movies-recommendation-project

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

```
[2]: credits = pd.read_csv("tmdb_5000_credits.csv")
     movies = pd.read_csv("tmdb_5000_movies.csv")
```

```
[3]: movies.head()
```

```
[3]:        budget                                               genres  \
     0   237000000   [{"id": 28, "name": "Action"}, {"id": 12, "nam…
     1   300000000   [{"id": 12, "name": "Adventure"}, {"id": 14, "…
     2   245000000   [{"id": 28, "name": "Action"}, {"id": 12, "nam…
     3   250000000   [{"id": 28, "name": "Action"}, {"id": 80, "nam…
     4   260000000   [{"id": 28, "name": "Action"}, {"id": 12, "nam…

                                             homepage      id  \
     0                     http://www.avatarmovie.com/   19995
     1   http://disney.go.com/disneypictures/pirates/     285
     2    http://www.sonypictures.com/movies/spectre/  206647
     3             http://www.thedarkknightrises.com/   49026
     4          http://movies.disney.com/john-carter   49529

                                         keywords original_language  \
     0  [{"id": 1463, "name": "culture clash"}, {"id":…                en
     1  [{"id": 270, "name": "ocean"}, {"id": 726, "na…                en
     2  [{"id": 470, "name": "spy"}, {"id": 818, "name…                en
     3  [{"id": 849, "name": "dc comics"}, {"id": 853,…                en
     4  [{"id": 818, "name": "based on novel"}, {"id":…                en

                            original_title  \
     0                                Avatar
     1  Pirates of the Caribbean: At World's End
     2                               Spectre
     3                 The Dark Knight Rises
     4                           John Carter

                                         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
2   A cryptic message from Bond's past sends him o…  107.376788
3   Following the death of District Attorney Harve…  112.312950
4   John Carter is a war-weary, former military ca…   43.926995

                             production_companies  \
0   [{"name": "Ingenious Film Partners", "id": 289…
1   [{"name": "Walt Disney Pictures", "id": 2}, {"…
2   [{"name": "Columbia Pictures", "id": 5}, {"nam…
3   [{"name": "Legendary Pictures", "id": 923}, {"…
4         [{"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
2   [{"iso_3166_1": "GB", "name": "United Kingdom"…   2015-10-26    880674609
3   [{"iso_3166_1": "US", "name": "United States o…   2012-07-16   1084939099
4   [{"iso_3166_1": "US", "name": "United States o…   2012-03-07    284139100

    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
2     148.0  [{"iso_639_1": "fr", "name": "Fran\u00e7ais"},…  Released
3     165.0          [{"iso_639_1": "en", "name": "English"}]  Released
4     132.0          [{"iso_639_1": "en", "name": "English"}]  Released

                             tagline  \
0                Enter the World of Pandora.
1   At the end of the world, the adventure begins.
2                        A Plan No One Escapes
3                             The Legend Ends
4            Lost in our world, found in another.

                             title  vote_average  vote_count
0                             Avatar           7.2       11800
1   Pirates of the Caribbean: At World's End      6.9        4500
2                            Spectre           6.3        4466
3                 The Dark Knight Rises          7.6        9106
4                        John Carter           6.1        2124
```

[4]: `credits.head()`

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

```
3     49026              The Dark Knight Rises
4     49529                       John Carter

                                      cast  \
0  [{"cast_id": 242, "character": "Jake Sully", "…
1  [{"cast_id": 4, "character": "Captain Jack Spa…
2  [{"cast_id": 1, "character": "James Bond", "cr…
3  [{"cast_id": 2, "character": "Bruce Wayne / Ba…
4  [{"cast_id": 5, "character": "John Carter", "c…

                                      crew
0  [{"credit_id": "52fe48009251416c750aca23", "de…
1  [{"credit_id": "52fe4232c3a36847f800b579", "de…
2  [{"credit_id": "54805967c3a36829b5002c41", "de…
3  [{"credit_id": "52fe4781c3a36847f81398c3", "de…
4  [{"credit_id": "52fe479ac3a36847f813eaa3", "de…
```

```python
movies = movies.merge(credits, on="title")         # merge data base on
  title column
```

```python
movies.head(1)
```

```
        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]

```
[7]: movies =␣
     ↪movies[["movie_id","title","overview","genres","keywords","cast","crew"]]   ␣
     ↪           # remove unneccesary columns
```

```
[8]: movies.head()
```

```
[8]:    movie_id                                      title  \
     0     19995                                      Avatar
     1       285  Pirates of the Caribbean: At World's End
     2    206647                                    Spectre
     3     49026                      The Dark Knight Rises
     4     49529                                John Carter

                                               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…
     3  Following the death of District Attorney Harve…
     4  John Carter is a war-weary, former military ca…

                                                 genres  \
     0  [{"id": 28, "name": "Action"}, {"id": 12, "nam…
     1  [{"id": 12, "name": "Adventure"}, {"id": 14, "…
     2  [{"id": 28, "name": "Action"}, {"id": 12, "nam…
     3  [{"id": 28, "name": "Action"}, {"id": 80, "nam…
     4  [{"id": 28, "name": "Action"}, {"id": 12, "nam…

                                               keywords  \
     0  [{"id": 1463, "name": "culture clash"}, {"id":…
     1  [{"id": 270, "name": "ocean"}, {"id": 726, "na…
     2  [{"id": 470, "name": "spy"}, {"id": 818, "name…
     3  [{"id": 849, "name": "dc comics"}, {"id": 853,…
     4  [{"id": 818, "name": "based on novel"}, {"id":…

                                                   cast  \
     0  [{"cast_id": 242, "character": "Jake Sully", "…
     1  [{"cast_id": 4, "character": "Captain Jack Spa…
     2  [{"cast_id": 1, "character": "James Bond", "cr…
     3  [{"cast_id": 2, "character": "Bruce Wayne / Ba…
     4  [{"cast_id": 5, "character": "John Carter", "c…

                                                   crew
     0  [{"credit_id": "52fe48009251416c750aca23", "de…
```

```
1  [{"credit_id": "52fe4232c3a36847f800b579", "de…
2  [{"credit_id": "54805967c3a36829b5002c41", "de…
3  [{"credit_id": "52fe4781c3a36847f81398c3", "de…
4  [{"credit_id": "52fe479ac3a36847f813eaa3", "de…
```

[9]: `movies.isnull().sum()`   *# check missing data*

[9]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

[10]: `movies.dropna(inplace=True)`   *# dlt missing row data*

[11]: `movies.duplicated().sum()`   *# for checking duplicate data*

[11]: `np.int64(0)`

[12]: `movies.iloc[0].genres`

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

[13]:
```python
import ast
def convert(obj):
    L=[]
    for i in ast.literal_eval(obj):
        L.append(i["name"])
    return L
```

[14]: `movies["genres"] = movies["genres"].apply(convert)`

[15]: `movies["genres"]`

[15]:
```
0            [Action, Adventure, Fantasy, Science Fiction]
1                         [Adventure, Fantasy, Action]
2                        [Action, Adventure, Crime]
3                   [Action, Crime, Drama, Thriller]
4              [Action, Adventure, Science Fiction]
                             …
4804                     [Action, Crime, Thriller]
4805                            [Comedy, Romance]
4806            [Comedy, Drama, Romance, TV Movie]
```

```
4807                                                        []
4808                                       [Documentary]
Name: genres, Length: 4806, dtype: object
```

[16]: `movies.head()`

[16]:
```
    movie_id                                      title  \
0      19995                                      Avatar
1        285  Pirates of the Caribbean: At World's End
2     206647                                     Spectre
3      49026                    The Dark Knight Rises
4      49529                                 John Carter

                                         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…
3  Following the death of District Attorney Harve…
4  John Carter is a war-weary, former military ca…

                                            genres  \
0  [Action, Adventure, Fantasy, Science Fiction]
1                    [Adventure, Fantasy, Action]
2                      [Action, Adventure, Crime]
3              [Action, Crime, Drama, Thriller]
4           [Action, Adventure, Science Fiction]

                                          keywords  \
0  [{"id": 1463, "name": "culture clash"}, {"id":…
1  [{"id": 270, "name": "ocean"}, {"id": 726, "na…
2  [{"id": 470, "name": "spy"}, {"id": 818, "name…
3  [{"id": 849, "name": "dc comics"}, {"id": 853,…
4  [{"id": 818, "name": "based on novel"}, {"id":…

                                              cast  \
0  [{"cast_id": 242, "character": "Jake Sully", "…
1  [{"cast_id": 4, "character": "Captain Jack Spa…
2  [{"cast_id": 1, "character": "James Bond", "cr…
3  [{"cast_id": 2, "character": "Bruce Wayne / Ba…
4  [{"cast_id": 5, "character": "John Carter", "c…

                                              crew
0  [{"credit_id": "52fe48009251416c750aca23", "de…
1  [{"credit_id": "52fe4232c3a36847f800b579", "de…
2  [{"credit_id": "54805967c3a36829b5002c41", "de…
3  [{"credit_id": "52fe4781c3a36847f81398c3", "de…
4  [{"credit_id": "52fe479ac3a36847f813eaa3", "de…
```

```
[17]: movies["keywords"] = movies["keywords"].apply(convert)
```

```
[18]: movies["keywords"]
```

```
[18]: 0         [culture clash, future, space war, space colon…
      1         [ocean, drug abuse, exotic island, east india …
      2         [spy, based on novel, secret agent, sequel, mi…
      3         [dc comics, crime fighter, terrorist, secret i…
      4         [based on novel, mars, medallion, space travel…
                                      …
      4804      [united states-mexico barrier, legs, arms, pap…
      4805                                                    []
      4806      [date, love at first sight, narration, investi…
      4807                                                    []
      4808              [obsession, camcorder, crush, dream girl]
      Name: keywords, Length: 4806, dtype: object
```

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

```
[19]:    movie_id                                        title  \
      0     19995                                        Avatar
      1       285  Pirates of the Caribbean: At World's End
      2    206647                                       Spectre
      3     49026                         The Dark Knight Rises
      4     49529                                   John Carter

                                                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…
      3  Following the death of District Attorney Harve…
      4  John Carter is a war-weary, former military ca…

                                                  genres  \
      0  [Action, Adventure, Fantasy, Science Fiction]
      1                    [Adventure, Fantasy, Action]
      2                       [Action, Adventure, Crime]
      3                 [Action, Crime, Drama, Thriller]
      4           [Action, Adventure, Science Fiction]

                                                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…
      3  [dc comics, crime fighter, terrorist, secret i…
      4  [based on novel, mars, medallion, space travel…
```

```
                                                            cast  \
0  [{"cast_id": 242, "character": "Jake Sully", "…
1  [{"cast_id": 4, "character": "Captain Jack Spa…
2  [{"cast_id": 1, "character": "James Bond", "cr…
3  [{"cast_id": 2, "character": "Bruce Wayne / Ba…
4  [{"cast_id": 5, "character": "John Carter", "c…


                                                            crew
0  [{"credit_id": "52fe48009251416c750aca23", "de…
1  [{"credit_id": "52fe4232c3a36847f800b579", "de…
2  [{"credit_id": "54805967c3a36829b5002c41", "de…
3  [{"credit_id": "52fe4781c3a36847f81398c3", "de…
4  [{"credit_id": "52fe479ac3a36847f813eaa3", "de…
```

[20]:
```python
import ast
def convert3(obj):
    L=[]
    counter = 0                                              #␣
↪only run 3 times because only need 3 famous characters
    for i in ast.literal_eval(obj):
        if counter != 3:
            L.append(i["name"])
            counter = counter+1
        else:
            break
    return L
```

[21]:
```python
movies["cast"] = movies["cast"].apply(convert3)
```

[22]:
```python
movies["cast"]
```

[22]:
```
0          [Sam Worthington, Zoe Saldana, Sigourney Weaver]
1              [Johnny Depp, Orlando Bloom, Keira Knightley]
2               [Daniel Craig, Christoph Waltz, Léa Seydoux]
3              [Christian Bale, Michael Caine, Gary Oldman]
4          [Taylor Kitsch, Lynn Collins, Samantha Morton]
                              …
4804    [Carlos Gallardo, Jaime de Hoyos, Peter Marqua…
4805          [Edward Burns, Kerry Bishé, Marsha Dietlein]
4806            [Eric Mabius, Kristin Booth, Crystal Lowe]
4807             [Daniel Henney, Eliza Coupe, Bill Paxton]
4808    [Drew Barrymore, Brian Herzlinger, Corey Feldman]
Name: cast, Length: 4806, dtype: object
```

[23]:
```python
movies.head()
```

```
[23]:    movie_id                                          title  \
      0     19995                                          Avatar
      1       285  Pirates of the Caribbean: At World's End
      2    206647                                         Spectre
      3     49026                           The Dark Knight Rises
      4     49529                                     John Carter

                                                    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…
      3  Following the death of District Attorney Harve…
      4  John Carter is a war-weary, former military ca…

                                                      genres  \
      0  [Action, Adventure, Fantasy, Science Fiction]
      1                    [Adventure, Fantasy, Action]
      2                       [Action, Adventure, Crime]
      3               [Action, Crime, Drama, Thriller]
      4          [Action, Adventure, Science Fiction]

                                                    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…
      3  [dc comics, crime fighter, terrorist, secret i…
      4  [based on novel, mars, medallion, space travel…

                                                        cast  \
      0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]
      1     [Johnny Depp, Orlando Bloom, Keira Knightley]
      2       [Daniel Craig, Christoph Waltz, Léa Seydoux]
      3       [Christian Bale, Michael Caine, Gary Oldman]
      4    [Taylor Kitsch, Lynn Collins, Samantha Morton]

                                                        crew
      0  [{"credit_id": "52fe48009251416c750aca23", "de…
      1  [{"credit_id": "52fe4232c3a36847f800b579", "de…
      2  [{"credit_id": "54805967c3a36829b5002c41", "de…
      3  [{"credit_id": "52fe4781c3a36847f81398c3", "de…
      4  [{"credit_id": "52fe479ac3a36847f813eaa3", "de…
```

```python
[24]: import ast
      def fetch_director(obj):
          L=[]
          for i in ast.literal_eval(obj):
              if i["job"]=="Director":
```

```
            L.append(i["name"])
            break
    return L
```

[25]: `movies["crew"] = movies["crew"].apply(fetch_director)`

[26]: `movies["crew"]`

[26]:
```
0                 [James Cameron]
1                [Gore Verbinski]
2                   [Sam Mendes]
3            [Christopher Nolan]
4               [Andrew Stanton]
                    …
4804          [Robert Rodriguez]
4805              [Edward Burns]
4806               [Scott Smith]
4807                [Daniel Hsia]
4808          [Brian Herzlinger]
Name: crew, Length: 4806, dtype: object
```

[27]: `movies.head()`

[27]:
```
   movie_id                                    title  \
0     19995                                    Avatar
1       285  Pirates of the Caribbean: At World's End
2    206647                                   Spectre
3     49026                     The Dark Knight Rises
4     49529                               John Carter

                                            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…
3  Following the death of District Attorney Harve…
4  John Carter is a war-weary, former military ca…

                                          genres  \
0  [Action, Adventure, Fantasy, Science Fiction]
1                    [Adventure, Fantasy, Action]
2                      [Action, Adventure, Crime]
3              [Action, Crime, Drama, Thriller]
4          [Action, Adventure, Science Fiction]

                                         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…
3  [dc comics, crime fighter, terrorist, secret i…
4  [based on novel, mars, medallion, space travel…
```

```
                                          cast                crew
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]      [James Cameron]
1      [Johnny Depp, Orlando Bloom, Keira Knightley]     [Gore Verbinski]
2        [Daniel Craig, Christoph Waltz, Léa Seydoux]        [Sam Mendes]
3        [Christian Bale, Michael Caine, Gary Oldman]  [Christopher Nolan]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]     [Andrew Stanton]
```

```
[28]: movies["overview"] = movies["overview"].apply(lambda x:x.split())
      ↪           # overview column convert into list
```

```
[29]: movies.head()
```

```
[29]:    movie_id                                   title  \
      0     19995                                   Avatar
      1       285  Pirates of the Caribbean: At World's End
      2    206647                                   Spectre
      3     49026                    The Dark Knight Rises
      4     49529                              John Carter

                                            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…
      3  [Following, the, death, of, District, Attorney…
      4  [John, Carter, is, a, war-weary,, former, mili…

                                              genres  \
      0  [Action, Adventure, Fantasy, Science Fiction]
      1                    [Adventure, Fantasy, Action]
      2                     [Action, Adventure, Crime]
      3              [Action, Crime, Drama, Thriller]
      4         [Action, Adventure, Science Fiction]

                                            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…
      3  [dc comics, crime fighter, terrorist, secret i…
      4  [based on novel, mars, medallion, space travel…

                                              cast                crew
      0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]      [James Cameron]
      1      [Johnny Depp, Orlando Bloom, Keira Knightley]     [Gore Verbinski]
```

```
2      [Daniel Craig, Christoph Waltz, Léa Seydoux]          [Sam Mendes]
3       [Christian Bale, Michael Caine, Gary Oldman]   [Christopher Nolan]
4      [Taylor Kitsch, Lynn Collins, Samantha Morton]       [Andrew Stanton]
```

```python
[30]: movies["genres"] = movies["genres"].apply(lambda x:[i.replace(" ", "") for i in
      ↪x])                  # space replace for without space
      movies["keywords"] = movies["keywords"].apply(lambda x:[i.replace(" ", "") for
      ↪i in x])
      movies["cast"] = movies["cast"].apply(lambda x:[i.replace(" ", "") for i in x])
      movies["crew"] = movies["crew"].apply(lambda x:[i.replace(" ", "") for i in x])
```

```python
[31]: movies.head(3)
```

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

                                         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…

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

                                        keywords  \
      0  [cultureclash, future, spacewar, spacecolony, …
      1  [ocean, drugabuse, exoticisland, eastindiatrad…
      2  [spy, basedonnovel, secretagent, sequel, mi6, …

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

```python
[32]: movies["tags"] = movies["overview"] + movies["genres"] + movies["keywords"] +
      ↪movies["cast"] + movies["crew"]
```

```python
[33]: movies["tags"]
```

```
[33]: 0        [In, the, 22nd, century,, a, paraplegic, Marin…
      1        [Captain, Barbossa,, long, believed, to, be, d…
      2        [A, cryptic, message, from, Bond's, past, send…
      3        [Following, the, death, of, District, Attorney…
```

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

[34]: `movies.head(3)`

[34]:
```
   movie_id                                   title  \
0     19995                                  Avatar
1       285  Pirates of the Caribbean: At World's End
2    206647                                 Spectre

                                            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…

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

                                          keywords  \
0  [cultureclash, future, spacewar, spacecolony, …
1  [ocean, drugabuse, exoticisland, eastindiatrad…
2  [spy, basedonnovel, secretagent, sequel, mi6, …

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

                                         tags
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…
```

[35]: `new_df = movies[["movie_id","title","tags"]]`   # new data ready

[36]: `new_df`

[36]:
```
    movie_id                                   title  \
0      19995                                  Avatar
```

```
1          285  Pirates of the Caribbean: At World's End
2       206647                                   Spectre
3        49026                     The Dark Knight Rises
4        49529                               John Carter
...          ...                                       ...
4804      9367                               El Mariachi
4805     72766                                 Newlyweds
4806    231617                 Signed, Sealed, Delivered
4807    126186                          Shanghai Calling
4808     25975                          My Date with Drew

                                              tags
0      [In, the, 22nd, century,, a, paraplegic, Marin…
1      [Captain, Barbossa,, long, believed, to, be, d…
2      [A, cryptic, message, from, Bond's, past, send…
3      [Following, the, death, of, District, Attorney…
4      [John, Carter, is, a, war-weary,, former, mili…
...                                                 …
4804   [El, Mariachi, just, wants, to, play, his, gui…
4805   [A, newlywed, couple's, honeymoon, is, upended…
4806   ["Signed,, Sealed,, Delivered", introduces, a,…
4807   [When, ambitious, New, York, attorney, Sam, is…
4808   [Ever, since, the, second, grade, when, he, fi…

[4806 rows x 3 columns]
```

[37]:
```python
new_df["tags"] = new_df["tags"].apply(lambda x: " ".join(x))      # list data
 ↪covert into string, all words add thorugh space
```

```
C:\Users\DELL\AppData\Local\Temp\ipykernel_7552\2119562113.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df["tags"] = new_df["tags"].apply(lambda x: " ".join(x))        # list
data covert into string, all words add thorugh space
```

[38]:
```python
new_df.head(3)
```

[38]:
```
   movie_id                                    title  \
0     19995                                   Avatar
1       285  Pirates of the Caribbean: At World's End
2    206647                                   Spectre

                                              tags
```

```
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…
```

[39]:
```python
new_df["tags"] = new_df["tags"].apply(lambda x:x.lower())
```

```
C:\Users\DELL\AppData\Local\Temp\ipykernel_7552\3481706753.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df["tags"] = new_df["tags"].apply(lambda x:x.lower())
```

[40]:
```python
new_df.head(3)
```

[40]:
```
   movie_id                                    title  \
0     19995                                   Avatar
1       285  Pirates of the Caribbean: At World's End
2    206647                                   Spectre


                                                tags
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…
```

[41]:
```python
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features = 5000, stop_words="english")
```

[42]:
```python
vectors = cv.fit_transform(new_df["tags"]).toarray()                    #␣
 ↪convert text data into numberical form (array)
```

[43]:
```python
vectors[0]
```

[43]:
```
array([0, 0, 0, …, 0, 0, 0], shape=(5000,))
```

[44]:
```python
cv.get_feature_names_out()
```

[44]:
```
array(['000', '007', '10', …, 'zone', 'zoo', 'zooeydeschanel'],
      shape=(5000,), dtype=object)
```

[45]:
```python
import nltk
```

[46]:
```python
from nltk.stem.porter import PorterStemmer
```

```
ps = PorterStemmer()                                # To remove repeating␣
 ↪same type of words (eg = love, loves, loved) their meaning is same so remove␣
 ↪it.
```

[47]:
```
ps.stem("loving")                                   # convert this example into root word
```

[47]: 'love'

[48]:
```
def stem(text):
    y = []
    for i in text.split():
        y.append(ps.stem(i))

    return " ".join(y)
```

[49]:
```
new_df["tags"] = new_df["tags"].apply(stem)
```

C:\Users\DELL\AppData\Local\Temp\ipykernel_7552\825579503.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  new_df["tags"] = new_df["tags"].apply(stem)

[50]:
```
new_df
```

[50]:
```
      movie_id                                   title  \
0        19995                                  Avatar
1          285  Pirates of the Caribbean: At World's End
2       206647                                 Spectre
3        49026                   The Dark Knight Rises
4        49529                             John Carter
...        ...                                     ...
4804      9367                             El Mariachi
4805     72766                               Newlyweds
4806    231617                 Signed, Sealed, Delivered
4807    126186                         Shanghai Calling
4808     25975                         My Date with Drew

                                                   tags
0      in the 22nd century, a parapleg marin is dispa…
1      captain barbossa, long believ to be dead, ha c…
2      a cryptic messag from bond' past send him on a…
3      follow the death of district attorney harvey d…
4      john carter is a war-weary, former militari ca…
```

```
 …                                                     …
4804   el mariachi just want to play hi guitar and ca…
4805   a newlyw couple' honeymoon is upend by the arr…
4806   "signed, sealed, delivered" introduc a dedic q…
4807   when ambiti new york attorney sam is sent to s…
4808   ever sinc the second grade when he first saw h…

[4806 rows x 3 columns]
```

[51]:
```python
from sklearn.metrics.pairwise import cosine_similarity
```

[52]:
```python
similarity = cosine_similarity(vectors)                    # To find out which
  ↪movie is at what distance from which other movie.
```

[53]:
```python
similarity[0]                              # Distance from every movie of the 0th
  ↪movie
```

[53]:
```
array([1.        , 0.08980265, 0.05986843, …, 0.02457366, 0.02777778,
       0.        ], shape=(4806,))
```

[54]:
```python
new_df.head(2)
```

[54]:
```
   movie_id                              title  \
0     19995                             Avatar
1       285  Pirates of the Caribbean: At World's End

                                            tags
0  in the 22nd century, a parapleg marin is dispa…
1  captain barbossa, long believ to be dead, ha c…
```

[55]:
```python
def recommand(movie):
    movie_index = new_df[new_df["title"] == movie].index[0]          #
  ↪This will indicate the index number and 0 means accessing the first index
  ↪value.
    distances = similarity[movie_index]
    movie_list = sorted(list(enumerate(similarity[0])), reverse = True, key=
  ↪lambda x:x[1])[1:6]  # enumerate function will show index as well as value
  ↪and after index first show and value then convert x[1] to lambda so that
  ↪sort on the basis of index not value i.e. 2nd item

    for i in movie_list:
        print(new_df.iloc[i[0]].title)
```

[56]:
```python
recommand("Avatar")
```

```
Titan A.E.
Small Soldiers
```

```
Ender's Game
Independence Day
Aliens vs Predator: Requiem
```

[57]: ```python
import pickle
```

[58]: ```python
pickle.dump(new_df.to_dict(), open("movie_dict.pkl","wb"))
        # Convert the data to dictationary, the file will be created
```

[59]: ```python
pickle.dump(similarity, open("similarity.pkl","wb"))
```