```
# 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 using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

```
import time, sys
from sklearn.cluster import KMeans
sess = pd.read_csv('tmdb_5000_credits.csv')
movies = pd.read_csv('tmdb_5000_movies.csv')
print('Loaded')
```

```
Matplotlib is building the font cache; this may take a moment.
```

```
Loaded
```

```
movies.shape
```

```
(4803, 20)
```

```
movies.sample(5)
```

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	release_date	revenue	runtime	spoken_languages	status	tagline	title	vote_average	vote_count
1	10000000	[{"id": 18, "name": "Drama"}, {"id": 35, "name": "Comedy"}, {"id": 10749, "name": "Romance"}]	http://www.sonypictures.com/movies/thegirlnextdoor/	10986	[{"id": 1416, "name": "teenager"}, {"id": 2101, "name": "sex"}, {"id": 4379, "name": "sexual awakening"}, {"id": 9712, "name": "high school"}, {"id": 10183, "name": "independent film"}]	en	The Girl Next Door	Matt Kemp, a 19-year-old high school senior, is a straight A student who has his future all planned out. He's going to graduate with his class, get into a top college, and then pursue a lucrative career. Everything changes when the girl next door, Danielle, moves in. Danielle is a 21-year-old who has been "kept" by her mobster boyfriend since she was 16. Matt and Danielle become friends and he soon discovers that she is not the only one in the family who has secrets. Matt's life is turned upside down when he falls in love with Danielle and must decide between the life he has planned and the life he wants.	10.137071	[{"name": "Franchise Pictures", "id": 52}]	2004-12-17	15000000	97.0	[{"iso_639_1": "en", "name": "English"}]	Released	Life's a party.	The Girl Next Door	6.5	1000
2	18000000	[{"id": 35, "name": "Comedy"}]	http://www.gravitypictures.com/Movies/Detail/Id/1000	14337	[{"id": 1378, "name": "beach"}, {"id": 2164, "name": "pirate"}, {"id": 4290, "name": "extraterrestrial"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10195, "name": "duringcreditsstinger"}]	en	Pirates of the Caribbean: Dead Man's Chest	Captain Barbossa, long believed to be dead, has come back to life and is haunting the Black Pearl with a crew of the living dead.	29.432601	[{"name": "Walt Disney Pictures", "id": 2}, {"name": "Jerry Bruckheimer Films", "id": 130}]	2006-07-19	1066179725	151.0	[{"iso_639_1": "en", "name": "English"}]	Released	Dead Men Tell No Tales.	Pirates of the Caribbean: Dead Man's Chest	7.0	6000
3	14000000	[{"id": 35, "name": "Comedy"}]	http://www.thehangoverlaws.com/	49026	[{"id": 1448, "name": "bachelor party"}, {"id": 2060, "name": "hangover"}, {"id": 3070, "name": "jealousy"}, {"id": 3250, "name": "partying"}, {"id": 4379, "name": "sexual awakening"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10513, "name": "bachelor"}]	en	The Hangover Part II	Two years after the events of the Las Vegas wedding, Doug, Stu, Phil and Alan meet up with Doug's fiancée and head to Thailand for a bachelor party. Unfortunately, trouble seems to follow the guys and they find themselves in a lot more than they bargained for on their journey.	29.395236	[{"name": "Legendary Pictures", "id": 923}, {"name": "Warner Bros.", "id": 6194}, {"name": "Green Hat Films (I)", "id": 10168}]	2011-05-26	586772172	102.0	[{"iso_639_1": "en", "name": "English"}]	Released	Bigger, longer, uncut	The Hangover Part II	6.4	4000
4	20000000	[{"id": 28, "name": "Action"}, {"id": 35, "name": "Comedy"}, {"id": 53, "name": "Thriller"}]	http://www.sonypictures.com/movies/jackass3d/	48223	[{"id": 1448, "name": "bachelor party"}, {"id": 2060, "name": "hangover"}, {"id": 3070, "name": "jealousy"}, {"id": 3250, "name": "partying"}, {"id": 4379, "name": "sexual awakening"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10513, "name": "bachelor"}]	en	Jackass 3D	Johnny Knoxville and the crew of the MTV series Jackass are back with a vengeance in this 3-D sequel. Expect the unexpected as they perform stunts, pranks and gross-out gags that are guaranteed to make you laugh, cringe and scream.	8.153988	[{"name": "Dickhouse Productions", "id": 4023}, {"name": "Relativity Media", "id": 9168}]	2010-10-14	117586000	92.0	[{"iso_639_1": "en", "name": "English"}]	Released	Expect the unexpected.	Jackass 3D	6.5	1000
5	12000000	[{"id": 28, "name": "Action"}, {"id": 35, "name": "Comedy"}, {"id": 53, "name": "Thriller"}]		48223	[{"id": 1448, "name": "bachelor party"}, {"id": 2060, "name": "hangover"}, {"id": 3070, "name": "jealousy"}, {"id": 3250, "name": "partying"}, {"id": 4379, "name": "sexual awakening"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10513, "name": "bachelor"}]	en	Jackass 3D	Johnny Knoxville and the crew of the MTV series Jackass are back with a vengeance in this 3-D sequel. Expect the unexpected as they perform stunts, pranks and gross-out gags that are guaranteed to make you laugh, cringe and scream.	8.153988	[{"name": "Dickhouse Productions", "id": 4023}, {"name": "Relativity Media", "id": 9168}]	2010-10-14	117586000	92.0	[{"iso_639_1": "en", "name": "English"}]	Released	Expect the unexpected.	Jackass 3D	6.5	1000

```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  ------                --------------  -----
0   budget                4803 non-null   int64
1   genres                4803 non-null   object
2   homepage              1712 non-null   object
3   id                    4803 non-null   int64
4   keywords              4803 non-null   object
5   original_language     4803 non-null   object
6   original_title        4803 non-null   object
7   overview              4800 non-null   object
8   popularity            4803 non-null   float64
9   production_companies  4803 non-null   object
10  production_countries  4803 non-null   object
11  release_date          4802 non-null   object
12  revenue               4803 non-null   int64
13  runtime               4801 non-null   float64
14  spoken_languages      4803 non-null   object
15  status                4803 non-null   object
16  tagline               3959 non-null   object
17  title                 4803 non-null   object
18  vote_average          4803 non-null   float64
19  vote_count            4803 non-null   int64
dtypes: float64(3), int64(6), object(11)
memory usage: 1.1 MB
```

we want to create a movie recommendation system so we will choose only those columns which are important from our project and remove others . eg . genres, id , keywords , title , overview , cast , crew

```
movies = movies[['id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]
movies.sample(5)
```

	id	title	overview	genres	keywords	cast	crew
1	10986	Jeepers Creepers II	In the 23rd century, a paleontologist discovers a cryogenically frozen spaceship containing a crew of 1st century Earth people.	[{"id": 28, "name": "Action"}, {"id": 878, "name": "Science Fiction"}]	[{"id": 1463, "name": "future"}, {"id": 2164, "name": "pirate"}, {"id": 4290, "name": "extraterrestrial"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10195, "name": "duringcreditsstinger"}]	[{"cast_id": 1, "character": "Jack Colvin", "credit_id": "52fe4a7dc3a368484e0fb7d3", "gender": 2, "id": 5176, "name": "Gordon Jump"}, {"cast_id": 2, "character": "Ms. Hoskins", "credit_id": "52fe4a7dc3a368484e0fb7d7", "gender": 1, "id": 5177, "name": "Karen Black"}, {"cast_id": 3, "character": "Jeepers Creeper", "credit_id": "52fe4a7dc3a368484e0fb7db", "gender": 2, "id": 5178, "name": "Eric Nenninger"}, {"cast_id": 4, "character": "Sheriff Jacobs", "credit_id": "52fe4a7dc3a368484e0fb7df", "gender": 2, "id": 5179, "name": "Ron Gilbert"}, {"cast_id": 5, "character": "Deputy Tom", "credit_id": "52fe4a7dc3a368484e0fb7e3", "gender": 2, "id": 5180, "name": "Tom Gallant"}]	[{"credit_id": "52fe4a7dc3a368484e0fb7e9", "department": "Writing", "gender": 2, "id": 5181, "name": "Victor Salva", "job": "Screenplay"}, {"credit_id": "52fe4a7dc3a368484e0fb7ef", "department": "Sound", "gender": 2, "id": 5182, "name": "Jim Dolan", "job": "Sound Mixer"}, {"credit_id": "52fe4a7dc3a368484e0fb7f5", "department": "Sound", "gender": 2, "id": 5183, "name": "Ronald Gilbert", "job": "Sound Mixer"}, {"credit_id": "52fe4a7dc3a368484e0fb7fb", "department": "Writing", "gender": 2, "id": 5184, "name": "Victor Salva", "job": "Story"}, {"credit_id": "52fe4a7dc3a368484e0fb801", "department": "Production", "gender": 2, "id": 5185, "name": "Barry Bernardi", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb807", "department": "Production", "gender": 2, "id": 5186, "name": "Lawrence Mortorano", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb80d", "department": "Production", "gender": 2, "id": 5187, "name": "Tim Zearfoss", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb813", "department": "Production", "gender": 2, "id": 5188, "name": "Frank Mancuso", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb819", "department": "Production", "gender": 2, "id": 5189, "name": "Frank Mancuso", "job": "Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb81f", "department": "Production", "gender": 2, "id": 5190, "name": "Michael Manasseri", "job": "Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb825", "department": "Production", "gender": 2, "id": 5191, "name": "Jeffrey Levy-Hinte", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb82b", "department": "Art", "gender": 2, "id": 5192, "name": "Leslie Pope", "job": "Production Design"}, {"credit_id": "52fe4a7dc3a368484e0fb831", "department": "Art", "gender": 1, "id": 5193, "name": "Edi Rhavens", "job": "Set Decorator"}, {"credit_id": "52fe4a7dc3a368484e0fb837", "department": "Art", "gender": 2, "id": 5194, "name": "Nancy Haigh", "job": "Set Decorator"}, {"credit_id": "52fe4a7dc3a368484e0fb83d", "department": "Art", "gender": 2, "id": 5195, "name": "Brian Stonestreet", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb843", "department": "Art", "gender": 2, "id": 5196, "name": "Antoinette Messina", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb849", "department": "Art", "gender": 1, "id": 5197, "name": "Anne Ahl", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb84f", "department": "Art", "gender": 2, "id": 5198, "name": "David Wasco", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb855", "department": "Sound", "gender": 2, "id": 5199, "name": "Sandy Berman", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb85b", "department": "Crew", "gender": 1, "id": 5200, "name": "Deborah Lurie", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb861", "department": "Crew", "gender": 2, "id": 5201, "name": "Maryann Brandon", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb867", "department": "Crew", "gender": 1, "id": 5202, "name": "Mitchell Dana", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb86d", "department": "Editing", "gender": 2, "id": 5203, "name": "Randy Roberts", "job": "Supervising Sound Editor"}, {"credit_id": "52fe4a7dc3a368484e0fb873", "department": "Editing", "gender": 2, "id": 5204, "name": "Steve Visher", "job": "Film Editor"}, {"credit_id": "52fe4a7dc3a368484e0fb879", "department": "Editing", "gender": 2, "id": 5205, "name": "Tracey Bannon", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb87f", "department": "Editing", "gender": 2, "id": 5206, "name": "Peter Miller", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb885", "department": "Editing", "gender": 2, "id": 5207, "name": "Bruce Miller", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb88b", "department": "Editing", "gender": 2, "id": 5208, "name": "Jeff Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb891", "department": "Editing", "gender": 2, "id": 5209, "name": "Jason Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb897", "department": "Editing", "gender": 2, "id": 5210, "name": "Scott Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb89d", "department": "Editing", "gender": 2, "id": 5211, "name": "Eric Hachikian", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8a3", "department": "Editing", "gender": 2, "id": 5212, "name": "Josh Taylor", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8a9", "department": "Editing", "gender": 2, "id": 5213, "name": "Ellen Somers", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8af", "department": "Editing", "gender": 2, "id": 5214, "name": "Aaron Hasz", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8b5", "department": "Editing", "gender": 2, "id": 5215, "name": "Matt Wheeler", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8bb", "department": "Editing", "gender": 2, "id": 5216, "name": "Bruce Wong", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8c1", "department": "Editing", "gender": 2, "id": 5217, "name": "Jonathan Dver", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8c7", "department": "Editing", "gender": 2, "id": 5218, "name": "Jennifer Bender", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8cd", "department": "Editing", "gender": 2, "id": 5219, "name": "Marc Jozefowicz", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8d3", "department": "Editing", "gender": 2, "id": 5220, "name": "Kaboom", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8d9", "department": "Editing", "gender": 2, "id": 5221, "name": "Matt Smith", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8df", "department": "Editing", "gender": 2, "id": 5222, "name": "Jay Prichard", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8e5", "department": "Editing", "gender": 2, "id": 5223, "name": "Tanner Pearson", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8eb", "department": "Editing", "gender": 2, "id": 5224, "name": "Dylan Highsmith", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8f1", "department": "Editing", "gender": 2, "id": 5225, "name": "Winnie Hsu", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8f7", "department": "Editing", "gender": 2, "id": 5226, "name": "Stephen Nemeth", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8fd", "department": "Editing", "gender": 2, "id": 5227, "name": "Gregory Plotkin", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb903", "department": "Editing", "gender": 2, "id": 5228, "name": "Tim Mirkovich", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb909", "department": "Editing", "gender": 2, "id": 5229, "name": "Cinna Dunn", "job": "Trailer"}]"}	[{"id": 28, "name": "Action"}, {"id": 878, "name": "Science Fiction"}]	[{"id": 1463, "name": "future"}, {"id": 2164, "name": "pirate"}, {"id": 4290, "name": "extraterrestrial"}, {"id": 5418, "name": "sequel"}, {"id": 10183, "name": "independent film"}, {"id": 10195, "name": "duringcreditsstinger"}]	[{"cast_id": 1, "character": "Jack Colvin", "credit_id": "52fe4a7dc3a368484e0fb7d3", "gender": 2, "id": 5176, "name": "Gordon Jump"}, {"cast_id": 2, "character": "Ms. Hoskins", "credit_id": "52fe4a7dc3a368484e0fb7d7", "gender": 1, "id": 5177, "name": "Karen Black"}, {"cast_id": 3, "character": "Jeepers Creeper", "credit_id": "52fe4a7dc3a368484e0fb7db", "gender": 2, "id": 5178, "name": "Eric Nenninger"}, {"cast_id": 4, "character": "Sheriff Jacobs", "credit_id": "52fe4a7dc3a368484e0fb7df", "gender": 2, "id": 5179, "name": "Ron Gilbert"}, {"cast_id": 5, "character": "Deputy Tom", "credit_id": "52fe4a7dc3a368484e0fb7e3", "gender": 2, "id": 5180, "name": "Tom Gallant"}]	[{"credit_id": "52fe4a7dc3a368484e0fb7e9", "department": "Writing", "gender": 2, "id": 5181, "name": "Victor Salva", "job": "Screenplay"}, {"credit_id": "52fe4a7dc3a368484e0fb7ef", "department": "Sound", "gender": 2, "id": 5182, "name": "Jim Dolan", "job": "Sound Mixer"}, {"credit_id": "52fe4a7dc3a368484e0fb7f5", "department": "Sound", "gender": 2, "id": 5183, "name": "Ronald Gilbert", "job": "Sound Mixer"}, {"credit_id": "52fe4a7dc3a368484e0fb7fb", "department": "Writing", "gender": 2, "id": 5184, "name": "Victor Salva", "job": "Story"}, {"credit_id": "52fe4a7dc3a368484e0fb801", "department": "Production", "gender": 2, "id": 5185, "name": "Barry Bernardi", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb807", "department": "Production", "gender": 2, "id": 5186, "name": "Lawrence Mortorano", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb80d", "department": "Production", "gender": 2, "id": 5187, "name": "Tim Zearfoss", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb813", "department": "Production", "gender": 2, "id": 5188, "name": "Frank Mancuso", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb819", "department": "Production", "gender": 2, "id": 5189, "name": "Frank Mancuso", "job": "Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb81f", "department": "Production", "gender": 2, "id": 5190, "name": "Michael Manasseri", "job": "Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb825", "department": "Production", "gender": 2, "id": 5191, "name": "Jeffrey Levy-Hinte", "job": "Executive Producer"}, {"credit_id": "52fe4a7dc3a368484e0fb82b", "department": "Art", "gender": 2, "id": 5192, "name": "Leslie Pope", "job": "Production Design"}, {"credit_id": "52fe4a7dc3a368484e0fb831", "department": "Art", "gender": 1, "id": 5193, "name": "Edi Rhavens", "job": "Set Decorator"}, {"credit_id": "52fe4a7dc3a368484e0fb837", "department": "Art", "gender": 2, "id": 5194, "name": "Nancy Haigh", "job": "Set Decorator"}, {"credit_id": "52fe4a7dc3a368484e0fb83d", "department": "Art", "gender": 2, "id": 5195, "name": "Brian Stonestreet", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb843", "department": "Art", "gender": 2, "id": 5196, "name": "Antoinette Messina", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb849", "department": "Art", "gender": 1, "id": 5197, "name": "Anne Ahl", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb84f", "department": "Art", "gender": 2, "id": 5198, "name": "David Wasco", "job": "Art Direction"}, {"credit_id": "52fe4a7dc3a368484e0fb855", "department": "Sound", "gender": 2, "id": 5199, "name": "Sandy Berman", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb85b", "department": "Crew", "gender": 1, "id": 5200, "name": "Deborah Lurie", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb861", "department": "Crew", "gender": 2, "id": 5201, "name": "Maryann Brandon", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb867", "department": "Crew", "gender": 1, "id": 5202, "name": "Mitchell Dana", "job": "Post Production Supervisor"}, {"credit_id": "52fe4a7dc3a368484e0fb86d", "department": "Editing", "gender": 2, "id": 5203, "name": "Randy Roberts", "job": "Supervising Sound Editor"}, {"credit_id": "52fe4a7dc3a368484e0fb873", "department": "Editing", "gender": 2, "id": 5204, "name": "Steve Visher", "job": "Film Editor"}, {"credit_id": "52fe4a7dc3a368484e0fb879", "department": "Editing", "gender": 2, "id": 5205, "name": "Tracey Bannon", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb87f", "department": "Editing", "gender": 2, "id": 5206, "name": "Peter Miller", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb885", "department": "Editing", "gender": 2, "id": 5207, "name": "Bruce Miller", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb88b", "department": "Editing", "gender": 2, "id": 5208, "name": "Jeff Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb891", "department": "Editing", "gender": 2, "id": 5209, "name": "Jason Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb897", "department": "Editing", "gender": 2, "id": 5210, "name": "Scott Bourque", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb89d", "department": "Editing", "gender": 2, "id": 5211, "name": "Eric Hachikian", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8a3", "department": "Editing", "gender": 2, "id": 5212, "name": "Josh Taylor", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8a9", "department": "Editing", "gender": 2, "id": 5213, "name": "Ellen Somers", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8af", "department": "Editing", "gender": 2, "id": 5214, "name": "Aaron Hasz", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8b5", "department": "Editing", "gender": 2, "id": 5215, "name": "Matt Wheeler", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8bb", "department": "Editing", "gender": 2, "id": 5216, "name": "Bruce Wong", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8c1", "department": "Editing", "gender": 2, "id": 5217, "name": "Jonathan Dver", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8c7", "department": "Editing", "gender": 2, "id": 5218, "name": "Jennifer Bender", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8cd", "department": "Editing", "gender": 2, "id": 5219, "name": "Marc Jozefowicz", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8d3", "department": "Editing", "gender": 2, "id": 5220, "name": "Kaboom", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8d9", "department": "Editing", "gender": 2, "id": 5221, "name": "Matt Smith", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8df", "department": "Editing", "gender": 2, "id": 5222, "name": "Jay Prichard", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8e5", "department": "Editing", "gender": 2, "id": 5223, "name": "Tanner Pearson", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8eb", "department": "Editing", "gender": 2, "id": 5224, "name": "Dylan Highsmith", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8f1", "department": "Editing", "gender": 2, "id": 5225, "name": "Winnie Hsu", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8f7", "department": "Editing", "gender": 2, "id": 5226, "name": "Stephen Nemeth", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb8fd", "department": "Editing", "gender": 2, "id": 5227, "name": "Gregory Plotkin", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb903", "department": "Editing", "gender": 2, "id": 5228, "name": "Tim Mirkovich", "job": "Trailer"}, {"credit_id": "52fe4a7dc3a368484e0fb909", "department": "Editing", "gender": 2, "id": 5229, "name": "Cinna Dunn", "job": "Trailer"}]"}

```
movies.shape
```

```
(4803, 7)
```

```
movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 7 columns):
#   Column    Non-Null Count  Dtype
---  ------    --------------  -----
0   id        4803 non-null   int64
1   title     4803 non-null   object
2   overview  4800 non-null   object
3   genres    4803 non-null   object
4   keywords  4803 non-null   object
5   cast      4803 non-null   object
6   crew      4803 non-null   object
dtype: int64
```

```
movies.isnull().sum()
```

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

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

everyone is important once we will create tags using these so we can drop null rows

```
movies.isnull().sum()
```

```
id          0
title       0
overview    0
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

```
movies.duplicated().sum()
```

```
0
```

```
import ast
ast.literal_eval
```

```
<function ast.literal_eval(node_or_string, *, mode='eval', include_type_comments=False)>
```

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

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

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

```
<ipykrnl_id_0>:2:KeyboardInterrupt:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

	id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but he soon becomes torn between following orders and protecting an alien civilization.	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, spacecraft, finance, spoof, manipulation, nanotechnology, pandora, paraplegic, alien, tribe, corporate greed, mercenary, marine, indigenous people, planet, exploitation, gunship, colonialism, aliens, alien planet, aerial combat, alien life form, resistance, sacrifice, greed, corporation, exploitation of nature, mine, military, technology, human versus nature, human versus alien, sacred site, pandora's box, unobtanium, hometree, dragon, helicopter, jungle, forest, ecosystem, biologist, tribal, mercenary, ambush, ceremony, ancient, ancient ruins, ancient tree, sacred tree, sacred place, sacred ground, ancient tree of souls, tsahik, tsahik kelutral, tsahik kelutral's ancient tree