



```
# overview
# cast
# crew

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

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

Out[10]:

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

```
In [11]: import ast
```

```
In [12]: def convert(text):
    L = []
    for i in ast.literal_eval(text):
        L.append(i['name'])
    return L
```

```
In [13]: movies.dropna(inplace=True)
```

```
In [14]: movies['genres'] = movies['genres'].apply(convert)
movies.head()
```

Out[14]:

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

In [15]:

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

Out[15]:

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

In [16]:

```
import ast
ast.literal_eval(' [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, ...
```

Out[16]:

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

In [17]:

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

In [18]:

```
movies['cast'] = movies['cast'].apply(convert)
movies.head()
```

Out[18]:

	movie_id	title	genres	overview	keywords	cast	
0	19995	Avatar	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di...	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	{{"credit_ "52fe48009251416c750aca",
1	285	Pirates of the Caribbean: At World's End	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha...	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	{{"credit_ "52fe4232c3a36847f800b5",
2	206647	Spectre	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o...	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	{{"credit_ "54805967c3a36829b5002c",
3	49026	The Dark Knight Rises	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve...	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman, A...	{{"credit_ "52fe4781c3a36847f81398",
4	49529	John Carter	[Action, Adventure, Science Fiction]	John Carter is a war-weary, former military ca...	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton,...	{{"credit_ "52fe479ac3a36847f813ea",

```
In [19]: movies['cast'] = movies['cast'].apply(lambda x:x[0:3])
```

```
In [20]: def fetch_director(text):
L = []
for i in ast.literal_eval(text):
    if i['job'] == 'Director':
        L.append(i['name'])
return L
```

```
In [21]: movies['crew'] = movies['crew'].apply(fetch_director)
```

```
In [22]: #movies['overview'] = movies['overview'].apply(lambda x:x.split())
movies.sample(5)
```

Out [22]:

	movie_id	title	genres	overview	keywords	cast	crew
<b>3064</b>	24034	Teacher's Pet	[Animation, Drama, Family, Music]	Spot (Lane) is a dog who can talk and read. Po...	[]	[Nathan Lane, Kelsey Grammer, Shaun Fleming]	[Timothy Bjorklund]
<b>2161</b>	107985	The World's End	[Comedy, Action, Science Fiction]	Five friends who reunite in an attempt to top ...	[end of the world, apocalypse, trilogy, homage]	[Simon Pegg, Nick Frost, Martin Freeman]	[Edgar Wright]
<b>3676</b>	11935	Capricorn One	[Drama, Action, Thriller, Science Fiction]	In order to protect the reputation of the Amer...	[helicopter, nasa, texas, spacecraft, beguilem...	[Elliott Gould, James Brolin, Brenda Vaccaro]	[Peter Hyams]
<b>2055</b>	17709	Dudley Do-Right	[Comedy, Family, Romance]	Based on the 60's-era cartoon of the same name...	[based on cartoon, mountie]	[Brendan Fraser, Sarah Jessica Parker, Alfred ...]	[Hugh Wilson]
<b>196</b>	38055	Megamind	[Animation, Action, Comedy, Family, Science Fi...	Bumbling supervillain Megamind finally defeats...	[saving the world, date, prison, secret identi...	[Will Ferrell, Brad Pitt, Tina Fey]	[Tom McGrath]

```
In [23]: def collapse(L):
          L1 = []
          for i in L:
              L1.append(i.replace(" ", ""))
          return L1
```

```
In [24]: movies['cast'] = movies['cast'].apply(collapse)
          movies['crew'] = movies['crew'].apply(collapse)
          movies['genres'] = movies['genres'].apply(collapse)
          movies['keywords'] = movies['keywords'].apply(collapse)
```

```
In [25]: movies.head()
```

Out[25]:

	movie_id	title	genres	overview	keywords	cast	
--	----------	-------	--------	----------	----------	------	--

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

```
In [26]: movies['overview'] = movies['overview'].apply(lambda x:x.split())

In [28]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords']

In [29]: new = movies.drop(columns=['overview','genres','keywords','cast','crew'])
#new.head()

In [30]: new['tags'] = new['tags'].apply(lambda x: " ".join(x))
new.head()
```

Out[30]:

	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 [31]: from sklearn.feature_extraction.text import CountVectorizer
```

```
cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
In [32]: vector = cv.fit_transform(new['tags']).toarray()
```

```
In [33]: vector.shape
```

```
Out[33]: (4806, 5000)
```

```
In [34]: from sklearn.metrics.pairwise import cosine_similarity
```

```
In [35]: similarity = cosine_similarity(vector)
```

```
In [36]: similarity
```

```
Out[36]: array([[1.          , 0.09583148, 0.06277648, ..., 0.02465568, 0.0270666 ,
        0.          ],
       [0.09583148, 1.          , 0.07018624, ..., 0.02756589, 0.          ,
        0.          ],
       [0.06277648, 0.07018624, 1.          , ..., 0.02708645, 0.          ,
        0.          ],
       ...,
       [0.02465568, 0.02756589, 0.02708645, ..., 1.          , 0.07007128,
        0.04732485],
       [0.0270666 , 0.          , 0.          , ..., 0.07007128, 1.          ,
        0.05195243],
       [0.          , 0.          , 0.          , ..., 0.04732485, 0.05195243,
        1.          ]])
```

```
In [37]: new[new['title'] == 'The Lego Movie'].index[0]
```

```
Out[37]: 744
```

```
In [38]: def recommend(movie):
        index = new[new['title'] == movie].index[0]
        distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda x:
        for i in distances[1:6]:
            print(new.iloc[i[0]].title)
```

```
In [39]: recommend('Gandhi')
```

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

```
In [40]: import pickle
```

```
In [41]: pickle.dump(new,open('movie_list.pkl','wb'))
        pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]:
```