

```

In [4]: import pandas as pd
import numpy as np

In [5]: movies = pd.read_csv('tmdb_5000_credits.csv')
credits = pd.read_csv('tmdb_5000_movies.csv')

In [6]: movies.head(2)

Out[6]:
  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": "52fe4232c3a36847800b579", "de...

In [7]: movies.shape

Out[7]:
(4893, 4)

In [8]: credits.head()

Out[8]:
  budget      genres      homepage      id      keywords      original_language      original_title      popularity      production_companies      production_countries      release_date      revenue      runtime      spr...
0    237000000  [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://www.avatarmovie.com/  19995  [{"id": 1463, "name": "culture clash"}, {"id": 14, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  en  Avatar  In the 22nd century, a paraplegic Marine is di...  150.437577  [{"name": "Ingenious Film Partners", "id": 289... [{"iso_3166_1": "US", "name": "United States o...  2009-12-10  2787965087  162.0  [{"t...
1    300000000  [{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://disney.go.com/disneypictures/pirates/  285  [{"id": 270, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  en  Pirates of the Caribbean: At World's End  Captain Barbossa, long believed to be dead, ha...  139.082615  [{"name": "Walt Disney Pictures", "id": 2}, {"name": "United States o...  2007-05-19  961000000  169.0  [{"t...
2    245000000  [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://www.sonypictures.com/movies/spectre/  206647  [{"id": 470, "name": "spy"}, {"id": 818, "name": "terrorism"}, {"id": 853, "name": "terrorism"}]  en  Spectre  A cryptic message from Bond's past sends him o...  107.376788  [{"name": "Columbia Pictures", "id": 5}, {"name": "United Kingdom...  2015-10-26  880674609  148.0  [{"t...
3    250000000  [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}, {"id": 90, "name": "Drama"}, {"id": 93, "name": "Thriller"}]  http://www.thedarkknightrises.com/  49026  [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "terrorism"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}]  en  The Dark Knight Rises  Following the death of District Attorney Harve...  112.312950  [{"name": "Legendary Pictures", "id": 923}, {"name": "United States o...  2012-07-16  1084939099  165.0  [{"t...
4    260000000  [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}, {"id": 90, "name": "Drama"}, {"id": 93, "name": "Thriller"}]  http://movies.disney.com/john-carter  49529  [{"id": 818, "name": "based on novel"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}, {"id": 883, "name": "terrorism"}]  en  John Carter  John Carter is a war-weary, former military ca...  43.926995  [{"name": "Walt Disney Pictures", "id": 2}] [{"iso_3166_1": "US", "name": "United States o...  2012-03-07  284139100  132.0  [{"t...

In [9]: movies = movies.merge(credits, on='title')

In [10]: movies.head()
# budget
# homepage
# id
# original_language
# original_title
# popularity
# production_company
# production_countries
# release-date(not sure)

Out[10]:
  movie_id      title      cast      crew      budget      genres      homepage      id      keywords      original_language      ...      production_companies      production_countries      release_date
0    19995      Avatar  [{"cast_id": 242, "character": "Jake Sully", "... [{"credit_id": "52fe48009251416c750aca23", "de...  237000000  [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://www.avatarmovie.com/  19995  [{"id": 1463, "name": "culture clash"}, {"id": 14, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  en  ...  [{"name": "Ingenious Film Partners", "id": 289... [{"iso_3166_1": "US", "name": "United States o...  2009-12-10
1    285  Pirates of the Caribbean: At World's End  [{"cast_id": 4, "character": "Captain Jack Spa... [{"credit_id": "52fe4232c3a36847800b579", "de...  300000000  [{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://disney.go.com/disneypictures/pirates/  285  [{"id": 270, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  en  ...  [{"name": "Walt Disney Pictures", "id": 2}, {"name": "United States o...  2007-05-19
2    206647      Spectre  [{"cast_id": 1, "character": "James Bond", "cr... [{"credit_id": "54805967c3a36829b5002c41", "de...  245000000  [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://www.sonypictures.com/movies/spectre/  206647  [{"id": 470, "name": "spy"}, {"id": 818, "name": "terrorism"}, {"id": 853, "name": "terrorism"}]  en  ...  [{"name": "Columbia Pictures", "id": 5}, {"name": "United Kingdom...  2015-10-26
3    49026      The Dark Knight Rises  [{"cast_id": 2, "character": "Bruce Wayne / Ba... [{"credit_id": "52fe4781c3a3684781398c3", "de...  250000000  [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}, {"id": 90, "name": "Drama"}, {"id": 93, "name": "Thriller"}]  http://www.thedarkknightrises.com/  49026  [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "terrorism"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}]  en  ...  [{"name": "Legendary Pictures", "id": 923}, {"name": "United States o...  2012-07-16
4    49529      John Carter  [{"cast_id": 5, "character": "John Carter", "c... [{"credit_id": "52fe479ac3a3684781398a3", "de...  260000000  [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  http://movies.disney.com/john-carter  49529  [{"id": 818, "name": "based on novel"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}, {"id": 883, "name": "terrorism"}]  en  ...  [{"name": "Walt Disney Pictures", "id": 2}] [{"iso_3166_1": "US", "name": "United States o...  2012-03-07

5 rows x 23 columns

In [11]: movies = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]

In [12]: movies.head()

Out[12]:
  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, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  [{"id": 1463, "name": "culture clash"}, {"id": 14, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  [{"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, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  [{"id": 270, "name": "ocean"}, {"id": 726, "name": "space"}, {"id": 818, "name": "spy"}, {"id": 853, "name": "terrorism"}]  [{"cast_id": 4, "character": "Captain Jack Spa... [{"credit_id": "52fe4232c3a36847800b579", "de...
2    206647      Spectre  A cryptic message from Bond's past sends him o... [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  [{"id": 470, "name": "spy"}, {"id": 818, "name": "terrorism"}, {"id": 853, "name": "terrorism"}]  [{"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, "name": "Crime"}, {"id": 90, "name": "Drama"}, {"id": 93, "name": "Thriller"}]  [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "terrorism"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}]  [{"cast_id": 2, "character": "Bruce Wayne / Ba... [{"credit_id": "52fe4781c3a3684781398c3", "de...
4    49529      John Carter  John Carter is a war-weary, former military ca... [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]  [{"id": 818, "name": "based on novel"}, {"id": 878, "name": "spy"}, {"id": 883, "name": "terrorism"}, {"id": 883, "name": "terrorism"}]  [{"cast_id": 5, "character": "John Carter", "c... [{"credit_id": "52fe479ac3a3684781398a3", "de...

In [15]: #removing null values
movies.isnull().sum()

Out[15]:
movie_id      0
title          0
overview       0
genres         0
keywords       0
cast           0
crew           0
dtype: int64

In [16]: movies.dropna(inplace=True)

In [19]: movies.duplicated().sum()

Out[19]:
0

In [20]: import ast

In [21]: #getting name from dictionary
def convert(text):
    L = []
    for i in ast.literal_eval(text):
        L.append(i['name'])
    return L

In [22]: movies['genres'] = movies['genres'].apply(convert)
movies.head()

Out[22]:
  movie_id      title      overview      genres      keywords      cast      crew
0    19995      Avatar  In the 22nd century, a paraplegic Marine is di... [Action
```