

```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

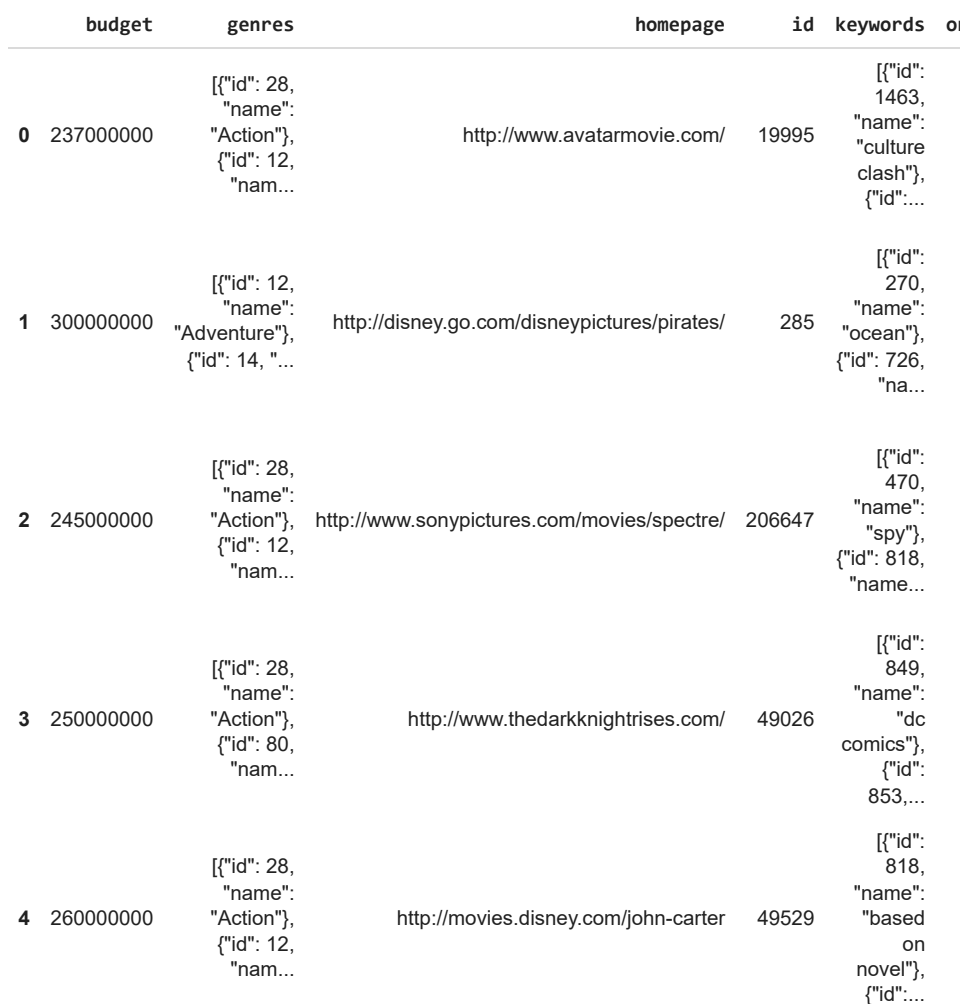
```
movies.head(2)
```

	budget	genres	homepage	id	keywords	orig
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 19995, "name": "Avatar"}]	Avatar
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "Pirates of the Caribbean"}]	Pirates of the Caribbean: The Curse of the Black Pearl

→ (4803, 20)

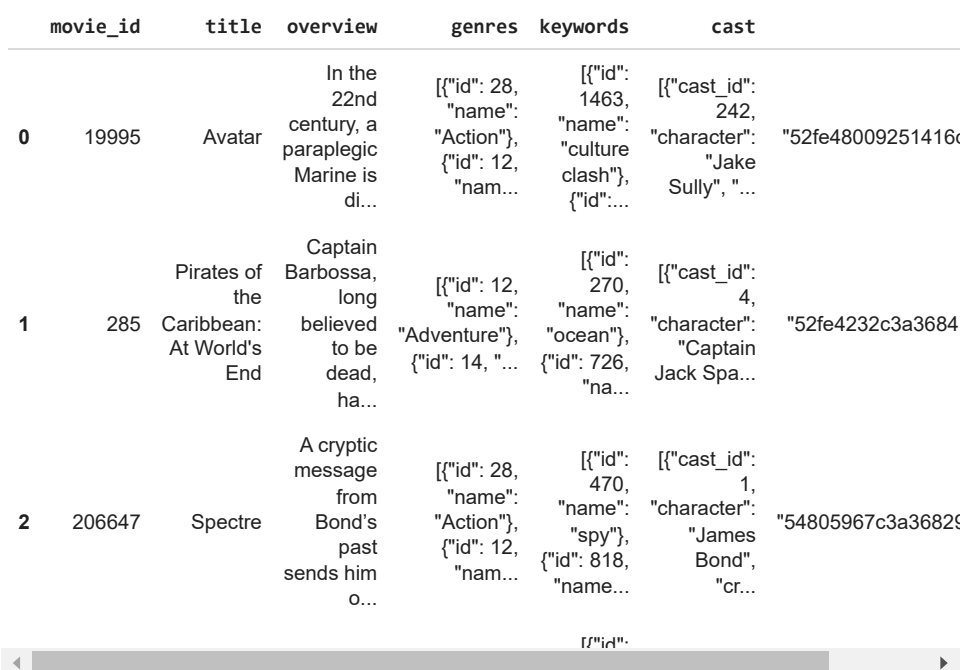
	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre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

```
movies.head()
```



◀ [REDACTED] ▶


```
movies.head()
```




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

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


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



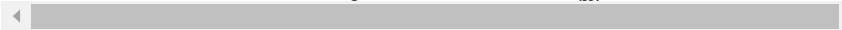
	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	{ "id": 1463, "name": "culture clash", "id": ... }	{ "cast_id": 242, "character": "Jake Sully", ... }
1	285	Pirates of the Caribbean: At World's 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...
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...




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



	movie_id	title	overview	genres	keywords	cast
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", ... }
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...
2	206647	Spectre	A cryptic message from Bond's past sends him	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel,	{ "cast_id": 1, "character": "James Bond", "cr...



```
import ast
ast.literal_eval('{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 87
```



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

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

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

↩

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...
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, ...
2	206647	Spectre	A cryptic message from Bond's past sends him	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel,	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...

◀ ▶

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

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

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

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

↩

	movie_id	title	overview	genres	keywords	cast	crew
631	425	Ice Age	[With, the, impending, ice, age, almost, upon,...	[Animation, Comedy, Family, Adventure]	[human evolution, parents kids relationship, s...	[Ray Romano, John Leguizamo, Denis Leary]	[Chris Wedge]
535	6520	First Knight	[The, timeless, tale, of, King, Arthur, and, t...	[Action, Adventure, Drama, Romance]	[camelot, knight, king arthur, excalibur, knig...	[Sean Connery, Richard Gere, Julia Ormond]	[Jerry Zucker]
596	8427	I Spy	[When, the, Switchblade,, the, most, sophistic	[Action, Adventure, Comedy, Thriller]	[budapest, kidnapping, boxer, secret agent li	[Eddie Murphy, Owen Wilson, Famke	[Betty Thomas]

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

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

```
movies.head()
```

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]
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]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]
		The Dark	[Following, the, death	[Action, Crime,	[dccomics, crimefighter	[ChristianBale,

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

```
new = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])
```

```
new.head()
```

	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...

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

	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

```
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')
```

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

```
vector.shape
```

```
➦ (937, 5000)
```

```
from sklearn.metrics.pairwise import cosine_similarity  
similarity = cosine_similarity(vector)
```

```
similarity
```

```
➦ array([[1.          , 0.08282364, 0.05437272, ..., 0.02602896, 0.09188458,  
          0.          ],  
        [0.08282364, 1.          , 0.06154575, ..., 0.02946278, 0.06933752,  
          0.          ],  
        [0.05437272, 0.06154575, 1.          , ..., 0.02901294, 0.03413944,  
          0.03711348],  
        ...,  
        [0.02602896, 0.02946278, 0.02901294, ..., 1.          , 0.13074409,  
          0.          ],  
        [0.09188458, 0.06933752, 0.03413944, ..., 0.13074409, 1.          ,  
          0.          ],  
        [0.          , 0.          , 0.03711348, ..., 0.          , 0.          ,  
          1.          ]])
```

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

```
➦ 744
```

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

```
recommend('Avatar')
```

```
➦ Titan A.E.  
  Ender's Game  
  Battle: Los Angeles  
  Independence Day  
  Edge of Tomorrow
```

```
import pickle
```

```
pickle.dump(new,open('movie_list.pkl','wb'))
```

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