

movie-recommendation-system

March 11, 2024

```
[1]: import os
      os.getcwd()
```

```
[1]: 'C:\\Users\\JoyBoy\\Desktop\\Code\\Learnbay\\Code\\ML\\Project\\Recommendation
      System'
```

```
[2]: import numpy as np
      import pandas as pd

      import matplotlib.pyplot as plt
      %matplotlib inline
      import seaborn as sns
      sns.set()

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

```
[ ]:
```

1 Load the data

```
[ ]:
```

```
[3]: movies = pd.read_csv("movies.csv")
      credits = pd.read_csv("credits.csv")
```

```
[4]: movies.head(2)
```

```
[4]:      budget      genres \
0  237000000 [{"id": 28, "name": "Action"}, {"id": 12, "nam...
1  300000000 [{"id": 12, "name": "Adventure"}, {"id": 14, "...

      homepage      id \
0      http://www.avatarmovie.com/  19995
1  http://disney.go.com/disneypictures/pirates/  285

      keywords original_language \
```

```

0 [{"id": 1463, "name": "culture clash"}, {"id":... en
1 [{"id": 270, "name": "ocean"}, {"id": 726, "na... en

                                original_title \
0                                Avatar
1 Pirates of the Caribbean: At World's End

                                overview popularity \
0 In the 22nd century, a paraplegic Marine is di... 150.437577
1 Captain Barbossa, long believed to be dead, ha... 139.082615

                                production_companies \
0 [{"name": "Ingenious Film Partners", "id": 289...
1 [{"name": "Walt Disney Pictures", "id": 2}, {""...

                                production_countries release_date revenue \
0 [{"iso_3166_1": "US", "name": "United States o... 2009-12-10 2787965087
1 [{"iso_3166_1": "US", "name": "United States o... 2007-05-19 961000000

runtime                                spoken_languages status \
0 162.0 [{"iso_639_1": "en", "name": "English"}, {"iso... Released
1 169.0 [{"iso_639_1": "en", "name": "English"}] Released

                                tagline \
0                                Enter the World of Pandora.
1 At the end of the world, the adventure begins.

                                title vote_average vote_count
0                                Avatar 7.2 11800
1 Pirates of the Caribbean: At World's End 6.9 4500

```

```
[5]: credits.head(2)
```

```

[5]: movie_id                                title \
0 19995                                Avatar
1 285 Pirates of the Caribbean: At World's End

                                cast \
0 [{"cast_id": 242, "character": "Jake Sully", "...
1 [{"cast_id": 4, "character": "Captain Jack Spa...

                                crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...
1 [{"credit_id": "52fe4232c3a36847f800b579", "de...

```

```
[6]: print(movies.shape)
      print(credits.shape)
```

```
(4803, 20)
(4803, 4)
```

```
[7]: print(movies.isnull().sum())
      print()
      print(credits.isnull().sum())
```

```
budget          0
genres          0
homepage        3091
id              0
keywords        0
original_language 0
original_title  0
overview        3
popularity      0
production_companies 0
production_countries 0
release_date    1
revenue         0
runtime         2
spoken_languages 0
status          0
tagline         844
title           0
vote_average    0
vote_count      0
dtype: int64
```

```
movie_id    0
title       0
cast        0
crew        0
dtype: int64
```

```
[ ]:
```

```
[8]: print(movies.columns)
      print(credits.columns)
```

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

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

```
[10]: df.shape
```

```
[10]: (4809, 23)
```

```
[11]: df.head(1)
```

```
[11]:      budget      genres \
0  237000000 [{"id": 28, "name": "Action"}, {"id": 12, "nam...

      homepage      id \
0  http://www.avatarmovie.com/  19995

      keywords original_language \
0 [{"id": 1463, "name": "culture clash"}, {"id":...      en

      original_title      overview \
0      Avatar  In the 22nd century, a paraplegic Marine is di...

      popularity      production_companies ... runtime \
0  150.437577 [{"name": "Ingenious Film Partners", "id": 289...      162.0

      spoken_languages      status \
0 [{"iso_639_1": "en", "name": "English"}, {"iso...      Released

      tagline      title vote_average vote_count movie_id \
0  Enter the World of Pandora.  Avatar      7.2      11800      19995

      cast \
0 [{"cast_id": 242, "character": "Jake Sully", "...

      crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...

[1 rows x 23 columns]
```

```
[12]: df.isnull().sum()
```

```
[12]: budget      0
      genres      0
      homepage    3096
      id          0
      keywords     0
      original_language  0
      original_title  0
      overview     3
```

popularity	0
production_companies	0
production_countries	0
release_date	1
revenue	0
runtime	2
spoken_languages	0
status	0
tagline	844
title	0
vote_average	0
vote_count	0
movie_id	0
cast	0
crew	0
dtype:	int64

2 Preprocessing data

```
[13]: df.isnull().sum()/len(df)*100
```

```
[13]: budget          0.000000
genres              0.000000
homepage           64.379289
id                 0.000000
keywords           0.000000
original_language  0.000000
original_title     0.000000
overview           0.062383
popularity         0.000000
production_companies 0.000000
production_countries 0.000000
release_date       0.020794
revenue            0.000000
runtime            0.041589
spoken_languages   0.000000
status             0.000000
tagline            17.550426
title              0.000000
vote_average       0.000000
vote_count         0.000000
movie_id           0.000000
cast               0.000000
crew               0.000000
dtype: float64
```

```
[14]: df.columns
```

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

```
[15]: ## Significant columns for my RC  
      'genres', 'keywords', 'overview', 'title', 'movie_id', 'cast', 'crew'
```

```
[15]: ('genres', 'keywords', 'overview', 'title', 'movie_id', 'cast', 'crew')
```

```
[16]: df['homepage'].value_counts()
```

```
[16]: homepage  
http://www.missionimpossible.com/      4  
http://www.thehungergames.movie/      4  
http://www.kungfupanda.com/           3  
http://www.transformersmovie.com/      3  
http://www.thehobbit.com/             3  
..  
http://www.ridealong.com/             1  
http://www.theinternmovie.com/        1  
http://www.miramax.com/movie/the-talented-mr-ripley/  1  
http://www.julieandjulia.com/         1  
http://shanghaicalling.com/          1  
Name: count, Length: 1691, dtype: int64
```

```
[17]: df['original_language'].value_counts()/len(df['original_language'])*100
```

```
[17]: original_language  
en      93.782491  
fr       1.455604  
es       0.665419  
zh       0.561447  
de       0.561447  
hi       0.395093  
ja       0.332710  
it       0.291121  
ko       0.249532  
cn       0.249532  
ru       0.228738  
pt       0.187149  
da       0.145560  
sv       0.103972
```

```

nl      0.083177
fa      0.083177
th      0.062383
he      0.062383
ta      0.041589
cs      0.041589
ro      0.041589
id      0.041589
ar      0.041589
vi      0.020794
sl      0.020794
ps      0.020794
no      0.020794
ky      0.020794
hu      0.020794
pl      0.020794
af      0.020794
nb      0.020794
tr      0.020794
is      0.020794
xx      0.020794
te      0.020794
el      0.020794
Name: count, dtype: float64

```

```
[18]: df1 = df.copy()
```

```
[19]: df1 = df1[['genres', 'keywords', 'overview', 'title', 'movie_id', 'cast',
↳ 'crew']]
```

```
[20]: df1.head(1)
```

```

[20]:                                     genres \
0  [{"id": 28, "name": "Action"}, {"id": 12, "nam...

                                     keywords \
0  [{"id": 1463, "name": "culture clash"}, {"id"::...

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

                                     cast \
0  [{"cast_id": 242, "character": "Jake Sully", "...

                                     crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...

```

```
[21]: df1.shape
```

```
[21]: (4809, 7)
```

```
[22]: # For Genres column:  
df1['genres'][0]
```

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

```
[23]: df1['genres'][2]
```

```
[23]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 80,  
"name": "Crime"}]'
```

```
[ ]:
```

```
[24]: import ast  
  
def conversion(text):  
    data = []  
    for _ in ast.literal_eval(text):  
        data.append(_['name'])  
    return data
```

```
[25]: df1['genres'] = df1['genres'].apply(conversion)
```

```
[ ]:
```

```
[26]: df1.head(1)
```

```
[26]:          genres \  
0  [Action, Adventure, Fantasy, Science Fiction]  
  
          keywords \  
0  [{"id": 1463, "name": "culture clash"}, {"id":...  
  
          overview  title  movie_id \  
0  In the 22nd century, a paraplegic Marine is di...  Avatar      19995  
  
          cast \  
0  [{"cast_id": 242, "character": "Jake Sully", "...  
  
          crew  
0  [{"credit_id": "52fe48009251416c750aca23", "de...
```

```
[27]: df1['genres'][0]
```



```
[27]: ['Action', 'Adventure', 'Fantasy', 'Science Fiction']
```

```
[28]: df1['keywords'][0]
```

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

```
[29]: df1['keywords'] = df1['keywords'].apply(conversion)
```

```
[30]: df1['keywords'][0]
```

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

```
[31]: df1.head(1)
```

```
[31]:          genres \  
0  [Action, Adventure, Fantasy, Science Fiction]  
  
          keywords \
```

```

0 [culture clash, future, space war, space colon...

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

                                cast  \
0 [{"cast_id": 242, "character": "Jake Sully", "...

                                crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...

```

```
[32]: df1['crew'][0]
```

```

[32]: '[{"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":
 "52fe48009251416c750aca11", "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": 0, "id": 89714,
 "job": "Stunts", "name": "Ilram 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": "54959cdffc3a3681153002729", "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 Karena"}, {"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": 2, "id": 1330562, "job": "Costume Supervisor", "name": "Robert Bavin"}]

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 Caira"}, {"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": "549596009251413af70028cc", "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 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 \Max\ 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":
"5592b0dbc3a368774b00112c", "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"}]
```

[]:

[33]: df1.head(3)


```

[33]:                                     genres \
0  [Action, Adventure, Fantasy, Science Fiction]
1                                     [Adventure, Fantasy, Action]
2                                     [Action, Adventure, Crime]

                                     keywords \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...
2  [spy, based on novel, secret agent, sequel, mi...

                                     overview \
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...

                                     title  movie_id \
0                                     Avatar      19995
1  Pirates of the Caribbean: At World's End      285
2                                     Spectre     206647

                                     cast \
0  [{"cast_id": 242, "character": "Jake Sully", "...
1  [{"cast_id": 4, "character": "Captain Jack Spa...
2  [{"cast_id": 1, "character": "James Bond", "cr...

                                     crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...
1  [{"credit_id": "52fe4232c3a36847f800b579", "de...
2  [{"credit_id": "54805967c3a36829b5002c41", "de...

```

```
[34]: df1['genres'][0]
```

```
[34]: ['Action', 'Adventure', 'Fantasy', 'Science Fiction']
```

```
[35]: df1['keywords'][0]
```

```
[35]: ['culture clash',
      'future',
      'space war',
      'space colony',
      'society',
      'space travel',
      'futuristic',
      'romance',
      'space',
      'alien',
      'tribe',
```

```
'alien planet',
'cgi',
'marine',
'soldier',
'battle',
'love affair',
'anti war',
'power relations',
'mind and soul',
'3d']
```

```
[36]: df1.head(3)
```

```
[36]:
          genres \
0  [Action, Adventure, Fantasy, Science Fiction]
1                [Adventure, Fantasy, Action]
2                [Action, Adventure, Crime]

          keywords \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...
2  [spy, based on novel, secret agent, sequel, mi...

          overview \
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...

          title  movie_id \
0                Avatar    19995
1  Pirates of the Caribbean: At World's End     285
2                Spectre    206647

          cast \
0  [{"cast_id": 242, "character": "Jake Sully", "...
1  [{"cast_id": 4, "character": "Captain Jack Spa...
2  [{"cast_id": 1, "character": "James Bond", "cr...

          crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...
1  [{"credit_id": "52fe4232c3a36847f800b579", "de...
2  [{"credit_id": "54805967c3a36829b5002c41", "de...
```

```
[37]: df1['cast'][0]
```

```
[37]: ' [{"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 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": "Sonia 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": "52fe48019251416c750acaeb", "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 Centreworker (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}]'
```

```
[38]: df1.head(2)
```

```
[38]:
      genres \
0  [Action, Adventure, Fantasy, Science Fiction]
1  [Adventure, Fantasy, Action]

      keywords \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...

      overview \
0  In the 22nd century, a paraplegic Marine is di...
1  Captain Barbossa, long believed to be dead, ha...

      title  movie_id \
0  Avatar  19995
1  Pirates of the Caribbean: At World's End  285
```

```

                                cast \
0 [{"cast_id": 242, "character": "Jake Sully", "...
1 [{"cast_id": 4, "character": "Captain Jack Spa...

```

```

                                crew
0 [{"credit_id": "52fe48009251416c750aca23", "de...
1 [{"credit_id": "52fe4232c3a36847f800b579", "de...

```

```
[39]: df1.isnull().sum()/len(df1)*100
```

```

[39]: genres      0.000000
      keywords    0.000000
      overview    0.062383
      title       0.000000
      movie_id    0.000000
      cast        0.000000
      crew        0.000000
      dtype: float64

```

```
[40]: df1.dropna(inplace=True)
```

```
[41]: df1.isnull().sum()/len(df1)*100
```

```

[41]: genres      0.0
      keywords    0.0
      overview    0.0
      title       0.0
      movie_id    0.0
      cast        0.0
      crew        0.0
      dtype: float64

```

```
[42]: df1.head(2)
```

```

[42]:                                genres \
0 [Action, Adventure, Fantasy, Science Fiction]
1 [Adventure, Fantasy, Action]

                                keywords \
0 [culture clash, future, space war, space colon...
1 [ocean, drug abuse, exotic island, east india ...

                                overview \
0 In the 22nd century, a paraplegic Marine is di...
1 Captain Barbossa, long believed to be dead, ha...

```

```

                                title  movie_id  \
0                                Avatar    19995
1  Pirates of the Caribbean: At World's End    285

```

```

                                cast  \
0  [{"cast_id": 242, "character": "Jake Sully", "...
1  [{"cast_id": 4, "character": "Captain Jack Spa...

```

```

                                crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...
1  [{"credit_id": "52fe4232c3a36847f800b579", "de...

```

```

[43]: def convert2(text):
      L = []
      counter = 0
      for i in ast.literal_eval(text):
          if counter < 4 :
              L.append(i['name'])
              counter +=1
      return L

```

```

[44]: df1['cast'] = df1['cast'].apply(convert2)

```

```

[45]: df1.head(2)

```

```

[45]:                                genres  \
0  [Action, Adventure, Fantasy, Science Fiction]
1                                [Adventure, Fantasy, Action]

```

```

                                keywords  \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...

```

```

                                overview  \
0  In the 22nd century, a paraplegic Marine is di...
1  Captain Barbossa, long believed to be dead, ha...

```

```

                                title  movie_id  \
0                                Avatar    19995
1  Pirates of the Caribbean: At World's End    285

```

```

                                cast  \
0  [Sam Worthington, Zoe Saldana, Sigourney Weave...
1  [Johnny Depp, Orlando Bloom, Keira Knightley, ...

```

```

                                crew
0  [{"credit_id": "52fe48009251416c750aca23", "de...

```



```
1 [{"credit_id": "52fe4232c3a36847f800b579", "de...
```

```
[46]: df1['cast'][0]
```

```
[46]: ['Sam Worthington', 'Zoe Saldana', 'Sigourney Weaver', 'Stephen Lang']
```

```
[47]: df1['crew'][0]
```

```
[47]: '[{"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": "52fe48009251416c750aca11", "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": 0, "id": 89714, "job": "Stunts", "name": "Ilram 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"}]

"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 Karena"}, {"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 Fawkner"}, {"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 Caira"}, {"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": "54491f500e0a26747c001c07", "department": "Crew", "gender": 0, "id": 1376909, "job": "Choreographer", "name": "Lula Washington"}]

{"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": "549596009251413af70028cc", "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 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 \\'Max\\' 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": 1452643, "job": "Hairstylist", "name": "Roxane Griffin"}]

```

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":
"5592b0dbc3a368774b00112c", "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"}]]'

```

```

[48]: def convert3(text):
      data = []
      for _ in ast.literal_eval(text):
          if _['job'] == 'Director':
              data.append(_['name'])
      return data

```

```

[49]: df1['crew'] = df1['crew'].apply(convert3)

```

```
[50]: df1['crew'][0]
```

```
[50]: ['James Cameron']
```

```
[51]: df1.head(2)
```

```
[51]:
```

	genres \	keywords \	overview \	title	movie_id \	cast	crew
0	[Action, Adventure, Fantasy, Science Fiction]		In the 22nd century, a paraplegic Marine is di...	Avatar	19995	[Sam Worthington, Zoe Saldana, Sigourney Weave...	[James Cameron]
1	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...]	Captain Barbossa, long believed to be dead, ha...	Pirates of the Caribbean: At World's End	285	[Johnny Depp, Orlando Bloom, Keira Knightley, ...]	[Gore Verbinski]

```
[52]: def removing_space(text):  
      data = []  
      for i in text:  
          data.append(i.replace(" ", ""))  
      return data
```

```
[53]: df1['crew'] = df1['crew'].apply(removing_space)
```

```
[54]: df1['crew'][0]
```

```
[54]: ['JamesCameron']
```

```
[55]: df1['cast'] = df1['cast'].apply(removing_space)
```

```
[56]: df1['cast'][0]
```

```
[56]: ['SamWorthington', 'ZoeSaldana', 'SigourneyWeaver', 'StephenLang']
```

```
[ ]:
```

```
[57]: df1['genres'] = df1['genres'].apply(removing_space)
```



```
[58]: df1.head(3)
```

```
[58]:                                     genres \
0  [Action, Adventure, Fantasy, ScienceFiction]
1                                     [Adventure, Fantasy, Action]
2                                     [Action, Adventure, Crime]

                                     keywords \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...
2  [spy, based on novel, secret agent, sequel, mi...

                                     overview \
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...

                                     title  movie_id \
0                                     Avatar      19995
1  Pirates of the Caribbean: At World's End      285
2                                     Spectre     206647

                                     cast          crew
0  [SamWorthington, ZoeSaldana, SigourneyWeaver, ...  [JamesCameron]
1  [JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste...  [GoreVerbinski]
2  [DanielCraig, ChristophWaltz, LéaSeydoux, Ralp...  [SamMendes]
```

```
[ ]:
```

```
[59]: def split_string_into_words(string):
      return string.split()
```

```
[60]: # Applying the function to the 'overview' column
      df1['overview'] = df1['overview'].apply(split_string_into_words)
```

```
[61]: df1.head(2)
```

```
[61]:                                     genres \
0  [Action, Adventure, Fantasy, ScienceFiction]
1                                     [Adventure, Fantasy, Action]

                                     keywords \
0  [culture clash, future, space war, space colon...
1  [ocean, drug abuse, exotic island, east india ...

                                     overview \
0  [In, the, 22nd, century,, a, paraplegic, Marin...
```

```
1 [Captain, Barbossa,, long, believed, to, be, d...
```

```
                                title  movie_id  \  
0                               Avatar    19995  
1  Pirates of the Caribbean: At World's End    285
```

```
                                cast                crew  
0 [SamWorthington, ZoeSaldana, SigourneyWeaver, ... [JamesCameron]  
1 [JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste... [GoreVerbinski]
```

```
[62]: df1.isnull().sum()
```

```
[62]: genres      0  
      keywords    0  
      overview    0  
      title       0  
      movie_id    0  
      cast        0  
      crew        0  
      dtype: int64
```

```
[ ]:
```

```
[63]: ### Merging The Columns
```

```
[64]: df1['context'] = df1['genres'] + df1['keywords'] + df1['overview'] +  
      ↪df1['cast'] + df1['crew']
```

```
[65]: df1.head(3)
```

```
[65]:                                genres  \  
0 [Action, Adventure, Fantasy, ScienceFiction]  
1                                [Adventure, Fantasy, Action]  
2                                [Action, Adventure, Crime]  
  
                                keywords  \  
0 [culture clash, future, space war, space colon...  
1 [ocean, drug abuse, exotic island, east india ...  
2 [spy, based on novel, secret agent, sequel, mi...  
  
                                overview  \  
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...  
  
                                title  movie_id  \  
0                               Avatar    19995
```

```

1 Pirates of the Caribbean: At World's End      285
2                               Spectre      206647

```

```

                                cast          crew \
0 [SamWorthington, ZoeSaldana, SigourneyWeaver, ... [JamesCameron]
1 [JohnnyDepp, OrlandoBloom, KeiraKnightley, Ste... [GoreVerbinski]
2 [DanielCraig, ChristophWaltz, LéaSeydoux, Ralp... [SamMendes]

```

```

                                context
0 [Action, Adventure, Fantasy, ScienceFiction, c...
1 [Adventure, Fantasy, Action, ocean, drug abuse...
2 [Action, Adventure, Crime, spy, based on novel...

```

```
[66]: del df1['genres']
```

```
[67]: df1.head(1)
```

```

[67]:                                keywords \
0 [culture clash, future, space war, space colon...

                                overview  title  movie_id \
0 [In, the, 22nd, century,, a, paraplegic, Marin... Avatar      19995

                                cast          crew \
0 [SamWorthington, ZoeSaldana, SigourneyWeaver, ... [JamesCameron]

                                context
0 [Action, Adventure, Fantasy, ScienceFiction, c...

```

```
[68]: del df1['overview']
```

```
[69]: del df1['cast']
```

```
[70]: del df1['crew']
```

```
[71]: del df1['keywords']
```

```
[72]: df1.head(3)
```

```

[72]:                                title  movie_id \
0                               Avatar      19995
1 Pirates of the Caribbean: At World's End      285
2                               Spectre      206647

                                context
0 [Action, Adventure, Fantasy, ScienceFiction, c...
1 [Adventure, Fantasy, Action, ocean, drug abuse...

```

2 [Action, Adventure, Crime, spy, based on novel...

[]:

[73]: df1['context'][0]

```
[73]: ['Action',
      'Adventure',
      'Fantasy',
      'ScienceFiction',
      'culture clash',
      'future',
      'space war',
      'space colony',
      'society',
      'space travel',
      'futuristic',
      'romance',
      'space',
      'alien',
      'tribe',
      'alien planet',
      'cgi',
      'marine',
      'soldier',
      'battle',
      'love affair',
      'anti war',
      'power relations',
      'mind and soul',
      '3d',
      'In',
      '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.',
'SamWorthington',
'ZoeSaldana',
'SigourneyWeaver',
'StephenLang',
'JamesCameron']
```

```
[74]: ## Joining Context to make a sentence
```

```
[ ]:
```

```
[75]: def joining(text):
      return " ".join(text)
```

```
[76]: df1['context'] = df1['context'].apply(joining)
```

```
[77]: df1['context'][0]
```

```
[77]: 'Action Adventure Fantasy ScienceFiction culture clash future space war space
colony society space travel futuristic romance space alien tribe alien planet
cgi marine soldier battle love affair anti war power relations mind and soul 3d
In 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. SamWorthington ZoeSaldana SigourneyWeaver StephenLang
JamesCameron'
```

```
[78]: df1.head(2)
```

```
[78]:
```

	title	movie_id	\
0	Avatar	19995	
1	Pirates of the Caribbean: At World's End	285	


```

context
0 Action Adventure Fantasy ScienceFiction cultur...
1 Adventure Fantasy Action ocean drug abuse exot...
```

```
[ ]:
```

```
[79]: # Converting all text value into lower case
```

```
[89]: df1['title']
```

```
[89]: 0          Avatar
      1  Pirates of the Caribbean: At World's End
      2          Spectre
      3    The Dark Knight Rises
      4    John Carter
      ...
      4804      El Mariachi
      4805    Newlyweds
      4806  Signed, Sealed, Delivered
      4807    Shanghai Calling
      4808    My Date with Drew
      Name: title, Length: 4806, dtype: object
```

```
[114]: text = df1['title'][0].lower()
      print(text)
```

avatar

```
[116]: data = []
      for _ in df1['title']:
          data.append(_)
```

```
[120]: counter = 0
      for i in data:
          df1['title'][counter] = i.lower()
          counter = counter + 1
```

```
[131]: df1.shape
```

```
[131]: (4806, 3)
```

```
[135]: df1['title'][4806] = df1['title'][4806].lower()
      df1['title'][4807] = df1['title'][4807].lower()
      df1['title'][4808] = df1['title'][4808].lower()
```

```
[139]: df1.head(2)
```

```
[139]:          title  movie_id \
0          avatar      19995
1  pirates of the caribbean: at world's end      285
```

```

                                context
0  Action Adventure Fantasy ScienceFiction cultur...
1  Adventure Fantasy Action ocean drug abuse exot...

```

```
[ ]:
```

```
[ ]:
```

3 Bag of Words (BOW) - Feature Extraction

```
[ ]:
```

```
[140]: from sklearn.feature_extraction.text import CountVectorizer
```

```
[141]: cv = CountVectorizer(stop_words='english')
```

```
[142]: vector = cv.fit_transform(df1['context']).toarray()
```

```
[143]: vector.shape
```

```
[143]: (4806, 32916)
```

```
[ ]:
```

```
[ ]:
```

4 Building Recommendation System basis similarity matrix

```
[ ]:
```

```
[144]: from sklearn.metrics.pairwise import cosine_similarity
```

```
[145]: similarity = cosine_similarity(vector)
```

```
[146]: similarity.shape
```

```
[146]: (4806, 4806)
```

```
[147]: similarity
```

```
[147]: array([[1.          , 0.06692745, 0.03151267, ..., 0.02823912, 0.01590712,
          0.          ],
          [0.06692745, 1.          , 0.04007219, ..., 0.03590948, 0.          ,
          0.          ],
          [0.03151267, 0.04007219, 1.          , ..., 0.01690792, 0.          ,
          ...]])
```

```

0.      ],
...,
[0.02823912, 0.03590948, 0.01690792, ..., 1.      , 0.05120916,
 0.03178209],
[0.01590712, 0.      , 0.      , ..., 0.05120916, 1.      ,
 0.03580574],
[0.      , 0.      , 0.      , ..., 0.03178209, 0.03580574,
 1.      ]])

```

```
[ ]:
```

```
[148]: df1.head(2)
```

```

[148]:
           title  movie_id \
0          avatar    19995
1  pirates of the caribbean: at world's end    285

           context
0  Action Adventure Fantasy ScienceFiction cultur...
1  Adventure Fantasy Action ocean drug abuse exot...

```

```
[ ]:
```

```
[151]: df1[df1['title'] == 'avatar']
```

```

[151]:
           title  movie_id
0  avatar    19995  Action Adventure Fantasy ScienceFiction cultur...

```

```
[152]: df1[df1['title'] == 'avatar'].index[0]
```

```
[152]: 0
```

```

[173]: def key_sort(item):
        return item[1]

        def recommendation(movie_name):
            movie_name = movie_name.lower()
            index = df1[df1['title']==movie_name].index[0]

            dist = []
            counter = 0
            for i in range(len(similarity)):
                dist.append((i,similarity[index][i]))  ## this will give tuple
                distance = sorted(dist , reverse=True , key=key_sort)

            while counter!=5:
                if counter == 0:

```



```

        counter = counter + 1
        continue
    print(df1.title[distance[counter][0]])
    counter = counter + 1

```

```
[174]: movie_name = 'Avatar'
```

```
[175]: recommendation(movie_name.lower())
```

```

aliens
mission to mars
moonraker
alien3

```

5 Final Prediction

```
[176]: recommendation('Captain America: The First Avenger')
```

```

captain america: the winter soldier
captain america: civil war
saints and soldiers
x-men: first class

```

```
[177]: recommendation('Spider-Man 3')
```

```

spider-man 2
spider-man
the amazing spider-man 2
the amazing spider-man

```

6 Saving the system

```
[ ]:
```

```
[178]: from joblib import dump
```

```
[179]: dump(similarity, 'recommendation.joblib')  ## This will save my cosine_
      ↪ similarity array
```

```
[179]: ['recommendation.joblib']
```

```
[180]: dump(df1, 'movie_dataset.joblib')
```

```
[180]: ['movie_dataset.joblib']
```

```
[ ]:
```

[]:

[]:

[]: