

Input output

```
In [1]: # This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

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

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version of your notebook
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session

/kaggle/input/tmdb-movie-metadata/tmdb_5000_movies.csv
/kaggle/input/tmdb-movie-metadata/tmdb_5000_credits.csv
```

```
In [30]: movies = pd.read_csv('/kaggle/input/tmdb-movie-metadata/tmdb_5000_movies.csv')
credits = pd.read_csv('/kaggle/input/tmdb-movie-metadata/tmdb_5000_credits.csv')
```

```
In [31]: movies.head(2)
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview
0	237000000	{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}	http://www.avatarmovie.com/	19995	{'id': 1463, 'name': 'culture clash'}, {'id': ...}	en	Avatar	On a remote world, a paraplegic Marine is dispatched with a military team as a courier of vital secrets to a hostile and terrifyingly advanced native tribe.
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'}	en	Pirates of the Caribbean: At World's End	As the world's most powerful ship approaches its final voyage, Captain Jack Sparrow and his crew must battle a powerful and terrifying enemy to save the world.

```
In [4]: movies.shape
```

```
Out[4]: (4803, 20)
```

```
In [5]: credits.head()
```

	movie_id	title	cast	crew
0	19995	Avatar	{'cast_id': 242, 'character': 'Jake Sully', 'order': 1}	{'credit_id': '52fe48009251416c750aca23', 'name': 'James Cameron', 'role': 'Director'}
1	285	Pirates of the Caribbean: At World's End	{'cast_id': 4, 'character': 'Captain Jack Sparrow', 'order': 1}	{'credit_id': '52fe4232c3a36847f800b579', 'name': 'Gore Verbinski', 'role': 'Director'}
2	206647	Spectre	{'cast_id': 1, 'character': 'James Bond', 'order': 1}	{'credit_id': '54805967c3a36829b5002c41', 'name': 'Sam Mendes', 'role': 'Director'}
3	49026	The Dark Knight Rises	{'cast_id': 2, 'character': 'Bruce Wayne / Batman', 'order': 1}	{'credit_id': '52fe4781c3a36847f81398c3', 'name': 'Christopher Nolan', 'role': 'Director'}
4	49529	John Carter	{'cast_id': 5, 'character': 'John Carter', 'order': 1}	{'credit_id': '52fe479ac3a36847f81398a3', 'name': 'Andrew A. Kosove', 'role': 'Director'}

```
In [32]: movies = movies.merge(credits, on='title')
```

```
In [3]: movies.head()
# budget
# homepage
# id
# original_language
# original_title
# popularity
# production_company
# production_countries
# release_date(not sure)
```

```
In [33]: movies = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
```

```
In [8]: movies.head()
```

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is dispatched with a military team as a courier of vital secrets to a hostile and terrifyingly advanced native tribe.	{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}	{'id': 1463, 'name': 'culture clash'}, {'id': ...}	{'cast_id': 242, 'character': 'Jake Sully', 'order': 1}	{'credit_id': '52fe48009251416c750aca23', 'name': 'James Cameron', 'role': 'Director'}
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, has returned to his crew and is now the most powerful man in the Caribbean.	{'id': 12, 'name': 'Adventure'}, {'id': 14, 'name': 'Fantasy'}	{'id': 270, 'name': 'ocean'}, {'id': 726, 'name': 'pirates'}	{'cast_id': 4, 'character': 'Captain Jack Sparrow', 'order': 1}	{'credit_id': '52fe4232c3a36847f800b579', 'name': 'Gore Verbinski', 'role': 'Director'}
2	206647	Spectre	A cryptic message from Bond's past sends him on a new adventure.	{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}	{'id': 470, 'name': 'spy'}, {'id': 818, 'name': 'thriller'}	{'cast_id': 1, 'character': 'James Bond', 'order': 1}	{'credit_id': '54805967c3a36829b5002c41', 'name': 'Sam Mendes', 'role': 'Director'}
3	49026	The Dark Knight Rises	Following the death of District Attorney Harvey Dent, Batman deduces a more powerful villain has manipulated the events of the last two years.	{'id': 28, 'name': 'Action'}, {'id': 80, 'name': 'Crime'}	{'id': 849, 'name': 'dc comics'}, {'id': 853, 'name': 'superhero'}	{'cast_id': 2, 'character': 'Bruce Wayne / Batman', 'order': 1}	{'credit_id': '52fe4781c3a36847f81398c3', 'name': 'Christopher Nolan', 'role': 'Director'}
4	49529	John Carter	John Carter is a war-weary, former military captain, plagued by traumatic experiences from a long and troubled past.	{'id': 28, 'name': 'Action'}, {'id': 12, 'name': 'Adventure'}	{'id': 818, 'name': 'western'}, {'id': ...}	{'cast_id': 5, 'character': 'John Carter', 'order': 1}	{'credit_id': '52fe479ac3a36847f81398a3', 'name': 'Andrew A. Kosove', 'role': 'Director'}

```
In [12]: import ast
```

Input output

```
In [14]: import ast

In [34]: def convert(text):
          L = []
          for i in ast.literal_eval(text):
              L.append(i['name'])
          return L

In [35]: movies.dropna(inplace=True)

In [36]: movies['genres'] = movies['genres'].apply(convert)
movies.head()
```

```
Out[36]:
```

	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": "52fe479ac3a36847f8139aa3", "de...

```
In [37]: movies['keywords'] = movies['keywords'].apply(convert)
movies.head()
```

```
Out[37]:
```

	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 l...	[{"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": "52fe479ac3a36847f8139aa3", "de...

```
In [ ]: import ast
ast.literal_eval('{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": ...

In [38]: 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 [39]: movies['cast'] = movies['cast'].apply(convert)
movies.head()
```

```
Out[39]:
```

	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 Weave...	[{"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, R...	[{"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 l...	[Christian Bale, Michael Caine, Gary Oldman, A...	[{"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": "52fe479ac3a36847f8139aa3", "de...

Input output

```
In [40]: movies['cast'] = movies['cast'].apply(lambda x:x[0:3])
```

```
In [41]: def fetch_director(text):
L = []
for i in ast.literal_eval(text):
    if i['job'] == 'Director':
        L.append(i['name'])
return L
```

```
In [42]: 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	overview	genres	keywords	cast	crew
1183	9889	Shallow Hal	A shallow man falls in love with a 300 pound w...	[Comedy, Romance]	[overweight, beauty, hypnosis, overweight man...	[Gwyneth Paltrow, Jack Black, Jason Alexander]	[Bobby Farrelly, Peter Farrelly]	
4215	15976	The Bubble	The movie follows a group of young friends in ...	[Drama, Romance]	[gay]	[Ohad Knoller, Youssef Sweid, Daniella Wicner]	[Eytan Fox]	
1571	22947	Up in the Air	George Clooney plays the dry cynical character...	[Drama, Romance]	[suitcase, business, omaha, on the road, downs...	[George Clooney, Vera Farmiga, Anna Kendrick]	[Jason Reitman]	
205	58574	Sherlock Holmes: A Game of Shadows	There is a new criminal mastermind at large (P...	[Adventure, Action, Crime, Mystery]	[detective inspector, steampunk, criminal mast...	[Robert Downey Jr., Jude Law, Jared Harris]	[Guy Ritchie]	
352	10674	Mulan	A tomboyish girl disguises herself as a young ...	[Animation, Family, Adventure]	[homeland, musical, training, daughter, cricke...	[Eddie Murphy, Jackie Chan, Ming-Na Wen]	[Tony Bancroft, Barry Cook]	

```
In [43]: def collapse(L):
L1 = []
for i in L:
    L1.append(i.replace(" ", ""))
return L1
```

```
In [44]: 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 [26]: movies.head()
```

Out[26]:		movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacemar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]	
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mif, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]	
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dcomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]	
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]	

```
In [45]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

```
In [46]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies
```

```
In [47]: new = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])
#new.head()
```

```
In [48]: new['tags'] = new['tags'].apply(lambda x: " ".join(x))
new.head()
```

Out[48]:		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 [49]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')
```

```
In [50]: vector = cv.fit_transform(new['tags']).toarray()
```

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

```
In [51]: from sklearn.metrics.pairwise import cosine_similarity
```

Input output

```
In [ ]: similarity
```

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

```
In [53]: 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)
```

```
In [54]: recommend('Gandhi')
```

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

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

```
In [56]: pickle.dump(new,open('movie_list.pkl','wb'))  
         pickle.dump(similarity,open('similarity.pkl','wb'))
```

```
In [ ]:
```