

```
In [76]: import numpy as np  
import pandas as pd
```

```
In [77]: movies= pd.read_csv('tmdb_5000_movies.csv')  
credits= pd.read_csv('tmdb_5000_credits.csv')
```

```
In [78]: movies.head(1)
```

```
Out[78]:      budget   genres           homepage    id  keywords  original_language  original_title  
0  237000000  [{"id": 28,  
  "name":  
  "Action"},  
  {"id": 12,  
  "nam..."]  http://www.avatarmovie.com/  1995  [{"id"::  
  1463,  
  "name":  
  "culture  
  clash"},  
  {"id":...]  
  en          Avatar
```

```
In [79]: credits.head(1)
```

```
Out[79]:      movie_id   title           cast           crew  
0      1995  Avatar  [{"cast_id": 242, "character": "Jake Sully", "..."}  [{"credit_id": "52fe48009251416c750aca23", "de..."]
```

```
In [80]: movies=movies.merge(credits,on='title')
```

```
In [81]: movies.shape
```

```
Out[81]: (4809, 23)
```

```
In [82]: movies.head(1)
```

```
Out[82]:      budget   genres           homepage    id  keywords  original_language  original_title  
0  237000000  [{"id": 28,  
  "name":  
  "Action"},  
  {"id": 12,  
  "nam..."]  http://www.avatarmovie.com/  1995  [{"id"::  
  1463,  
  "name":  
  "culture  
  clash"},  
  {"id":...]  
  en          Avatar
```

1 rows × 23 columns

```
In [83]: 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 [84]: #columns to keep
```

```
...
genres
id
keywords
title
overview
cast
crew
...
```

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

```
In [85]: movies.head()
```

Out[85]:	movie_id	title	overview	genres	keywords	cast	cre
	0 19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...	[{"id": 1463, "name": "culture clash"}, {"id": ...	[{"cast_id": 242, "character": "Jake Sully", "na...}	[{"credit_id": "52fe48009251416c750aca2", "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, "name": "...	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "...	[{"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, "name": "...	[{"id": 470, "name": "spy"}, {"id": 818, "name": "...	[{"cast_id": 1, "character": "James Bond", "name..."}]	[{"credit_id": "54805967c3a36829b5002c4", "de...
	3 49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[{"id": 28, "name": "Action"}, {"id": 80, "name": "...	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "...	[{"cast_id": 2, "character": "Bruce Wayne / Ba..."}]	[{"credit_id": "52fe4781c3a36847f81398c", "de...
	4 49529	John Carter	John Carter is a war-weary, former military ca...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "...	[{"id": 818, "name": "based on novel"}, {"id": ...	[{"cast_id": 5, "character": "John Carter", "name..."}]	[{"credit_id": "52fe479ac3a36847f813eaa", "de...

In [86]: `movies.isnull().sum()`

Out[86]:

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

In [87]: `movies.dropna(inplace=True)`

In [88]: `movies.duplicated().sum()`

Out[88]: 0

In [89]: `movies.iloc[0].genres`

```
Out[89]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [90]: # to bring in form - ['Action', 'Adventure', 'Fantasy', 'SciFi']
```

```
In [91]: import ast
def process(obj):
    L=[]
    for i in ast.literal_eval(obj):
        L.append(i['name'])
    return L
```

```
In [92]: movies['genres']=movies['genres'].apply(process)
```

```
In [93]: movies.head()
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action]	[{"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]	[{"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]	[{"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]	[{"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]	[{"id": 818, "name": "based on novel"}, {"id": ...}	[{"cast_id": 5, "character": "John Carter", "c...}	[{"credit_id": "52fe479ac3a36847f813eaa3", "de..

```
In [94]: movies['keywords']=movies['keywords'].apply(process)
```

In [95]: `movies.head()`

	<code>movie_id</code>	<code>title</code>	<code>overview</code>	<code>genres</code>	<code>keywords</code>	<code>cast</code>	<code>crew</code>
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action]	[culture clash]	[{"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]	[ocean]	[{"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]	[spy]	[{"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]	[dc comics]	[{"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]	[based on novel]	[{"cast_id": 5, "character": "John Carter", "c...}	[{"credit_id": "52fe479ac3a36847f813ea3", "de..}

In [96]: `def processcast(obj):`  
`L=[ ]`  
`counter=0`  
`for i in ast.literal_eval(obj):`  
 `if counter !=3:`  
 `L.append(i['name'])`  
 `counter+=1`  
 `else:`  
 `break`  
`return L`

In [99]: `movies['cast']= movies['cast'].apply(processcast)`

In [100...]: `movies.head()`

	movie_id	title	overview	genres	keywords	cast	cre
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action]	[culture clash]	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	{"credit_id": "52fe48009251416c750aca2", "de...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure]	[ocean]	[Johnny Depp, Orlando Bloom, Keira Knightley]	{"credit_id": "52fe4232c3a36847f800b57", "de...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action]	[spy]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	{"credit_id": "54805967c3a36829b5002c4", "de...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action]	[dc comics]	[Christian Bale, Michael Caine, Gary Oldman]	{"credit_id": "52fe4781c3a36847f81398c", "de...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action]	[based on novel]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	{"credit_id": "52fe479ac3a36847f813eaa", "de...

In [101...]

```
def processdirect(obj):
    L=[]
    for i in ast.literal_eval(obj):
        if i['job']=='Director':
            L.append(i['name'])
            break
    return L
```

In [102...]

```
movies['crew']=movies['crew'].apply(processdirect)
```

In [103...]

```
movies.head()
```

Out[103]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action]	[culture clash]	[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]	[ocean]	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action]	[spy]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action]	[dc comics]	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action]	[based on novel]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

In [105...]

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

In [106...]

```
movies.head()
```

Out[106]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...]	[Action]	[culture clash]	[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]	[ocean]	[Johnny Depp, Orlando Bloom, Keira Knightley]	[Gore Verbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...]	[Action]	[spy]	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[Sam Mendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...]	[Action]	[dc comics]	[Christian Bale, Michael Caine, Gary Oldman]	[Christopher Nolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...]	[Action]	[based on novel]	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[Andrew Stanton]

In [ ]: # transform:

'''

Zoe Saldana to ZoeSaldana

'''

In [111...]

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

In [112...]

movies.head()

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

```
In [114... movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['cast']
```

```
In [115... movies.head()
```

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

movies.info(tags)

In [120]: DF=movies[['movie\_id','title','tags']]

In [121]: DF

Out[121]:

	movie_id	title	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...
...	...	...	...
4804	9367	El Mariachi	[El, Mariachi, just, wants, to, play, his, gui...
4805	72766	Newlyweds	NaN
4806	231617	Signed, Sealed, Delivered	["Signed,, Sealed,, Delivered", introduces, a,...
4807	126186	Shanghai Calling	NaN
4808	25975	My Date with Drew	[Ever, since, the, second, grade, when, he, fi...

4806 rows × 3 columns

In [126...]

```
DF['tags'] = DF['tags'].apply(lambda x: " ".join(x) if isinstance(x, list) else str)

C:\Users\HP\AppData\Local\Temp\ipykernel_4996\258494434.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
    DF['tags'] = DF['tags'].apply(lambda x: " ".join(x) if isinstance(x, list) else str(x))
```

In [127...]

DF

Out[127]:

	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...
...	...	...	...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...
4805	72766	Newlyweds	nan
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...
4807	126186	Shanghai Calling	nan
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...

4806 rows × 3 columns

In [128...]

DF.head()

Out[128]:

	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 [132...]

DF['tags'][0]

Out[132]:

```
'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 cultureclash samworthington zoesaldana sigourneyweaver jamescameron'
```

In [130...]

DF['tags']=DF['tags'].apply(lambda x:x.lower())

```
C:\Users\HP\AppData\Local\Temp\ipykernel_4996\3760184172.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  
DF['tags']=DF['tags'].apply(lambda x:x.lower())
```

In [133...]: `DF.head()`

	<code>movie_id</code>	<code>title</code>	<code>tags</code>
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 [ ]: `# our work is to find the similarity between two tags but this in in words that is  
#Bag of words- combine all words now it will find most freq words  
#remove stop words also - which is are and or etc...`

In [153...]: `import nltk  
from nltk.stem.porter import PorterStemmer  
  
ps = PorterStemmer()`

In [154...]: `def stem(text):  
 y=[]  
 for i in text.split():  
 y.append(ps.stem(i))  
 return " ".join(y)`

In [155...]: `DF['tags']= DF['tags'].apply(stem)`

```
C:\Users\HP\AppData\Local\Temp\ipykernel_4996\1127512851.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  
DF['tags']= DF['tags'].apply(stem)
```

In [156...]: `from sklearn.feature_extraction.text import CountVectorizer  
cv=CountVectorizer(max_features=5000, stop_words='english')`

In [157...]: `vectors=cv.fit_transform(DF['tags']).toarray()`

```
In [158...]: vectors
```

```
Out[158]: 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 [160...]: cv.get_feature_names()
```

```
C:\Users\HP\anaconda3\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[160]: ['000',
 '007',
 '10',
 '100',
 '10th',
 '11',
 '12',
 '12th',
 '13',
 '14',
 '15',
 '150',
 '16',
 '16th',
 '17',
 '17th',
 '18',
 '1890',
 '18th',
 '19',
 '1910',
 '1920',
 '1930',
 '1930s',
 '1937',
 '1941',
 '1944',
 '1945',
 '1950',
 '1950s',
 '1955',
 '1959',
 '1960',
 '1960s',
 '1962',
 '1964',
 '1965',
 '1967',
 '1970',
 '1970s',
 '1971',
 '1972',
 '1973',
 '1974',
 '1976',
 '1977',
 '1979',
 '1980',
 '1980s',
 '1984',
 '1985',
 '1990',
 '1999',
 '19th',
 '20',
 '200',
 '2000',
 '2001',
 '2003',
```

'2009',  
'20th',  
'21st',  
'23',  
'24',  
'25',  
'30',  
'300',  
'35',  
'3d',  
'40',  
'47',  
'50',  
'500',  
'60',  
'60s',  
'70',  
'80',  
'aaron',  
'aaroneckhart',  
'aarontaylor',  
'abandon',  
'abduct',  
'abigailbreslin',  
'abil',  
'abl',  
'aboard',  
'abov',  
'abroad',  
'absolut',  
'abu',  
'academ',  
'academi',  
'accept',  
'access',  
'accid',  
'accident',  
'acclaim',  
'accompani',  
'accomplish',  
'account',  
'accu',  
'ace',  
'achiev',  
'acquaint',  
'acquir',  
'act',  
'action',  
'activ',  
'activist',  
'activities',  
'actor',  
'actress',  
'actual',  
'adam',  
'adammckay',  
'adamsandl',  
'adamshankman',  
'adapt',

'add',  
'addict',  
'adjust',  
'admir',  
'admit',  
'adolesc',  
'adopt',  
'ador',  
'adrienbrodi',  
'adrift',  
'adult',  
'adulteri',  
'adulthood',  
'advanc',  
'adventur',  
'adventure',  
'adventures',  
'adverti',  
'advi',  
'advic',  
'affair',  
'affect',  
'afghanistan',  
'afraid',  
'africa',  
'african',  
'aftermath',  
'ag',  
'age',  
'agenc',  
'agency',  
'agenda',  
'agent',  
'agents',  
'aggress',  
'ago',  
'agr',  
'ahead',  
'aid',  
'ail',  
'aim',  
'air',  
'airborn',  
'aircraft',  
'airplan',  
'airport',  
'aka',  
'al',  
'alabama',  
'alan',  
'alanarkin',  
'alaska',  
'alaskan',  
'albert',  
'albertbrook',  
'albertfinney',  
'alcohol',  
'alecbaldwin',  
'alejandrogonzálezíñárritu',

'alex',  
'alexanderpayn',  
'alexanderskarsgård',  
'alexapenavega',  
'alexia',  
'alexisbledel',  
'alexkendrick',  
'alexpettyf',  
'alexproya',  
'alfredhitchcock',  
'alfredmolina',  
'ali',  
'alic',  
'alice',  
'alicebraga',  
'alien',  
'aliens',  
'alik',  
'alike',  
'alilart',  
'aliv',  
'alive',  
'alleg',  
'allen',  
'alli',  
'allianc',  
'allies',  
'allow',  
'alon',  
'alongsid',  
'alpacino',  
'alpha',  
'already',  
'alter',  
'altern',  
'alway',  
'alyssa',  
'alzheimer',  
'amanda',  
'amandabyn',  
'amandapeet',  
'amandaseyfri',  
'amateur',  
'amaz',  
'amazon',  
'ambassador',  
'amberheard',  
'ambit',  
'ambiti',  
'ambush',  
'america',  
'american',  
'americanfootb',  
'amid',  
'amidst',  
'amnesia',  
'amp',  
'amsterdam',  
'amu',

'amy',  
'amyadam',  
'amysmart',  
'ana',  
'anakin',  
'analyst',  
'ancient',  
'anderson',  
'andi',  
'andiemandowel',  
'andrew',  
'andrewadamson',  
'andrewniccol',  
'android',  
'andrzejbartkowiak',  
'andy',  
'andyfickman',  
'andygarcía',  
'andyserki',  
'andytenn',  
'angel',  
'angela',  
'angelabassett',  
'angeles',  
'angelinajoli',  
'anger',  
'angl',  
'angri',  
'ani',  
'anim',  
'animal',  
'animals',  
'anjelicahuston',  
'ann',  
'anna',  
'annafari',  
'annakendrick',  
'anne',  
'annefletch',  
'annehathaway',  
'annehech',  
'annemoss',  
'annetteben',  
'anni',  
'annie',  
'anniversari',  
'announc',  
'annual',  
'anonym',  
'anoth',  
'answer',  
'ant',  
'anthoni',  
'anthonyanderson',  
'anthonyhopkin',  
'anthonymacki',  
'anti',  
'antic',  
'antoinefuqua',

'antoniobandera',  
'antonyelchin',  
'anyon',  
'anyth',  
'apart',  
'apartment',  
'ape',  
'apes',  
'apocalypse',  
'apocalypt',  
'appar',  
'appear',  
'appl',  
'apple',  
'appli',  
'appoint',  
'appreci',  
'apprehend',  
'apprent',  
'approach',  
'april',  
'arab',  
'arch',  
'archaeologist',  
'architect',  
'arctic',  
'area',  
'aren',  
'arena',  
'ari',  
'aristocrat',  
'arm',  
'armi',  
'armor',  
'army',  
'arnoldschwarzenegg',  
'arrang',  
'array',  
'arrest',  
'arriv',  
'arrives',  
'arrog',  
'art',  
'arthur',  
'artifici',  
'artificialintellig',  
'artist',  
'ash',  
'ashley',  
'ashleyjudd',  
'ashtonkutch',  
'asid',  
'ask',  
'aspect',  
'aspir',  
'assassin',  
'assault',  
'assembl',  
'assign',

'assist',  
'assistant',  
'associ',  
'assum',  
'asteroid',  
'astronaut',  
'asylum',  
'athlet',  
'atlant',  
'atom',  
'atomegoyan',  
'attach',  
'attack',  
'attacks',  
'attempt',  
'attend',  
'attent',  
'attitud',  
'attorney',  
'attract',  
'audienc',  
'audit',  
'audreytaut',  
'august',  
'aunt',  
'austin',  
'australia',  
'australian',  
'author',  
'auto',  
'autobiograph',  
'avail',  
'aveng',  
'averag',  
'avoid',  
'await',  
'awaken',  
'awar',  
'award',  
'away',  
'awesom',  
'awri',  
'awry',  
'axel',  
'ba',  
'babe',  
'bab'i',  
'baby',  
'bachelor',  
'backdrop',  
'background',  
'backyard',  
'bad',  
'bag',  
'bail',  
'balanc',  
'ball',  
'ballet',  
'baltimor',

'ban',  
'band',  
'bandit',  
'banish',  
'bank',  
'banker',  
'bar',  
'barbara',  
'barbrastreisand',  
'bare',  
'bargain',  
'barn',  
'barney',  
'baron',  
'barri',  
'barrier',  
'barry',  
'barrylevinson',  
'barrypepp',  
'barrysonnenfeld',  
'bars',  
'base',  
'baseb',  
'basedonnovel',  
'basement',  
'basketb',  
'basketball',  
'batman',  
'battalion',  
'battl',  
'battle',  
'bay',  
'beach',  
'beam',  
'bear',  
'beast',  
'beat',  
'beau',  
'beauti',  
'beautiful',  
'beauty',  
'becam',  
'becau',  
'becom',  
'becominganadult',  
'bed',  
'bedroom',  
'bee',  
'befor',  
'befriend',  
'beg',  
'began',  
'begin',  
'begins',  
'behavior',  
'belief',  
'beliefs',  
'believ',  
'bell',

'bella',  
'belong',  
'belov',  
'ben',  
'benaffleck',  
'bend',  
'beneath',  
'benefit',  
'benfost',  
'beniciodeltoro',  
'benjamin',  
'benjaminbratt',  
'benkingsley',  
'bennett',  
'benstil',  
'bent',  
'berlin',  
'berniemac',  
'best',  
'bet',  
'beth',  
'betray',  
'betrayal',  
'bettemidl',  
'better',  
'bettythoma',  
'beverli',  
'bibl',  
'bicker',  
'bid',  
'big',  
'bigger',  
'biggest',  
'bike',  
'biker',  
'bilbo',  
'billcondon',  
'billhad',  
'billi',  
'billion',  
'billionair',  
'billmurray',  
'billnighi',  
'billpaxton',  
'billpullman',  
'billybobthornton',  
'billyconnolli',  
'billycrudup',  
'billycryst',  
'biograph',  
'biographi',  
'biolog',  
'biopic',  
'bird',  
'birth',  
'birthday',  
'bishop',  
'bit',  
'bitter',

'bizzarr',  
'black',  
'blackmail',  
'blackpeopl',  
'blade',  
'blake',  
'blame',  
'blart',  
'blaze',  
'blend',  
'bless',  
'blind',  
'bliss',  
'block',  
'blofeld',  
'blood',  
'bloodi',  
'bloodthirsti',  
'bloom',  
'blossom',  
'blow',  
'blu',  
'blue',  
'board',  
'boat',  
'bob',  
'bobbi',  
'bobby',  
'bobbyfarrelli',  
'bobhoskin',  
'bodi',  
'body',  
'bodyguard',  
'bohemian',  
'bold',  
'bomb',  
'bond',  
'bone',  
'bonniehunt',  
'boo',  
'book',  
'books',  
'boot',  
'border',  
'bore',  
'boring',  
'born',  
'boss',  
'boston',  
'botch',  
'bound',  
'boundari',  
'bounti',  
'bountyhunt',  
'bourn',  
'bout',  
'bowl',  
'box',  
'boxer',

'boy',  
'boyd',  
'boyfriend',  
'boys',  
'brad',  
'bradanderson',  
'bradleycoop',  
'bradpitt',  
'brain',  
'brand',  
'brandon',  
'brave',  
'brazil',  
'brazilian',  
'break',  
'breakdown',  
'breakup',  
'breath',  
'breed',  
'brendan',  
'brendanfra',  
'brendangleeson',  
'brent',  
'brentspin',  
'brettratn',  
'brian',  
'briancox',  
'briandepalma',  
'brianhelgeland',  
'brianlev',  
'brianrobbin',  
'bride',  
'bridg',  
'bridge',  
'bridgetfonda',  
'brief',  
'brielarson',  
'bright',  
'brilliant',  
'bring',  
'brink',  
'britain',  
'british',  
'brittanymurphi',  
'broadcast',  
'broadway',  
'broke',  
'broken',  
'broker',  
'bronson',  
'brook',  
'brooklyn',  
'brother',  
'brotherbrotherrelationship',  
'brothers',  
'brothersisterrelationship',  
'brought',  
'brown',  
'bruce',

'brucecampbel',  
'brucedern',  
'brucegreenwood',  
'brucewilli',  
'brutal',  
'bryancranston',  
'bryans',  
'bu',  
'buck',  
'bud',  
'budapest',  
'buddi',  
'buddy',  
'budget',  
'bug',  
'build',  
'building',  
'built',  
'bullet',  
'bulli',  
'bumbl',  
'bump',  
'bunch',  
'bureaucraci',  
'buri',  
'burn',  
'burlancast',  
'burton',  
'burtreynold',  
'bush',  
'busi',  
'business',  
'businessman',  
'businesswoman',  
'bust',  
'butch',  
'butcher',  
'butler',  
'butt',  
'button',  
'buy',  
'buzz',  
'cabin',  
'caesar',  
'cain',  
'cal',  
'caleb',  
'california',  
'calvin',  
'came',  
'camera',  
'cameraman',  
'cameron',  
'cameroncrow',  
'camerondiaz',  
'camgigandet',  
'camp',  
'campaign',  
'campbell',

'campu',  
'canada',  
'canadian',  
'cancer',  
'candid',  
'canin',  
'capabl',  
'capit',  
'capt',  
'captain',  
'captiv',  
'captur',  
'capture',  
'car',  
'card',  
'care',  
'career',  
'careymulligan',  
'cargo',  
'caribbean',  
'carl',  
'carlagugino',  
'carlo',  
'carmen',  
'carol',  
'carolina',  
'carrac',  
'carri',  
'carrie',  
'carriefish',  
'cartel',  
'carter',  
'caryelw',  
'case',  
'caseyaffleck',  
'cash',  
'casino',  
'cast',  
'castl',  
'castle',  
'cat',  
'cataclysm',  
'catastroph',  
'catch',  
'cateblanchett',  
'catherin',  
'catherinedeneuv',  
'catherinehardwick',  
'catherinekeen',  
'catherinezeta',  
'cathol',  
'cattl',  
'cau',  
'caught',  
'cave',  
'celebr',  
'celebration',  
'center',  
'centr',

'central',  
'centuri',  
'centuries',  
'century',  
'certain',  
'chad',  
'chain',  
'challeng',  
'chamber',  
'champion',  
'championship',  
'chanc',  
'chance',  
'chang',  
'change',  
'changed',  
'changes',  
'channel',  
'channingtatum',  
'chao',  
'chaos',  
'chaotic',  
'chapter',  
'charact',  
'character',  
'characters',  
'charg',  
'charismat',  
'charl',  
'charless',  
'charli',  
'charlie',  
'charliesheen',  
'charlizetheron',  
'charlott',  
'charm',  
'chart',  
'chase',  
'cheat',  
'check',  
'cheerlead',  
'chef',  
'chemic',  
'cher',  
'chicago',  
'chicken',  
'chief',  
'child',  
'childabu',  
'childhood',  
'children',  
'chill',  
'china',  
'chine',  
'chip',  
'chipmunk',  
'chiwetelejiofor',  
'chloe',  
'chloëgracemoretz',

'chloësevigni',  
'chocol',  
'choic',  
'choice',  
'chon',  
'choo',  
'chosen',  
'chowyun',  
'chri',  
'chriscolumbu',  
'chriscoop',  
'chrisevan',  
'chrishemsworth',  
'chrisklein',  
'chrispin',  
'chrisrock',  
'christ',  
'christian',  
'christianb',  
'christianslat',  
'christin',  
'christinaapple',  
'christinaricci',  
'christma',  
'christmas',  
'christoph',  
'christopherlambert',  
'christopherlloyd',  
'christophernolan',  
'christopherplumm',  
'christopherwalken',  
'christophwaltz',  
'chrisweitz',  
'chronicl',  
'chuck',  
'chuckrussel',  
'church',  
'cia',  
'ciaráhind',  
'cillianmurphi',  
'cindi',  
'cinema',  
'circl',  
'circu',  
'circuit',  
'circumst',  
'circumstances',  
'citi',  
'citizen',  
'city',  
'civil',  
'civilian',  
'civilwar',  
'claim',  
'clair',  
'clairedan',  
'claireforlani',  
'clan',  
'clark',

'clash',  
'class',  
'classic',  
'classmat',  
'clau',  
'claudevandamm',  
'clay',  
'clean',  
'clear',  
'clerk',  
'client',  
'clients',  
'climat',  
'climb',  
'clinteastwood',  
'cliveowen',  
'clock',  
'clone',  
'close',  
'closer',  
'cloud',  
'clown',  
'club',  
'clue',  
'clueless',  
'clutch',  
'coach',  
'coast',  
'cocain',  
'cocki',  
'code',  
'codi',  
'cohen',  
'col',  
'cold',  
'cole',  
'colin',  
'colinfarrel',  
'colinfirth',  
'collap',  
'collapse',  
'colleagu',  
'collect',  
'collector',  
'colleg',  
'college',  
'colli',  
'collid',  
'colonel',  
'coloni',  
'color',  
'colorado',  
'coma',  
'combat',  
'combin',  
'come',  
'comeback',  
'comedi',  
'comedian',

```
'comedy',
'comfort',
'comic',
'coming',
'command',
'commerci',
'commiss',
'commit',
'common',
'commun',
'communist',
'community',
'compani',
'companion',
'company',
'compet',
'competit',
'competition',
'complet',
'complex',
'complic',
'compo',
'compound',
'compton',
'compul',
'comput',
'conan',
'concern',
'concert',
'concoct',
'condit',
'conduct',
'confess',
'confid',
'confin',
'conflict',
'confront',
'confu',
'connect',
'connecticut',
'connel',
'conni',
'connienielsen',
'connor',
'conquer',
'conrad',
'consequ',
'consequences',
'conserv',
'consid',
'conspir',
'conspiraci',
'conspiracy',
'constant',
'constantli',
'construct',
...]
```

```
In [149...]: #now we have find distance between thw movies so we will usw cosine similarity

In [161...]: from sklearn.metrics.pairwise import cosine_similarity

In [164...]: similarity=cosine_similarity(vectors)

In [171...]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]

Out[171]: [(529, 0.24595492912420727),
 (3608, 0.2333333333333336),
 (3083, 0.19611613513818407),
 (2786, 0.18898223650461363),
 (2148, 0.1814436846506058)]

In [179...]: def recommend(movie):
    movie_index=DF[DF['title']== movie].index[0]
    distances=similarity[movie_index]
    movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1]

    for i in movies_list:
        print(DF.iloc[i[0]].title)

In [180...]: recommend('Avatar')
Tears of the Sun
Apollo 18
Bucky Larson: Born to Be a Star
Attack the Block
Risen

In [182...]: import pickle

In [185...]: DF['title'].values

Out[185]: array(['Avatar', "Pirates of the Caribbean: At World's End", 'Spectre',
       ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
       'My Date with Drew'], dtype=object)

In [186...]: pickle.dump(DF.to_dict(),open('movie_dict.pkl','wb'))

In [187...]: pickle.dump(similarity,open('similarity.pkl','wb'))

In [ ]:
```