]:	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity
0	237000000	{{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	{{"id": 1463, "name": "culture clash"}, {"id": "...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577
1	300000000	{{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	{{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbosa, long believed to be dead, ha...	139.082615
2	245000000	{{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	{{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788
3	250000000	{{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	{{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950
4	260000000	{{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	{{"id": 818, "name": "based on novel"}, {"id": "...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995

```
In [4]: #merge the two different cvs dataset
movies = movies.merge(credits,on='title')
```

```
In [5]: movies.head(1)
```

```
Out[5]:
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_compan
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "marine"}]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "Ingeni Film Partners", "id": 28}]

1 rows × 23 columns

```
In [6]: # genres
# id
#keyword
#title
#overview
#cast
#crew

movies = movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

```
In [7]: #movies.shape
#movies['original_language'].value_counts()
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Int64Index: 4809 entries, 0 to 4808
```

Data columns (total 7 columns):

#	Column	Non-Null	Count	Dtype
0	movie_id	4809	non-null	int64
1	title	4809	non-null	object
2	overview	4806	non-null	object
3	genres	4809	non-null	object
4	keywords	4809	non-null	object
5	cast	4809	non-null	object
6	crew	4809	non-null	object

```
dtypes: int64(1), object(6)
```

```
memory usage: 300.6+ KB
```

```
movies.head()
```

Out[8]:	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...

```
# find the missing values
movies.isnull().sum()
```

```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

```
movies.dropna(inplace=True)
```

```
# find the duplicated in the dataset code is given below
movies.duplicated().sum()
```

0

```
#chick the location any data
movies.iloc[0].genres
```

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

```
import ast
def convert(object):
    list=[]
    for i in ast.literal_eval(object):
        list.append(i['name'])
    return list
```

```
movies['genres']=movies['genres'].apply(convert)
```

In [15]:

movies.head()

Out[15]:

	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 [16]:

movies['keywords']=movies['keywords'].apply(convert)

In [17]:

movies.head()

Out[17]:

	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", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

In [18]:

```
import ast
def convert3(obj):
    list=[]
    count=0
    for i in ast.literal_eval(obj):
        if count != 3:
            list.append(i['name'])
            count+=1
        else:
            break
    return list
```

In [19]:

movies['cast'] = movies['cast'].apply(convert3)

In [20]:

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...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[{"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]	[{"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]	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
			John Carter is a war-		[based on novel,	[Taylor Kitsch, Lynn	[{"credit_id":

4	49529	John Carter	weary, former military ca...	[Action, Adventure, Science Fiction]	mars, medallion, space travel...	Collins, Samantha Morton]	"52fe479ac3a36847f813eaa3", "de...
---	-------	-------------	------------------------------	--------------------------------------	----------------------------------	---------------------------	------------------------------------

```
In [21]: def director(text):
          L = []
          for i in ast.literal_eval(text):
              if i['job'] == 'Director':
                  L.append(i['name'])
          return L
```

```
In [22]: movies['crew'] = movies['crew'].apply(director)
```

```
In [23]: movies.head()
```

Out[23]:	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 Weaver]	[James Cameron]
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]	[Gore Verbinski]
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]	[Sam Mendes]
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]	[Christopher Nolan]
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]	[Andrew Stanton]

```
In [24]: movies['overview'][0]
```

```
Out[24]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization.'
```

```
In [25]: movies['overview']=movies['overview'].apply(lambda x:x.split())
```

```
In [26]: movies.head()
```

Out[26]:	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

```
In [27]: movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])
          movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ", "") for i in x])
          movies['crew']=movies['crew'].apply(lambda x:[i.replace(" ", "") for i in x])
          movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ", "") for i in x])
```

```
In [28]: movies.head()
```

Out[28]:	movie_id	title	overview	genres	keywords	cast	crew
				[Action, Adventure,		[SamWorthington,	

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

In [29]:

```
movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

In [30]:

```
movies.head()
```

Out[30]:

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

In [31]:

```
new_df=movies[['movie_id','title','tags']]
```

In [32]:

```
new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

C:\Users\KIRANS~1\AppData\Local\Temp\ipykernel_11604\487797088.py:1: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

In [33]:

```
new_df.head()
```

Out[33]:

	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 [34]: import nltk
```

```
In [35]: from nltk.stem.porter import PorterStemmer  
ps=PorterStemmer()
```

```
In [36]: def stem (text):  
    y=[]  
    for i in text.split():  
        y.append( ps.stem(i))  
    return " ".join(y)
```

```
In [37]: new_df['tags']=new_df['tags'].apply(stem)
```

C:\Users\KIRANS~1\AppData\Local\Temp\ipykernel_11604\3514595201.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
new_df['tags']=new_df['tags'].apply(stem)

```
In [38]: new_df['tags'][0]
```

```
Out[38]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn between  
follow order and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar s  
pacecoloni societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair a  
ntiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'
```

```
In [39]: new_df['tags'][1]
```

```
Out[39]: "captain barbossa, long believ to be dead, ha come back to life and is head to the edg of the earth with will tur  
ner and elizabeth swann. but noth is quit as it seems. adventur fantasi action ocean drugabus exoticisland eastin  
diatradingcompani loveofone'slif traitor shipwreck strongwoman ship allianc calypso afterlif fighter pirat swashb  
uckl aftercreditssting johnnydepp orlandobloom keiraknightley goreverbinski"
```

```
In [40]: new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
```

C:\Users\KIRANS~1\AppData\Local\Temp\ipykernel_11604\4224080999.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
new_df['tags']=new_df['tags'].apply(lambda x:x.lower())

```
In [41]: new_df.head()
```

```
Out[41]:
```

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a parapleg marin is dispa...
1	285	Pirates of the Caribbean: At World's End	captain barbossa, long believ to be dead, ha c...
2	206647	Spectre	a cryptic messag from bond' past send him on a...
3	49026	The Dark Knight Rises	follow the death of district attorney harvey d...
4	49529	John Carter	john carter is a war-weary, former militari ca...

```
In [42]: from sklearn.feature_extraction.text import CountVectorizer  
cv= CountVectorizer(max_features=5000,stop_words='english')
```

```
In [43]: vectors=cv.fit_transform(new_df['tags']).toarray()
```

```
In [44]: vectors
```

```
Out[44]: array([[0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               ...,
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0],
               [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

```
In [45]: vectors[0]
```

```
Out[45]: array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
```

```
In [46]: cv.get_feature_names()
```

```
C:\Users\kiransindam\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use get_feature_names_out instead.
  warnings.warn(msg, category=FutureWarning)
```

```
Out[46]: ['000',
          '007',
          '10',
          '100',
          '11',
          '12',
          '13',
          '14',
          '15',
          '16',
          '17',
          '17th',
          '18',
          '18th',
          '18thcenturi',
          '19',
          '1910',
          '1920',
          '1930',
          '1940',
          '1944',
          '1950',
          '1950s',
          '1960',
          '1960s',
          '1970',
          '1970s',
          '1971',
          '1974',
          '1976',
          '1980',
          '1985',
          '1990',
          '1999',
          '19th',
          '19thcenturi',
          '20',
          '200',
          '2003',
          '2009',
          '20th',
          '21st',
          '23',
          '24',
          '25',
          '30',
          '300',
          '3d',
          '40',
          '50',
          '500',
          '60',
          '70',
          '80',
          'aaron',
          'aaroneckhart',
          'abandon',
          'abduct',
          'abigailbreslin',
          'abil',
```

'abl',
'aboard',
'abov',
'abus',
'academ',
'academi',
'accept',
'access',
'accid',
'accident',
'acclaim',
'accompani',
'accomplish',
'account',
'accus',
'ace',
'achiev',
'acquaint',
'act',
'action',
'actionhero',
'activ',
'activist',
'activities',
'actor',
'actress',
'actual',
'ad',
'adam',
'adamsandl',
'adamshankman',
'adapt',
'add',
'addict',
'adjust',
'admir',
'admit',
'adolesc',
'adopt',
'ador',
'adrienbrodi',
'adult',
'adultanim',
'adulteri',
'adulthood',
'advanc',
'adventur',
'adventure',
'adventures',
'advertis',
'advic',
'advis',
'affair',
'affect',
'afghanistan',
'africa',
'african',
'africanamerican',
'aftercreditssting',
'afterlif',
'aftermath',
'ag',
'age',
'agediffer',
'agenc',
'agency',
'agenda',
'agent',
'agents',
'aggress',
'ago',
'agre',
'ahead',
'aid',
'aidanquinn',
'ail',
'aim',
'air',
'airplan',
'airplanecrash',
'airport',
'aka',
'al',
'alabama',
'alan',
'alaska',
'albert',
'alcatraz',
'alcohol',

'alecbaldwin',
'alex',
'alexxkendrick',
'alfredhitchcock',
'alfredmolina',
'ali',
'alic',
'alice',
'alien',
'alieninvas',
'alienlife',
'alienplanet',
'aliens',
'alik',
'aliv',
'alive',
'allen',
'alli',
'allianc',
'allow',
'alon',
'alongsid',
'alpacino',
'alpha',
'alreadi',
'alter',
'altern',
'alway',
'alyssa',
'alzheimer',
'amanda',
'amandapeet',
'amandaseyfri',
'amateur',
'amaz',
'amazon',
'ambassador',
'ambit',
'ambiti',
'ambul',
'ambush',
'america',
'american',
'americanabroad',
'americancivilwar',
'americanfootbal',
'americanfootballplay',
'amid',
'amidst',
'amnesia',
'amp',
'amsterdam',
'amus',
'amusementpark',
'amy',
'amyadam',
'amysmart',
'ana',
'anakin',
'analyst',
'anarchiccomedi',
'ancient',
'ancientrom',
'ancientworld',
'anderson',
'andi',
'andiemacdowel',
'andrew',
'android',
'andy',
'andygarcía',
'angel',
'angela',
'angelabassett',
'angeles',
'angelinajoli',
'anger',
'angle',
'angri',
'ani',
'anim',
'animalattack',
'animalhorror',
'animals',
'anjelicaheston',
'ann',
'anna',
'annafari',
'annakendrick',

'anne',
'annehathaway',
'annemoss',
'annetteben',
'anni',
'annie',
'anniversari',
'announc',
'annual',
'anonym',
'anoth',
'answer',
'ant',
'antholog',
'anthoni',
'anthonyanderson',
'anthonyhopkin',
'anthonymacki',
'anthropomorph',
'anti',
'antic',
'antihero',
'antiqu',
'antoinefuqua',
'antoniobandera',
'antonyelchin',
'anyon',
'anyth',
'apart',
'apartheid',
'apartment',
'ape',
'apocalyps',
'apocalypse',
'apocalypt',
'appar',
'appear',
'appl',
'apple',
'appoint',
'appreci',
'apprentic',
'approach',
'april',
'aquarium',
'arab',
'arch',
'archaeologist',
'archeolog',
'archer',
'architect',
'arctic',
'area',
'aren',
'arena',
'argument',
'aris',
'aristocrat',
'arm',
'armi',
'armor',
'armsdeal',
'army',
'arnold',
'arnoldschwarzenegg',
'arrang',
'arrangedmarriag',
'arrest',
'arriv',
'arrog',
'art',
'arthur',
'artifact',
'artifici',
'artificialintellig',
'artist',
'ash',
'ashley',
'ashleyjudd',
'ashtonkutch',
'asia',
'asian',
'asid',
'ask',
'aspect',
'aspir',
'assassin',
'assault',
'assembl',

'assign',
'assist',
'assistant',
'associ',
'assum',
'asteroid',
'astronaut',
'asylum',
'atheist',
'athlet',
'atom',
'atomicbomb',
'attack',
'attacks',
'attempt',
'attempttoescap',
'attend',
'attent',
'attic',
'attitud',
'attorney',
'attract',
'auction',
'audienc',
'audit',
'august',
'aunt',
'austin',
'australia',
'australian',
'author',
'autism',
'auto',
'automobilerac',
'aveng',
'averag',
'avoid',
'await',
'awak',
'awaken',
'award',
'away',
'awkward',
'awri',
'awry',
'ax',
'babe',
'babi',
'baby',
'bachelor',
'backdrop',
'background',
'backpack',
'bad',
'bag',
'bahama',
'bail',
'balanc',
'ball',
'ballet',
'balloon',
'ban',
'band',
'bandit',
'bangkok',
'banish',
'bank',
'banker',
'bankrobb',
'bankrobber',
'bar',
'barbrastreisand',
'bare',
'bargain',
'barn',
'barney',
'baron',
'barri',
'barrylevinson',
'barrysonnenfeld',
'bas',
'base',
'basebal',
'basedoncomicbook',
'basedongraphicnovel',
'basedonnovel',
'basedonplay',
'basedonstagemus',
'basedontrueev',

'basedontruestori',
'basedontvseri',
'basedonvideogam',
'basedonyoungadultnovel',
'basement',
'basketbal',
'bat',
'batman',
'battl',
'battle',
'battlefield',
'bay',
'beach',
'beam',
'bear',
'beard',
'beast',
'beat',
'beauti',
'beautiful',
'beautifulwoman',
'beauty',
'becam',
'becaus',
'becki',
'becom',
'becominganadult',
'bed',
'bedroom',
'beer',
'befor',
'befriend',
'began',
'begin',
'begins',
'behavior',
'belief',
'believ',
'bell',
'bella',
'belong',
'belov',
'ben',
'benaffleck',
'bend',
'beneath',
'benefit',
'benfost',
'beniciodeltoro',
'benjamin',
'benjaminbratt',
'benkingsley',
'bennett',
'benstil',
'bent',
'berlin',
'best',
'bestfriend',
'bestfriendsinlov',
'bet',
'beth',
'betray',
'bettemidl',
'better',
'betti',
'beverli',
'bibl',
'bid',
'big',
'bigger',
'biggest',
'bike',
'biker',
'bikini',
'billhad',
'billi',
'billionair',
'billmurray',
'billnighi',
'billpaxton',
'billpullman',
'billybobthornton',
'billycrudup',
'billycryst',
'biographi',
'biolog',
'bird',
'birth',
'birthday',

'bisexu',
'bishop',
'bit',
'bite',
'bitter',
'bizarr',
'black',
'blackmag',
'blackmail',
'blackpeopl',
'blacksmith',
'blade',
'blame',
'blend',
'blind',
'bliss',
'blizzard',
'block',
'blond',
'blood',
'bloodi',
'bloodsplatt',
'bloodthirsti',
'blow',
'blue',
'board',
'boardingschool',
'boat',
'bob',
'bobbi',
'bobbyfarrelli',
'bobhoskin',
'bodi',
'body',
'bodyguard',
'bold',
'bollywood',
'bomb',
'bond',
'bone',
'book',
'border',
'bore',
'boredom',
'born',
'boss',
'boston',
'botch',
'bound',
'boundari',
'bounti',
'bountyhunt',
'box',
'boxer',
'boy',
'boyfriend',
'boys',
'bradleycoop',
'bradpitt',
'brain',
'brand',
'brandon',
'brave',
'braveri',
'brazil',
'brazilian',
'break',
'breakdown',
'breakup',
'breast',
'breath',
'breed',
'brendanfras',
'brendangleeson',
'brent',
'brettratn',
'brian',
'briandepalma',
'bride',
'bridesmaid',
'bridg',
'brief',
'brien',
'bright',
'brilliant',
'bring',
'brink',
'britain',
'british',

'britishsecretservic',
'brittanymurphi',
'broadcast',
'broadway',
'broke',
'broken',
'broker',
'bronx',
'brook',
'brooklyn',
'broom',
'brothel',
'brother',
'brotherbrotherrelationship',
'brothers',
'brothersisterrelationship',
'brought',
'brown',
'bruce',
'brucegreenwood',
'brucewilli',
'brutal',
'bryansing',
'bu',
'buck',
'bud',
'buddi',
'buddy',
'buddycomedi',
'buddycop',
'budget',
'build',
'building',
'built',
'bullet',
'bulli',
'bumbl',
'bunch',
'bunker',
'bunni',
'burglar',
'buri',
'burn',
'bush',
'busi',
'business',
'businessman',
'bust',
'butcher',
'butler',
'butt',
'button',
'buy',
'buzz',
'cabin',
'caesar',
'cage',
'cairo',
'cal',
'california',
'calvin',
'camcord',
'came',
'camera',
'cameraman',
'camerondiaz',
'camp',
'campaign',
'campbell',
'campu',
'canada',
'canadian',
'cancer',
'candi',
'candid',
'canin',
'cannib',
'canuxploit',
'capabl',
'caper',
'capit',
'capt',
'captain',
'captiv',
'captur',
'capture',
'car',
'caraccid',
'carchas',

'carcrash',
'card',
'care',
'career',
'carefre',
'caretak',
'careymulligan',
'caribbean',
'carjourney',
'carl',
'carlagugino',
'carmen',
'carol',
'carolina',
'carrac',
'carri',
'carrie',
'cartel',
'carter',
'cartoon',
'caryelw',
'case',
'caseyaffleck',
'cash',
'casino',
'cast',
'castl',
'cat',
'cataclysm',
'catastroph',
'catch',
'cateblanchett',
'catherinedeneuv',
'catherinekeen',
'catherinezeta',
'cathol',
'catholic',
'cattl',
'caught',
'caus',
'cavalri',
'cave',
'cavemen',
'celebr',
'celebration',
'cell',
'cellphon',
'cemeteri',
'center',
'centr',
'central',
'centuri',
'centuries',
'century',
'ceo',
'certain',
'chad',
'chain',
'chainsaw',
'challeng',
'chamber',
'champion',
'championship',
'chanc',
'chance',
'chang',
'change',
'changed',
'changes',
'channingtatum',
'chao',
'chaos',
'chaotic',
'chapter',
'charact',
'character',
'characters',
'charg',
'charismat',
'charl',
'charli',
'charlie',
'charliesheen',
'charlizetheron',
'charm',
'chart',
'chase',
'chauffeur',
'chazzpalminteri',

'cheat',
'check',
'cheerlead',
'chef',
'chemic',
'cher',
'chevychas',
'chicago',
'chicken',
'chief',
'child',
'childabus',
'childhero',
'childhood',
'childprodigi',
'children',
'chill',
'chimp',
'china',
'chines',
'chip',
'chipmunk',
'chiwetelejiofor',
'chloe',
'chloëgracemoretz',
'chloësevigni',
'chocol',
'choic',
'choice',
'choos',
'chosen',
'chowyun',
'chri',
'chriscolumnbu',
'chriscoop',
'chrisevan',
'chrishemsworth',
'chrisklein',
'chrispin',
'chrisrock',
'christ',
'christian',
'christianbal',
'christianslat',
'christin',
'christinaappleg',
'christinaricci',
'christma',
'christmas',
'christmasparti',
'christmastre',
'christoph',
'christopherlambert',
'christopherlloyd',
'christophernolan',
'christopherplumm',
'christopherwalken',
'christophwaltz',
'chrisweitz',
'chronicl',
'chuck',
'church',
'cia',
'ciaránhind',
'cigarettesmok',
'cillianmurphi',
'cinema',
'circl',
'circu',
'circuit',
'circumst',
'citi',
'citizen',
'city',
'civil',
'civilian',
'civilwar',
'claim',
'clair',
'clairedan',
'claireforlani',
'clan',
'clark',
'clash',
'class',
'classdiffer',
'classic',
'classmat',
'classroom',

'claudevandamm',
'clay',
'clean',
'clear',
'clerk',
'clever',
'client',
'clients',
'cliff',
'climat',
'climb',
'clinteastwood',
'cliveowen',
'clock',
'clone',
'close',
'closer',
'cloud',
'clown',
'club',
'clue',
'clueless',
'clutch',
'coach',
'coast',
'cocain',
'code',
'coffin',
'cohen',
'col',
'cold',
'coldwar',
'cole',
'colin',
'colinfarrel',
'colinfirth',
'collaps',
'colleagu',
'collect',
'collector',
'colleg',
'college',
'collid',
'collis',
'colombia',
'colonel',
'coloni',
'color',
'colorado',
'coma',
'combat',
'combin',
'come',
'comeback',
'comed',
'comedi',
'comedian',
'comedy',
'comet',
'comfort',
'comic',
'coming',
'comingofag',
'comingout',
'command',
'commando',
'commerci',
'commiss',
'commit',
'common',
'commun',
'communist',
'community',
'compani',
'companion',
'company',
'compet',
'competit',
'competition',
'complet',
'complex',
'complic',
'compos',
'compuls',
'comput',
'computerviru',
'conan',
'concern',
'concert',

```
'concoct',
'condit',
'condition',
'conduct',
'confeder',
'confess',
'confid',
'confin',
'conflict',
'confront',
'confus',
'congress',
'conman',
'connect',
'connecticut',
'connel',
'connor',
'conquer',
'consequ',
'consequences',
'conserv',
'consid',
'conspir',
'conspiraci',
'conspiracy',
'constant',
'constantli',
'construct',
'consum',
'contact',
'contain',
'contemporari',
'contend',
'content',
'contest',
'contin',
'contract',
'contractor',
'control',
'controversi',
'convent',
'converg',
'convers',
'convict',
'convinc',
'cook',
'cool',
'cooper',
'cop',
'cope',
...]
```

```
In [47]: ps.stem('dancing')
```

```
Out[47]: 'danc'
```

```
In [48]: stem('In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization. action adventure fantasy sciencefiction culture clash futuristic space war science fiction colonial society space travel futuristic romance space alien tribe alien planet cgi marine soldier battle love affair action war power religion mind and soul 3d sam worthington zoe seldana sigourney weaver james cameron')
```

```
Out[48]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn between follow order and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur spacewar science fiction coloniali societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair action ntiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'
```

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

```
In [50]: similarity=cosine_similarity(vectors)
```

```
In [51]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]
#sorted(similarity[0],reverse=True)
#similarity
```

```
Out[51]: [(1216, 0.28676966733820225),
(2409, 0.26901379342448517),
(3730, 0.2605130246476754),
(507, 0.255608593705383),
```

```
(539, 0.2503866978335957)]
```

```
In [52]: def recommend(movie):
         movie_index = new_df[new_df['title'] == movie].index[0]
         distances= similarity[movie_index]
         movies_list = sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1:6]
         for i in movies_list:
             print(new_df.iloc[i[0]].title)
```

```
In [53]: recommend('Batman Begins')
```

```
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf
```

```
In [54]: new_df.iloc[1216].title
```

```
Out[54]: 'Aliens vs Predator: Requiem'
```

```
In [55]: import pickle
```

```
In [56]: pickle.dump(new_df,open("recommended_movies.pkl",'wb'))
```

```
In [57]: import bz2
         import os
         pickle.dump(similarity,open("similarity.pkl",'wb'))

         os.path.getsize("similarity.pkl")
```

```
Out[57]: 184781251
```

```
pickle.dump(new_df,open('movies.pkl','wb'))
```

```
In [60]: ofile = bz2.BZ2File("similarity.pkl","wb")
         pickle.dump(similarity,ofile)
         ofile.close()

         os.path.getsize("similarity.pkl")
```

```
Out[60]: 42656464
```

```
In [61]: new_df['title'].values
```

```
Out[61]: array(['Avatar', 'Pirates of the Caribbean: At World's End', 'Spectre',
        ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
        'My Date with Drew'], dtype=object)
```

```
pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))
```

```
pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]:
```