recommendersystem

November 5, 2023

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

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

[3]:
```
movies.head(1)
```

[3]:
```
      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  \
0  150.437577  [{"name": "Ingenious Film Partners", "id": 289…

                               production_countries release_date     revenue  \
0  [{"iso_3166_1": "US", "name": "United States o…   2009-12-10  2787965087

   runtime                                     spoken_languages    status  \
0    162.0  [{"iso_639_1": "en", "name": "English"}, {"iso…  Released

                      tagline   title  vote_average  vote_count
0  Enter the World of Pandora.  Avatar           7.2       11800
```

[4]:
```
movies.columns
```

[4]:
```
Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
       'original_title', 'overview', 'popularity', 'production_companies',
       'production_countries', 'release_date', 'revenue', 'runtime',
```

```
       'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
       'vote_count'],
      dtype='object')
```

[5]:
```
credits.head(1)
```

[5]:
```
   movie_id   title                                                cast  \
0     19995  Avatar  [{"cast_id": 242, "character": "Jake Sully", "…

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

[6]:
```
credits.head(1)['cast']
```

[6]:
```
0    [{"cast_id": 242, "character": "Jake Sully", "…
Name: cast, dtype: object
```

[7]:
```
credits.columns
```

[7]:
```
Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')
```

[8]:
```
movies.shape
```

[8]:
```
(4803, 20)
```

[9]:
```
credits.shape
```

[9]:
```
(4803, 4)
```

[10]:
```
movies=movies.merge(credits,on='title')
```

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

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

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

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

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

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

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

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

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

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

[1 rows x 23 columns]
```

[12]:
```
movies =␣
↪movies[['movie_id','title','overview','genres','keywords','cast','crew']]
```

[13]:
```
movies.head(1)
```

[13]:
```
   movie_id   title                                            overview  \
0     19995  Avatar  In the 22nd century, a paraplegic Marine is di…

                                               genres  \
0  [{"id": 28, "name": "Action"}, {"id": 12, "nam…

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

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

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

[14]:
```
movies.isnull().sum()
```

[14]:
```
movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

[15]:
```
#3 records with no overview

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

[15]:
```
movie_id    0
title       0
overview    0
genres      0
keywords    0
cast        0
crew        0
dtype: int64
```

[16]:
```
movies.duplicated().sum()
```

[16]:
```
0
```

[17]:
```
movies.iloc[0].genres  #gives genres of 0th index movie
```

[17]:
```
'[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14,
"name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
```

[18]:
```
#returns a list of all the values associated with the key "name"

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

[19]:
```
#convert('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id":
 ↪ 14, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]')
movies['genres'] = movies['genres'].apply(convert)
```

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

[20]:
```
   movie_id   title                                                overview  \
0     19995  Avatar  In the 22nd century, a paraplegic Marine is di…

                                        genres  \
0  [Action, Adventure, Fantasy, Science Fiction]

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

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

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

[21]:
```
movies['keywords'] = movies['keywords'].apply(convert)
```

[22]:
```
movies.head(1)
```

[22]:
```
   movie_id   title                                              overview  \
0     19995  Avatar  In the 22nd century, a paraplegic Marine is di…

                                        genres  \
0  [Action, Adventure, Fantasy, Science Fiction]

                                      keywords  \
0  [culture clash, future, space war, space colon…

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

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

[23]:
```
movies['cast'][0]
```

[23]:
```
'[{"cast_id": 242, "character": "Jake Sully", "credit_id":
"5602a8a7c3a3685532001c9a", "gender": 2, "id": 65731, "name": "Sam Worthington",
"order": 0}, {"cast_id": 3, "character": "Neytiri", "credit_id":
"52fe48009251416c750ac9cb", "gender": 1, "id": 8691, "name": "Zoe Saldana",
"order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustine", "credit_id":
"52fe48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigourney
Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch", "credit_id":
"52fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang",
"order": 3}, {"cast_id": 5, "character": "Trudy Chacon", "credit_id":
"52fe48009251416c750ac9d3", "gender": 1, "id": 17647, "name": "Michelle
Rodriguez", "order": 4}, {"cast_id": 8, "character": "Selfridge", "credit_id":
"52fe48009251416c750ac9e1", "gender": 2, "id": 1771, "name": "Giovanni Ribisi",
"order": 5}, {"cast_id": 7, "character": "Norm Spellman", "credit_id":
"52fe48009251416c750ac9dd", "gender": 2, "id": 59231, "name": "Joel David
Moore", "order": 6}, {"cast_id": 9, "character": "Moat", "credit_id":
"52fe48009251416c750ac9e5", "gender": 1, "id": 30485, "name": "CCH Pounder",
"order": 7}, {"cast_id": 11, "character": "Eytukan", "credit_id":
"52fe48009251416c750ac9ed", "gender": 2, "id": 15853, "name": "Wes Studi",
"order": 8}, {"cast_id": 10, "character": "Tsu\'Tey", "credit_id":
"52fe48009251416c750ac9e9", "gender": 2, "id": 10964, "name": "Laz Alonso",
```

"order": 9}, {"cast_id": 12, "character": "Dr. Max Patel", "credit_id":
"52fe48009251416c750ac9f1", "gender": 2, "id": 95697, "name": "Dileep Rao",
"order": 10}, {"cast_id": 13, "character": "Lyle Wainfleet", "credit_id":
"52fe48009251416c750ac9f5", "gender": 2, "id": 98215, "name": "Matt Gerald",
"order": 11}, {"cast_id": 32, "character": "Private Fike", "credit_id":
"52fe48009251416c750aca5b", "gender": 2, "id": 154153, "name": "Sean Anthony
Moran", "order": 12}, {"cast_id": 33, "character": "Cryo Vault Med Tech",
"credit_id": "52fe48009251416c750aca5f", "gender": 2, "id": 397312, "name":
"Jason Whyte", "order": 13}, {"cast_id": 34, "character": "Venture Star Crew
Chief", "credit_id": "52fe48009251416c750aca63", "gender": 2, "id": 42317,
"name": "Scott Lawrence", "order": 14}, {"cast_id": 35, "character": "Lock Up
Trooper", "credit_id": "52fe48009251416c750aca67", "gender": 2, "id": 986734,
"name": "Kelly Kilgour", "order": 15}, {"cast_id": 36, "character": "Shuttle
Pilot", "credit_id": "52fe48009251416c750aca6b", "gender": 0, "id": 1207227,
"name": "James Patrick Pitt", "order": 16}, {"cast_id": 37, "character":
"Shuttle Co-Pilot", "credit_id": "52fe48009251416c750aca6f", "gender": 0, "id":
1180936, "name": "Sean Patrick Murphy", "order": 17}, {"cast_id": 38,
"character": "Shuttle Crew Chief", "credit_id": "52fe48009251416c750aca73",
"gender": 2, "id": 1019578, "name": "Peter Dillon", "order": 18}, {"cast_id":
39, "character": "Tractor Operator / Troupe", "credit_id":
"52fe48009251416c750aca77", "gender": 0, "id": 91443, "name": "Kevin Dorman",
"order": 19}, {"cast_id": 40, "character": "Dragon Gunship Pilot", "credit_id":
"52fe48009251416c750aca7b", "gender": 2, "id": 173391, "name": "Kelson
Henderson", "order": 20}, {"cast_id": 41, "character": "Dragon Gunship Gunner",
"credit_id": "52fe48009251416c750aca7f", "gender": 0, "id": 1207236, "name":
"David Van Horn", "order": 21}, {"cast_id": 42, "character": "Dragon Gunship
Navigator", "credit_id": "52fe48009251416c750aca83", "gender": 0, "id": 215913,
"name": "Jacob Tomuri", "order": 22}, {"cast_id": 43, "character": "Suit #1",
"credit_id": "52fe48009251416c750aca87", "gender": 0, "id": 143206, "name":
"Michael Blain-Rozgay", "order": 23}, {"cast_id": 44, "character": "Suit #2",
"credit_id": "52fe48009251416c750aca8b", "gender": 2, "id": 169676, "name": "Jon
Curry", "order": 24}, {"cast_id": 46, "character": "Ambient Room Tech",
"credit_id": "52fe48009251416c750aca8f", "gender": 0, "id": 1048610, "name":
"Luke Hawker", "order": 25}, {"cast_id": 47, "character": "Ambient Room Tech /
Troupe", "credit_id": "52fe48009251416c750aca93", "gender": 0, "id": 42288,
"name": "Woody Schultz", "order": 26}, {"cast_id": 48, "character": "Horse Clan
Leader", "credit_id": "52fe48009251416c750aca97", "gender": 2, "id": 68278,
"name": "Peter Mensah", "order": 27}, {"cast_id": 49, "character": "Link Room
Tech", "credit_id": "52fe48009251416c750aca9b", "gender": 0, "id": 1207247,
"name": "Sonia Yee", "order": 28}, {"cast_id": 50, "character": "Basketball
Avatar / Troupe", "credit_id": "52fe48009251416c750aca9f", "gender": 1, "id":
1207248, "name": "Jahnel Curfman", "order": 29}, {"cast_id": 51, "character":
"Basketball Avatar", "credit_id": "52fe48009251416c750acaa3", "gender": 0, "id":
89714, "name": "Ilram Choi", "order": 30}, {"cast_id": 52, "character": "Na\'vi
Child", "credit_id": "52fe48009251416c750acaa7", "gender": 0, "id": 1207249,
"name": "Kyla Warren", "order": 31}, {"cast_id": 53, "character": "Troupe",
"credit_id": "52fe48009251416c750acaab", "gender": 0, "id": 1207250, "name":

"Lisa Roumain", "order": 32}, {"cast_id": 54, "character": "Troupe",
"credit_id": "52fe48009251416c750acaaf", "gender": 1, "id": 83105, "name":
"Debra Wilson", "order": 33}, {"cast_id": 57, "character": "Troupe",
"credit_id": "52fe48009251416c750acabb", "gender": 0, "id": 1207253, "name":
"Chris Mala", "order": 34}, {"cast_id": 55, "character": "Troupe", "credit_id":
"52fe48009251416c750acab3", "gender": 0, "id": 1207251, "name": "Taylor Kibby",
"order": 35}, {"cast_id": 56, "character": "Troupe", "credit_id":
"52fe48009251416c750acab7", "gender": 0, "id": 1207252, "name": "Jodie Landau",
"order": 36}, {"cast_id": 58, "character": "Troupe", "credit_id":
"52fe48009251416c750acabf", "gender": 0, "id": 1207254, "name": "Julie Lamm",
"order": 37}, {"cast_id": 59, "character": "Troupe", "credit_id":
"52fe48009251416c750acac3", "gender": 0, "id": 1207257, "name": "Cullen B.
Madden", "order": 38}, {"cast_id": 60, "character": "Troupe", "credit_id":
"52fe48009251416c750acac7", "gender": 0, "id": 1207259, "name": "Joseph Brady
Madden", "order": 39}, {"cast_id": 61, "character": "Troupe", "credit_id":
"52fe48009251416c750acacb", "gender": 0, "id": 1207262, "name": "Frankie
Torres", "order": 40}, {"cast_id": 62, "character": "Troupe", "credit_id":
"52fe48009251416c750acacf", "gender": 1, "id": 1158600, "name": "Austin Wilson",
"order": 41}, {"cast_id": 63, "character": "Troupe", "credit_id":
"52fe48019251416c750acad3", "gender": 1, "id": 983705, "name": "Sara Wilson",
"order": 42}, {"cast_id": 64, "character": "Troupe", "credit_id":
"52fe48019251416c750acad7", "gender": 0, "id": 1207263, "name": "Tamica
Washington-Miller", "order": 43}, {"cast_id": 65, "character": "Op Center
Staff", "credit_id": "52fe48019251416c750acadb", "gender": 1, "id": 1145098,
"name": "Lucy Briant", "order": 44}, {"cast_id": 66, "character": "Op Center
Staff", "credit_id": "52fe48019251416c750acadf", "gender": 2, "id": 33305,
"name": "Nathan Meister", "order": 45}, {"cast_id": 67, "character": "Op Center
Staff", "credit_id": "52fe48019251416c750acae3", "gender": 0, "id": 1207264,
"name": "Gerry Blair", "order": 46}, {"cast_id": 68, "character": "Op Center
Staff", "credit_id": "52fe48019251416c750acae7", "gender": 2, "id": 33311,
"name": "Matthew Chamberlain", "order": 47}, {"cast_id": 69, "character": "Op
Center Staff", "credit_id": "52fe48019251416c750acaeb", "gender": 0, "id":
1207265, "name": "Paul Yates", "order": 48}, {"cast_id": 70, "character": "Op
Center Duty Officer", "credit_id": "52fe48019251416c750acaef", "gender": 0,
"id": 1207266, "name": "Wray Wilson", "order": 49}, {"cast_id": 71, "character":
"Op Center Staff", "credit_id": "52fe48019251416c750acaf3", "gender": 2, "id":
54492, "name": "James Gaylyn", "order": 50}, {"cast_id": 72, "character":
"Dancer", "credit_id": "52fe48019251416c750acaf7", "gender": 0, "id": 1207267,
"name": "Melvin Leno Clark III", "order": 51}, {"cast_id": 73, "character":
"Dancer", "credit_id": "52fe48019251416c750acafb", "gender": 0, "id": 1207268,
"name": "Carvon Futrell", "order": 52}, {"cast_id": 74, "character": "Dancer",
"credit_id": "52fe48019251416c750acaff", "gender": 0, "id": 1207269, "name":
"Brandon Jelkes", "order": 53}, {"cast_id": 75, "character": "Dancer",
"credit_id": "52fe48019251416c750acb03", "gender": 0, "id": 1207270, "name":
"Micah Moch", "order": 54}, {"cast_id": 76, "character": "Dancer", "credit_id":
"52fe48019251416c750acb07", "gender": 0, "id": 1207271, "name": "Hanniyah
Muhammad", "order": 55}, {"cast_id": 77, "character": "Dancer", "credit_id":

"52fe48019251416c750acb0b", "gender": 0, "id": 1207272, "name": "Christopher Nolen", "order": 56}, {"cast_id": 78, "character": "Dancer", "credit_id": "52fe48019251416c750acb0f", "gender": 0, "id": 1207273, "name": "Christa Oliver", "order": 57}, {"cast_id": 79, "character": "Dancer", "credit_id": "52fe48019251416c750acb13", "gender": 0, "id": 1207274, "name": "April Marie Thomas", "order": 58}, {"cast_id": 80, "character": "Dancer", "credit_id": "52fe48019251416c750acb17", "gender": 0, "id": 1207275, "name": "Bravita A. Threatt", "order": 59}, {"cast_id": 81, "character": "Mining Chief (uncredited)", "credit_id": "52fe48019251416c750acb1b", "gender": 0, "id": 1207276, "name": "Colin Bleasdale", "order": 60}, {"cast_id": 82, "character": "Veteran Miner (uncredited)", "credit_id": "52fe48019251416c750acb1f", "gender": 0, "id": 107969, "name": "Mike Bodnar", "order": 61}, {"cast_id": 83, "character": "Richard (uncredited)", "credit_id": "52fe48019251416c750acb23", "gender": 0, "id": 1207278, "name": "Matt Clayton", "order": 62}, {"cast_id": 84, "character": "Nav\'i (uncredited)", "credit_id": "52fe48019251416c750acb27", "gender": 1, "id": 147898, "name": "Nicole Dionne", "order": 63}, {"cast_id": 85, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2b", "gender": 0, "id": 1207280, "name": "Jamie Harrison", "order": 64}, {"cast_id": 86, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb2f", "gender": 0, "id": 1207281, "name": "Allan Henry", "order": 65}, {"cast_id": 87, "character": "Ground Technician (uncredited)", "credit_id": "52fe48019251416c750acb33", "gender": 2, "id": 1207282, "name": "Anthony Ingruber", "order": 66}, {"cast_id": 88, "character": "Flight Crew Mechanic (uncredited)", "credit_id": "52fe48019251416c750acb37", "gender": 0, "id": 1207283, "name": "Ashley Jeffery", "order": 67}, {"cast_id": 14, "character": "Samson Pilot", "credit_id": "52fe48009251416c750ac9f9", "gender": 0, "id": 98216, "name": "Dean Knowsley", "order": 68}, {"cast_id": 89, "character": "Trooper (uncredited)", "credit_id": "52fe48019251416c750acb3b", "gender": 0, "id": 1201399, "name": "Joseph Mika-Hunt", "order": 69}, {"cast_id": 90, "character": "Banshee (uncredited)", "credit_id": "52fe48019251416c750acb3f", "gender": 0, "id": 236696, "name": "Terry Notary", "order": 70}, {"cast_id": 91, "character": "Soldier (uncredited)", "credit_id": "52fe48019251416c750acb43", "gender": 0, "id": 1207287, "name": "Kai Pantano", "order": 71}, {"cast_id": 92, "character": "Blast Technician (uncredited)", "credit_id": "52fe48019251416c750acb47", "gender": 0, "id": 1207288, "name": "Logan Pithyou", "order": 72}, {"cast_id": 93, "character": "Vindum Raah (uncredited)", "credit_id": "52fe48019251416c750acb4b", "gender": 0, "id": 1207289, "name": "Stuart Pollock", "order": 73}, {"cast_id": 94, "character": "Hero (uncredited)", "credit_id": "52fe48019251416c750acb4f", "gender": 0, "id": 584868, "name": "Raja", "order": 74}, {"cast_id": 95, "character": "Ops Centreworker (uncredited)", "credit_id": "52fe48019251416c750acb53", "gender": 0, "id": 1207290, "name": "Gareth Ruck", "order": 75}, {"cast_id": 96, "character": "Engineer (uncredited)", "credit_id": "52fe48019251416c750acb57", "gender": 0, "id": 1062463, "name": "Rhian Sheehan", "order": 76}, {"cast_id": 97, "character": "Col. Quaritch\'s Mech Suit (uncredited)", "credit_id": "52fe48019251416c750acb5b", "gender": 0, "id": 60656, "name": "T. J. Storm", "order": 77}, {"cast_id": 98, "character": "Female Marine (uncredited)",

```
"credit_id": "52fe48019251416c750acb5f", "gender": 0, "id": 1207291, "name":
"Jodie Taylor", "order": 78}, {"cast_id": 99, "character": "Ikran Clan Leader
(uncredited)", "credit_id": "52fe48019251416c750acb63", "gender": 1, "id":
1186027, "name": "Alicia Vela-Bailey", "order": 79}, {"cast_id": 100,
"character": "Geologist (uncredited)", "credit_id": "52fe48019251416c750acb67",
"gender": 0, "id": 1207292, "name": "Richard Whiteside", "order": 80},
{"cast_id": 101, "character": "Na\'vi (uncredited)", "credit_id":
"52fe48019251416c750acb6b", "gender": 0, "id": 103259, "name": "Nikie Zambo",
"order": 81}, {"cast_id": 102, "character": "Ambient Room Tech / Troupe",
"credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name":
"Julene Renee", "order": 82}]'
```

[24]:
```
#only return top 3 actors relevant to the recommendation system
import ast
def convert3(obj):
    actors = ast.literal_eval(obj)
    l = []

    for i in actors[:3]:
        l.append(i['name'])

    return l
```

[25]:
```
movies['cast']=movies['cast'].apply(convert3)
```

[26]:
```
movies['cast'][0]
```

[26]:
```
['Sam Worthington', 'Zoe Saldana', 'Sigourney Weaver']
```

[27]:
```
movies.head(1)
```

[27]:
```
   movie_id   title                                                overview  \
0     19995  Avatar  In the 22nd century, a paraplegic Marine is di…

                                        genres  \
0  [Action, Adventure, Fantasy, Science Fiction]

                                          keywords  \
0  [culture clash, future, space war, space colon…

                                             cast  \
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]

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

[28]:
```
#to extract just the director name from the crew
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
```

[29]:
```
movies['crew']=movies['crew'].apply(fetch_director)
```

[30]:
```
movies.head(1)
```

[30]:
```
   movie_id   title                                                overview  \
0     19995  Avatar  In the 22nd century, a paraplegic Marine is di…

                                        genres  \
0  [Action, Adventure, Fantasy, Science Fiction]

                                          keywords  \
0  [culture clash, future, space war, space colon…

                                           cast              crew
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]  [James Cameron]
```

[31]:
```
movies['overview']=movies['overview'].apply(lambda x:x.split()) #overview is␣
 ↪converted from string to list of words
```

[32]:
```
movies.head(1)
```

[32]:
```
   movie_id   title                                                overview  \
0     19995  Avatar  [In, the, 22nd, century,, a, paraplegic, Marin…

                                        genres  \
0  [Action, Adventure, Fantasy, Science Fiction]

                                          keywords  \
0  [culture clash, future, space war, space colon…

                                           cast              crew
0  [Sam Worthington, Zoe Saldana, Sigourney Weaver]  [James Cameron]
```

[33]:
```
movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ","")for i in x])
movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ","")for i in␣
 ↪x])
movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ","")for i in x])
```

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

[34]:
```
movies.head(1)
```

[34]:
```
   movie_id   title                                                overview  \
0     19995  Avatar  [In, the, 22nd, century,, a, paraplegic, Marin…

                                        genres  \
0  [Action, Adventure, Fantasy, ScienceFiction]

                                          keywords  \
0  [cultureclash, future, spacewar, spacecolony, …

                                          cast            crew
0  [SamWorthington, ZoeSaldana, SigourneyWeaver]  [JamesCameron]
```

[35]:
```
movies['tags']=movies['overview']+movies['keywords']+movies['cast']+movies['crew']␣
 ↪#a singular string for all the required details for the movies
```

[36]:
```
new_df=movies[['movie_id','title','tags']]
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]:
```
new_df['tags']=new_df['tags'].apply(lambda x:" ".join(x)) #converts list to␣
↪string by joining each list with a space
```

```
/var/folders/tn/5qwnwx_j20b_5qhwx5vg07840000gn/T/ipykernel_27645/232006812.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)) #converts list to
string by joining each list with a space
```

[38]:
```
new_df
```

[38]:

	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 Marine is di…
1	Captain Barbossa, long believed to be dead, ha…
2	A cryptic message from Bond’s past sends him o…
3	Following the death of District Attorney Harve…
4	John Carter is a war-weary, former military ca…
…	…
4804	El Mariachi just wants to play his guitar and …
4805	A newlywed couple's honeymoon is upended by th…
4806	"Signed, Sealed, Delivered" introduces a dedic…
4807	When ambitious New York attorney Sam is sent t…
4808	Ever since the second grade when he first saw …

```
[4806 rows x 3 columns]
```

[39]:
```
new_df['tags']=new_df['tags'].apply(lambda x:x.lower())
new_df[0:1]
```

```
/var/folders/tn/5qwnwx_j20b_5qhwx5vg07840000gn/T/ipykernel_27645/2588369410.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())
```

[39]:

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

[40]:
```
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')
```

[41]:
```
vectors= cv.fit_transform(new_df['tags']).toarray()
```

[42]:
```
vectors.shape #transfomed vectors (ie 5000 words for each movie)
```

[42]:
```
(4806, 5000)
```

[43]:
```
vectors
```

[43]:
```
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]])
```

[44]:
```
import nltk
from nltk.stem.porter import *
ps=PorterStemmer()
```

[45]:
```
def stem(text):
  y=[]
  for i in text.split():
    y.append(ps.stem(i))
  return " ".join(y) #returns string after stemming
```

[46]:
```
new_df['tags']= new_df['tags'].apply(stem)
```

```
/var/folders/tn/5qwnwx_j20b_5qhwx5vg07840000gn/T/ipykernel_27645/3021978705.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)
```

[47]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

[48]:
```
similarity = cosine_similarity(vectors) #calculates the distance btw each␣
↪vector with another vector
```

[49]:
```
cosine_similarity(vectors)# for all diagonal elements means movie wrt same␣
↪movie thus angle 0 thus cos0=1
```

[49]:
```
array([[1.        , 0.        , 0.        , …, 0.        , 0.02752409,
        0.        ],
       [0.        , 1.        , 0.        , …, 0.02865341, 0.        ,
        0.        ],
       [0.        , 0.        , 1.        , …, 0.02865341, 0.        ,
        0.        ],
       …,
       [0.        , 0.02865341, 0.02865341, …, 1.        , 0.048795  ,
        0.05006262],
       [0.02752409, 0.        , 0.        , …, 0.048795  , 1.        ,
        0.05129892],
       [0.        , 0.        , 0.        , …, 0.05006262, 0.05129892,
        1.        ]])
```

[50]:
```
new_df[new_df['title'] =='Avatar'].index[0]
```

[50]:
```
0
```

[51]:
```
list(enumerate(similarity[0]))[0:10]
#creates a list of tuples mentioning distance w.r.t index as well, only first␣
↪10 displayed here.
```

[51]:
```
[(0, 1.0000000000000002),
 (1, 0.0),
 (2, 0.0),
 (3, 0.021110016546037454),
 (4, 0.11295649894498103),
 (5, 0.05025189076296061),
 (6, 0.0),
 (7, 0.05647824947249051),
 (8, 0.0),
 (9, 0.0)]
```

[52]:
```
sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[0:10]
#sort in reverse order w.r.t 1st index, only first 10 displayed here.
```

[52]:
```
[(0, 1.0000000000000002),
 (3608, 0.21320071635561044),
 (1216, 0.20769510081357428),
 (1920, 0.2059714602177749),
 (582, 0.20533080093573816),
 (539, 0.20100756305184245),
 (2409, 0.20032733246124987),
 (507, 0.19500597976723483),
 (74, 0.18582615562066462),
 (3538, 0.17654696590094993)]
```

[53]:
```
def recommend(movie):
  movie_index=new_df[new_df['title'] ==movie].index[0]
  distances=similarity[movie_index]#gives a list of all distances for a movie
  movies_list=sorted(list(enumerate(distances)),reverse=True,key=lambda x:
  ↪x[1])[1:6]#since we want top 5 recommendations

  #movies_list will return tuple with index of movie and dist corres
  for i in movies_list:
    print(new_df.iloc[i[0]].title)
    #print(i[0])
```

[54]:
```
recommend("Pirates of the Caribbean: At World's End")
```

```
Pirates of the Caribbean: Dead Man's Chest
Pirates of the Caribbean: The Curse of the Black Pearl
20,000 Leagues Under the Sea
Pirates of the Caribbean: On Stranger Tides
The Pirates! In an Adventure with Scientists!
```

[55]:
```
new_df.iloc[1149].title
```

[55]:
```
'Man on a Ledge'
```

[56]:
```
import pickle
```

[57]:
```
pickle.dump(new_df.to_dict(),open('movie_dict.pkl','wb'))#transferring as dict
```

[58]:
```
pickle.dump(similarity,open('similarity.pkl','wb'))
```