

# movie-recommender-systems

September 9, 2024

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import ast
from ast import literal_eval

import warnings
warnings.filterwarnings("ignore", message="numpy.dtype size changed")
warnings.filterwarnings("ignore", message="numpy.ufunc size changed")
warnings.filterwarnings('ignore')

import wordcloud
import nltk
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from nltk.tokenize import word_tokenize
from wordcloud import WordCloud, STOPWORDS

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import cosine_similarity

import pickle
import os

[2]: creditsdf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\credits.csv')
keywordsdf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\keywords.csv')
linksdf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\links.csv')
links_smallidf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\links_small.csv')
movies_metadatadf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\movies_metadata.csv', low_memory=False )
```

```
ratingsdf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\ratings.csv')
ratings_smallidf = pd.read_csv('C:\\Users\\Srijan\\Desktop\\Internship\\Movie_
↳Recommender Systems\\DataSet\\ratings_small.csv')
```

```
[3]: movies_metadatadf.head(2)
```

```
[3]:      adult      belongs_to_collection      budget \
0  False  {'id': 10194, 'name': 'Toy Story Collection', ...  30000000
1  False                                     NaN  65000000

      genres \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, 'name': 'Comedy'}]
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, 'name': 'Fantasy'}]

      homepage      id      imdb_id original_language \
0  http://toystory.disney.com/toy-story      862      tt0114709      en
1                                     NaN      8844      tt0113497      en

      original_title      overview ... \
0      Toy Story  Led by Woody, Andy's toys live happily in his world.
1      Jumanji  When siblings Judy and Peter discover an enchanted book,

      release_date      revenue runtime \
0  1995-10-30  373554033.0      81.0
1  1995-12-15  262797249.0      104.0

      spoken_languages      status \
0  [{'iso_639_1': 'en', 'name': 'English'}]  Released
1  [{'iso_639_1': 'en', 'name': 'English'}, {'iso_639_1': 'es', 'name': 'Spanish'}]  Released

      tagline      title  video vote_average \
0      NaN  Toy Story  False      7.7
1  Roll the dice and unleash the excitement!  Jumanji  False      6.9

      vote_count
0      5415.0
1      2413.0

[2 rows x 24 columns]
```

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

```
[4]:      id      keywords
0      862  [{'id': 931, 'name': 'jealousy'}, {'id': 4290, 'name': 'love'}]
1      8844  [{'id': 10090, 'name': 'board game'}, {'id': 10091, 'name': 'strategy'}]
```

```
[5]: creditsdf.head(5)
```

```
[5]:                                     cast \
0  [{'cast_id': 14, 'character': 'Woody (voice)',...
1  [{'cast_id': 1, 'character': 'Alan Parrish', '...'
2  [{'cast_id': 2, 'character': 'Max Goldman', 'c...'
3  [{'cast_id': 1, 'character': 'Savannah 'Vannah...'
4  [{'cast_id': 1, 'character': 'George Banks', '...'

                                     crew      id
0  [{'credit_id': '52fe4284c3a36847f8024f49', 'de...    862
1  [{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...   8844
2  [{'credit_id': '52fe466a9251416c75077a89', 'de...  15602
3  [{'credit_id': '52fe44779251416c91011acb', 'de...  31357
4  [{'credit_id': '52fe44959251416c75039ed7', 'de...  11862
```

```
[6]: creditsdf.isnull().sum()
```

```
[6]: cast      0
crew      0
id      0
dtype: int64
```

```
[7]: keywordsdf.isnull().sum()
```

```
[7]: id      0
keywords  0
dtype: int64
```

```
[8]: movies_metadatadf.isnull().sum()
```

```
[8]: adult      0
belongs_to_collection  40972
budget      0
genres      0
homepage    37684
id      0
imdb_id    17
original_language    11
original_title      0
overview    954
popularity      5
poster_path    386
production_companies      3
production_countries      3
release_date    87
revenue      6
```

```
runtime                263
spoken_languages       6
status                 87
tagline               25054
title                  6
video                  6
vote_average           6
vote_count             6
dtype: int64
```

```
[9]: movies_metadatadf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45466 entries, 0 to 45465
Data columns (total 24 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   adult                                45466 non-null  object
1   belongs_to_collection                4494 non-null   object
2   budget                              45466 non-null  object
3   genres                              45466 non-null  object
4   homepage                            7782 non-null   object
5   id                                   45466 non-null  object
6   imdb_id                             45449 non-null  object
7   original_language                   45455 non-null  object
8   original_title                       45466 non-null  object
9   overview                            44512 non-null  object
10  popularity                           45461 non-null  object
11  poster_path                         45080 non-null  object
12  production_companies                 45463 non-null  object
13  production_countries                 45463 non-null  object
14  release_date                        45379 non-null  object
15  revenue                             45460 non-null  float64
16  runtime                             45203 non-null  float64
17  spoken_languages                     45460 non-null  object
18  status                              45379 non-null  object
19  tagline                             20412 non-null  object
20  title                               45460 non-null  object
21  video                               45460 non-null  object
22  vote_average                         45460 non-null  float64
23  vote_count                          45460 non-null  float64
dtypes: float64(4), object(20)
memory usage: 8.3+ MB
```

```
[10]: movies_metadatadf.head(2)
```

```

[10]:      adult                                belongs_to_collection    budget \
0  False  {'id': 10194, 'name': 'Toy Story Collection', ... 30000000
1  False                                     NaN  65000000

                                genres \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, '...'
1  [{'id': 12, 'name': 'Adventure'}, {'id': 14, '...'

                                homepage    id    imdb_id original_language \
0  http://toystory.disney.com/toy-story    862    tt0114709                en
1                                     NaN    8844    tt0113497                en

    original_title                                overview ... \
0      Toy Story  Led by Woody, Andy's toys live happily in his ... ...
1      Jumanji   When siblings Judy and Peter discover an encha... ...

    release_date    revenue runtime \
0  1995-10-30    373554033.0    81.0
1  1995-12-15    262797249.0    104.0

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

                                tagline    title    video vote_average \
0                                     NaN    Toy Story    False        7.7
1  Roll the dice and unleash the excitement!    Jumanji    False        6.9

    vote_count
0      5415.0
1      2413.0

[2 rows x 24 columns]

```

```

[11]: movies_metadatadf['id'] = movies_metadatadf['id'].astype(str)
      creditsdf['id'] = creditsdf['id'].astype(str)
      movie_creditsdf = pd.merge(movies_metadatadf, creditsdf, left_on='id',
      ↪right_on='id')

```

```

[12]: movie_creditsdf.head(2)

```

```

[12]:      adult                                belongs_to_collection    budget \
0  False  {'id': 10194, 'name': 'Toy Story Collection', ... 30000000
1  False                                     NaN  65000000

                                genres \
0  [{'id': 16, 'name': 'Animation'}, {'id': 35, '...'

```



```

2  [{'id': 10749, 'name': 'Romance'}, {'id': 35, ...
3  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...
4      [{'id': 35, 'name': 'Comedy'}]

```

```

                                homepage      id      imdb_id original_language \
0  http://toystory.disney.com/toy-story      862    tt0114709                en
1                                NaN      8844    tt0113497                en
2                                NaN     15602    tt0113228                en
3                                NaN     31357    tt0114885                en
4                                NaN     11862    tt0113041                en

```

```

                                original_title \
0                                Toy Story
1                                Jumanji
2                                Grumpier Old Men
3                                Waiting to Exhale
4  Father of the Bride Part II

```

```

                                overview ... \
0  Led by Woody, Andy's toys live happily in his ...
1  When siblings Judy and Peter discover an encha...
2  A family wedding reignites the ancient feud be...
3  Cheated on, mistreated and stepped on, the wom...
4  Just when George Banks has recovered from his ...

```

```

                                spoken_languages      status \
0      [{'iso_639_1': 'en', 'name': 'English'}] Released
1  [{'iso_639_1': 'en', 'name': 'English'}, {'iso... Released
2      [{'iso_639_1': 'en', 'name': 'English'}] Released
3      [{'iso_639_1': 'en', 'name': 'English'}] Released
4      [{'iso_639_1': 'en', 'name': 'English'}] Released

```

```

                                tagline \
0                                NaN
1      Roll the dice and unleash the excitement!
2  Still Yelling. Still Fighting. Still Ready for...
3  Friends are the people who let you be yourself...
4  Just When His World Is Back To Normal... He's ...

```

```

                                title  video  vote_average  vote_count \
0      Toy Story  False           7.7         5415.0
1      Jumanji   False           6.9         2413.0
2  Grumpier Old Men  False           6.5           92.0
3  Waiting to Exhale  False           6.1           34.0
4  Father of the Bride Part II  False           5.7          173.0

```

```

                                cast \

```

```

0  [{'cast_id': 14, 'character': 'Woody (voice)',...
1  [{'cast_id': 1, 'character': 'Alan Parrish', '...'
2  [{'cast_id': 2, 'character': 'Max Goldman', 'c...'
3  [{'cast_id': 1, 'character': "Savannah 'Vannah..."
4  [{'cast_id': 1, 'character': 'George Banks', '...'

                                crew \
0  [{'credit_id': '52fe4284c3a36847f8024f49', 'de...'
1  [{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...'
2  [{'credit_id': '52fe466a9251416c75077a89', 'de...'
3  [{'credit_id': '52fe44779251416c91011acb', 'de...'
4  [{'credit_id': '52fe44959251416c75039ed7', 'de...'

                                keywords
0  [{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1  [{'id': 10090, 'name': 'board game'}, {'id': 1...
2  [{'id': 1495, 'name': 'fishing'}, {'id': 12392...
3  [{'id': 818, 'name': 'based on novel'}, {'id': ...
4  [{'id': 1009, 'name': 'baby'}, {'id': 1599, 'n...

[5 rows x 27 columns]

```

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

```

[16]: adult                0
      belongs_to_collection  42054
      budget                0
      genres                0
      homepage              38619
      id                   0
      imdb_id              17
      original_language     11
      original_title        0
      overview              995
      popularity            4
      poster_path           399
      production_companies   4
      production_countries   4
      release_date           88
      revenue                4
      runtime               268
      spoken_languages       4
      status                 86
      tagline               25845
      title                  4
      video                  4
      vote_average           4

```



```

vote_count          4
cast                0
crew               0
keywords            0
dtype: int64

```

```
[17]: df.shape
```

```
[17]: (46628, 27)
```

```
[18]: df = df[["id", "genres", "cast", "crew", "keywords", "original_title", "title", "overview"]]
```

```
[19]: df.head()
```

```
[19]:
      id                                     genres \
0    862  [{'id': 16, 'name': 'Animation'}, {'id': 35, '...
1   8844  [{'id': 12, 'name': 'Adventure'}, {'id': 14, '...
2  15602  [{'id': 10749, 'name': 'Romance'}, {'id': 35, ...
3  31357  [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'nam...
4  11862                                     [{'id': 35, 'name': 'Comedy'}]

                                     cast \
0  [{'cast_id': 14, 'character': 'Woody (voice)', ...
1  [{'cast_id': 1, 'character': 'Alan Parrish', '...
2  [{'cast_id': 2, 'character': 'Max Goldman', 'c...
3  [{'cast_id': 1, 'character': "Savannah 'Vannah...
4  [{'cast_id': 1, 'character': 'George Banks', '...

                                     crew \
0  [{'credit_id': '52fe4284c3a36847f8024f49', 'de...
1  [{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...
2  [{'credit_id': '52fe466a9251416c75077a89', 'de...
3  [{'credit_id': '52fe44779251416c91011acb', 'de...
4  [{'credit_id': '52fe44959251416c75039ed7', 'de...

                                     keywords \
0  [{'id': 931, 'name': 'jealousy'}, {'id': 4290,...
1  [{'id': 10090, 'name': 'board game'}, {'id': 1...
2  [{'id': 1495, 'name': 'fishing'}, {'id': 12392...
3  [{'id': 818, 'name': 'based on novel'}, {'id': ...
4  [{'id': 1009, 'name': 'baby'}, {'id': 1599, 'n...

      original_title      title \
0      Toy Story      Toy Story
1      Jumanji      Jumanji
2  Grumpier Old Men  Grumpier Old Men

```

```

3           Waiting to Exhale           Waiting to Exhale
4  Father of the Bride Part II  Father of the Bride Part II

```

```

                                overview
0  Led by Woody, Andy's toys live happily in his ...
1  When siblings Judy and Peter discover an encha...
2  A family wedding reignites the ancient feud be...
3  Cheated on, mistreated and stepped on, the wom...
4  Just when George Banks has recovered from his ...

```

```
[20]: df.shape
```

```
[20]: (46628, 8)
```

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

```

[21]: id                0
      genres            0
      cast             0
      crew             0
      keywords         0
      original_title    0
      title            4
      overview        995
      dtype: int64

```

```
[22]: df.dropna(inplace=True)
```

```
[23]: df.shape
```

```
[23]: (45629, 8)
```

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

```

[24]: id                0
      genres            0
      cast             0
      crew             0
      keywords         0
      original_title    0
      title            0
      overview         0
      dtype: int64

```

```

[25]: df.duplicated().sum()
      print(f"Number of duplicates: {df.duplicated().sum()}")

```

Number of duplicates: 1147

```
[26]: df.drop_duplicates(inplace=True)
      duplicates_after_dropping = df.duplicated().sum()
      print(f"Number of duplicates after dropping: {duplicates_after_dropping}")
```

Number of duplicates after dropping: 0

```
[27]: print(df.iloc[0]['genres'])
```

```
[{'id': 16, 'name': 'Animation'}, {'id': 35, 'name': 'Comedy'}, {'id': 10751,
'name': 'Family'}]
```

```
[28]: # genres is a string. So my work i have to convert this in a list
      def extract(genre_str):
          genre_list = ast.literal_eval(genre_str)
          genre_names = [genre['name'] for genre in genre_list]
          return genre_names
```

```
[29]: df['genres'] = df['genres'].apply(extract)
```

```
[30]: df['keywords'] = df['keywords'].apply(extract)
```

```
[31]: df.head()
```

```
[31]:      id      genres \
0    862  [Animation, Comedy, Family]
1   8844  [Adventure, Fantasy, Family]
2  15602      [Romance, Comedy]
3  31357  [Comedy, Drama, Romance]
4  11862      [Comedy]

      cast \
0  [{'cast_id': 14, 'character': 'Woody (voice)',...
1  [{'cast_id': 1, 'character': 'Alan Parrish', '...
2  [{'cast_id': 2, 'character': 'Max Goldman', 'c...
3  [{'cast_id': 1, 'character': "Savannah 'Vannah...
4  [{'cast_id': 1, 'character': 'George Banks', '...

      crew \
0  [{'credit_id': '52fe4284c3a36847f8024f49', 'de...
1  [{'credit_id': '52fe44bfc3a36847f80a7cd1', 'de...
2  [{'credit_id': '52fe466a9251416c75077a89', 'de...
3  [{'credit_id': '52fe44779251416c91011acb', 'de...
4  [{'credit_id': '52fe44959251416c75039ed7', 'de...

      keywords \
0  [jealousy, toy, boy, friendship, friends, riva...
```

```

1 [board game, disappearance, based on children'...
2 [fishing, best friend, duringcreditsstinger, o...
3 [based on novel, interracial relationship, sin...
4 [baby, midlife crisis, confidence, aging, daug...

```

|   | original_title              | title \                     |
|---|-----------------------------|-----------------------------|
| 0 | Toy Story                   | Toy Story                   |
| 1 | Jumanji                     | Jumanji                     |
| 2 | Grumpier Old Men            | Grumpier Old Men            |
| 3 | Waiting to Exhale           | Waiting to Exhale           |
| 4 | Father of the Bride Part II | Father of the Bride Part II |

```

                                overview
0  Led by Woody, Andy's toys live happily in his ...
1  When siblings Judy and Peter discover an encha...
2  A family wedding reignites the ancient feud be...
3  Cheated on, mistreated and stepped on, the wom...
4  Just when George Banks has recovered from his ...

```

```
[32]: df.shape
```

```
[32]: (44482, 8)
```

```
[33]: # in cast we want to first 3 character. so we make a function
```

```

def cast_conver(obj):
    l = []
    c = 0
    for i in ast.literal_eval(obj):
        if c != 3:
            l.append(i['name'])
            c+=1
        else:
            break
    return l

```

```
[34]: df['cast'] =df['cast'].apply(cast_conver)
```

```
[35]: df['crew'][0]
```

```

[35]: '[{\''credit_id\': \''52fe4284c3a36847f8024f49\'', \''department\': \''Directing\'',
\'gender\': 2, \''id\': 7879, \''job\': \''Director\'', \''name\': \''John Lasseter\'',
\'profile_path\': \''/7EdqiNbr4FRjIhKHYPdFfEEFEG.jpg\'}, {\''credit_id\':
\'52fe4284c3a36847f8024f4f\'', \''department\': \''Writing\'', \''gender\': 2,
\'id\': 12891, \''job\': \''Screenplay\'', \''name\': \''Joss Whedon\'',
\'profile_path\': \''/dTIVsuaTVTeGmvkhcyJvKp2A5kr.jpg\'}, {\''credit_id\':
\'52fe4284c3a36847f8024f55\'', \''department\': \''Writing\'', \''gender\': 2,

```

\id\': 7, \job\': \Screenplay\, \name\': \Andrew Stanton\,  
 \profile\_path\': \'/pvQWsu0qc8JFQhMVJkTHuexUAa1.jpg\'}, {\credit\_id\':  
 \52fe4284c3a36847f8024f5b\, \department\': \Writing\, \gender\': 2,  
 \id\': 12892, \job\': \Screenplay\, \name\': \Joel Cohen\,  
 \profile\_path\': \'/dAubAiZcvKFbboWlj7oXOkZnTSu.jpg\'}, {\credit\_id\':  
 \52fe4284c3a36847f8024f61\, \department\': \Writing\, \gender\': 0,  
 \id\': 12893, \job\': \Screenplay\, \name\': \Alec Sokolow\,  
 \profile\_path\': \'/v79vLRyI94BZUQnkkyznbgUZLjT.jpg\'}, {\credit\_id\':  
 \52fe4284c3a36847f8024f67\, \department\': \Production\, \gender\': 1,  
 \id\': 12894, \job\': \Producer\, \name\': \Bonnie Arnold\,  
 \profile\_path\': None}, {\credit\_id\': \52fe4284c3a36847f8024f6d\,  
 \department\': \Production\, \gender\': 0, \id\': 12895, \job\':  
 \Executive Producer\, \name\': \Ed Catmull\, \profile\_path\': None},  
 {\credit\_id\': \52fe4284c3a36847f8024f73\, \department\': \Production\,  
 \gender\': 2, \id\': 12896, \job\': \Producer\, \name\': \Ralph  
 Guggenheim\, \profile\_path\': None}, {\credit\_id\':  
 \52fe4284c3a36847f8024f79\, \department\': \Production\, \gender\': 2,  
 \id\': 12897, \job\': \Executive Producer\, \name\': \Steve Jobs\,  
 \profile\_path\': \'/mOMP3SwD5qWQSR0ldCIByd3guTV.jpg\'}, {\credit\_id\':  
 \52fe4284c3a36847f8024f8b\, \department\': \Editing\, \gender\': 2,  
 \id\': 8, \job\': \Editor\, \name\': \Lee Unkrich\, \profile\_path\':  
 \'/bdTCCXjgOV3YyaNmLGYG0xFQM0c.jpg\'}, {\credit\_id\':  
 \52fe4284c3a36847f8024f91\, \department\': \Art\, \gender\': 2, \id\':  
 7883, \job\': \Art Direction\, \name\': \Ralph Eggleston\,  
 \profile\_path\': \'/uUfcGKDsK01aROMpXR67Hn6RvR.jpg\'}, {\credit\_id\':  
 \598331bf925141421201044b\, \department\': \Editing\, \gender\': 2,  
 \id\': 1168870, \job\': \Editor\, \name\': \Robert Gordon\,  
 \profile\_path\': None}, {\credit\_id\': \5892168cc3a36809660095f9\,  
 \department\': \Sound\, \gender\': 0, \id\': 1552883, \job\': \Foley  
 Editor\, \name\': \Mary Helen Leasman\, \profile\_path\': None},  
 {\credit\_id\': \5531824d9251415289000945\, \department\': \Visual  
 Effects\, \gender\': 0, \id\': 1453514, \job\': \Animation\, \name\':  
 \Kim Blanchette\, \profile\_path\': None}, {\credit\_id\':  
 \589215969251412dcb009bf6\, \department\': \Sound\, \gender\': 0, \id\':  
 1414182, \job\': \ADR Editor\, \name\': \Marilyn McCoppen\,  
 \profile\_path\': None}, {\credit\_id\': \589217099251412dc500a018\,  
 \department\': \Sound\, \gender\': 2, \id\': 7885, \job\':  
 \Orchestrator\, \name\': \Randy Newman\, \profile\_path\':  
 \'/w0JzfoiM25nrnxY0zosPHRq6mlE.jpg\'}, {\credit\_id\':  
 \5693e6b29251417b0e0000e3\, \department\': \Editing\, \gender\': 0,  
 \id\': 1429549, \job\': \Color Timer\, \name\': \Dale E. Grahn\,  
 \profile\_path\': None}, {\credit\_id\': \572e2522c3a36869e6001a9c\,  
 \department\': \Visual Effects\, \gender\': 0, \id\': 7949, \job\': \CG  
 Painter\, \name\': \Robin Cooper\, \profile\_path\': None}, {\credit\_id\':  
 \574f12309251415ca1000012\, \department\': \Writing\, \gender\': 2,  
 \id\': 7879, \job\': \Original Story\, \name\': \John Lasseter\,  
 \profile\_path\': \'/7EdqiNbr4FRjIhKHYPdFfEEFEG.jpg\'}, {\credit\_id\':

\574f1240c3a3682e7300001c\, \department\': \Writing\, \gender\': 2,  
 \id\': 12890, \job\': \Original Story\, \name\': \Pete Docter\,  
 \profile\_path\': \r6ngPgnReA3RHmKjmSoVsc6Awjp.jpg\}, {\credit\_id\':  
 \574f12519251415c92000015\, \department\': \Writing\, \gender\': 0,  
 \id\': 7911, \job\': \Original Story\, \name\': \Joe Ranft\,  
 \profile\_path\': \f1BoWC2JbCcfP1e5hKfGsxkHhVU.jpg\}, {\credit\_id\':  
 \574f12cec3a3682e82000022\, \department\': \Crew\, \gender\': 0, \id\':  
 1629419, \job\': \Post Production Supervisor\, \name\': \Patsy Bouge\,  
 \profile\_path\': None}, {\credit\_id\': \574f14f19251415ca1000082\,  
 \department\': \Art\, \gender\': 0, \id\': 7961, \job\': \Sculptor\,  
 \name\': \Norm DeCarlo\, \profile\_path\': None}, {\credit\_id\':  
 \5751ae4bc3a3683772002b7f\, \department\': \Visual Effects\, \gender\': 2,  
 \id\': 12905, \job\': \Animation Director\, \name\': \Ash Brannon\,  
 \profile\_path\': \6ueWgPEEBHvS3De2BHYQnYjRTig.jpg\}, {\credit\_id\':  
 \5891edbe9251412dc5007cd6\, \department\': \Sound\, \gender\': 2, \id\':  
 7885, \job\': \Music\, \name\': \Randy Newman\, \profile\_path\':  
 \w0JzfoiM25nrnxY0zosPHRq6mlE.jpg\}, {\credit\_id\':  
 \589213d39251412dc8009832\, \department\': \Directing\, \gender\': 0,  
 \id\': 1748707, \job\': \Layout\, \name\': \Roman Figun\,  
 \profile\_path\': None}, {\credit\_id\': \5892173dc3a3680968009351\,  
 \department\': \Sound\, \gender\': 2, \id\': 4949, \job\':  
 \Orchestrator\, \name\': \Don Davis\, \profile\_path\': None},  
 {\credit\_id\': \589217cec3a3686b0a0052ba\, \department\': \Sound\,  
 \gender\': 0, \id\': 1372885, \job\': \Music Editor\, \name\': \James  
 Flamberg\, \profile\_path\': None}, {\credit\_id\':  
 \58921831c3a3686348004a64\, \department\': \Editing\, \gender\': 0,  
 \id\': 1739962, \job\': \Negative Cutter\, \name\': \Mary Beth Smith\,  
 \profile\_path\': None}, {\credit\_id\': \58921838c3a36809700096c0\,  
 \department\': \Editing\, \gender\': 0, \id\': 1748513, \job\': \Negative  
 Cutter\, \name\': \Rick Mackay\, \profile\_path\': None}, {\credit\_id\':  
 \589218429251412dd1009d1b\, \department\': \Art\, \gender\': 0, \id\':  
 1458006, \job\': \Title Designer\, \name\': \Susan Bradley\,  
 \profile\_path\': None}, {\credit\_id\': \5891ed99c3a3680966007670\,  
 \department\': \Crew\, \gender\': 0, \id\': 1748557, \job\': \Supervising  
 Technical Director\, \name\': \William Reeves\, \profile\_path\': None},  
 {\credit\_id\': \5891edcec3a3686b0a002eb2\, \department\': \Sound\,  
 \gender\': 2, \id\': 7885, \job\': \Songs\, \name\': \Randy Newman\,  
 \profile\_path\': \w0JzfoiM25nrnxY0zosPHRq6mlE.jpg\}, {\credit\_id\':  
 \5891edf9c3a36809700075e6\, \department\': \Writing\, \gender\': 2,  
 \id\': 7, \job\': \Original Story\, \name\': \Andrew Stanton\,  
 \profile\_path\': \pvQWsu0qc8JFQhMVJkTHuexUAa1.jpg\}, {\credit\_id\':  
 \58920f0b9251412dd7009104\, \department\': \Crew\, \gender\': 2, \id\':  
 12890, \job\': \Supervising Animator\, \name\': \Pete Docter\,  
 \profile\_path\': \r6ngPgnReA3RHmKjmSoVsc6Awjp.jpg\}, {\credit\_id\':  
 \58920f1fc3a3680977009021\, \department\': \Sound\, \gender\': 2, \id\':  
 2216, \job\': \Sound Designer\, \name\': \Gary Rydstrom\,  
 \profile\_path\': \jZpr1nVf071ldWIOYtmP1FGw7Rj.jpg\}, {\credit\_id\':

```

\'58920f389251412dd700912d\', \'department\': \'Production\', \'gender\': 0,
\'id\': 12909, \'job\': \'Production Supervisor\', \'name\': \'Karen Robert
Jackson\', \'profile_path\': None}, {\'credit_id\':
\'58920fbd9251412dcb00969c\', \'department\': \'Crew\', \'gender\': 0, \'id\':
953331, \'job\': \'Executive Music Producer\', \'name\': \'Chris Montan\',
\'profile_path\': None}, {\'credit_id\': \'589210069251412dd7009219\',
\'department\': \'Visual Effects\', \'gender\': 0, \'id\': 7893, \'job\':
\'Animation Director\', \'name\': \'Rich Quade\', \'profile_path\': None},
{\'credit_id\': \'589210329251412dcd00943b\', \'department\': \'Visual
Effects\', \'gender\': 0, \'id\': 8025, \'job\': \'Animation\', \'name\':
\'Michael Berenstein\', \'profile_path\': None}, {\'credit_id\':
\'5892103bc3a368096a009180\', \'department\': \'Visual Effects\', \'gender\': 0,
\'id\': 78009, \'job\': \'Animation\', \'name\': \'Colin Brady\',
\'profile_path\': None}, {\'credit_id\': \'5892105dc3a3680968008db2\',
\'department\': \'Visual Effects\', \'gender\': 0, \'id\': 1748682, \'job\':
\'Animation\', \'name\': \'Davey Crockett Feiten\', \'profile_path\': None},
{\'credit_id\': \'589210669251412dcd009466\', \'department\': \'Visual
Effects\', \'gender\': 0, \'id\': 1454030, \'job\': \'Animation\', \'name\':
\'Angie Glocka\', \'profile_path\': None}, {\'credit_id\':
\'5892107c9251412dd1009613\', \'department\': \'Visual Effects\', \'gender\': 0,
\'id\': 1748683, \'job\': \'Animation\', \'name\': \'Rex Grignon\',
\'profile_path\': None}, {\'credit_id\': \'5892108ac3a3680973008d3f\',
\'department\': \'Visual Effects\', \'gender\': 0, \'id\': 1748684, \'job\':
\'Animation\', \'name\': \'Tom K. Gurney\', \'profile_path\': None},
{\'credit_id\': \'58921093c3a3686348004477\', \'department\': \'Visual
Effects\', \'gender\': 2, \'id\': 8029, \'job\': \'Animation\', \'name\':
\'Jimmy Hayward\', \'profile_path\': \'/lTDRpudEY7BDwTefXbXzMlmbOui.jpg\'},
{\'credit_id\': \'5892109b9251412dcd0094b0\', \'department\': \'Visual
Effects\', \'gender\': 0, \'id\': 1426773, \'job\': \'Animation\', \'name\':
\'Hal T. Hickel\', \'profile_path\': None}, {\'credit_id\':
\'589210a29251412dc5009a29\', \'department\': \'Visual Effects\', \'gender\': 0,
\'id\': 8035, \'job\': \'Animation\', \'name\': \'Karen Kiser\',
\'profile_path\': None}, {\'credit_id\': \'589210ccc3a3680977009191\',
\'department\': \'Visual Effects\', \'gender\': 0, \'id\': 1748688, \'job\':
\'Animation\', \'name\': \'Anthony B. LaMolinara\', \'profile_path\': None},
{\'credit_id\': \'589210d7c3a3686b0a004c1f\', \'department\': \'Visual
Effects\', \'gender\': 0, \'id\': 587314, \'job\': \'Animation\', \'name\':
\'Guionne Leroy\', \'profile_path\': None}, {\'credit_id\':
\'589210e1c3a36809770091a7\', \'department\': \'Visual Effects\', \'gender\': 2,
\'id\': 7918, \'job\': \'Animation\', \'name\': \'Bud Luckey\',
\'profile_path\': \'/pcCh7G19FKMNijmPQg1PMH1btic.jpg\'}, {\'credit_id\':
\'589210ee9251412dc200978a\', \'department\': \'Visual Effects\', \'gender\': 0,
\'id\': 1748689, \'job\': \'Animation\', \'name\': \'Les Major\',
\'profile_path\': None}, {\'credit_id\': \'589210fa9251412dc8009595\',
\'department\': \'Visual Effects\', \'gender\': 2, \'id\': 7892, \'job\':
\'Animation\', \'name\': \'Glenn McQueen\', \'profile_path\': None},
{\'credit_id\': \'589211029251412dc8009598\', \'department\': \'Visual

```

Effects\','gender': 0, 'id': 555795, 'job': 'Animation\','name': 'Mark Oftedal\','profile\_path': None}, {'credit\_id': '5892110b9251412dc800959d\','department': 'Visual Effects\','gender': 2, 'id': 7882, 'job': 'Animation\','name': 'Jeff Pidgeon\','profile\_path': '/yLddkg5HcgbJg00cS13GVbNPOHY.jpg\'}, {'credit\_id': '58921113c3a36863480044e4\','department': 'Visual Effects\','gender': 0, 'id': 8017, 'job': 'Animation\','name': 'Jeff Pratt\','profile\_path': None}, {'credit\_id': '5892111c9251412dc0097e9\','department': 'Visual Effects\','gender': 0, 'id': 1184140, 'job': 'Animation\','name': 'Steve Rabatch\','profile\_path': None}, {'credit\_id': '58921123c3a36809700090f6\','department': 'Visual Effects\','gender': 0, 'id': 8049, 'job': 'Animation\','name': 'Roger Rose\','profile\_path': None}, {'credit\_id': '5892112b9251412dc0097fb\','department': 'Visual Effects\','gender': 0, 'id': 1509559, 'job': 'Animation\','name': 'Steve Segal\','profile\_path': None}, {'credit\_id': '589211349251412dc80095c3\','department': 'Visual Effects\','gender': 0, 'id': 1748691, 'job': 'Animation\','name': 'Doug Sheppeck\','profile\_path': None}, {'credit\_id': '5892113cc3a3680970009106\','department': 'Visual Effects\','gender': 0, 'id': 8050, 'job': 'Animation\','name': 'Alan Sperling\','profile\_path': None}, {'credit\_id': '58921148c3a3686b0a004c99\','department': 'Visual Effects\','gender': 0, 'id': 8010, 'job': 'Animation\','name': 'Doug Sweetland\','profile\_path': None}, {'credit\_id': '58921150c3a3680966009125\','department': 'Visual Effects\','gender': 0, 'id': 8044, 'job': 'Animation\','name': 'David Tart\','profile\_path': None}, {'credit\_id': '589211629251412dc5009b00\','department': 'Visual Effects\','gender': 0, 'id': 1454034, 'job': 'Animation\','name': 'Ken Willard\','profile\_path': None}, {'credit\_id': '589211c1c3a3686b0a004d28\','department': 'Visual Effects\','gender': 0, 'id': 7887, 'job': 'Visual Effects Supervisor\','name': 'Thomas Porter\','profile\_path': None}, {'credit\_id': '589211d4c3a3680968008ed9\','department': 'Visual Effects\','gender': 0, 'id': 1406878, 'job': 'Visual Effects\','name': 'Mark Thomas Henne\','profile\_path': None}, {'credit\_id': '589211f59251412dd4008e65\','department': 'Visual Effects\','gender': 0, 'id': 1748698, 'job': 'Visual Effects\','name': 'Oren Jacob\','profile\_path': None}, {'credit\_id': '58921242c3a368096a00939b\','department': 'Visual Effects\','gender': 0, 'id': 1748699, 'job': 'Visual Effects\','name': 'Darwyn Peachey\','profile\_path': None}, {'credit\_id': '5892124b9251412dc5009bd2\','department': 'Visual Effects\','gender': 0, 'id': 1748701, 'job': 'Visual Effects\','name': 'Mitch Prater\','profile\_path': None}, {'credit\_id': '58921264c3a3686b0a004dbf\','department': 'Visual Effects\','gender': 0, 'id': 1748703, 'job': 'Visual Effects\','name': 'Brian M. Rosen\','profile\_path': None}, {'credit\_id': '589212709251412dcd009676\','department': 'Lighting\','gender': 1, 'id': 12912, 'job': 'Lighting Supervisor\','name': 'Sharon Calahan\','



```

\ 'profile_path\': None}, {\ 'credit_id\': \ '5892127fc3a3686b0a004de5\ ',
\ 'department\': \ 'Lighting\ ', \ 'gender\': 0, \ 'id\': 7899, \ 'job\': \ 'Lighting
Supervisor\ ', \ 'name\': \ 'Galyn Susman\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589212cdc3a3680970009268\ ', \ 'department\': \ 'Visual
Effects\ ', \ 'gender\': 0, \ 'id\': 12915, \ 'job\': \ 'CG Painter\ ', \ 'name\':
\ 'William Cone\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '5892130f9251412dc8009791\ ', \ 'department\': \ 'Art\ ', \ 'gender\': 0, \ 'id\':
1748705, \ 'job\': \ 'Sculptor\ ', \ 'name\': \ 'Shelley Daniels Lekven\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '5892131c9251412dd4008f4c\ ',
\ 'department\': \ 'Visual Effects\ ', \ 'gender\': 2, \ 'id\': 7889, \ 'job\':
\ 'Character Designer\ ', \ 'name\': \ 'Bob Pauley\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589213249251412dd100987b\ ', \ 'department\': \ 'Visual
Effects\ ', \ 'gender\': 2, \ 'id\': 7918, \ 'job\': \ 'Character Designer\ ',
\ 'name\': \ 'Bud Luckey\ ', \ 'profile_path\':
\ '/pcCh7G19FKMNijmPQg1PMH1btic.jpg\ '}, {\ 'credit_id\':
\ '5892132b9251412dc80097b1\ ', \ 'department\': \ 'Visual Effects\ ', \ 'gender\': 2,
\ 'id\': 7, \ 'job\': \ 'Character Designer\ ', \ 'name\': \ 'Andrew Stanton\ ',
\ 'profile_path\': \ '/pvQWsu0qc8JFQhMVJkTHuexUAa1.jpg\ '}, {\ 'credit_id\':
\ '58921332c3a368634800467b\ ', \ 'department\': \ 'Visual Effects\ ', \ 'gender\': 0,
\ 'id\': 12915, \ 'job\': \ 'Character Designer\ ', \ 'name\': \ 'William Cone\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '5892135f9251412dd4008f90\ ',
\ 'department\': \ 'Visual Effects\ ', \ 'gender\': 0, \ 'id\': 1748706, \ 'job\':
\ 'Character Designer\ ', \ 'name\': \ 'Steve Johnson\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '58921384c3a3680973008fd4\ ', \ 'department\': \ 'Visual
Effects\ ', \ 'gender\': 0, \ 'id\': 1176752, \ 'job\': \ 'Character Designer\ ',
\ 'name\': \ 'Dan Haskett\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '5892138e9251412dc20099fc\ ', \ 'department\': \ 'Visual Effects\ ', \ 'gender\': 0,
\ 'id\': 1088034, \ 'job\': \ 'Character Designer\ ', \ 'name\': \ 'Tom Holloway\ ',
\ 'profile_path\': \ '/aOr0T2usTBpgMI5aZbRBDW1fTl8.jpg\ '}, {\ 'credit_id\':
\ '58921395c3a368097700942f\ ', \ 'department\': \ 'Visual Effects\ ', \ 'gender\': 0,
\ 'id\': 1447465, \ 'job\': \ 'Character Designer\ ', \ 'name\': \ 'Jean Gillmore\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '589213e2c3a3680973009026\ ',
\ 'department\': \ 'Directing\ ', \ 'gender\': 0, \ 'id\': 1748709, \ 'job\':
\ 'Layout\ ', \ 'name\': \ 'Desirée Mourad\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589214099251412dc5009d57\ ', \ 'department\': \ 'Art\ ',
\ 'gender\': 0, \ 'id\': 1748710, \ 'job\': \ 'Set Dresser\ ', \ 'name\': "Kelly
O'Connell", \ 'profile_path\': None}, {\ 'credit_id\':
\ '58921411c3a3686b0a004f70\ ', \ 'department\': \ 'Art\ ', \ 'gender\': 0, \ 'id\':
1443471, \ 'job\': \ 'Set Dresser\ ', \ 'name\': \ 'Sonoko Konishi\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '58921434c3a368096a00956e\ ',
\ 'department\': \ 'Art\ ', \ 'gender\': 0, \ 'id\': 1748711, \ 'job\': \ 'Set
Dresser\ ', \ 'name\': \ 'Ann M. Rockwell\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '5892144ac3a36809680090de\ ', \ 'department\': \ 'Editing\ ',
\ 'gender\': 0, \ 'id\': 1748712, \ 'job\': \ 'Editorial Manager\ ', \ 'name\':
\ 'Julie M. McDonald\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '58921479c3a368096800910f\ ', \ 'department\': \ 'Editing\ ', \ 'gender\': 0,
\ 'id\': 1589729, \ 'job\': \ 'Assistant Editor\ ', \ 'name\': \ 'Robin Lee\ ',

```

```

\ 'profile_path\': None}, {\ 'credit_id\': \ '5892148b9251412dd10099cc\ ',
\ 'department\': \ 'Editing\ ', \ 'gender\': 0, \ 'id\': 1748716, \ 'job\':
\ 'Assistant Editor\ ', \ 'name\': \ 'Tom Freeman\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589214959251412dcb009b1f\ ', \ 'department\': \ 'Editing\ ',
\ 'gender\': 0, \ 'id\': 1748717, \ 'job\': \ 'Assistant Editor\ ', \ 'name\': \ 'Ada
Cochavi\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '5892149ec3a3686348004798\ ', \ 'department\': \ 'Editing\ ', \ 'gender\': 0,
\ 'id\': 1336438, \ 'job\': \ 'Assistant Editor\ ', \ 'name\': \ 'Dana Mulligan\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '589214adc3a368096a0095db\ ',
\ 'department\': \ 'Editing\ ', \ 'gender\': 0, \ 'id\': 1748718, \ 'job\':
\ 'Editorial Coordinator\ ', \ 'name\': \ 'Deirdre Morrison\ ', \ 'profile_path\':
None}, {\ 'credit_id\': \ '589214c7c3a368097700952b\ ', \ 'department\':
\ 'Production\ ', \ 'gender\': 0, \ 'id\': 1748719, \ 'job\': \ 'Production
Coordinator\ ', \ 'name\': \ 'Lori Lombardo\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589214cec3a368096a009603\ ', \ 'department\': \ 'Production\ ',
\ 'gender\': 0, \ 'id\': 1748720, \ 'job\': \ 'Production Coordinator\ ', \ 'name\':
\ 'Ellen Devine\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '589214e39251412dc8009904\ ', \ 'department\': \ 'Crew\ ', \ 'gender\': 0, \ 'id\':
1468014, \ 'job\': \ 'Unit Publicist\ ', \ 'name\': \ 'Lauren Beth Strogoff\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '58921544c3a3686b0a00507d\ ',
\ 'department\': \ 'Sound\ ', \ 'gender\': 2, \ 'id\': 2216, \ 'job\': \ 'Sound Re-
Recording Mixer\ ', \ 'name\': \ 'Gary Rydstrom\ ', \ 'profile_path\':
\ '/jZpr1nVf071ldWIOYtmP1FGw7Rj.jpg\ '}, {\ 'credit_id\':
\ '5892154c9251412dd1009a56\ ', \ 'department\': \ 'Sound\ ', \ 'gender\': 0, \ 'id\':
1425978, \ 'job\': \ 'Sound Re-Recording Mixer\ ', \ 'name\': \ 'Gary Summers\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '58921555c3a36809680091bd\ ',
\ 'department\': \ 'Sound\ ', \ 'gender\': 2, \ 'id\': 8276, \ 'job\': \ 'Supervising
Sound Editor\ ', \ 'name\': \ 'Tim Holland\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589215c39251412dcb009c12\ ', \ 'department\': \ 'Sound\ ',
\ 'gender\': 0, \ 'id\': 7069, \ 'job\': \ 'Sound Effects Editor\ ', \ 'name\': \ 'Pat
Jackson\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '58921698c3a368096a009788\ ', \ 'department\': \ 'Crew\ ', \ 'gender\': 2, \ 'id\':
15894, \ 'job\': \ 'Sound Design Assistant\ ', \ 'name\': \ 'Tom Myers\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '589216a89251412dc2009ca4\ ',
\ 'department\': \ 'Sound\ ', \ 'gender\': 0, \ 'id\': 1414177, \ 'job\': \ 'Assistant
Sound Editor\ ', \ 'name\': \ 'J.R. Grubbs\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589216c19251412dc2009cb9\ ', \ 'department\': \ 'Sound\ ',
\ 'gender\': 1, \ 'id\': 1748724, \ 'job\': \ 'Assistant Sound Editor\ ', \ 'name\':
\ 'Susan Sanford\ ', \ 'profile_path\': None}, {\ 'credit_id\':
\ '589216ccc3a3680973009274\ ', \ 'department\': \ 'Sound\ ', \ 'gender\': 0, \ 'id\':
1748725, \ 'job\': \ 'Assistant Sound Editor\ ', \ 'name\': \ 'Susan Popovic\ ',
\ 'profile_path\': None}, {\ 'credit_id\': \ '589216d79251412dc8009aa0\ ',
\ 'department\': \ 'Sound\ ', \ 'gender\': 0, \ 'id\': 8067, \ 'job\': \ 'Assistant
Sound Editor\ ', \ 'name\': \ 'Dan Engstrom\ ', \ 'profile_path\': None},
{\ 'credit_id\': \ '589216e49251412dcd009a4f\ ', \ 'department\': \ 'Production\ ',
\ 'gender\': 1, \ 'id\': 7902, \ 'job\': \ 'Casting Consultant\ ', \ 'name\': \ 'Ruth
Lambert\ ', \ 'profile_path\': None}, {\ 'credit_id\':

```

```
\'589216f39251412dc2009cf3\\', \'department\\': \'Production\\', \'gender\\': 0,
\'id\\': 84493, \'job\\': \'ADR Voice Casting\\', \'name\\': \'Mickie McGowan\\',
\'profile_path\\': \'/k7TjJBfINsg8vLQxJwos6XObAD6.jpg\\'}]'
```

```
[36]: # in crew i want to job director. so we make a function
```

```
def fetch_directer(obj):
    l = []
    for i in ast.literal_eval(obj):
        if i['job'] == 'Director':
            l.append(i['name'])
            break
    return l
```

```
[37]: df['crew'] = df['crew'].apply(fetch_directer)
```

```
[38]: df.head(5)
```

```
[38]:      id      genres \
0    862  [Animation, Comedy, Family]
1   8844  [Adventure, Fantasy, Family]
2  15602      [Romance, Comedy]
3  31357  [Comedy, Drama, Romance]
4  11862      [Comedy]

      cast      crew \
0  [Tom Hanks, Tim Allen, Don Rickles]  [John Lasseter]
1  [Robin Williams, Jonathan Hyde, Kirsten Dunst]  [Joe Johnston]
2  [Walter Matthau, Jack Lemmon, Ann-Margret]  [Howard Deutch]
3  [Whitney Houston, Angela Bassett, Loretta Devine]  [Forest Whitaker]
4  [Steve Martin, Diane Keaton, Martin Short]  [Charles Shyer]

      keywords \
0  [jealousy, toy, boy, friendship, friends, riva...
1  [board game, disappearance, based on children'...
2  [fishing, best friend, duringcreditsstinger, o...
3  [based on novel, interracial relationship, sin...
4  [baby, midlife crisis, confidence, aging, daug...

      original_title      title \
0      Toy Story      Toy Story
1      Jumanji      Jumanji
2      Grumpier Old Men      Grumpier Old Men
3      Waiting to Exhale      Waiting to Exhale
4  Father of the Bride Part II  Father of the Bride Part II

      overview
```

```

0 Led by Woody, Andy's toys live happily in his ...
1 When siblings Judy and Peter discover an encha...
2 A family wedding reignites the ancient feud be...
3 Cheated on, mistreated and stepped on, the wom...
4 Just when George Banks has recovered from his ...

```

```
[39]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 44482 entries, 0 to 46627
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   id              44482 non-null  object
1   genres          44482 non-null  object
2   cast            44482 non-null  object
3   crew            44482 non-null  object
4   keywords        44482 non-null  object
5   original_title  44482 non-null  object
6   title           44482 non-null  object
7   overview        44482 non-null  object
dtypes: object(8)
memory usage: 4.1+ MB

```

```
[40]: def changingLabels(number):

      return str(number/10**6) + "M"
```

```
[41]: df_exploded = df.explode('genres')

genres_df = pd.DataFrame(df_exploded['genres'].value_counts()).reset_index()
genres_df.columns = ["Genre", "Count"]

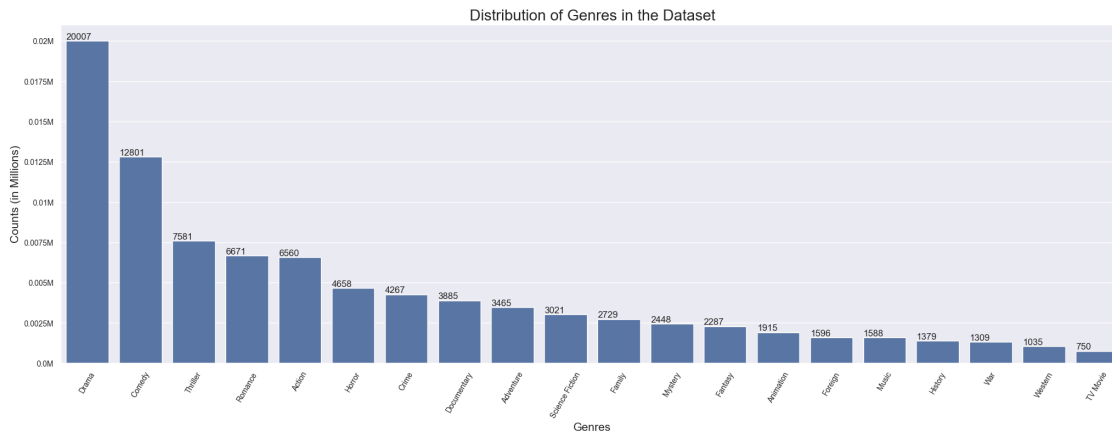
sns.set(style="darkgrid")
fig, axes = plt.subplots(1, 1, figsize=(25, 8), sharey=True)

sns.barplot(y="Count", x="Genre", data=genres_df, ax=axes)
axes.set_yticklabels([changingLabels(num) for num in axes.get_yticks()])
for p in axes.patches:
    axes.annotate('{}' .format(int(p.get_height())) , (p.get_x(), p.
    ↳get_height()+100))

plt.tick_params(labelsiz e = 15)
plt.title("Distribution of Genres in the Dataset", fontsize = 20)
plt.xlabel("Genres", fontsize = 15)
plt.xticks(rotation=60, fontsize=10)
plt.yticks(fontsize=10)

```

```
plt.ylabel("Counts (in Millions)", fontsize = 15)
plt.show()
```



```
[42]: df['genres']=df['genres'].apply(lambda x:[i.replace(" ", "") for i in x] )
df['cast']=df['cast'].apply(lambda x:[i.replace(" ", "") for i in x] )
df['crew']=df['crew'].apply(lambda x:[i.replace(" ", "") for i in x] )
df['keywords']=df['keywords'].apply(lambda x:[i.replace(" ", "") for i in x] )
```

```
[43]: df.head()
```

```
[43]:      id      genres \
0    862  [Animation, Comedy, Family]
1   8844  [Adventure, Fantasy, Family]
2  15602      [Romance, Comedy]
3  31357  [Comedy, Drama, Romance]
4  11862      [Comedy]

      cast      crew \
0  [TomHanks, TimAllen, DonRickles]  [JohnLasseter]
1  [RobinWilliams, JonathanHyde, KirstenDunst]  [JoeJohnston]
2  [WalterMatthau, JackLemmon, Ann-Margret]  [HowardDeutch]
3  [WhitneyHouston, AngelaBassett, LorettaDevine]  [ForestWhitaker]
4  [SteveMartin, DianeKeaton, MartinShort]  [CharlesShyer]

      keywords \
0  [jealousy, toy, boy, friendship, friends, riva...
1  [boardgame, disappearance, basedonchildren'sbo...
2  [fishing, bestfriend, duringcreditsstinger, ol...
3  [basedonnovel, interracialrelationship, single...
4  [baby, midlifecrisis, confidence, aging, daugh...
```

```
original_title      title \
```

|   |                             |                             |
|---|-----------------------------|-----------------------------|
| 0 | Toy Story                   | Toy Story                   |
| 1 | Jumanji                     | Jumanji                     |
| 2 | Grumpier Old Men            | Grumpier Old Men            |
| 3 | Waiting to Exhale           | Waiting to Exhale           |
| 4 | Father of the Bride Part II | Father of the Bride Part II |

```

                                overview
0  Led by Woody, Andy's toys live happily in his ...
1  When siblings Judy and Peter discover an encha...
2  A family wedding reignites the ancient feud be...
3  Cheated on, mistreated and stepped on, the wom...
4  Just when George Banks has recovered from his ...

```

```
[44]: df['overview'][0]
```

```
[44]: "Led by Woody, Andy's toys live happily in his room until Andy's birthday brings
Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody
plots against Buzz. But when circumstances separate Buzz and Woody from their
owner, the duo eventually learns to put aside their differences."
```

```
[45]: df['overview'] = df['overview'].apply(lambda x:x.split())
```

```
[46]: df['overview'][0]
```

```
[46]: ['Led',
      'by',
      'Woody,',
      'Andy's',
      'toys',
      'live',
      'happily',
      'in',
      'his',
      'room',
      'until',
      'Andy's',
      'birthday',
      'brings',
      'Buzz',
      'Lightyear',
      'onto',
      'the',
      'scene.',
      'Afraid',
      'of',
      'losing',
      'his',
```

```
'place',
'in',
"Andy's",
'heart,',
'Woody',
'plots',
'against',
'Buzz.',
'But',
'when',
'circumstances',
'separate',
'Buzz',
'and',
'Woody',
'from',
'their',
'owner,',
'the',
'duo',
'eventually',
'learns',
'to',
'put',
'aside',
'their',
'differences.']
```

```
[47]: df['tag'] = df['overview'] + df['genres'] + df['keywords'] + df['cast'] +
      ↪df['crew']
```

```
[48]: df.head()
```

```
[48]:
```

|   | id    | genres \                     | cast   | crew \           |
|---|-------|------------------------------|--|------------------|
| 0 | 862   | [Animation, Comedy, Family]  | [TomHanks, TimAllen, DonRickles]               | [JohnLasseter]   |
| 1 | 8844  | [Adventure, Fantasy, Family] | [RobinWilliams, JonathanHyde, KirstenDunst]    | [JoeJohnston]    |
| 2 | 15602 | [Romance, Comedy]            | [WalterMatthau, JackLemmon, Ann-Margret]       | [HowardDeutch]   |
| 3 | 31357 | [Comedy, Drama, Romance]     | [WhitneyHouston, AngelaBassett, LorettaDevine] | [ForestWhitaker] |
| 4 | 11862 | [Comedy]                     | [SteveMartin, DianeKeaton, MartinShort]        | [CharlesShyer]   |

```

                                keywords \
0 [jealousy, toy, boy, friendship, friends, riva...
1 [boardgame, disappearance, basedonchildren'sbo...
2 [fishing, bestfriend, duringcreditsstinger, ol...
3 [basedonnovel, interracialrelationship, single...
4 [baby, midlifecrisis, confidence, aging, daugh...

                                original_title                                title \
0                                Toy Story                                Toy Story
1                                Jumanji                                Jumanji
2                                Grumpier Old Men                        Grumpier Old Men
3                                Waiting to Exhale                      Waiting to Exhale
4 Father of the Bride Part II  Father of the Bride Part II

                                overview \
0 [Led, by, Woody,, Andy's, toys, live, happily,...
1 [When, siblings, Judy, and, Peter, discover, a...
2 [A, family, wedding, reignites, the, ancient, ...
3 [Cheated, on,, mistreated, and, stepped, on,, ...
4 [Just, when, George, Banks, has, recovered, fr...

                                tag
0 [Led, by, Woody,, Andy's, toys, live, happily,...
1 [When, siblings, Judy, and, Peter, discover, a...
2 [A, family, wedding, reignites, the, ancient, ...
3 [Cheated, on,, mistreated, and, stepped, on,, ...
4 [Just, when, George, Banks, has, recovered, fr...

```

```
[49]: df['tag'][0]
```

```
[49]: ['Led',
      'by',
      'Woody,',
      "Andy's",
      'toys',
      'live',
      'happily',
      'in',
      'his',
      'room',
      'until',
      "Andy's",
      'birthday',
      'brings',
      'Buzz',
      'Lightyear',
      'onto',
```



'the',  
'scene.',  
'Afraid',  
'of',  
'losing',  
'his',  
'place',  
'in',  
"Andy's",  
'heart',  
'Woody',  
'plots',  
'against',  
'Buzz.',  
'But',  
'when',  
'circumstances',  
'separate',  
'Buzz',  
'and',  
'Woody',  
'from',  
'their',  
'owner',  
'the',  
'duo',  
'eventually',  
'learns',  
'to',  
'put',  
'aside',  
'their',  
'differences.',  
'Animation',  
'Comedy',  
'Family',  
'jealousy',  
'toy',  
'boy',  
'friendship',  
'friends',  
'rivalry',  
'boynextdoor',  
'newtoy',  
'toycomestolife',  
'TomHanks',  
'TimAllen',

```
'DonRickles',
'JohnLasseter']
```

```
[50]: df['tag'] = df['tag'].apply(lambda x: " ".join(x))
```

```
[51]: df.head()
```

```
[51]:
```

|   | id    | genres \                     | cast   | crew \           | keywords \  | original_title              | title \                     | overview \  | tag   |
|---|-------|------------------------------|--|------------------|---|-----------------------------|-----------------------------|---|---|
| 0 | 862   | [Animation, Comedy, Family]  | [TomHanks, TimAllen, DonRickles]               | [JohnLasseter]   | [jealousy, toy, boy, friendship, friends, riva... | Toy Story                   | Toy Story                   | [Led, by, Woody,, Andy's, toys, live, happily,... | Led by Woody, Andy's toys live happily in his ... |
| 1 | 8844  | [Adventure, Fantasy, Family] | [RobinWilliams, JonathanHyde, KirstenDunst]    | [JoeJohnston]    | [boardgame, disappearance, basedonchildren'sbo... | Jumanji                     | Jumanji                     | [When, siblings, Judy, and, Peter, discover, a... | When siblings Judy and Peter discover an encha... |
| 2 | 15602 | [Romance, Comedy]            | [WalterMatthau, JackLemmon, Ann-Margret]       | [HowardDeutch]   | [fishing, bestfriend, duringcreditsstinger, ol... | Grumpier Old Men            | Grumpier Old Men            | [A, family, wedding, reignites, the, ancient, ... | A family wedding reignites the ancient feud be... |
| 3 | 31357 | [Comedy, Drama, Romance]     | [WhitneyHouston, AngelaBassett, LorettaDevine] | [ForestWhitaker] | [basedonnovel, interracialrelationship, single... | Waiting to Exhale           | Waiting to Exhale           | [Cheated, on,, mistreated, and, stepped, on,, ... | Cheated on, mistreated and stepped on, the wom... |
| 4 | 11862 | [Comedy]                     | [SteveMartin, DianeKeaton, MartinShort]        | [CharlesShyer]   | [baby, midlifecrisis, confidence, aging, daugh... | Father of the Bride Part II | Father of the Bride Part II | [Just, when, George, Banks, has, recovered, fr... |   |

4 Just when George Banks has recovered from his ...

```
[52]: new_df = df[['id','title','tag']]
```

```
[53]: new_df.head()
```

```
[53]:
```

|   | id    | title \                     | tag   |
|---|-------|-----------------------------|---|
| 0 | 862   | Toy Story                   | Led by Woody, Andy's toys live happily in his ... |
| 1 | 8844  | Jumanji                     | When siblings Judy and Peter discover an encha... |
| 2 | 15602 | Grumpier Old Men            | A family wedding reignites the ancient feud be... |
| 3 | 31357 | Waiting to Exhale           | Cheated on, mistreated and stepped on, the wom... |
| 4 | 11862 | Father of the Bride Part II | Just when George Banks has recovered from his ... |

```
[54]: new_df['tag'][0]
```

```
[54]: "Led by Woody, Andy's toys live happily in his room until Andy's birthday brings Buzz Lightyear onto the scene. Afraid of losing his place in Andy's heart, Woody plots against Buzz. But when circumstances separate Buzz and Woody from their owner, the duo eventually learns to put aside their differences. Animation Comedy Family jealousy toy boy friendship friends rivalry boynextdoor newtoy toycomestolife TomHanks TimAllen DonRickles JohnLasseter"
```

```
[55]: new_df['tag'] = new_df['tag'].apply(lambda x:x.lower())
```

```
[56]: new_df['tag'][0]
```

```
[56]: "led by woody, andy's toys live happily in his room until andy's birthday brings buzz lightyear onto the scene. afraid of losing his place in andy's heart, woody plots against buzz. but when circumstances separate buzz and woody from their owner, the duo eventually learns to put aside their differences. animation comedy family jealousy toy boy friendship friends rivalry boynextdoor newtoy toycomestolife tomhanks timallen donrickles johnlasseter"
```

```
[57]: nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Srijan\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
[nlTK_data] Downloading package punkt to
[nlTK_data]      C:\Users\Srijan\AppData\Roaming\nlTK_data...
[nlTK_data] Package punkt is already up-to-date!
[nlTK_data] Downloading package wordnet to
[nlTK_data]      C:\Users\Srijan\AppData\Roaming\nlTK_data...
[nlTK_data] Package wordnet is already up-to-date!
```

```
[57]: True
```

```
[58]: stop_words = set(stopwords.words('english'))
```

```
[59]: def remove_stopwords(text):
      """Remove common English words using NLTK's list of stopwords"""
      words = word_tokenize(text)
      filtered = [w for w in words if not w in stop_words]
      return ' '.join(filtered)
```

```
[60]: new_df['tag'].apply(lambda x: remove_stopwords(x))
```

```
[60]: 0      led woody , andy 's toys live happily room and...
      1      siblings judy peter discover enchanted board g...
      2      family wedding reignites ancient feud next-doo...
      3      cheated , mistreated stepped , women holding b...
      4      george banks recovered daughter 's wedding , r...
      ...
      46623    rising falling man woman . drama family tragic...
      46624    artist struggles finish work storyline cult pl...
      46625    one hits goes wrong , professional assassin en...
      46626    small town live two brothers , one minister on...
      46627    50 years decriminalisation homosexuality uk , ...
      Name: tag, Length: 44482, dtype: object
```

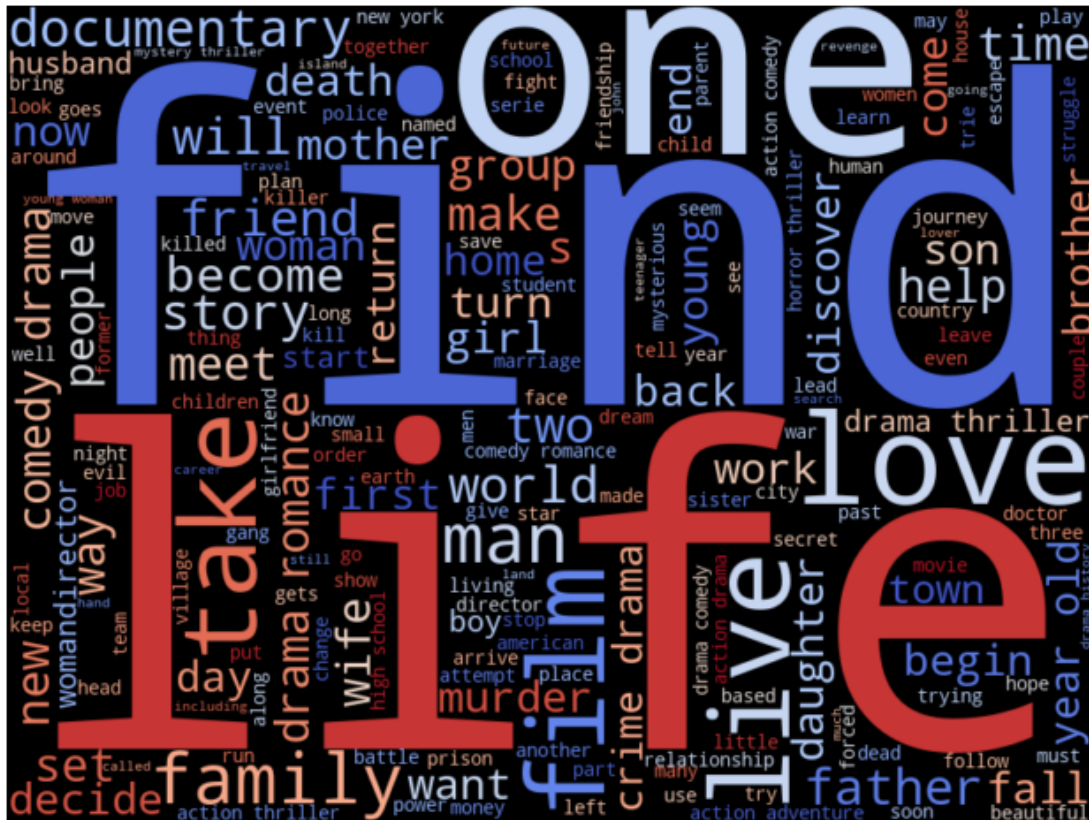
```
[61]: unique_categories = new_df['tag'].nunique()
      print("Number of unique categories:", unique_categories)
```

```
Number of unique categories: 44474
```

```
[62]: def generate_wordcloud_from_dataframe(df, column_name):
      # Concatenate all text in the specified column into a single string
      text = ' '.join(df[column_name].fillna(''))
      # Generate word cloud
      stopwords = set(STOPWORDS)
      diag = wordcloud.WordCloud(stopwords=stopwords, background_color='black',
      colormap='coolwarm', width=800, height=600).generate(text)
      # Display the word cloud using matplotlib
      plt.figure(figsize=(10, 6))
      plt.imshow(diag, interpolation='bilinear')
```

```
plt.axis('off')
plt.show()
```

```
generate_wordcloud_from_dataframe(new_df, 'tag')
```



```
cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
cv.fit_transform(new_df['tag']).toarray()
```

```
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]], dtype=int64)
```

```
cv.fit_transform(new_df['tag']).toarray().shape
```

(44482, 5000)

```
[67]: vector = cv.fit_transform(new_df['tag']).toarray()
```

```
[68]: vector
```

```
[68]: 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]], dtype=int64)
```

```
[69]: vector.shape
```

```
[69]: (44482, 5000)
```

```
[70]: cv.get_stop_words()
```

```
[70]: frozenset({'a',
               'about',
               'above',
               'across',
               'after',
               'afterwards',
               'again',
               'against',
               'all',
               'almost',
               'alone',
               'along',
               'already',
               'also',
               'although',
               'always',
               'am',
               'among',
               'amongst',
               'amoungst',
               'amount',
               'an',
               'and',
               'another',
               'any',
               'anyhow',
               'anyone',
               'anything',
               'anyway',
```

'anywhere',  
'are',  
'around',  
'as',  
'at',  
'back',  
'be',  
'became',  
'because',  
'become',  
'becomes',  
'becoming',  
'been',  
'before',  
'beforehand',  
'behind',  
'being',  
'below',  
'beside',  
'besides',  
'between',  
'beyond',  
'bill',  
'both',  
'bottom',  
'but',  
'by',  
'call',  
'can',  
'cannot',  
'cant',  
'co',  
'con',  
'could',  
'couldnt',  
'cry',  
'de',  
'describe',  
'detail',  
'do',  
'done',  
'down',  
'due',  
'during',  
'each',  
'eg',  
'eight',

'either',  
'eleven',  
'else',  
'elsewhere',  
'empty',  
'enough',  
'etc',  
'even',  
'ever',  
'every',  
'everyone',  
'everything',  
'everywhere',  
'except',  
'few',  
'fifteen',  
'fifty',  
'fill',  
'find',  
'fire',  
'first',  
'five',  
'for',  
'former',  
'formerly',  
'forty',  
'found',  
'four',  
'from',  
'front',  
'full',  
'further',  
'get',  
'give',  
'go',  
'had',  
'has',  
'hasnt',  
'have',  
'he',  
'hence',  
'her',  
'here',  
'hereafter',  
'hereby',  
'herein',  
'hereupon',



'hers',  
'herself',  
'him',  
'himself',  
'his',  
'how',  
'however',  
'hundred',  
'i',  
'ie',  
'if',  
'in',  
'inc',  
'indeed',  
'interest',  
'into',  
'is',  
'it',  
'its',  
'itself',  
'keep',  
'last',  
'latter',  
'latterly',  
'least',  
'less',  
'ltd',  
'made',  
'many',  
'may',  
'me',  
'meanwhile',  
'might',  
'mill',  
'mine',  
'more',  
'moreover',  
'most',  
'mostly',  
'move',  
'much',  
'must',  
'my',  
'myself',  
'name',  
'namely',  
'neither',

'never',  
'nevertheless',  
'next',  
'nine',  
'no',  
'nobody',  
'none',  
'noone',  
'nor',  
'not',  
'nothing',  
'now',  
'nowhere',  
'of',  
'off',  
'often',  
'on',  
'once',  
'one',  
'only',  
'onto',  
'or',  
'other',  
'others',  
'otherwise',  
'our',  
'ours',  
'ourselves',  
'out',  
'over',  
'own',  
'part',  
'per',  
'perhaps',  
'please',  
'put',  
'rather',  
're',  
'same',  
'see',  
'seem',  
'seemed',  
'seeming',  
'seems',  
'serious',  
'several',  
'she',

'should',  
'show',  
'side',  
'since',  
'sincere',  
'six',  
'sixty',  
'so',  
'some',  
'somehow',  
'someone',  
'something',  
'sometime',  
'sometimes',  
'somewhere',  
'still',  
'such',  
'system',  
'take',  
'ten',  
'than',  
'that',  
'the',  
'their',  
'them',  
'themselves',  
'then',  
'thence',  
'there',  
'thereafter',  
'thereby',  
'therefore',  
'therein',  
'thereupon',  
'these',  
'they',  
'thick',  
'thin',  
'third',  
'this',  
'those',  
'though',  
'three',  
'through',  
'throughout',  
'thru',  
'thus',

'to',  
'together',  
'too',  
'top',  
'toward',  
'towards',  
'twelve',  
'twenty',  
'two',  
'un',  
'under',  
'until',  
'up',  
'upon',  
'us',  
'very',  
'via',  
'was',  
'we',  
'well',  
'were',  
'what',  
'whatever',  
'when',  
'whence',  
'whenever',  
'where',  
'whereafter',  
'whereas',  
'whereby',  
'wherein',  
'whereupon',  
'wherever',  
'whether',  
'which',  
'while',  
'whither',  
'who',  
'whoever',  
'whole',  
'whom',  
'whose',  
'why',  
'will',  
'with',  
'within',  
'without',

```

        'would',
        'yet',
        'you',
        'your',
        'yours',
        'yourself',
        'yourselves'})

```

```
[71]: cv.get_feature_names_out()
```

```
[71]: array(['000', '10', '100', ..., 'zombies', 'zone', 'zoo'], dtype=object)
```

```
[72]: ps = PorterStemmer()
```

```
[73]: def stem(text):
        y = []
        for i in text.split():
            y.append(ps.stem(i))
        return " ".join(y)
```

```
[74]: new_df['tag'][0]
```

```
[74]: "led by woody, andy's toys live happily in his room until andy's birthday brings
buzz lightyear onto the scene. afraid of losing his place in andy's heart, woody
plots against buzz. but when circumstances separate buzz and woody from their
owner, the duo eventually learns to put aside their differences. animation
comedy family jealousy toy boy friendship friends rivalry boynextdoor newtoy
toycomestolife tomhanks timallen donrickles johnlasseter"
```

```
[75]: result = stem("led by woody, andy's toys live happily in his room until andy's
↳birthday brings buzz lightyear onto the scene. afraid of losing his place in
↳andy's heart, woody plots against buzz. but when circumstances separate buzz
↳and woody from their owner, the duo eventually learns to put aside their
↳differences. animation comedy family jealousy toy boy friendship friends
↳rivalry boynextdoor newtoy toycomestolife tomhanks timallen donrickles
↳johnlasseter")
```

```
[76]: result
```

```
[76]: "led by woody, andy' toy live happili in hi room until andy' birthday bring buzz
lightyear onto the scene. afraid of lose hi place in andy' heart, woodi plot
against buzz. but when circumst separ buzz and woodi from their owner, the duo
eventu learn to put asid their differences. anim comedi famili jealousy toy boy
friendship friend rivalri boynextdoor newtoy toycomestolif tomhank timallen
donrickl johnlasset"
```

```
[77]: new_df['tag']=new_df['tag'].apply(stem)
```

```
[78]: cv = CountVectorizer(max_features=5000,stop_words='english')
```

```
[79]: cv.fit_transform(new_df['tag']).toarray()
```

```
[79]: 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]], dtype=int64)
```

```
[80]: vector = cv.fit_transform(new_df['tag']).toarray()
```

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

```
[82]: similarity
```

```
[82]: array([[1.          , 0.07216878, 0.05025189, ..., 0.          , 0.17099639,
          0.          ],
          [0.07216878, 1.          , 0.06527912, ..., 0.          , 0.05923489,
          0.03608439],
          [0.05025189, 0.06527912, 1.          , ..., 0.          , 0.01546721,
          0.          ],
          ...,
          [0.          , 0.          , 0.          , ..., 1.          , 0.          ,
          0.          ],
          [0.17099639, 0.05923489, 0.01546721, ..., 0.          , 1.          ,
          0.          ],
          [0.          , 0.03608439, 0.          , ..., 0.          , 0.          ,
          1.          ]])
```

```
[83]: cosine_similarity(vector).shape
```

```
[83]: (44482, 44482)
```

```
[84]: # lambda x:x[1] second number basis short
      # enumerate use to hold the list position
      # list is use to enumerate to see

      sorted(list(enumerate(similarity[0])), reverse=True, key= lambda x:x[1])[1:11]
```

```
[84]: [(2979, 0.4341215710622296),
      (44453, 0.3779644730092272),
      (25548, 0.3726779962499649),
      (7184, 0.3698001308168194),
      (24304, 0.3698001308168194),
```

```
(42246, 0.3578300267477955),
(33701, 0.34020690871988585),
(2603, 0.3370999312316211),
(9179, 0.33471934069760145),
(1906, 0.3333333333333333)]
```

```
[85]: def recommended(movie):
        movie_index = new_df[new_df['title'] == movie].index[0]
        distance = similarity[movie_index]
        movie_list = sorted(list(enumerate(distance)), reverse=True, key= lambda x:
↪x[1])[1:11]
        for i in movie_list:
            new_df.iloc[i[0]].id
            print(new_df.iloc[i[0]].title)
```

```
[86]: new_df.tail()
```

```
[86]:      id      title \
46623  439050      Subdue
46624  111109  Century of Birthing
46625   67758      Betrayal
46626  227506    Satan Triumphant
46627  461257      Queerama

tag
46623  rise and fall between a man and woman. drama f...
46624  an artist struggl to finish hi work while a st...
46625  when one of her hit goe wrong, a profession as...
46626  in a small town live two brothers, one a minis...
46627  50 year after decriminalis of homosexu in the ...
```

```
[87]: new_df[new_df['title'] == 'Century of Birthing'].index[0]
```

```
[87]: 46624
```

```
[88]: df[df['title'] == 'Jumanji']
```

```
[88]:      id      genres \
1  8844  [Adventure, Fantasy, Family]

cast      crew \
1  [RobinWilliams, JonathanHyde, KirstenDunst]  [JoeJohnston]

keywords original_title  title \
1  [boardgame, disappearance, basedonchildren'sbo...  Jumanji  Jumanji

overview \
```

```
1 [When, siblings, Judy, and, Peter, discover, a...
```

```
tag
```

```
1 When siblings Judy and Peter discover an encha...
```

```
[89]: recommended('Jumanji')
```

```
Table No. 21
```

```
Enter the Battlefield: Life on the Magic - The Gathering Pro Tour
```

```
Liar Game: Reborn
```

```
The Games Maker
```

```
100 Yen: The Japanese Arcade Experience
```

```
Beta Test
```

```
Indie Game: The Movie
```

```
The Last Starfighter
```

```
Wreck-It Ralph
```

```
Brainscan
```

```
[90]: new_df.iloc[21465]
```

```
[90]: id                157129
      title                Table No. 21
      tag      a coupl live a mediocr life and are thrill to ...
      Name: 21766, dtype: object
```

```
[91]: pickle.dump(new_df.to_dict(), open('movies_name.pkl', 'wb'))
```

```
[94]: pickle.dump(similarity, open('recommended.pkl', 'wb'))
```

```
[96]: with open('C:\\Users\\Srijan\\Desktop\\Internship\\Movie Recommender_
      ↪Systems\\recommended.pkl', 'rb') as file:
      data = pickle.load(file)
```

```
[97]: data.size
```

```
[97]: 1978648324
```

```
[98]: import joblib
```

```
[99]: joblib.dump(similarity , 'model_save')
```

```
[99]: ['model_save']
```

```
[100]: job = joblib.load('C:\\Users\\Srijan\\Desktop\\Internship\\Movie Recommender_
      ↪Systems\\model_save')
```



```
[101]: def recommended1(movie):
        movie_index = new_df[new_df['title'] == movie].index[0]
        distance = job[movie_index]
        movie_list = sorted(list(enumerate(distance)), reverse=True, key= lambda x:
↪x[1])[1:11]
        for i in movie_list:
            new_df.iloc[i[0]].id
            print(new_df.iloc[i[0]].title)
```

```
[102]: recommended1('Jumanji')
```

Table No. 21

Enter the Battlefield: Life on the Magic - The Gathering Pro Tour

Liar Game: Reborn

The Games Maker

100 Yen: The Japanese Arcade Experience

Beta Test

Indie Game: The Movie

The Last Starfighter

Wreck-It Ralph

Brainscan

```
[ ]:
```