

```
In [1]: ┌─▶ import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt

      import ast
      import warnings
      warnings.filterwarnings('ignore')
```

```
In [2]: ┌─▶ movies = pd.read_csv('tmdb_5000_movies.csv')
      credits = pd.read_csv('tmdb_5000_credits.csv')
```

```
In [3]: ┌─▶ movies.head()
```

Out[3]:

	budget	genres	homepage	id	keywords	original_language	original_title	overv...
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": ...]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...]	en	Avatar	In 2 centu...
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": ...]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...]	en	Pirates of the Caribbean: At World's End	Cap Barbo...
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": ...]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": ...]	en	Spectre	A cry mess...

```
In [ ]: ┌─▶
```

In [4]: ⏷ credits.head()

Out[4]:

	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 Sparrow", "credit_id": "52fe4232c3a36847f800b579", "de...}	
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "credit_id": "54805967c3a36829b5002c41", "de...}	
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Batman", "credit_id": "52fe4781c3a36847f81398c3", "de...}	
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "credit_id": "52fe479ac3a36847f813eaa3", "de...}	

In [ ]: ⏷

In [5]: ⏷ movies.shape

Out[5]: (4803, 20)

In [6]: ⏷ credits.shape

Out[6]: (4803, 4)

In [ ]: ⏷

## Merge the dataframe

```
In [7]: ┌─ movies.columns
```

```
Out[7]: Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
   'original_title', 'overview', 'popularity', 'production_companies',
   'production_countries', 'release_date', 'revenue', 'runtime',
   'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
   'vote_count'],
  dtype='object')
```

```
In [8]: ┌─ credits.columns
```

```
Out[8]: Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')
```

```
In [9]: ┌─ movies['title']
```

```
Out[9]: 0          Avatar
1      Pirates of the Caribbean: At World's End
2          Spectre
3      The Dark Knight Rises
4          John Carter
...
4798          El Mariachi
4799          Newlyweds
4800      Signed, Sealed, Delivered
4801          Shanghai Calling
4802          My Date with Drew
Name: title, Length: 4803, dtype: object
```

In [10]: credits['title']

```
Out[10]: 0           Avatar
1   Pirates of the Caribbean: At World's End
2           Spectre
3   The Dark Knight Rises
4       John Carter
...
4798      El Mariachi
4799      Newlyweds
4800 Signed, Sealed, Delivered
4801      Shanghai Calling
4802 My Date with Drew
Name: title, Length: 4803, dtype: object
```

In [ ]:

In [11]: movies = movies.merge(credits, on='title')

In [12]: movies.head(2)

Out[12]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	...
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...	1:
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...	1:

2 rows × 23 columns

In [13]: ⏎ movies.shape

Out[13]: (4809, 23)

In [ ]: ⏎

## Analysing the Dataset

In [14]: ► movies.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 4809 entries, 0 to 4808
Data columns (total 23 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   budget            4809 non-null    int64  
 1   genres             4809 non-null    object  
 2   homepage          1713 non-null    object  
 3   id                4809 non-null    int64  
 4   keywords          4809 non-null    object  
 5   original_language 4809 non-null    object  
 6   original_title    4809 non-null    object  
 7   overview          4806 non-null    object  
 8   popularity         4809 non-null    float64 
 9   production_companies 4809 non-null    object  
 10  production_countries 4809 non-null    object  
 11  release_date      4808 non-null    object  
 12  revenue            4809 non-null    int64  
 13  runtime            4807 non-null    float64 
 14  spoken_languages  4809 non-null    object  
 15  status              4809 non-null    object  
 16  tagline            3965 non-null    object  
 17  title               4809 non-null    object  
 18  vote_average       4809 non-null    float64 
 19  vote_count          4809 non-null    int64  
 20  movie_id            4809 non-null    int64  
 21  cast                4809 non-null    object  
 22  crew                4809 non-null    object  
dtypes: float64(3), int64(5), object(15)
memory usage: 901.7+ KB
```

In [ ]: ►

In [15]: ⏷ movies.isna().sum()

Out[15]:

budget	0
genres	0
homepage	3096
id	0
keywords	0
original_language	0
original_title	0
overview	3
popularity	0
production_companies	0
production_countries	0
release_date	1
revenue	0
runtime	2
spoken_languages	0
status	0
tagline	844
title	0
vote_average	0
vote_count	0
movie_id	0
cast	0
crew	0

dtype: int64

In [17]: ► movies.isna().mean()\*100

Out[17]:

budget	0.000000
genres	0.000000
homepage	64.379289
id	0.000000
keywords	0.000000
original_language	0.000000
original_title	0.000000
overview	0.062383
popularity	0.000000
production_companies	0.000000
production_countries	0.000000
release_date	0.020794
revenue	0.000000
runtime	0.041589
spoken_languages	0.000000
status	0.000000
tagline	17.550426
title	0.000000
vote_average	0.000000
vote_count	0.000000
movie_id	0.000000
cast	0.000000
crew	0.000000

dtype: float64

In [ ]: ►

In [18]: ► movies.duplicated().sum()

Out[18]: 0

In [ ]: ►

In [ ]:

## Choosing relevant features for recommendation engine

In [19]:

movies.columns

```
Out[19]: Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
       'original_title', 'overview', 'popularity', 'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
       'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
       'vote_count', 'movie_id', 'cast', 'crew'],
      dtype='object')
```

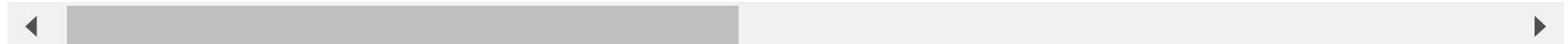
In [21]:

movies[:1]

Out[21]:

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	product
237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Sci-Fi"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": ...}]	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"na..."]	

rows × 23 columns



## Final Selected Features

- movie\_id
- title
- overview
- genres
- keywords
- cast

- crew

In [ ]:

In [29]:

```
df = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
df.head(3)
```

Out[29]:

	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": "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, "name": "..."}]	[{"id": 470, "name": "spy"}, {"id": 818, "name..."}]	[{"cast_id": 1, "character": "James Bond", "cr..."}]	[{"credit_id": "54805967c3a36829b5002c41", "de..."}]

In [31]:

Out[31]:

```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

In [38]:

Out[38]:

(4809, 7)

In [39]: ⏎ df.dropna(inplace=True)

In [40]: ⏎ df.shape

Out[40]: (4806, 7)

In [41]: ⏎ df.isna().sum()

Out[41]:

	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...]

In [42]: ⏎ df.head(3)

Out[42]:

	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...]

In [ ]: ⏎

## Final DF - 'movie\_id' + 'title' + 'tags'

### Exploring every single feature data

```
In [51]: df['overview']
```

```
Out[51]: 0      In the 22nd century, a paraplegic Marine is di...
 1      Captain Barbosa, long believed to be dead, ha...
 2      A cryptic message from Bond's past sends him o...
 3      Following the death of District Attorney Harve...
 4      John Carter is a war-weary, former military ca...
 ...
 4804    El Mariachi just wants to play his guitar and ...
 4805    A newlywed couple's honeymoon is upended by th...
 4806    "Signed, Sealed, Delivered" introduces a dedic...
 4807    When ambitious New York attorney Sam is sent t...
 4808    Ever since the second grade when he first saw ...
Name: overview, Length: 4806, dtype: object
```

```
In [52]: df['overview'][0]
```

```
Out[52]: '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 [53]: df['genres']

```
Out[53]: 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...  
4      [{"id": 28, "name": "Action"}, {"id": 12, "nam...  
     ...  
4804     [{"id": 28, "name": "Action"}, {"id": 80, "nam...  
4805     [{"id": 35, "name": "Comedy"}, {"id": 10749, "...  
4806     [{"id": 35, "name": "Comedy"}, {"id": 18, "nam...  
4807           []  
4808           [{"id": 99, "name": "Documentary"}]  
Name: genres, Length: 4806, dtype: object
```

In [54]: df['genres'][0]

```
Out[54]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

In [ ]: df['genres'][0]

In [55]: df['keywords']

```
Out[55]: 0      [{"id": 1463, "name": "culture clash"}, {"id":...  
1      [{"id": 270, "name": "ocean"}, {"id": 726, "na...  
2      [{"id": 470, "name": "spy"}, {"id": 818, "name...  
3      [{"id": 849, "name": "dc comics"}, {"id": 853,...  
4      [{"id": 818, "name": "based on novel"}, {"id":...  
     ...  
4804     [{"id": 5616, "name": "united states\u2013mexi...  
4805           []  
4806     [{"id": 248, "name": "date"}, {"id": 699, "nam...  
4807           []  
4808     [{"id": 1523, "name": "obsession"}, {"id": 224...  
Name: keywords, Length: 4806, dtype: object
```

In [56]: df['keywords'][0]

Out[56]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldier"}, {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165431, "name": "anti war"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "name": "mind and soul"}, {"id": 209714, "name": "3d"}]'

In [ ]: df

In [57]: df['cast']

Out[57]: 0 [{"cast\_id": 242, "character": "Jake Sully", "...}  
1 [{"cast\_id": 4, "character": "Captain Jack Spa...}  
2 [{"cast\_id": 1, "character": "James Bond", "cr...}  
3 [{"cast\_id": 2, "character": "Bruce Wayne / Ba...}  
4 [{"cast\_id": 5, "character": "John Carter", "c...}  
...  
4804 [{"cast\_id": 1, "character": "El Mariachi", "c...}  
4805 [{"cast\_id": 1, "character": "Buzzy", "credit\_...}  
4806 [{"cast\_id": 8, "character": "Oliver O\u2019To...}  
4807 [{"cast\_id": 3, "character": "Sam", "credit\_id...}  
4808 [{"cast\_id": 3, "character": "Herself", "credi...}  
Name: cast, Length: 4806, dtype: object

In [58]: ► df['cast'][0]

In [1]:

In [59]: ► df['crew']

```
Out[59]: 0      [{"credit_id": "52fe48009251416c750aca23", "de...  
1      [{"credit_id": "52fe4232c3a36847f800b579", "de...  
2      [{"credit_id": "54805967c3a36829b5002c41", "de...  
3      [{"credit_id": "52fe4781c3a36847f81398c3", "de...  
4      [{"credit_id": "52fe479ac3a36847f813eaa3", "de...  
       ...  
4804     [{"credit_id": "52fe44eec3a36847f80b280b", "de...  
4805     [{"credit_id": "52fe487dc3a368484e0fb013", "de...  
4806     [{"credit_id": "52fe4df3c3a36847f8275ecf", "de...  
4807     [{"credit_id": "52fe4ad9c3a368484e16a36b", "de...  
4808     [{"credit_id": "58ce021b9251415a390165d9", "de...  
Name: crew, Length: 4806, dtype: object
```

In [60]: df['crew'][0]

```
Out[60]: '[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "539c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "5544ee3b925141499f0008fc", "department": "Sound", "gender": 2, "id": 1729, "job": "Original Music Composer", "name": "James Horner"}, {"credit_id": "52fe48009251416c750ac9c3", "department": "Directing", "gender": 2, "id": 2710, "job": "Director", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750ac9d9", "department": "Writing", "gender": 2, "id": 2710, "job": "Writer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca17", "department": "Editing", "gender": 2, "id": 2710, "job": "Editor", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca29", "department": "Production", "gender": 2, "id": 2710, "job": "Producer", "name": "James Cameron"}, {"credit_id": "52fe48009251416c750aca3f", "department": "Writing", "gender": 2, "id": 2710, "job": "Screenplay", "name": "James Cameron"}, {"credit_id": "539c4987c3a36810ba0021a4", "department": "Art", "gender": 2, "id": 7236, "job": "Art Direction", "name": "Andrew Menzies"}, {"credit_id": "549598c3c3a3686ae9004383", "department": "Visual Effects", "gender": 0, "id": 6690, "job": "Visual Effects Producer", "name": "Jill Brooks"}, {"credit_id": "52fe48009251416c750aca4b", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "539c4a4cc3a36810c9002101", "department": "Production", "gender": 1, "id": 1262, "job": "Casting", "name": "Mali Finn"}, {"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "id": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a36810e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Sound", "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900, "job": "Supervising Sound Editor", "name": "Christopher Boyes"}]
```

In [ ]:

In [61]: df[:2]

Out[61]:

	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...

# Genres

```
In [62]: ┏ df['genres']
```

```
Out[62]: 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...
 4      [{"id": 28, "name": "Action"}, {"id": 12, "nam...
 ...
 4804     [{"id": 28, "name": "Action"}, {"id": 80, "nam...
 4805     [{"id": 35, "name": "Comedy"}, {"id": 10749, "...
 4806     [{"id": 35, "name": "Comedy"}, {"id": 18, "nam...
 4807             []
 4808             [{"id": 99, "name": "Documentary"}]
Name: genres, Length: 4806, dtype: object
```

```
In [63]: ┏ df['genres'][0]
```

```
Out[63]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

```
In [64]: ┏ type(df['genres'][0])
```

```
Out[64]: str
```

```
In [67]: ┏ import ast
```

```
In [68]: ┏ ast.literal_eval(df['genres'][0])
```

```
Out[68]: [{'id': 28, 'name': 'Action'},
 {'id': 12, 'name': 'Adventure'},
 {'id': 14, 'name': 'Fantasy'},
 {'id': 878, 'name': 'Science Fiction'}]
```

```
In [69]: ┏━ for i in ast.literal_eval(df['genres'][0]):  
      print(i)
```

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

```
In [ ]: ┏━
```

```
In [70]: ┏━ lst = ast.literal_eval(df['genres'][0])  
lst
```

```
Out[70]: [ {'id': 28, 'name': 'Action'},  
          {'id': 12, 'name': 'Adventure'},  
          {'id': 14, 'name': 'Fantasy'},  
          {'id': 878, 'name': 'Science Fiction'}]
```

```
In [71]: ┏━ type(lst)
```

```
Out[71]: list
```

```
In [73]: ┏━ d = lst[0]  
d
```

```
Out[73]: {'id': 28, 'name': 'Action'}
```

```
In [74]: ┏━ type(d)
```

```
Out[74]: dict
```

```
In [75]: ┏━ d['name']
```

```
Out[75]: 'Action'
```

In [ ]:

```
for i in ast.literal_eval(df['genres'][0]):  
    print(i['name'])
```

Action  
Adventure  
Fantasy  
Science Fiction

In [83]:

```
def fetch_genres(data):  
  
    lst = []  
    for i in ast.literal_eval(data):  
        lst.append(i['name'])  
  
    return lst
```

In [88]:

```
fetch_genres(df['genres'][0])
```

Out[88]: ['Action', 'Adventure', 'Fantasy', 'Science Fiction']

In [ ]:

In [89]: df['genres'].apply(fetch\_genres)

```
Out[89]: 0      [Action, Adventure, Fantasy, Science Fiction]
1          [Adventure, Fantasy, Action]
2          [Action, Adventure, Crime]
3          [Action, Crime, Drama, Thriller]
4          [Action, Adventure, Science Fiction]
...
4804      [Action, Crime, Thriller]
4805      [Comedy, Romance]
4806      [Comedy, Drama, Romance, TV Movie]
4807      []
4808      [Documentary]
Name: genres, Length: 4806, dtype: object
```

In [90]: df[:2]

Out[90]:

	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..."}]

In [91]: df['genres'] = df['genres'].apply(fetch\_genres)

In [93]: ⏷ df[:4]

Out[93]:

	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...]

In [ ]: ⏷

## Keywords

In [94]: df['keywords']

```
Out[94]: 0      [{"id": 1463, "name": "culture clash"}, {"id":...  
1      [{"id": 270, "name": "ocean"}, {"id": 726, "na...  
2      [{"id": 470, "name": "spy"}, {"id": 818, "name...  
3      [{"id": 849, "name": "dc comics"}, {"id": 853,...  
4      [{"id": 818, "name": "based on novel"}, {"id":...  
     ...  
4804     [{"id": 5616, "name": "united states\u2013mexi...  
4805             []  
4806     [{"id": 248, "name": "date"}, {"id": 699, "nam...  
4807             []  
4808     [{"id": 1523, "name": "obsession"}, {"id": 224...  
Name: keywords, Length: 4806, dtype: object
```

In [95]: df['keywords'][0]

```
Out[95]: '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386, "name": "space wa  
r"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "society"}, {"id": 3801, "name": "space  
travel"}, {"id": 9685, "name": "futuristic"}, {"id": 9840, "name": "romance"}, {"id": 9882, "name": "spa  
ce"}, {"id": 9951, "name": "alien"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien plane  
t"}, {"id": 10987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldier"},  
 {"id": 14643, "name": "battle"}, {"id": 14720, "name": "love affair"}, {"id": 165431, "name": "anti wa  
r"}, {"id": 193554, "name": "power relations"}, {"id": 206690, "name": "mind and soul"}, {"id": 209714,  
"name": "3d"}]'
```

In [ ]:

In [96]: def fetch\_keywords(data):

```
lst = []  
for i in ast.literal_eval(data):  
    lst.append(i['name'])  
  
return lst
```

In [98]: ► fetch\_keywords(df['keywords'][0])

Out[98]: ['culture clash',  
 'future',  
 'space war',  
 'space colony',  
 'society',  
 'space travel',  
 'futuristic',  
 'romance',  
 'space',  
 'alien',  
 'tribe',  
 'alien planet',  
 'cgi',  
 'marine',  
 'soldier',  
 'battle',  
 'love affair',  
 'anti war',  
 'power relations',  
 'mind and soul',  
 '3d']

In [ ]: ►

In [100]: ► df['keywords'].apply(fetch\_keywords)

Out[100]: 0 [culture clash, future, space war, space colon...  
 1 [ocean, drug abuse, exotic island, east india ...  
 2 [spy, based on novel, secret agent, sequel, mi...  
 3 [dc comics, crime fighter, terrorist, secret i...  
 4 [based on novel, mars, medallion, space travel...  
 ...  
 4804 [united states-mexico barrier, legs, arms, pap...  
 4805 []  
 4806 [date, love at first sight, narration, investi...  
 4807 []  
 4808 [obsession, camcorder, crush, dream girl]  
 Name: keywords, Length: 4806, dtype: object

In [101]: df['keywords'] = df['keywords'].apply(fetch\_keywords)

In [102]: df[:3]

Out[102]:

	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 Barbosa, 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..."}

In [ ]: df

In [ ]: