

Recommendation system for movies

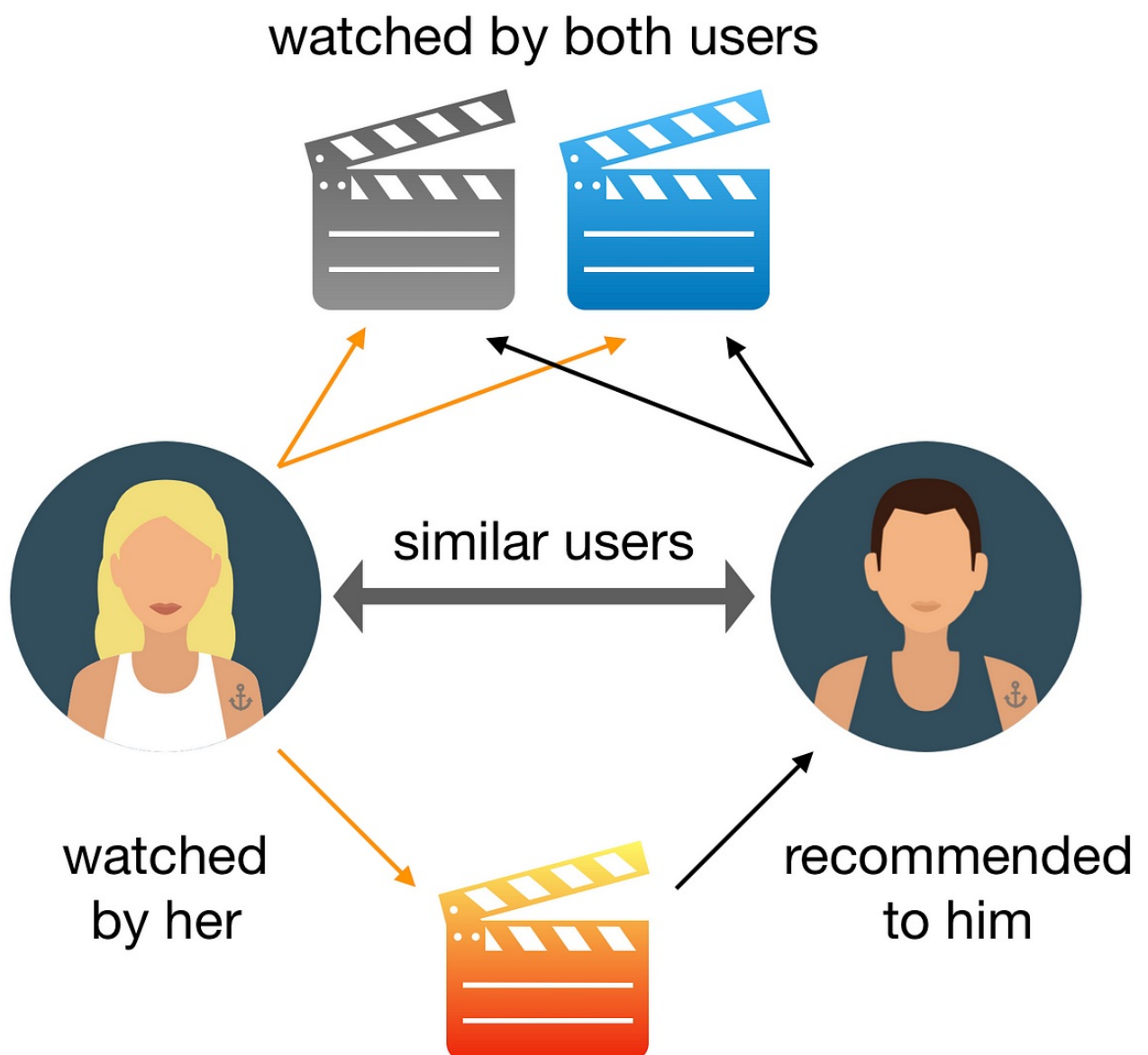
What Are Recommendation Systems in Machine Learning?

Today, we are facing a very rapid growth in the volume and structure of the Internet. Users are more often found to be lost in this complex and messy environment of websites due to their complex structure and large amounts of information. So personalizing and simplifying the web is more important than ever before for users and owners of e-commerce websites.

Machine learning-based recommendation systems are powerful engines using machine learning algorithms to segment customers based on their user data and behavioral patterns (such as purchase and browsing history, likes, or reviews) and target them with personalized product and content suggestions

Who is using them?

Today, it plays a very important role in sites that have a lot of hits, users or products, in the fields of entertainment, content-based, e-commerce, advertising and social networks, etc., such as Netflix, youtube, amazon, lastfm, imdb, Yahoo, Spotify and so on.



```
In [2]: import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")
```

```
In [3]: movies = pd.read_csv('../input/movies-recomandation/tmdb_5000_movies.csv')
movies.head()
```

Out[3]:	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "ocean"}]	en	Avatar	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 those who have become his family.	150.437571
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 28, "name": "Action"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, has returned to the Caribbean Sea.	139.082615
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name": "international relations"}]	en	Spectre	A cryptic message from Bond's past sends him on a mission to揭开神秘面纱。	107.376788
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}, {"id": 18, "name": "Drama"}]	http://www.thedarkknighttrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}]	en	The Dark Knight Rises	Following the death of District Attorney Harvey Dent, Batman deduces that the only person capable of bringing the city back to order is the man who has become a legend.	112.312950
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id": 1464, "name": "ocean"}]	en	John Carter	John Carter is a war-weary, former military captain, plagued by traumatic experiences from a long and troubled past.	43.926995

```
In [4]: print(f'\033[94m')
movies.shape
```

Out[4]: (4803, 20)

```
In [5]: credits = pd.read_csv('../input/movies-recomandation/tmdb_5000_credits.csv')
credits.head()
```

Out[5]:	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 [6]: print(f'\033[94m')
credits.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4803 entries, 0 to 4802
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   movie_id    4803 non-null   int64
1   title       4803 non-null   object
2   cast        4803 non-null   object
3   crew        4803 non-null   object
dtypes: int64(1), object(3)
memory usage: 150.2+ KB
```

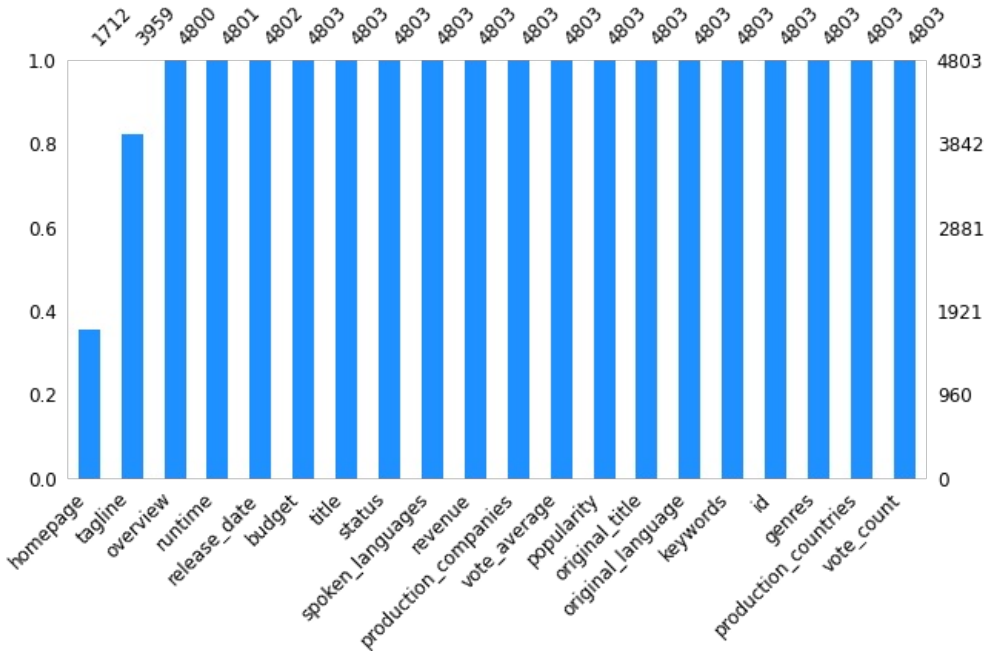
```
In [7]: movies.shape
```

```
Out[7]: (4803, 20)
```

```
In [8]: print(f'\033[94m')
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(4), object(13)
memory usage: 750.6+ KB
```

```
In [9]: import missingno
missingno.bar(movies, color="dodgerblue", sort="ascending", figsize=(10,5), fontsize=12);
```



```
In [10]: import missingno as msno
missingno.matrix(movies, figsize=(10,5), fontsize=12, color=(1, 0.38, 0.27));
```



```
Out[18]: 0 [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1 [{"id": 12, "name": "Adventure"}, {"id": 14, "...
2 [{"id": 28, "name": "Action"}, {"id": 12, "nam...
Name: genres, dtype: object
```

```
In [19]: movies.loc[0:3]['genres']
```

```
Out[19]: 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...
Name: genres, dtype: object
```

```
In [20]: movies.iloc[2:7,0:3] # [Rows , Columns]
```

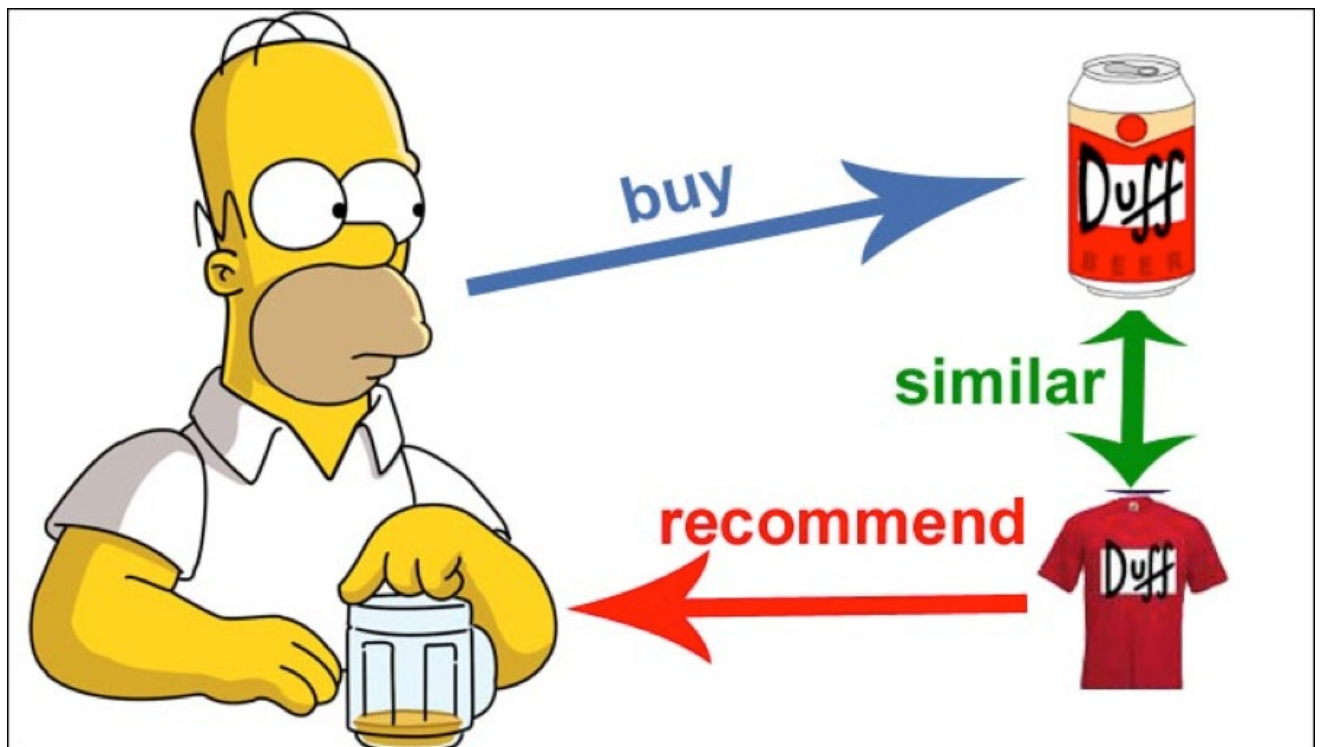
```
Out[20]:
```

	movie_id	title	overview
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...
5	559	Spider-Man 3	The seemingly invincible Spider-Man goes up ag...
6	38757	Tangled	When the kingdom's most wanted-and most charmi...

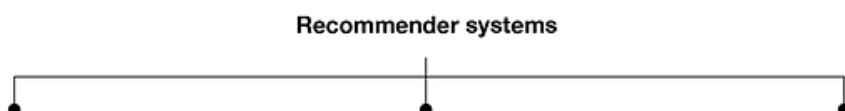
Movie Recommender System Building

Recommender System

Recommender Systems simply put, are AI algorithms that utilize Big Data to suggest additional products to consumers based on a variety of features. These recommendations can be based on factors such as past purchases, demographic info, their search history, time spent reviewing the product or a like, dislike or a comment left behind by these consumers. The idea of Recommender Systems is that if you can narrow down the pool of selection options for your customers to a few meaningful and relevant choices, they are more likely to make a purchase now, as well as come back for more down the road.



Different Types of Recommender Systems



Content based methods

Define a model for user-item interactions where users and/or items representations are given (explicit features).

Collaborative filtering methods

Model based

Define a model for user-item interactions where users and items representations have to be learned from interactions matrix.

Memory based

Define no model for user-item interactions and rely on similarities between users or items in terms of observed interactions.

Hybrid methods

Mix content based and collaborative filtering approaches.



Collaborative Filtering

The Collaborative filtering method for recommender systems is a method that is solely based on the past interactions that have been recorded between users and items, in order to produce new recommendations.

Collaborative Filtering tends to find what similar users would like and the recommendations to be provided and in order to classify the users into clusters of similar types and recommend each user according to the preference of its cluster.

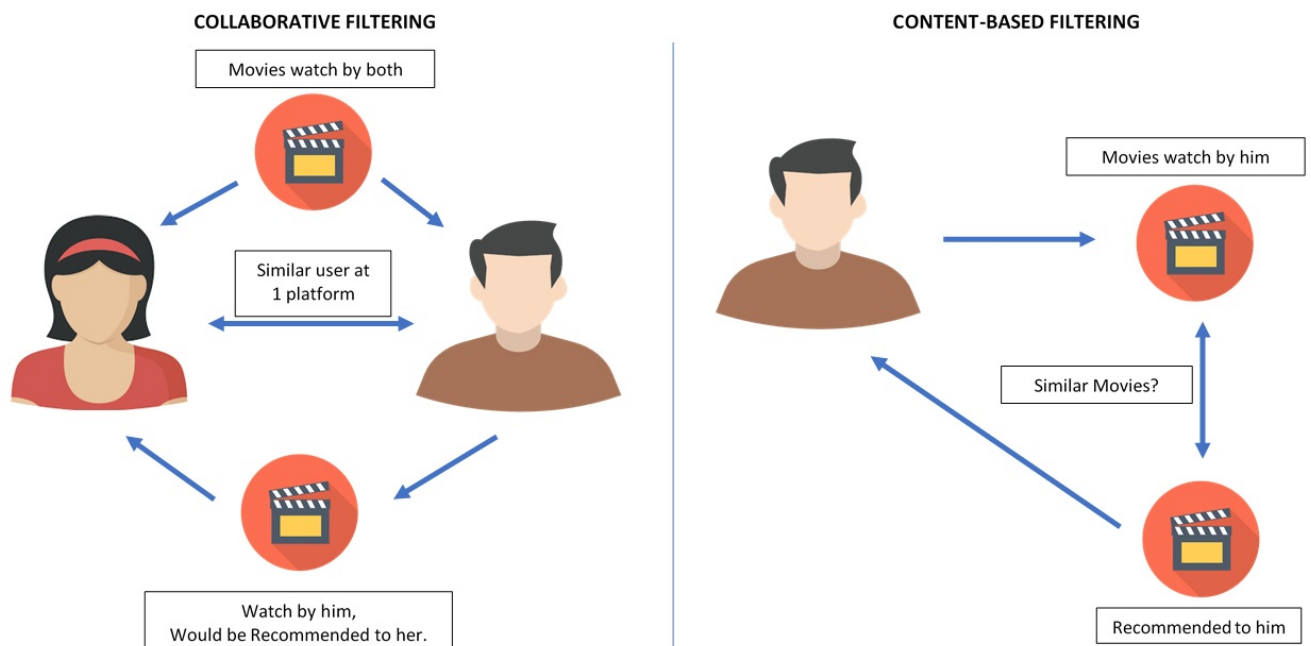
The main idea that governs the collaborative methods is that through past user-item interactions when processed through the system, it becomes sufficient to detect similar users or similar items to make predictions based on these estimated facts and insights.

Content Based Filtering

The content-based approach uses additional information about users and/or items. This filtering method uses item features to recommend other items similar to what the user likes and also based on their previous actions or explicit feedback.

The main idea of content-based methods is to try to build a model, based on the available "features", that explain the observed user-item interactions.

Such a model helps us in making new predictions for a user pretty easily, with just a look at the profile of this user and based on its information, to determine relevant movies to suggest.



We will be using Content Based Filtering for this dataset

```
Out[29]: '[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id": "52fe48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch", "credit_id": "52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order": 3}, {"cast_id": 5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Michelle Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge", "credit_id": "52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi", "order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id": "52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel David Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id": "52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder", "order": 7}, {"cast_id": 11, "character": "Eytukan", "credit_id": "52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "Wes Studi", "order": 8}, {"cast_id": 10, "character": "Tsu\\'Tey", "credit_id": "52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso", "order": 9}, {"cast_id": 12, "character": "Dr. Max Patel", "credit_id": "52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "name": "Dileep Rao", "order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id": "52fe48009251416c750ac9f5", "gender": 2, "id": 98215, "name": "Matt Gerald", "order": 11}, {"cast_id": 32, "character": "Private Fike", "credit_id": "52fe48009251416c750aca5b", "gender": 2, "id": 154153, "name": "Sean Anthony Moran", "order": 12}, {"cast_id": 33, "character": "Cryo Vault Med Tech", "credit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name": "Jason Whyte", "order": 13}, {"cast_id": 34, "character": "Venture Star Crew Chief", "credit_id": "52fe48009251416c750aca63", "gender": 2, "id": 42317, "name": "Scott Lawrence", "order": 14}, {"cast_id": 35, "character": "Lock Up Trooper", "credit_id": "52fe48009251416c750aca67", "gender": 2, "id": 986734, "name": "Kelly Kilgour", "order": 15}, {"cast_id": 36, "character": "Shuttle Pilot", "credit_id": "52fe48009251416c750aca6b", "gender": 0, "id": 1207227, "name": "James Patrick Pitt", "order": 16}, {"cast_id": 37, "character": "Shuttle Co-Pilot", "credit_id": "52fe48009251416c750aca6f", "gender": 0, "id": 1180936, "name": "Sean Patrick Murphy", "order": 17}, {"cast_id": 38, "character": "Shuttle Crew Chief", "credit_id": "52fe48009251416c750aca73", "gender": 2, "id": 1019578, "name": "Peter Dillon", "order": 18}, {"cast_id": 39, "character": "Tractor Operator / Troupe", "credit_id": "52fe48009251416c750aca77", "gender": 0, "id": 91443, "name": "Kevin Dorman", "order": 19}, {"cast_id": 40, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7b", "gender": 2, "id": 173391, "name": "Kelson Henderson", "order": 20}, {"cast_id": 41, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7c", "gender": 2, "id": 173392, "name": "Kelson Henderson", "order": 21}, {"cast_id": 42, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7d", "gender": 2, "id": 173393, "name": "Kelson Henderson", "order": 22}, {"cast_id": 43, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7e", "gender": 2, "id": 173394, "name": "Kelson Henderson", "order": 23}, {"cast_id": 44, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca7f", "gender": 2, "id": 173395, "name": "Kelson Henderson", "order": 24}, {"cast_id": 45, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca80", "gender": 2, "id": 173396, "name": "Kelson Henderson", "order": 25}, {"cast_id": 46, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca81", "gender": 2, "id": 173397, "name": "Kelson Henderson", "order": 26}, {"cast_id": 47, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca82", "gender": 2, "id": 173398, "name": "Kelson Henderson", "order": 27}, {"cast_id": 48, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca83", "gender": 2, "id": 173399, "name": "Kelson Henderson", "order": 28}, {"cast_id": 49, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca84", "gender": 2, "id": 173400, "name": "Kelson Henderson", "order": 29}, {"cast_id": 50, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca85", "gender": 2, "id": 173401, "name": "Kelson Henderson", "order": 30}, {"cast_id": 51, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca86", "gender": 2, "id": 173402, "name": "Kelson Henderson", "order": 31}, {"cast_id": 52, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca87", "gender": 2, "id": 173403, "name": "Kelson Henderson", "order": 32}, {"cast_id": 53, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca88", "gender": 2, "id": 173404, "name": "Kelson Henderson", "order": 33}, {"cast_id": 54, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca89", "gender": 2, "id": 173405, "name": "Kelson Henderson", "order": 34}, {"cast_id": 55, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8a", "gender": 2, "id": 173406, "name": "Kelson Henderson", "order": 35}, {"cast_id": 56, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 173407, "name": "Kelson Henderson", "order": 36}, {"cast_id": 57, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8c", "gender": 2, "id": 173408, "name": "Kelson Henderson", "order": 37}, {"cast_id": 58, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8d", "gender": 2, "id": 173409, "name": "Kelson Henderson", "order": 38}, {"cast_id": 59, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8e", "gender": 2, "id": 173410, "name": "Kelson Henderson", "order": 39}, {"cast_id": 60, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca8f", "gender": 2, "id": 173411, "name": "Kelson Henderson", "order": 40}, {"cast_id": 61, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca90", "gender": 2, "id": 173412, "name": "Kelson Henderson", "order": 41}, {"cast_id": 62, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca91", "gender": 2, "id": 173413, "name": "Kelson Henderson", "order": 42}, {"cast_id": 63, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca92", "gender": 2, "id": 173414, "name": "Kelson Henderson", "order": 43}, {"cast_id": 64, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca93", "gender": 2, "id": 173415, "name": "Kelson Henderson", "order": 44}, {"cast_id": 65, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca94", "gender": 2, "id": 173416, "name": "Kelson Henderson", "order": 45}, {"cast_id": 66, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca95", "gender": 2, "id": 173417, "name": "Kelson Henderson", "order": 46}, {"cast_id": 67, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca96", "gender": 2, "id": 173418, "name": "Kelson Henderson", "order": 47}, {"cast_id": 68, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 173419, "name": "Kelson Henderson", "order": 48}, {"cast_id": 69, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca98", "gender": 2, "id": 173420, "name": "Kelson Henderson", "order": 49}, {"cast_id": 70, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca99", "gender": 2, "id": 173421, "name": "Kelson Henderson", "order": 50}, {"cast_id": 71, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca9a", "gender": 2, "id": 173422, "name": "Kelson Henderson", "order": 51}, {"cast_id": 72, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca9b", "gender": 2, "id": 173423, "name": "Kelson Henderson", "order": 52}, {"cast_id": 73, "character": "Dragon Gunship Pilot", "credit_id": "52fe48009251416c750aca9c", "gender": 2, "id": 173424, "name": "Kelson Henderson", "order": 53},
```


ast_id": 41, "character": "Dragon Gunship Gunner", "credit_id": "52fe48009251416c750aca7f", "gender": 0, "id": 1207236, "name": "David Van Horn", "order": 21}, {"cast_id": 42, "character": "Dragon Gunship Navigator", "credit_id": "52fe48009251416c750aca83", "gender": 0, "id": 215913, "name": "Jacob Tomuri", "order": 22}, {"cast_id": 43, "character": "Suit #1", "credit_id": "52fe48009251416c750aca87", "gender": 0, "id": 143206, "name": "Michael Blain-Rozgay", "order": 23}, {"cast_id": 44, "character": "Suit #2", "credit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 169676, "name": "Jon Curry", "order": 24}, {"cast_id": 46, "character": "Ambient Room Tech", "credit_id": "52fe48009251416c750aca8f", "gender": 0, "id": 1048610, "name": "Luke Hawker", "order": 25}, {"cast_id": 47, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48009251416c750aca93", "gender": 0, "id": 42288, "name": "Woody Schultz", "order": 26}, {"cast_id": 48, "character": "Horse Clan Leader", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 68278, "name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room Tech", "credit_id": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247, "name": "Sonnia Yee", "order": 28}, {"cast_id": 50, "character": "Basketball Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender": 1, "id": 1207248, "name": "Jahnel Curfman", "order": 29}, {"cast_id": 51, "character": "Basketball Avatar", "credit_id": "52fe48009251416c750acaa3", "gender": 0, "id": 89714, "name": "Ilram Choi", "order": 30}, {"cast_id": 52, "character": "Na\\vi Child", "credit_id": "52fe48009251416c750acaa7", "gender": 0, "id": 1207249, "name": "Kyla Warren", "order": 31}, {"cast_id": 53, "character": "Troupe", "credit_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name": "Lisa Roumain", "order": 32}, {"cast_id": 54, "character": "Troupe", "credit_id": "52fe48009251416c750acaaf", "gender": 1, "id": 83105, "name": "Debra Wilson", "order": 33}, {"cast_id": 57, "character": "Troupe", "credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name": "Chris Mala", "order": 34}, {"cast_id": 55, "character": "Troupe", "credit_id": "52fe48009251416c750acab3", "gender": 0, "id": 1207251, "name": "Taylor Kibby", "order": 35}, {"cast_id": 56, "character": "Troupe", "credit_id": "52fe48009251416c750acab7", "gender": 0, "id": 1207252, "name": "Jodie Landau", "order": 36}, {"cast_id": 58, "character": "Troupe", "credit_id": "52fe48009251416c750acabf", "gender": 0, "id": 1207254, "name": "Julie Lamm", "order": 37}, {"cast_id": 59, "character": "Troupe", "credit_id": "52fe48009251416c750acac3", "gender": 0, "id": 1207257, "name": "Cullen B. Madden", "order": 38}, {"cast_id": 60, "character": "Troupe", "credit_id": "52fe48009251416c750acac7", "gender": 0, "id": 1207259, "name": "Joseph Brady Madden", "order": 39}, {"cast_id": 61, "character": "Troupe", "credit_id": "52fe48009251416c750acacb", "gender": 0, "id": 1207262, "name": "Frankie Torres", "order": 40}, {"cast_id": 62, "character": "Troupe", "credit_id": "52fe48009251416c750acacf", "gender": 1, "id": 1158600, "name": "Austin Wilson", "order": 41}, {"cast_id": 63, "character": "Troupe", "credit_id": "52fe48019251416c750acad3", "gender": 1, "id": 983705, "name": "Sara Wilson", "order": 42}, {"cast_id": 64, "character": "Troupe", "credit_id": "52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica Washington-Miller", "order": 43}, {"cast_id": 65, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadb", "gender": 1, "id": 1145098, "name": "Lucy Briant", "order": 44}, {"cast_id": 66, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acadf", "gender": 2, "id": 33305, "name": "Nathan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae3", "gender": 0, "id": 1207264, "name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae7", "gender": 2, "id": 33311, "name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acae8", "gender": 0, "id": 1207265, "name": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0, "id": 1207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character": "Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "gender": 2, "id": 54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character": "Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "id": 1207267, "name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character": "Dancer", "credit_id": "52fe48019251416c750acafb", "gender": 0, "id": 1207268, "name": "Carvon Futrell", "order": 52}, {"cast_id": 74, "character": "Dancer", "credit_id": "52fe48019251416c750acaff", "gender": 0, "id": 1207269, "name": "Brandon Jelkes", "order": 53}, {"cast_id": 75, "character": "Dancer", "credit_id": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "name": "Micah Moch", "order": 54}, {"cast_id": 76, "character": "Dancer", "credit_id": "52fe48019251416c750acb07", "gender": 0, "id": 1207271, "name": "Hanniyah Muhammad", "order": 55}, {"cast_id": 77, "character": "Dancer", "credit_id": "52fe48019251416c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"cast_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncredited)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav\\i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 14, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 92, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "character": "Ops Centroworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 96, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 97, "character": "Col. Quaritch\\s Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 98, "character": "Female Marine (uncredited)", "credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name": "Jodie Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader (uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id": 1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 100, "character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67", "gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80}, {"cast_id": 101, "character": "Na\\vi (uncredited)", "credit_id": "52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Nikie Zambo", "order": 81}, {"cast_id": 102, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}]'


```
In [30]: def convert3(obj):
        L = []
        counter = 0
        for i in ast.literal_eval(obj):
            if counter !=3:
                L.append(i['name'])
                counter += 1
            else:
                break
        return L
```

```
In [31]: movies['cast'].apply(convert3)
```

```
Out[31]: 0      [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1      [Johnny Depp, Orlando Bloom, Keira Knightley]
2      [Daniel Craig, Christoph Waltz, Léa Seydoux]
3      [Christian Bale, Michael Caine, Gary Oldman]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]
...
4804   [Carlos Gallardo, Jaime de Hoyos, Peter Marqua...
4805   [Edward Burns, Kerry Bishé, Marsha Dietlein]
4806   [Eric Mabius, Kristin Booth, Crystal Lowe]
4807   [Daniel Henney, Eliza Coupe, Bill Paxton]
4808   [Drew Barrymore, Brian Herzlinger, Corey Feldman]
Name: cast, Length: 4806, dtype: object
```

```
In [32]: movies['cast'] = movies['cast'].apply(convert3)
movies.head(1)
```

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 Weaver]	{"credit_id": "52fe48009251416c750aca23", "de...

```
In [33]: movies['crew'][0]
```

```
Out[33]: '{"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": 6347, "job": "Casting", "name": "Margery Simkin"}, {"credit_id": "570b6f419251417da70032fe", "department": "Art", "gender": 2, "id": 6878, "job": "Supervising Art Director", "name": "Kevin Ishioka"}, {"credit_id": "5495a0fac3a3686ae9004468", "department": "Sound", "gender": 0, "id": 6883, "job": "Music Editor", "name": "Dick Bernstein"}, {"credit_id": "54959706c3a3686af3003e81", "department": "Sound", "gender": 0, "id": 8159, "job": "Sound Effects Editor", "name": "Shannon Mills"}, {"credit_id": "54491d58c3a3680fb1001ccb", "department": "Sound", "gender": 0, "id": 8160, "job": "Foley", "name": "Dennie Thorpe"}, {"credit_id": "54491d6cc3a3680fa5001b2c", "department": "Sound", "gender": 0, "id": 8163, "job": "Foley", "name": "Jana Vance"}, {"credit_id": "52fe48009251416c750aca57", "department": "Costume & Make-Up", "gender": 1, "id": 8527, "job": "Costume Design", "name": "Deborah Lynn Scott"}, {"credit_id": "52fe48009251416c750aca2f", "department": "Production", "gender": 2, "id": 8529, "job": "Producer", "name": "Jon Landau"}, {"credit_id": "539c4937c3a36810ba002194", "department": "Art", "gender": 0, "id": 9618, "job": "Art Direction", "name": "Sean Haworth"}, {"credit_id": "539c49b6c3a36810c10020e6", "department": "Art", "gender": 1, "id": 12653, "job": "Set Decoration", "name": "Kim Sinclair"}, {"credit_id": "570b6f2f9251413a0e00020d", "department": "Art", "gender": 1, "id": 12653, "job": "Supervising Art Director", "name": "Kim Sinclair"}, {"credit_id": "54491a6c0e0a26748c001b19", "department": "Art", "gender": 2, "id": 14350, "job": "Set Designer", "name": "Richard F. Mays"}, {"credit_id": "56928cf4c3a3684cff0025c4", "department": "Production", "gender": 1, "id": 20294, "job": "Executive Producer", "name": "Laeta Kalogridis"}, {"credit_id": "52fe48009251416c750aca51", "department": "Costume & Make-Up", "gender": 0, "id": 17675, "job": "Costume Design", "name": "Mayes C. Rubeo"}, {"credit_id": "52fe48009251416c750acall", "department": "Camera", "gender": 2, "id": 18265, "job": "Director of Photography", "name": "Mauro Fiore"}, {"credit_id": "5449194d0e0a26748f001b39", "department": "Art", "gender": 0, "id": 42281, "job": "Set Designer", "name": "Scott Herbertson"}, {"credit_id": "52fe48009251416c750aca05", "department": "Crew", "gender": 0, "id": 42288, "job": "Stunts", "name": "Woody Schultz"}, {"credit_id": "5592aefb92514152de0010f5", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Makeup Artist", "name": "Linda DeVetta"}, {"credit_id": "5592afa492514152de00112c", "department": "Costume & Make-Up", "gender": 0, "id": 29067, "job": "Hairstylist", "name": "Linda DeVetta"}, {"credit_id": "54959ed592514130fc002e5d", "department": "Camera", "gender": 2, "id": 33302, "job": "Camera Operator", "name": "Richard Bluck"}, {"credit_id": "539c4891c3a36810ba002147", "department": "Art", "gender": 2, "id": 33303, "job": "Art Direction", "name": "Simon Bright"}, {"credit_id": "54959c069251417a81001f3a", "department": "Visual Effects", "gender": 0, "id": 113145, "job": "Visual Effects Supervisor", "name": "Richard Martin"}, {"credit_id": "54959a0dc3a3680ff5002c8d", "department": "Crew", "gender": 2, "id": 58188, "job": "Visual Effects Editor", "name": "Steve R. Moore"}, {"credit_id": "52fe48009251416c750aca1d", "department": "Editing", "gender": 2, "id": 58871, "job": "Editor", "name": "John Refoua"}, {"credit_id": "54491a4dc3a3680fc30018ca", "department": "Art", "gender": 0, "id": 92359, "job": "Set Designer", "name": "Karl J. Martin"}, {"credit_id": "52fe48009251416c750aca35", "department": "Camera", "gender": 1, "id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}, {"credit_id": "52fe48009251416c750ac9ff", "department": "Crew", "gender": 2, "id": 72201, "job": "Director of Photography", "name": "Chiling Lin"}]
```

nder": 0, "id": 89714, "job": "Stunts", "name": "Iram Choi"}, {"credit_id": "54959c529251416e2b004394", "department": "Visual Effects", "gender": 2, "id": 93214, "job": "Visual Effects Supervisor", "name": "Steven Quale"}, {"credit_id": "54491edf0e0a267489001c37", "department": "Crew", "gender": 1, "id": 122607, "job": "Dialect Coach", "name": "Carla Meyer"}, {"credit_id": "539c485bc3a368653d001a3a", "department": "Art", "gender": 2, "id": 132585, "job": "Art Direction", "name": "Nick Bassett"}, {"credit_id": "539c4903c3a368653d001a74", "department": "Art", "gender": 0, "id": 132596, "job": "Art Direction", "name": "Jill Cormack"}, {"credit_id": "539c4967c3a368653d001a94", "department": "Art", "gender": 0, "id": 132604, "job": "Art Direction", "name": "Andy McLaren"}, {"credit_id": "52fe48009251416c750aca45", "department": "Crew", "gender": 0, "id": 236696, "job": "Motion Capture Artist", "name": "Terry Notary"}, {"credit_id": "54959e02c3a3680fc60027d2", "department": "Crew", "gender": 2, "id": 956198, "job": "Stunt Coordinator", "name": "Garrett Warren"}, {"credit_id": "54959ca3c3a3686ae300438c", "department": "Visual Effects", "gender": 2, "id": 957874, "job": "Visual Effects Supervisor", "name": "Jonathan Rothbart"}, {"credit_id": "570b6f519251412c74001b2f", "department": "Art", "gender": 0, "id": 957889, "job": "Supervising Art Director", "name": "Stefan Dechant"}, {"credit_id": "570b6f62c3a3680b77007460", "department": "Art", "gender": 2, "id": 959555, "job": "Supervising Art Director", "name": "Todd Cherniawsky"}, {"credit_id": "539c4a3ac3a36810da0021cc", "department": "Production", "gender": 0, "id": 1016177, "job": "Casting", "name": "Miranda Rivers"}, {"credit_id": "539c482cc3a36810c1002062", "department": "Art", "gender": 0, "id": 1032536, "job": "Production Design", "name": "Robert Stromberg"}, {"credit_id": "539c4b65c3a36810c9002125", "department": "Costume & Make-Up", "gender": 2, "id": 1071680, "job": "Costume Design", "name": "John Harding"}, {"credit_id": "54959e6692514130fc002e4e", "department": "Camera", "gender": 0, "id": 1177364, "job": "Steadicam Operator", "name": "Roberto De Angelis"}, {"credit_id": "539c49f1c3a368653d001aac", "department": "Costume & Make-Up", "gender": 2, "id": 1202850, "job": "Makeup Department Head", "name": "Mike Smithson"}, {"credit_id": "5495999ec3a3686ae100460c", "department": "Visual Effects", "gender": 0, "id": 1204668, "job": "Visual Effects Producer", "name": "Alain Lalanne"}, {"credit_id": "54959cdfc3a3681153002729", "department": "Visual Effects", "gender": 0, "id": 1206410, "job": "Visual Effects Supervisor", "name": "Lucas Salton"}, {"credit_id": "549596239251417a81001eae", "department": "Crew", "gender": 0, "id": 1234266, "job": "Post Production Supervisor", "name": "Janace Tashjian"}, {"credit_id": "54959c859251416e1e003efe", "department": "Visual Effects", "gender": 0, "id": 1271932, "job": "Visual Effects Supervisor", "name": "Stephen Rosenbaum"}, {"credit_id": "5592af28c3a368775a00105f", "department": "Costume & Make-Up", "gender": 0, "id": 1310064, "job": "Makeup Artist", "name": "Frankie Karen"}, {"credit_id": "539c4adfc3a36810e300203b", "department": "Costume & Make-Up", "gender": 1, "id": 1319844, "job": "Costume Supervisor", "name": "Lisa Lovaas"}, {"credit_id": "54959b579251416e2b004371", "department": "Visual Effects", "gender": 0, "id": 1327028, "job": "Visual Effects Supervisor", "name": "Jonathan Fawcner"}, {"credit_id": "539c48a7c3a36810b5001fa7", "department": "Art", "gender": 0, "id": 1330561, "job": "Art Direction", "name": "Robert Bavin"}, {"credit_id": "539c4a71c3a36810da0021e0", "department": "Costume & Make-Up", "gender": 0, "id": 1330567, "job": "Costume Supervisor", "name": "Anthony Almaraz"}, {"credit_id": "539c4a8ac3a36810ba0021e4", "department": "Costume & Make-Up", "gender": 0, "id": 1330570, "job": "Costume Supervisor", "name": "Carolyn M. Fenton"}, {"credit_id": "539c4ab6c3a36810da0021f0", "department": "Costume & Make-Up", "gender": 0, "id": 1330574, "job": "Costume Supervisor", "name": "Beth Koenigsberg"}, {"credit_id": "54491ab70e0a267480001ba2", "department": "Art", "gender": 0, "id": 1336191, "job": "Set Designer", "name": "Sam Page"}, {"credit_id": "544919d9c3a3680fc30018bd", "department": "Art", "gender": 0, "id": 1339441, "job": "Set Designer", "name": "Tex Kadonaga"}, {"credit_id": "54491cf50e0a267483001b0c", "department": "Editing", "gender": 0, "id": 1352422, "job": "Dialogue Editor", "name": "Kim Foscatto"}, {"credit_id": "544919f40e0a26748c001b09", "department": "Art", "gender": 0, "id": 1352962, "job": "Set Designer", "name": "Tammy S. Lee"}, {"credit_id": "5495a115c3a3680ff5002d71", "department": "Crew", "gender": 0, "id": 1357070, "job": "Transportation Coordinator", "name": "Denny Cairra"}, {"credit_id": "5495a12f92514130fc002e94", "department": "Crew", "gender": 0, "id": 1357071, "job": "Transportation Coordinator", "name": "James Waitkus"}, {"credit_id": "5495976fc3a36811530026b0", "department": "Sound", "gender": 0, "id": 1360103, "job": "Supervising Sound Editor", "name": "Addison Teague"}, {"credit_id": "54491837c3a3680fb1001c5a", "department": "Art", "gender": 2, "id": 1376887, "job": "Set Designer", "name": "C. Scott Baker"}, {"credit_id": "54491878c3a3680fb4001c9d", "department": "Art", "gender": 0, "id": 1376888, "job": "Set Designer", "name": "Luke Caska"}, {"credit_id": "544918dac3a3680fa5001ae0", "department": "Art", "gender": 0, "id": 1376889, "job": "Set Designer", "name": "David Chow"}, {"credit_id": "544919110e0a267486001b68", "department": "Art", "gender": 0, "id": 1376890, "job": "Set Designer", "name": "Jonathan Dyer"}, {"credit_id": "54491967c3a3680faa001b5e", "department": "Art", "gender": 0, "id": 1376891, "job": "Set Designer", "name": "Joseph Hiura"}, {"credit_id": "54491997c3a3680fb1001c8a", "department": "Art", "gender": 0, "id": 1376892, "job": "Art Department Coordinator", "name": "Rebecca Jellie"}, {"credit_id": "544919ba0e0a26748f001b42", "department": "Art", "gender": 0, "id": 1376893, "job": "Set Designer", "name": "Robert Andrew Johnson"}, {"credit_id": "54491b1dc3a3680faa001b8c", "department": "Art", "gender": 0, "id": 1376895, "job": "Assistant Art Director", "name": "Mike Stassi"}, {"credit_id": "54491b79c3a3680fbb001826", "department": "Art", "gender": 0, "id": 1376897, "job": "Construction Coordinator", "name": "John Villarino"}, {"credit_id": "54491baec3a3680fb4001ce6", "department": "Art", "gender": 2, "id": 1376898, "job": "Assistant Art Director", "name": "Jeffrey Wisniewski"}, {"credit_id": "54491d2fc3a3680fb4001d07", "department": "Editing", "gender": 0, "id": 1376899, "job": "Dialogue Editor", "name": "Cheryl Nardi"}, {"credit_id": "54491d86c3a3680fa5001b2f", "department": "Editing", "gender": 0, "id": 1376901, "job": "Dialogue Editor", "name": "Marshall Winn"}, {"credit_id": "54491d9dc3a3680faa001bb0", "department": "Sound", "gender": 0, "id": 1376902, "job": "Supervising Sound Editor", "name": "Gwendolyn Yates Whittle"}, {"credit_id": "54491dc10e0a267486001bce", "department": "Sound", "gender": 0, "id": 1376903, "job": "Sound Re-Recording Mixer", "name": "William Stein"}, {"credit_id": "54491f500e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}, {"credit_id": "549599239251412c4e002a2e", "department": "Visual Effects", "gender": 0, "id": 1391692, "job": "Visual Effects Producer", "name": "Chris Del Conte"}, {"credit_id": "54959d54c3a36831b8001d9a", "department": "Visual Effects", "gender": 2, "id": 1391695, "job": "Visual Effects Supervisor", "name": "R. Christopher White"}, {"credit_id": "54959bdf9251412c4e002a66", "department": "Visual Effects", "gender": 0, "id": 1394070, "job": "Visual Effects Supervisor", "name": "Dan Lemmon"}, {"credit_id": "5495971d92514132ed002922", "department": "Sound", "gender": 0, "id": 1394129, "job": "Sound Effects Editor", "name": "Tim Nielsen"}, {"credit_id": "5592b25792514152cc0011aa", "department": "Crew", "gender": 0, "id": 1394286, "job": "CG Supervisor", "name": "Michael Mulholland"}, {"credit_id": "54959a329251416e2b004355", "department": "Crew", "gender": 0, "id": 1394750, "job": "Visual Effects Editor", "name": "Thomas Nittmann"}, {"credit_id": "54959d6dc3a3686ae9004401", "department": "Visual Effects", "gender": 0, "id": 1394755, "job": "Visual Effects Supervisor", "name": "Edson Williams"}, {"credit_id": "5495a08fc3a3686ae300441c", "department": "Editing", "gender": 0, "id": 1394953, "job": "Digital Intermediate", "name": "Christine Carr"}, {"credit_id": "55402d659251413d6d000249", "department": "Visual Effects", "gender": 0, "id": 1395269, "job": "Visual Effects Supervisor", "name": "John Bruno"}, {"credit_id": "54959e7b9251416e1e003f3e", "department": "Camera", "gender": 0, "id": 1398970, "job": "Steadicam Operator", "name": "David Emmerichs"}, {"credit_id": "54959734c3a3686ae10045e0", "department": "Sound", "gender": 0, "id": 1400906, "job": "Sound Effects Editor", "name": "Christopher Scarabosio"}, {"credit_id": "549595dd92514130fc002d79", "department": "Production", "gender": 0, "id": 1401784, "job": "Production Supervisor", "name": "Jennifer Teves"}, {"credit_id": "549596009251413af7028cc", "department": "Production", "gender": 0, "id": 1401785, "job": "Production Manager", "name": "Brigitte Yorke"}, {"credit_id": "549596e892514130fc002d99", "department": "Sound", "gender": 0, "id": 1401786, "job": "Sound Effects Editor", "name": "Ken Fischer"}, {"credit_id": "549598229251412c4e002a1c", "department": "Crew", "gender": 0, "id": 1401787, "job": "Special Effects Coordinator", "name": "Iain Hutton"}, {"credit_id": "549598349251416e2b00432b", "department": "Crew", "gender": 0, "id": 1401788, "job": "Special Effects Coordinator", "name": "Steve Ingram"}, {"credit_id": "54959905c3a3686ae3004324", "department": "Visual Effects", "gender": 0, "id": 1401789, "job": "Visual Effects Coordinator", "name": "Steve Ingram"}]

```
d": 1401789, "job": "Visual Effects Producer", "name": "Joyce Cox"}, {"credit_id": "5495994b92514132ed002951",
"department": "Visual Effects", "gender": 0, "id": 1401790, "job": "Visual Effects Producer", "name": "Jenny Foster"}, {"credit_id": "549599cbc3a3686ae1004613", "department": "Crew", "gender": 0, "id": 1401791, "job": "Visual Effects Editor", "name": "Christopher Marino"}, {"credit_id": "549599f2c3a3686ae100461e", "department": "Crew", "gender": 0, "id": 1401792, "job": "Visual Effects Editor", "name": "Jim Milton"}, {"credit_id": "54959a51c3a3686af3003eb5", "department": "Visual Effects", "gender": 0, "id": 1401793, "job": "Visual Effects Producer", "name": "Cyndi Ochs"}, {"credit_id": "54959a7cc3a36811530026f4", "department": "Crew", "gender": 0, "id": 1401794, "job": "Visual Effects Editor", "name": "Lucas Putnam"}, {"credit_id": "54959b91c3a3680ff5002cb4", "department": "Visual Effects", "gender": 0, "id": 1401795, "job": "Visual Effects Supervisor", "name": "Anthony \Ma x\ Ivins"}, {"credit_id": "54959bb69251412c4e002a5f", "department": "Visual Effects", "gender": 0, "id": 1401796, "job": "Visual Effects Supervisor", "name": "John Knoll"}, {"credit_id": "54959cbbc3a3686ae3004391", "department": "Visual Effects", "gender": 2, "id": 1401799, "job": "Visual Effects Supervisor", "name": "Eric Saindon"}, {"credit_id": "54959d06c3a3686ae90043f6", "department": "Visual Effects", "gender": 0, "id": 1401800, "job": "Visual Effects Supervisor", "name": "Wayne Stables"}, {"credit_id": "54959d259251416e1e003f11", "department": "Visual Effects", "gender": 0, "id": 1401801, "job": "Visual Effects Supervisor", "name": "David Stinnett"}, {"credit_id": "54959db49251413af7002975", "department": "Visual Effects", "gender": 0, "id": 1401803, "job": "Visual Effects Supervisor", "name": "Guy Williams"}, {"credit_id": "54959de4c3a3681153002750", "department": "Crew", "gender": 0, "id": 1401804, "job": "Stunt Coordinator", "name": "Stuart Thorp"}, {"credit_id": "54959ef2c3a3680fc60027f2", "department": "Lighting", "gender": 0, "id": 1401805, "job": "Best Boy Electric", "name": "Giles Coburn"}, {"credit_id": "54959f07c3a3680fc60027f9", "department": "Camera", "gender": 2, "id": 1401806, "job": "Still Photographer", "name": "Mark Fellman"}, {"credit_id": "54959f47c3a3681153002774", "department": "Lighting", "gender": 0, "id": 1401807, "job": "Lighting Technician", "name": "Scott Sprague"}, {"credit_id": "54959f8cc3a36831b8001df2", "department": "Visual Effects", "gender": 0, "id": 1401808, "job": "Animation Director", "name": "Jeremy Hollobon"}, {"credit_id": "54959fa0c3a36831b8001dfb", "department": "Visual Effects", "gender": 0, "id": 1401809, "job": "Animation Director", "name": "Orlando Meunier"}, {"credit_id": "54959fb6c3a3686af3003f54", "department": "Visual Effects", "gender": 0, "id": 1401810, "job": "Animation Director", "name": "Taisuke Tanimura"}, {"credit_id": "54959fd2c3a36831b8001e02", "department": "Costume & Make-Up", "gender": 0, "id": 1401812, "job": "Set Costumer", "name": "Lilia Mishel Acevedo"}, {"credit_id": "54959ff9c3a3686ae300440c", "department": "Costume & Make-Up", "gender": 0, "id": 1401814, "job": "Set Costumer", "name": "Alejandro M. Hernandez"}, {"credit_id": "5495a0ddc3a3686ae10046fe", "department": "Editing", "gender": 0, "id": 1401815, "job": "Digital Intermediate", "name": "Marvin Hall"}, {"credit_id": "5495a1f7c3a3686ae3004443", "department": "Production", "gender": 0, "id": 1401816, "job": "Publicist", "name": "Judy Alley"}, {"credit_id": "5592b29fc3a36869d100002f", "department": "Crew", "gender": 0, "id": 1418381, "job": "CG Supervisor", "name": "Mike Perry"}, {"credit_id": "5592b23a9251415df8001081", "department": "Crew", "gender": 0, "id": 1426854, "job": "CG Supervisor", "name": "Andrew Morley"}, {"credit_id": "55491e1192514104c40002d8", "department": "Art", "gender": 0, "id": 1438901, "job": "Conceptual Design", "name": "Seth Engstrom"}, {"credit_id": "5525d5809251417276002b06", "department": "Crew", "gender": 0, "id": 1447362, "job": "Visual Effects Art Director", "name": "Eric Oliver"}, {"credit_id": "554427ca925141586500312a", "department": "Visual Effects", "gender": 0, "id": 1447503, "job": "Modeling", "name": "Matsune Suzuki"}, {"credit_id": "551906889251415aab001c88", "department": "Art", "gender": 0, "id": 1447524, "job": "Art Department Manager", "name": "Paul Tobin"}, {"credit_id": "5592af8492514152cc0010de", "department": "Costume & Make-Up", "gender": 0, "id": 1452643, "job": "Hairstylist", "name": "Roxane Griffin"}, {"credit_id": "553d3c109251415852001318", "department": "Lighting", "gender": 0, "id": 1453938, "job": "Lighting Artist", "name": "Arun Ram-Mohan"}, {"credit_id": "5592af4692514152d5001355", "department": "Costume & Make-Up", "gender": 0, "id": 1457305, "job": "Makeup Artist", "name": "Georgia Lockhart-Adams"}, {"credit_id": "5592b2eac3a36877470012a5", "department": "Crew", "gender": 0, "id": 1466035, "job": "CG Supervisor", "name": "Thrain Shadbolt"}, {"credit_id": "5592b032c3a36877450015f1", "department": "Crew", "gender": 0, "id": 1483220, "job": "CG Supervisor", "name": "Brad Alexander"}, {"credit_id": "5592b05592514152d80012f6", "department": "Crew", "gender": 0, "id": 1483221, "job": "CG Supervisor", "name": "Shadi Almassizadeh"}, {"credit_id": "5592b090c3a36877570010b5", "department": "Crew", "gender": 0, "id": 1483222, "job": "CG Supervisor", "name": "Simon Clutterbuck"}, {"credit_id": "5592b0db3a368774b00112c", "department": "Crew", "gender": 0, "id": 1483223, "job": "CG Supervisor", "name": "Graeme Demmocks"}, {"credit_id": "5592b0fe92514152db0010c1", "department": "Crew", "gender": 0, "id": 1483224, "job": "CG Supervisor", "name": "Adrian Fernandes"}, {"credit_id": "5592b11f9251415df8001059", "department": "Crew", "gender": 0, "id": 1483225, "job": "CG Supervisor", "name": "Mitch Gates"}, {"credit_id": "5592b15dc3a3687745001645", "department": "Crew", "gender": 0, "id": 1483226, "job": "CG Supervisor", "name": "Jerry Kung"}, {"credit_id": "5592b18e925141645a0004ae", "department": "Crew", "gender": 0, "id": 1483227, "job": "CG Supervisor", "name": "Andy Lomas"}, {"credit_id": "5592b1bfc3a368775d0010e7", "department": "Crew", "gender": 0, "id": 1483228, "job": "CG Supervisor", "name": "Sebastian Marino"}, {"credit_id": "5592b2049251415df8001078", "department": "Crew", "gender": 0, "id": 1483229, "job": "CG Supervisor", "name": "Matthias Menz"}, {"credit_id": "5592b27b92514152d800136a", "department": "Crew", "gender": 0, "id": 1483230, "job": "CG Supervisor", "name": "Sergei Nevshupov"}, {"credit_id": "5592b2c3c3a36869e800003c", "department": "Crew", "gender": 0, "id": 1483231, "job": "CG Supervisor", "name": "Philippe Rebours"}, {"credit_id": "5592b317c3a36877470012af", "department": "Crew", "gender": 0, "id": 1483232, "job": "CG Supervisor", "name": "Michael Takarangi"}, {"credit_id": "5592b345c3a36877470012bb", "department": "Crew", "gender": 0, "id": 1483233, "job": "CG Supervisor", "name": "David Weitzberg"}, {"credit_id": "5592b37cc3a368775100113b", "department": "Crew", "gender": 0, "id": 1483234, "job": "CG Supervisor", "name": "Ben White"}, {"credit_id": "573c8e2f9251413f5d000094", "department": "Crew", "gender": 1, "id": 1621932, "job": "Stunts", "name": "Min Windle"}]
```

```
In [34]: def fetch_director(obj):
L = []
for i in ast.literal_eval(obj):
    if i['job'] == "Director":

        L.append(i['name'])
return L
```

```
In [35]: movies['crew'].apply(fetch_director)
```

```
Out[35]: 0 [James Cameron]
1 [Gore Verbinski]
2 [Sam Mendes]
3 [Christopher Nolan]
4 [Andrew Stanton]

...
4804 [Robert Rodriguez]
4805 [Edward Burns]
4806 [Scott Smith]
4807 [Daniel Hsia]
4808 [Brian Herzlinger, Jon Gunn, Brett Winn]
Name: crew, Length: 4806, dtype: object
```

```
In [36]: movies['crew'] = movies['crew'].apply(fetch_director)
movies.head(1)
```

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]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]

```
In [37]: movies['overview'][0]
```

```
Out[37]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes to rn between following orders and protecting an alien civilization.'
```

```
In [38]: movies['overview'].apply(lambda x:x.split())
```

```
Out[38]: 0 [In, the, 22nd, century,, a, paraplegic, Marin...
1 [Captain, Barbossa,, long, believed, to, be, d...
2 [A, cryptic, message, from, Bond's, past, send...
3 [Following, the, death, of, District, Attorney...
4 [John, Carter, is, a, war-weary,, former, mili...

...
4804 [El, Mariachi, just, wants, to, play, his, gui...
4805 [A, newlywed, couple's, honeymoon, is, upended...
4806 ["Signed,, Sealed,, Delivered", introduces, a,...
4807 [When, ambitious, New, York, attorney, Sam, is...
4808 [Ever, since, the, second, grade, when, he, fi...
Name: overview, Length: 4806, dtype: object
```

```
In [39]: movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

```
In [40]: movies.head(1)
```

Out[40]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[James Cameron]

```
In [41]: # Sam Worthington

# Convert

# SamWorthington
```

```
In [42]: movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])
```

```
Out[42]: 0 [Action, Adventure, Fantasy, ScienceFiction]
1 [Adventure, Fantasy, Action]
2 [Action, Adventure, Crime]
3 [Action, Crime, Drama, Thriller]
4 [Action, Adventure, ScienceFiction]

...
4804 [Action, Crime, Thriller]
4805 [Comedy, Romance]
4806 [Comedy, Drama, Romance, TVMovie]
4807 []
4808 [Documentary]
Name: genres, Length: 4806, dtype: object
```

```
In [43]: movies['genres'] = movies['genres'].apply(lambda x:[i.replace(" ", "") for i in x])

movies['keywords'] = movies['keywords'].apply(lambda x:[i.replace(" ", "") for i in x])

movies['cast'] = movies['cast'].apply(lambda x:[i.replace(" ", "") for i in x])

movies['crew'] = movies['crew'].apply(lambda x:[i.replace(" ", "") for i in x])
```

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

Out[44]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]
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]	[GoreVerbinski]
2	206647	Spectre	[A, cryptic, message, from, Bond's, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]
4	49529	John Carter	[John, Carter, is, a, war-weary,, former, mili...	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p...	[TaylorKitsch, LynnCollins, SamanthaMorton]	[AndrewStanton]

```
In [45]: movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords'] + movies['cast'] + movies['crew']
```

```
In [46]: movies.head(1)
```

Out[46]:

	movie_id	title	overview	genres	keywords	cast	crew	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[In, the, 22nd, century,, a, paraplegic, Marin...

```
In [47]: new_df = movies[["movie_id","title","tags"]]
new_df.head(2)
```

Out[47]:

	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...

```
In [48]: # This is called steaming

["loved","loving","love"]

["love","love","love"]
```

Out[48]: ['love', 'love', 'love']

```
In [49]: import nltk

from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
```

```
In [50]: # Example

def stem1(text):
    y = []

    for i in text.split():
        ps.stem()
```

```
In [51]: def stem(text):
    y = []

    for i in text.split():
        y.append(ps.stem(i))

    return " ".join(y)
```

```
In [52]: new_df['tags'][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.',
          'Action',
          'Adventure',
          'Fantasy',
          'ScienceFiction',
          'cultureclash',
          'future',
          'spacewar',
          'spacecolony',
          'society',
          'spacetravel',
          'futuristic',
          'romance',
          'space',
          'alien',
          'tribe',
          'alienplanet',
          'cgi',
          'marine',
          'soldier',
          'battle',
          'loveaffair',
          'antiwar',
          'powerrelations',
          'mindandsoul',
          '3d',
          'SamWorthington',
          'ZoeSaldana',
          'SigourneyWeaver',
          'JamesCameron']
```

```
In [53]: new_df['tags'] = new_df["tags"].apply(lambda x:" ".join(x))
```

```
In [54]: new_df.head(1)
```

Out[54]:

	movie_id	title	tags
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...

```
In [55]: new_df['tags'][0]
```

```
Out[55]: 'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes to
rn between following orders and protecting an alien civilization. Action Adventure Fantasy ScienceFiction cultu
reclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi ma
rine soldier battle loveaffair antiwar powerrelations mindandsoul 3d SamWorthington ZoeSaldana SigourneyWeaver
JamesCameron'
```

```
In [56]: new_df['tags'].apply(lambda x:x.lower())
```

```
Out[56]: 0      in the 22nd century, a paraplegic marine is di...
1      captain barbossa, 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: tags, Length: 4806, dtype: object
```



```
In [57]: # Convert all in lower case
new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
In [58]: new_df.head(1)
```

```
Out[58]:
```

	movie_id	title	tags
0	19995	Avatar	in the 22nd century, a paraplegic marine is di...

```
In [59]: new_df['tags'][0]
```

```
Out[59]: 'in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes to
rn between following orders and protecting an alien civilization. action adventure fantasy sciencefiction cultu
reclash future spacewar spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi ma
rine soldier battle loveaffair antiwar powerrelations mindandsoul 3d samworthington zoesaldana sigourneyweaver
jamescameron'
```

```
In [60]: new_df['tags'][1]
```

```
Out[60]: "captain barbossa, long believed to be dead, has come back to life and is headed to the edge of the earth with
will turner and elizabeth swann. but nothing is quite as it seems. adventure fantasy action ocean drugabuse exo
ticisland eastindiatradingcompany loveofone'slife traitor shipwreck strongwoman ship alliance calypso afterlife
fighter pirate swashbuckler aftercreditsstinger johnnydepp orlandobloom keiraknightley goreverbinski"
```

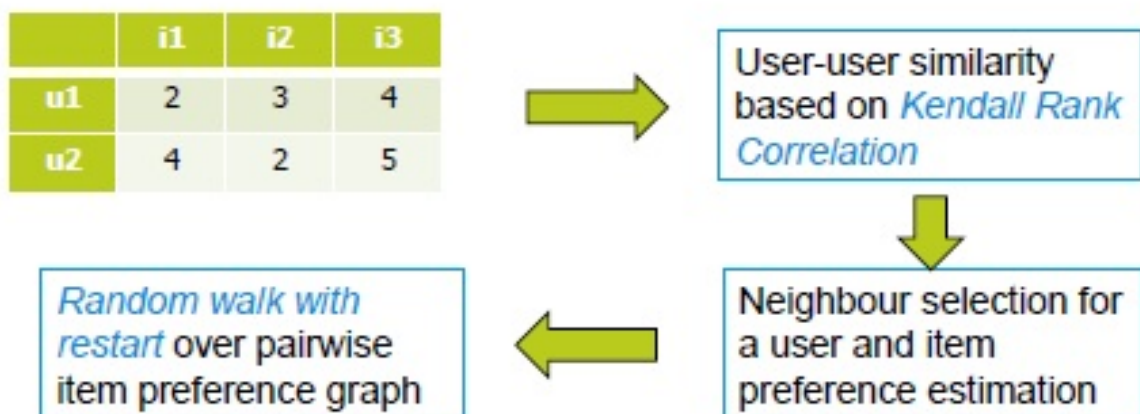
Ranking vs. Recommendation

People sometimes confuse between ranking (or search ranking) system and recommender system, and some may even think they are interchangeable. While both algorithms are trying to present items in a sorted way, there are some key differences between these two:

Ranking algorithms normally put more relevant items closer to the top of the showing list whereas recommender systems sometimes try to avoid overspecialization. A good recommender system should not recommend items that are too similar to what users have seen before, and should diversify its recommendations.

Ranking algorithms rely on search query provided by users, who know what they are looking for. Recommender systems, on the other hand, without any explicit inputs from users, aim to discovering things they might not have found otherwise.

Ranking-based similarity instead of rating based similarity



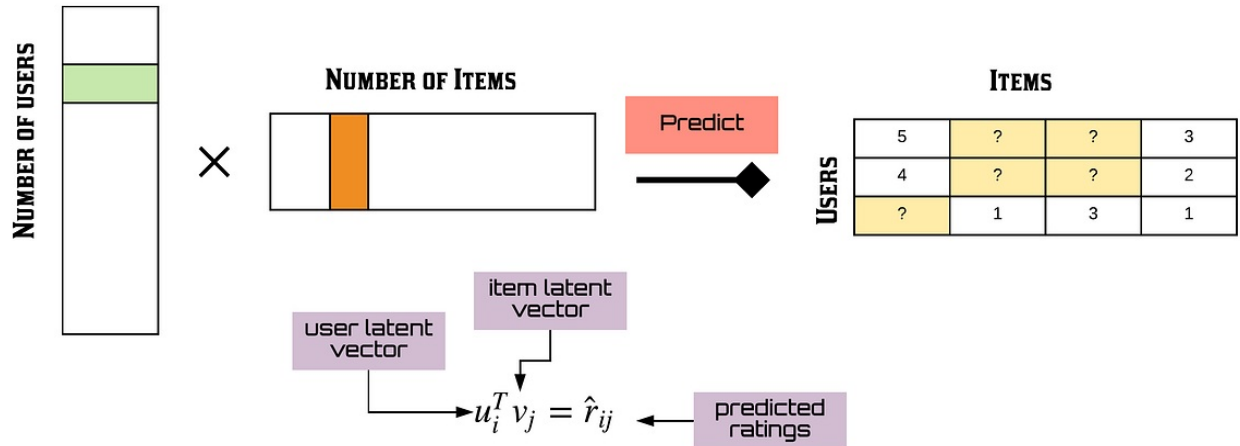
Matrix Factorization

Latent factor models compress user-item matrix into a low-dimensional representation in terms of latent factors. One advantage of using this approach is that instead of having a high dimensional matrix containing abundant number of missing values we will be dealing with a much smaller matrix in lower-dimensional space.

A reduced presentation could be utilized for either user-based or item-based neighborhood algorithms that are presented in the previous section. There are several advantages with this paradigm. It handles the sparsity of the original matrix better than memory based ones. Also comparing similarity on the resulting matrix is much more scalable especially in dealing with large sparse datasets.

An important decision is the number of factors to factor the user-item matrix. The higher the number of factors, the more precise is the factorization in the original matrix reconstructions. Therefore, if the model is allowed to memorize too much details of the original matrix, it may not generalize well for data it was not trained on. Reducing the number of factors increases the model generalization.

MATRIX FACTORIZATION



```
In [61]: from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000, stop_words='english')
```

```
In [62]: vector = cv.fit_transform(new_df['tags']).toarray()
```

```
In [63]: vector
```

```
Out[63]: array([[0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 ...,
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0],
 [0, 0, 0, ..., 0, 0, 0]])
```

```
In [64]: vector[0]
```

```
Out[64]: array([0, 0, 0, ..., 0, 0, 0])
```

```
In [65]: vector.shape
```

```
Out[65]: (4806, 5000)
```

```
In [66]: cv.get_feature_names()
```

```
Out[66]: ['000',
 '007',
 '10',
 '100',
 '11',
 '12',
 '13',
 '14',
 '15',
 '16',
 '17',
 '18',
 '18th',
 '19',
 '1930s',
 '1940s',
 '1944',
 '1950',
 '1950s',
 '1960s',
 '1970s',
 '1971',
 '1974',
 '1976',
 '1980',
 '1980s',
```

'1985',
'1990s',
'1999',
'19th',
'19thcentury',
'20',
'200',
'2003',
'2009',
'20th',
'24',
'25',
'30',
'300',
'3d',
'40',
'50',
'500',
'60',
'60s',
'70',
'aaron',
'aaroneckhart',
'abandoned',
'abducted',
'abigailbreslin',
'abilities',
'ability',
'able',
'aboard',
'abuse',
'abusive',
'academic',
'academy',
'accept',
'accepted',
'accepts',
'access',
'accident',
'accidental',
'accidentally',
'accompanied',
'accomplish',
'account',
'accountant',
'accused',
'ace',
'achieve',
'act',
'acting',
'action',
'actionhero',
'actions',
'activist',
'activities',
'activity',
'actor',
'actors',
'actress',
'acts',
'actual',
'actually',
'adam',
'adams',
'adamsandler',
'adamshankman',
'adaptation',
'adapted',
'addict',
'addicted',
'addiction',
'adolescence',
'adopt',
'adopted',
'adoption',
'adopts',
'adrienbrody',
'adult',
'adultery',
'adulthood',
'adults',
'advantage',
'adventure',
'adventures',
'advertising',
'advice',
'affair',
'affairs',
'affection',

'affections',
'afghanistan',
'africa',
'african',
'africanamerican',
'aftercreditsstinger',
'afterlife',
'aftermath',
'age',
'aged',
'agedifference',
'agency',
'agenda',
'agent',
'agents',
'aggressive',
'aging',
'ago',
'agree',
'agrees',
'ahead',
'aid',
'aided',
'aids',
'ailing',
'air',
'airplane',
'airplanecrash',
'airport',
'aka',
'al',
'alabama',
'alan',
'alaska',
'albert',
'alcohol',
'alcoholic',
'alcoholism',
'alecbaldwin',
'alex',
'alfredhitchcock',
'ali',
'alice',
'alien',
'alieninvasion',
'alienlife',
'aliens',
'alike',
'alive',
'allen',
'alliance',
'allied',
'allies',
'allow',
'allows',
'ally',
'alongside',
'alpacino',
'alter',
'alternate',
'alternative',
'alzheimer',
'amanda',
'amandapeet',
'amandaseyfried',
'amateur',
'amazing',
'ambassador',
'ambition',
'ambitious',
'ambulance',
'ambush',
'america',
'american',
'americanabroad',
'americanfootball',
'americans',
'amid',
'amidst',
'amnesia',
'amp',
'amsterdam',
'amusementpark',
'amy',
'amyadams',
'amysmart',
'analyst',
'anarchiccomedy',
'ancient',

'ancientrome',
'ancientworld',
'anderson',
'andiemacdowell',
'andrew',
'android',
'andy',
'andygarcía',
'angel',
'angelabassett',
'angeles',
'angelinajolie',
'angels',
'anger',
'anglee',
'angry',
'animal',
'animalattack',
'animalhorror',
'animals',
'animated',
'animation',
'ann',
'anna',
'annabelle',
'annafaris',
'annakendrick',
'anne',
'annehathaway',
'annemoss',
'annettebening',
'annie',
'anniversary',
'announces',
'annual',
'anonymity',
'anonymous',
'answer',
'answers',
'ant',
'antarctic',
'anthology',
'anthony',
'anthonyanderson',
'anthonyhopkins',
'anthonymackie',
'anthropomorphism',
'anti',
'antics',
'antihero',
'antoinefuqua',
'antoniobanderas',
'antonyelchin',
'apart',
'apartheid',
'apartment',
'ape',
'apes',
'apocalypse',
'apocalyptic',
'apparent',
'apparently',
'appear',
'appears',
'apple',
'appointed',
'apprentice',
'approach',
'approaches',
'approaching',
'april',
'aquarium',
'arab',
'arch',
'archaeologist',
'archeology',
'archer',
'architect',
'arctic',
'area',
'aren',
'arena',
'argument',
'arise',
'aristocrat',
'armed',
'arms',
'army',
'arnold',

'arnoldschwarzenegger',
'arrangedmarriage',
'arrangement',
'arrest',
'arrested',
'arrival',
'arrive',
'arrives',
'arriving',
'arrogant',
'art',
'arthur',
'artificialintelligence',
'artist',
'artistic',
'artists',
'arts',
'ashley',
'ashleyjudd',
'ashtonkutcher',
'asia',
'aside',
'ask',
'asked',
'asking',
'asks',
'aspirations',
'aspiring',
'assassin',
'assassinate',
'assassination',
'assassins',
'assault',
'assigned',
'assignment',
'assistant',
'assumes',
'asteroid',
'astronaut',
'astronauts',
'asylum',
'atheist',
'athlete',
'atlantic',
'atomicbomb',
'attack',
'attacked',
'attacks',
'attempt',
'attempting',
'attempts',
'attending',
'attends',
'attention',
'attic',
'attitude',
'attorney',
'attracted',
'attraction',
'attractive',
'audience',
'audiences',
'audition',
'august',
'aunt',
'austin',
'australia',
'australian',
'author',
'authorities',
'authority',
'autism',
'auto',
'avenge',
'average',
'avoid',
'awaits',
'awakens',
'award',
'away',
'awry',
'ax',
'babe',
'baby',
'bachelor',
'backdrop',
'background',
'backgrounds',
'bad',

'bag',
'bahamas',
'bail',
'balance',
'ball',
'ballet',
'baltimore',
'band',
'bandits',
'bangkok',
'banished',
'bank',
'banker',
'bankrobber',
'bankrobbery',
'bar',
'barely',
'bargained',
'barn',
'barney',
'barry',
'barrylevinson',
'barrysonnenfeld',
'bars',
'base',
'baseball',
'based',
'basedoncomicbook',
'basedongraphicnovel',
'basedonnovel',
'basedonplay',
'basedonstagemusical',
'basedontrueevents',
'basedontruestory',
'basedontvseries',
'basedonvideogame',
'basedonyoungadultnovel',
'basement',
'basketball',
'batman',
'battle',
'battlefield',
'battles',
'battling',
'bay',
'beach',
'bear',
'bears',
'beast',
'beasts',
'beat',
'beating',
'beautiful',
'beautifulwoman',
'beauty',
'becky',
'becominganadult',
'bed',
'bedroom',
'beer',
'befriends',
'began',
'begin',
'beginning',
'begins',
'behavior',
'beings',
'beliefs',
'believe',
'believed',
'believes',
'believing',
'beloved',
'ben',
'benaffleck',
'beneath',
'benfoster',
'beniciodeltoro',
'benjamin',
'benjaminbratt',
'benkingsley',
'bennett',
'benstilller',
'bent',
'berlin',
'best',
'bestfriend',
'bet',
'beth',

'betrayal',
'betrayed',
'bettemidler',
'better',
'betty',
'beverly',
'bible',
'big',
'bigger',
'biggest',
'biker',
'bikini',
'billhader',
'billionaire',
'billmurray',
'billnighy',
'billpaxton',
'billpullman',
'billy',
'billybobthornton',
'billycrudup',
'billycrystal',
'biography',
'bird',
'birth',
'birthday',
'bisexual',
'bishop',
'bit',
'bite',
'bitter',
'bizarre',
'black',
'blackmagic',
'blackmail',
'blackpeople',
'blade',
'blame',
'blind',
'bliss',
'block',
'blonde',
'blood',
'bloodsplatter',
'bloodthirsty',
'bloody',
'blow',
'blue',
'board',
'boarding',
'boardingschool',
'boat',
'bob',
'bobby',
'bobbyfarrelly',
'bobhoskins',
'bodies',
'body',
'bodyguard',
'bold',
'bollywood',
'bomb',
'bombing',
'bond',
'bonds',
'bone',
'book',
'books',
'border',
'bored',
'boredom',
'boring',
'born',
'boss',
'boston',
'botched',
'bound',
'boundaries',
'bounty',
'bountyhunter',
'box',
'boxer',
'boxing',
'boy',
'boyfriend',
'boys',
'bradleycooper',
'bradpitt',
'brain',

'brand',
'brandon',
'brave',
'bravery',
'brazil',
'brazilian',
'break',
'breakdown',
'breaking',
'breaks',
'breast',
'brendanfraser',
'brendangleeson',
'brent',
'brettratner',
'brian',
'briandepalma',
'bride',
'bridge',
'brief',
'brien',
'bright',
'brilliant',
'bring',
'bringing',
'brings',
'brink',
'britain',
'british',
'britishsecretservice',
'brittanymurphy',
'broadway',
'broke',
'broken',
'broker',
'brooklyn',
'brooks',
'brothel',
'brother',
'brotherbrotherrelationship',
'brothers',
'brothersisterrelationship',
'brought',
'brown',
'bruce',
'brucewillis',
'brutal',
'brutality',
'brutally',
'bryansinger',
'buck',
'buddies',
'buddy',
'buddycomedy',
'budget',
'build',
'building',
'built',
'bully',
'bullying',
'bumbling',
'bunny',
'burglar',
'buried',
'bus',
'bush',
'business',
'businessman',
'bust',
'busy',
'butler',
'buy',
'cabin',
'caesar',
'cage',
'cairo',
'cal',
'california',
'called',
'calls',
'calvin',
'camcorder',
'came',
'camera',
'cameraman',
'cameras',
'camerondiaz',
'camp',
'campaign',

'campbell',
'camping',
'campus',
'canada',
'canadian',
'cancer',
'candidate',
'candy',
'cannibal',
'capable',
'capital',
'capitalism',
'capt',
'captain',
'captive',
'capture',
'captured',
'captures',
'car',
'caraccident',
'carchase',
'carcrash',
'card',
'care',
'career',
'carefree',
'caretaker',
'caribbean',
'carjourney',
'carl',
'carlagugino',
'carmen',
'carol',
'carolina',
'carrace',
'carrie',
'carry',
'carrying',
'cars',
'cartel',
'carter',
'cartoon',
'caryelwes',
'case',
'caseyaffleck',
'cash',
'casino',
'cast',
'castle',
'cat',
'cataclysm',
'catastrophe',
'catch',
'catches',
'cateblanchett',
'catherinedeneuve',
'catherinekeener',
'catherinezeta',
'catholic',
'catholicism',
'cattle',
'caught',
'cause',
'caused',
'causes',
'causing',
'cavalry',
'cave',
'cavemen',
'celebrate',
'celebrated',
'celebration',
'celebrity',
'cell',
'cellphone',
'cemetery',
'center',
'centered',
'centers',
'central',
'centuries',
'century',
'ceo',
'ceremony',
'certain',
'chad',
'chain',
'chainsaw',
'challenge',

'challenged',
'challenges',
'champion',
'championship',
'chance',
'change',
'changed',
'changes',
'changing',
'channingtatum',
'chaos',
'chaotic',
'chapter',
'character',
'characters',
'charge',
'charged',
'charismatic',
'charles',
'charlie',
'charliesheen',
'charlizetheron',
'charlotte',
'charm',
'charming',
'chase',
'chased',
'chauffeur',
'cheating',
'cheerleader',
'chef',
'chemical',
'cher',
'chicago',
'chicken',
'chief',
'child',
'childabuse',
'childhero',
'childhood',
'children',
'chilling',
'china',
'chinese',
'chip',
'chiwetelejofor',
'chloëgracemoretz',
'chloësevigny',
'chocolate',
'choice',
'choices',
'choose',
'chosen',
'chris',
'chris columbus',
'chriscooper',
'chrisevans',
'chrishemsworth',
'chrisklein',
'chrispine',
'chrisrock',
'christ',
'christian',
'christianbale',
'christianity',
'christianslater',
'christinaapplegate',
'christinaricci',
'christine',
'christmas',
'christmasparty',
'christmastree',
'christopher',
'christopherlambert',
'christopherlloyd',
'christophernolan',
'christopherplummer',
'christopherwalken',
'christophwaltz',
'chronicle',
'chronicles',
'chuck',
'church',
'cia',
'cigarettesmoking',
'cillianmurphy',
'cinema',
'circle',
'circuit',

'circumstances',
'circus',
'cities',
'citizens',
'city',
'civil',
'civilization',
'civilwar',
'claim',
'claims',
'claire',
'clairedanes',
'clan',
'clark',
'clash',
'class',
'classes',
'classic',
'classmate',
'classmates',
'classroom',
'claudevandamme',
'clay',
'clean',
'clear',
'clerk',
'client',
'clients',
'climate',
'climbing',
'clinteastwood',
'cliveowen',
'clock',
'clone',
'cloning',
'close',
'closer',
'club',
'clubs',
'clues',
'clutches',
'coach',
'coast',
'cocaine',
'code',
'cody',
'coffin',
'cohen',
'col',
'cold',
'coldwar',
'cole',
'colin',
'colinfarrell',
'colinfirth',
'collapse',
'colleague',
'colleagues',
'collect',
'collection',
'collector',
'college',
'collide',
'collins',
'collision',
'colombia',
'colonel',
'colony',
'color',
'colorado',
'colorful',
'coma',
'combat',
'combination',
'combined',
'come',
'comeback',
'comedian',
'comedic',
'comedy',
'comes',
'comet',
'comfort',
'comic',
'comics',
'coming',
'comingofage',
'comingout',
'command',

```

'commander',
'commercial',
'commit',
'commitment',
'committed',
'common',
'communication',
'communist',
'community',
'companion',
'company',
'compete',
'competing',
'competition',
'competitive',
'complete',
'completely',
'complex',
'complicated',
'complications',
'composer',
'computer',
'computervirus',
'conan',
'concert',
'conclusion',
'condition',
'conditions',
'confession',
'confidence',
'confident',
'conflict',
'confront',
'confronted',
'confused',
'congress',
'conman',
'connected',
'connection',
'connell',
'connor',
'conquer',
'conscience',
'consequences',
'conservative',
'considered',
'conspiracy',
'constant',
'constantly',
'construction',
'contact',
'contain',
'contemporary',
'contend',
'contest',
'continue',
'continues',
'continuing',
'contract',
'control',
'controlled',
'controlling',
'controversial',
'convention',
'converge',
'convict',
'convicted',
'convince',
'convinced',
'convinces',
'cook',
'cooking',
'cool',
'cooper',
'cop',
'cope',
'cops',
'core',
'corner',
'corners',
'corporate',
'corporation',
'corpse',
'corrupt',
...]
```

In [67]: # Steaming

```
["loved", "loving", "love"]
```

```

["love","love","love"]

Out[67]: ['love', 'love', 'love']

In [68]: ps.stem('loved')

Out[68]: 'love'

In [69]: ps.stem('loving')

Out[69]: 'love'

In [70]: ps.stem('love')

Out[70]: 'love'

In [71]: stem('in the 22nd century, a paraplegic marine is dispatched to the moon pandora on a unique mission, but becom
Out[71]: 'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu mission, but becom torn betwe
en follow order and protect an alien civilization. action adventur fantasi sciencefict cultureclash futur space
war spacecoloni societi spacetravel futurist romanc space alien tribe alienplanet cgi marin soldier battl lovea
ffair antiwar powerrel mindandsoul 3d samworthington zoesaldana sigourneyweav jamescameron'

In [72]: from sklearn.metrics.pairwise import cosine_similarity

In [73]: similarity = cosine_similarity(vector)

In [74]: similarity

Out[74]: array([[1.          , 0.08964215, 0.06071767, ..., 0.02519763, 0.0277885 ,
        [0.08964215, 1.          , 0.06350006, ..., 0.02635231, 0.          ,
        [0.06071767, 0.06350006, 1.          , ..., 0.02677398, 0.          ,
        ...,
        [0.02519763, 0.02635231, 0.02677398, ..., 1.          , 0.07352146,
        0.04774099],
        [0.0277885 , 0.          , 0.          , ..., 0.07352146, 1.          ,
        0.05264981],
        [0.          , 0.          , 0.          , ..., 0.04774099, 0.05264981,
        1.          ]])

In [75]: similarity[0]

Out[75]: array([1.          , 0.08964215, 0.06071767, ..., 0.02519763, 0.0277885 ,
        [0.          ]])

In [76]: new_df[new_df['title'] == 'The Lego Movie'].index[0]

Out[76]: 744

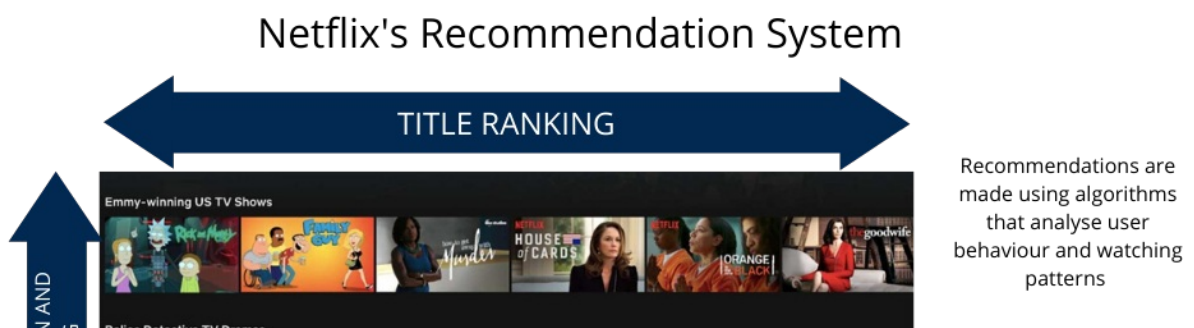
In [77]: sorted(list(enumerate(similarity[0])),reverse=True,key= lambda x:x[1])[1:6]

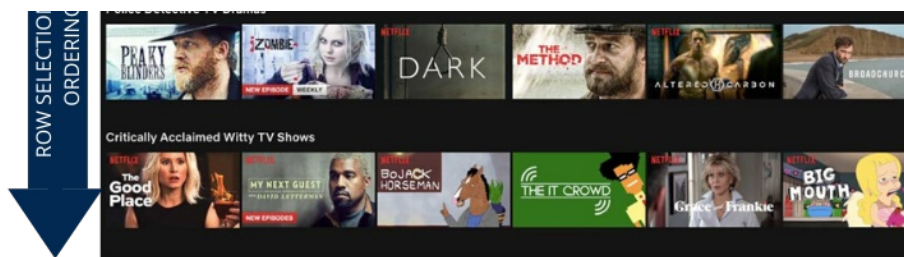
Out[77]: [(539, 0.26089696604360174),
        (1194, 0.2581988897471611),
        (260, 0.25110592822973776),
        (1216, 0.24944382578492943),
        (507, 0.24846467329894412)]

In [78]: def recommend(movie):
        index = new_df[new_df['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_df.iloc[i[0]].title)

```

Movies recommended according to similarity





80% of content streamed on **Netflix** comes from the recommendation algorithm

```
In [79]: print(f'\033[94m')
         recommend('Spectre')
```

Quantum of Solace
Never Say Never Again
Skyfall
Thunderball
From Russia with Love

```
In [80]: print(f'\033[94m')
         recommend('John Carter')
```

Star Trek: Insurrection
Mission to Mars
Captain America: The First Avenger
Escape from Planet Earth
Ghosts of Mars

References

<https://www.zeolearn.com/magazine/recommendation-systems-in-machine-learning>

<https://www.analyticssteps.com/blogs/what-are-recommendation-systems-machine-learning>

<https://medium.com/@zeolearn/recommendation-systems-in-machine-learning-96444fd90702>

<https://www.itransition.com/machine-learning/recommendation-systems>

<https://muitnoida.edu.in/recommendation-system-in-machine-learning/>

https://www.researchgate.net/figure/Types-of-recommendation-algorithms_fig5_322360704

```
In [ ]:
```