Steps of project:-

1. Loading Data
2. Data Cleaning
3. Exploratory Data Analysis (EDA)
4. Text Preprocessing
5. Model Building
6. Evaluation
7. Convert into Website
8. Deploy

1. Loading Data

```python
In [76]:  #importing necessary library
          import numpy as np
          import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns
          import ast
          import warnings
          warnings.filterwarnings('ignore')
```

```python
In [77]:  #Load datasets
          df_movie = pd.read_csv('tmdb_5000_movies.csv')
          df_credict = pd.read_csv('tmdb_5000_credits.csv')
```

```python
In [78]:  df_movie.head()
```

Out[78]:

	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 Barbossa, 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

```python
In [79]:  df_credict.head(1)
```

Out[79]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...

```python
In [80]:  data=pd.merge(df_movie,df_credict,on='title')
          data.head(2)
```

	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 Barbossa, long believed to be dead, ha...	139.082615

2 rows × 23 columns

2. Data Cleaning

In [81]:
```python
#feature selection  that will we important in our project
df = data[["movie_id","title","overview","genres","keywords","cast","crew"]]
```

In [82]:
```python
df
```

Out[82]:

	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...
...	...	...	...	...	...	...	...
4804	9367	El Mariachi	El Mariachi just wants to play his guitar and ...	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	[{"id": 5616, "name": "united states\u2013mexi...	[{"cast_id": 1, "character": "El Mariachi", "c...	[{"credit_id": "52fe44eec3a36847f80b280b", "de...
4805	72766	Newlyweds	A newlywed couple's honeymoon is upended by th...	[{"id": 35, "name": "Comedy"}, {"id": 10749, "...	[]	[{"cast_id": 1, "character": "Buzzy", "credit_...	[{"credit_id": "52fe487dc3a368484e0fb013", "de...
4806	231617	Signed, Sealed, Delivered	"Signed, Sealed, Delivered" introduces a dedic...	[{"id": 35, "name": "Comedy"}, {"id": 18, "nam...	[{"id": 248, "name": "date"}, {"id": 699, "nam...	[{"cast_id": 8, "character": "Oliver O\u2019To...	[{"credit_id": "52fe4df3c3a36847f8275ecf", "de...
4807	126186	Shanghai Calling	When ambitious New York attorney Sam is sent t...	[]	[]	[{"cast_id": 3, "character": "Sam", "credit_id...	[{"credit_id": "52fe4ad9c3a368484e16a36b", "de...
4808	25975	My Date with Drew	Ever since the second grade when he first saw ...	[{"id": 99, "name": "Documentary"}]	[{"id": 1523, "name": "obsession"}, {"id": 224...	[{"cast_id": 3, "character": "Herself", "credi...	[{"credit_id": "58ce021b9251415a390165d9", "de...

4809 rows × 7 columns

In [83]:
```python
pd.isnull(df).sum()
```

Out[83]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [84]:
```python
df.dropna(inplace=True)
```

In [85]:
```python
pd.isnull(df).sum()
```

```
Out[85]:  movie_id     0
          title        0
          overview     0
          genres       0
          keywords     0
          cast         0
          crew         0
          dtype: int64
```

```
In [86]:  print(df.duplicated().sum())
```

```
          0
```

3. Exploratory Data Analysis (EDA)

```
In [87]:  df.iloc[0].genres
```

```
Out[87]:  '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "na
          me": "Science Fiction"}]'
```

```
In [88]:  df['genres'].head(4)
```

```
Out[88]:  0    [{"id": 28, "name": "Action"}, {"id": 12, "nam...
          1    [{"id": 12, "name": "Adventure"}, {"id": 14, "...
          2    [{"id": 28, "name": "Action"}, {"id": 12, "nam...
          3    [{"id": 28, "name": "Action"}, {"id": 80, "nam...
          Name: genres, dtype: object
```

We need to convert these string of list into list==>>>'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'#we convert these string of list into list by using ast module in python and it provide a function literal_eval thai function is exactely done these kind of work

4. Text Preprocessing

```
In [89]:  def convert(obj):
              L = []
              for i in ast.literal_eval(obj):
                  L.append(i['name'])
              return L
```

```
In [90]:  df['genres']=df['genres'].apply(convert)
```

```
In [91]:  df.head()
```

Out[91]:

	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 [92]:  df['keywords']=df['keywords'].apply(convert)
```

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

	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...

```python
In [94]: #in the cast feature we retrive top three actor information
         #df['cast'][0]
         def convert3(obj):
             L = []
             counter=0
             for i in ast.literal_eval(obj):
                 if counter!=3:
                     L.append(i["name"])
                     counter+=1
                 else:
                     break
             return L
```

```python
In [95]: df['cast']=df['cast'].apply(convert3)#Now these way we got top three actor name by each movie
```

```python
In [96]: df.head()
```

	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...
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...

```python
In [97]: # df['crew'][0]#Now in crew feature we retrive information where the value of job=director
         def featch_director(obj):
             L = []
             counter=0
             for i in ast.literal_eval(obj):
                 if i['job'] == 'Director':
                     L.append(i["name"])
                     break
             return L
```

```python
In [98]: df['crew']=df['crew'].apply(featch_director)
```

```python
In [99]: df.head()
```

`Out[99]:`

	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 [100…`
```python
df['overview'].dtypes
```

`Out[100]:` `dtype('O')`

`In [101…`
```python
df['overview'][0]
```

`Out[101]:` `'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 [102…`
```python
# df['overview'].dtypes==>>>>>Then we converted into list
df['overview']=df['overview'].apply(lambda x:x.split())
```

`In [103…`
```python
df.head()
```

`Out[103]:`

	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 [104…`
```python
df['cast']
```

`Out[104]:`
```
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 [105…`
```python
df['cast']=df['cast'].apply(lambda x:[i.replace(" ","") for i in x])
df['genres']=df['genres'].apply(lambda x:[i.replace (" ","")for i in x])
df['keywords']=df['keywords'].apply(lambda x:[i.replace(" ","") for i in x])
df['crew']=df['crew'].apply(lambda x:[i.replace(" ","") for i in x])
```

`In [106…`
```python
df.head()
```

```
Out[106]:
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[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, 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 [107…  df['tags'] = df['overview']+df['genres']+df["keywords"]+df['cast']+df['crew']
```

```
In [108…  df.head()
```

```
Out[108]:
```

	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 [109…  new_df = df[["movie_id","title","tags"]]
```

```
In [110…  new_df
```

```
Out[110]:
```

	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	[A, newlywed, couple's, honeymoon, is, upended...
4806	231617	Signed, Sealed, Delivered	["Signed,, Sealed,, Delivered", introduces, a,...
4807	126186	Shanghai Calling	[When, ambitious, New, York, attorney, Sam, is...
4808	25975	My Date with Drew	[Ever, since, the, second, grade, when, he, fi...

4806 rows × 3 columns

```
In [111…  new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x))
```

```
In [112…  new_df.head()
```

Out[112]:

	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 [113]:
```python
new_df['tags'][0]
```

Out[113]: '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 cultureclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver JamesCameron'

In [114]:
```python
#Then we try to convert tags into lowerstate
new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
```

In [115]:
```python
new_df
```

Out[115]:

	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	a newlywed couple's honeymoon is upended by th...
4806	231617	Signed, Sealed, Delivered	"signed, sealed, delivered" introduces a dedic...
4807	126186	Shanghai Calling	when ambitious new york attorney sam is sent t...
4808	25975	My Date with Drew	ever since the second grade when he first saw ...

4806 rows × 3 columns

In [116]:
```python
#We applying steaming#Staming means ['loved','loving','love'] these kind of word and after applying steaming it
# ['love','love','love']
```

In [117]:
```python
import nltk
from nltk.stem.porter import PorterStemmer
post = PorterStemmer()
```

In [118]:
```python
def stem(text):
    y=[]
    for i in text.split():
        y.append(post.stem(i))
    return " ".join(y)
```

In [119]:
```python
new_df['tags']=new_df['tags'].apply(stem)
```

Now we create a website on these website the user Enter name of a movie & on the bases of that movie the website suggest user another

5 similar kind of movie.

In [120]:
```python
#We need to suggest similiraty on the bases of tags
new_df['tags'][0]
```

Out[120]: '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 spacecoloni societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl loveaffair antiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'

In [121]:
```python
new_df['tags'][1]#We need to calculate similarity on the bases of new_df[0] and new_df[1] text data
```

Out[121]: "captain barbossa, long believ to be dead, ha come back to life and is head to the edg of the earth with will turner and elizabeth swann. but noth is quit as it seems. adventur fantasi action ocean drugabus exoticisland eastindiatradingcompani loveofone'slif traitor shipwreck strongwoman ship allianc calypso afterlif fighter pir at swashbuckl aftercreditssting johnnydepp orlandobloom keiraknightley goreverbinski"

Now we try to do text vectorization by the calculating similarity

We try to convert each movie tags text data into vector and after converting each movie has tags text into vertor. These the the following technique that is uses we easily convert text data into vector that is called text vectorization

1. bag of words
2. tf-idf
3. word to vectore # In the case of these project we will using the technique which is bag of word In the case bag of word first of all we try to combine all tags data then we need to get most 5 common tags on those 12 tags. & During the performing text vectorization we try to ignore stop word

What is the stop word

Stop word is use only formating of word in english sentence. Example==>>>>>> are and in to from him==>>>>>>these kind of word is called stop word

5.Model Building

```python
In [122...   #Now we performing these task using countvectorizer
             from sklearn.feature_extraction.text import CountVectorizer
             cv = CountVectorizer(max_features=5000,stop_words="english")
             vector=cv.fit_transform(new_df['tags']).toarray()
```

```python
In [123...   vector
```
```
Out[123]:   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]])
```

```python
In [124...   vector[0]
```
```
Out[124]:   array([0, 0, 0, ..., 0, 0, 0])
```

```python
In [125...   vector
```
```
Out[125]:   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]])
```

Notes:-

the scikit-learn library does not have the get_feature_names() method. Instead, it has been replaced with get_feature_names_out().

```python
In [126...   cv.get_feature_names_out()
```
```
Out[126]:   array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
                  dtype=object)
```

```python
In [127...   len(cv.get_feature_names_out())#Now these 5000 word is most occurance word in our corpus
```
```
Out[127]:   5000
```

```python
In [128...   cv.get_feature_names_out()
```
```
Out[128]:   array(['000', '007', '10', ..., 'zone', 'zoo', 'zooeydeschanel'],
                  dtype=object)
```

```python
In [129...   post.stem('kops')
```
```
Out[129]:   'kop'
```

```python
In [130...   post.stem('paraplegic')
```
```
Out[130]:   'parapleg'
```

```python
In [131...   post.stem('in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but
```

'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 cultureclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle loveaffair antiwar powerrelations mindandsoul 3d samworthington zoesaldana sigourneyweaver jamescameron'

In [132]:
```python
from sklearn.metrics.pairwise import cosine_similarity
```

Working of next steps:-

Now in these case we calculating distance by using cosine-similarity method and pass our vector of the object of cosine-similarity object and it calculating distance each movie to 4806 movie which means 1-4806,2-4806,3-4806 then last movie is also calculating distance 4806 movie.

In [133]:
```python
cosine_similarity(vector)
```
Out[133]:
```
array([[1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
        0.        ],
       [0.08346223, 1.        , 0.06063391, ..., 0.02378257, 0.        ,
        0.02615329],
       [0.0860309 , 0.06063391, 1.        , ..., 0.02451452, 0.        ,
        0.        ],
       ...,
       [0.04499213, 0.02378257, 0.02451452, ..., 1.        , 0.03962144,
        0.04229549],
       [0.        , 0.        , 0.        , ..., 0.03962144, 1.        ,
        0.08714204],
       [0.        , 0.02615329, 0.        , ..., 0.04229549, 0.08714204,
        1.        ]])
```

In [134]:
```python
cosine_similarity(vector).shape#We calculating total distance 4806*4806
```
Out[134]: (4806, 4806)

In [135]:
```python
similarity=cosine_similarity(vector)
```

In [136]:
```python
similarity
```
Out[136]:
```
array([[1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
        0.        ],
       [0.08346223, 1.        , 0.06063391, ..., 0.02378257, 0.        ,
        0.02615329],
       [0.0860309 , 0.06063391, 1.        , ..., 0.02451452, 0.        ,
        0.        ],
       ...,
       [0.04499213, 0.02378257, 0.02451452, ..., 1.        , 0.03962144,
        0.04229549],
       [0.        , 0.        , 0.        , ..., 0.03962144, 1.        ,
        0.08714204],
       [0.        , 0.02615329, 0.        , ..., 0.04229549, 0.08714204,
        1.        ]])
```

In [137]:
```python
similarity[0].shape
```
Out[137]: (4806,)

Working:-

first movie on the first movie is a similiraty score is 1#Similary second movie to second movie similarity score is 1 Basically these matrix have digonal will always 1.

In [138]:
```python
similarity[0]
```
Out[138]:
```
array([1.        , 0.08346223, 0.0860309 , ..., 0.04499213, 0.        ,
       0.        ])
```

In [139]:
```python
# similarity[1]
# list(enumerate(similarity[0]))
# sorted(list(enumerate(similarity[0])),reverse=True)
# sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:6]
```

In [144]:
```python
# sorted(similarity[0])[-10:-1]#we need to sorting it in descending orders
sorted(similarity[0],reverse=True)[5]
```
Out[144]: np.float64(0.25038669783359574)

In [145]:
```python
def recommend(movie):
    movie_index=new_df[new_df['title']==movie].index[0]
    distances = similarity[movie_index]
```

```python
        movie_list = sorted(list(enumerate(distances)),reverse=True,key=lambda x:x[1])[1:6]
        for i in movie_list:
            print(i)
#           print(new_df.iloc[i[0]].title)
```

In [146… `recommend('Batman Begins')`

```
(65, np.float64(0.40218090755486674))
(1361, np.float64(0.35434169344615046))
(1360, np.float64(0.3340765523905305))
(3, np.float64(0.3177444546511212))
(3295, np.float64(0.3120099844792576))
```

In [147… `new_df.iloc[1216].title`*#iloc stand for integer location bases indexing and loc stand for label based indexing*

Out[147]: `'Autumn in New York'`

In [ ]:

6. Evaluation

In [148…
```python
print(recommend('The Dark Knight'))
print(recommend('Inception'))
```

```
(3, np.float64(0.42256142372325745))
(119, np.float64(0.40218090755486674))
(428, np.float64(0.33142090599966445))
(299, np.float64(0.2941494555969643))
(1361, np.float64(0.27665837249659325))
None
(2209, np.float64(0.3036661238170935))
(711, np.float64(0.2821440468234174))
(1427, np.float64(0.2795084971874737))
(3091, np.float64(0.2746751327867679))
(2025, np.float64(0.27456258919345766))
None
```

In [ ]:

7. Convert into Website

In [149… `import pickle`

In [150… `pickle.dump(new_df,open('movies.pkl','wb'))`*#wb means write binary mode*

In [151… `new_df['title'].values`

Out[151]:
```
array(['Avatar', "Pirates of the Caribbean: At World's End", 'Spectre',
       ..., 'Signed, Sealed, Delivered', 'Shanghai Calling',
       'My Date with Drew'], dtype=object)
```

In [152… `pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))`

In [160… `# new_df.to_dict()`

In [ ]:

In [80]: `pickle.dump(similarity,open('similarity.pkl','wb'))`

In [ ]:

8. Deploy

Deploy the application on platforms like Heroku or Streamlit Cloud to make it accessible to users.

In [ ]:

In [ ]:

In [ ]:

In [ ]:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js