```
In [1]:
           import numpy as np
           import pandas as pd
In [2]:
           movies=pd.read_csv("tmdb_5000_movies.csv")
           credits=pd.read_csv("tmdb_5000_credits.csv")
In [3]:
           movies=movies.merge(credits,on="title")
In [4]:
           movies.shape
Out[4]:
           (4809, 23)
           movies = movies[['movie_id','title','overview','genres','keywords','cast',
In [5]:
In [6]:
           movies.head()
Out[6]:
                movie id
                                  title
                                         overview
                                                          genres
                                                                   keywords
                                                                                       cast
                                             In the
                                                                         [{"id":
                                                                                 [{"cast_id":
                                                         [{"id": 28.
                                              22nd
                                                                        1463,
                                                          "name":
                                                                                       242,
                                                                                                                 [{"cr
                                         century, a
                                                                      "name":
            0
                   19995
                                                                                "character":
                                                                                              "52fe48009251416c75(
                                Avatar
                                                         "Action"},
                                                                       "culture
                                        paraplegic
                                                         {"id": 12,
                                                                                      "Jake
                                         Marine is
                                                                       clash"},
                                                           "nam...
                                                                                  Sully", "...
                                                                       {"id":...
                                           Captain
                                                                    [{"id": 270,
                                                                                 [{"cast id":
                             Pirates of
                                        Barbossa,
                                                         [{"id": 12,
                                   the
                                                                      "name":
                                                                                                                 [{"cr
                                              long
                                                          "name":
            1
                           Caribbean:
                                                                     "ocean"},
                                                                                "character":
                                                                                               "52fe4232c3a36847f8(
                                          believed
                                                     "Adventure"},
                            At World's
                                             to be
                                                                     {"id": 726,
                                                                                   "Captain
                                                      {"id": 14, "...
                                                                         "na...
                                                                                 Jack Spa...
                                  Fnd
                                             dead,
                                              ha...
                                          A cryptic
                                                                                 [{"cast_id":
                                                                    [{"id": 470,
                                                         [{"id": 28,
                                          message
                                                                                         1,
                                              from
                                                          "name":
                                                                      "name":
                                                                                                                 [{"cr
                                                                                "character":
                                                                                              "54805967c3a36829b50
                                                         "Action"},
            2
                  206647
                               Spectre
                                            Bond's
                                                                       "spy"},
                                                                                    "James
                                                                    {"id": 818,
                                                         {"id": 12,
                                              past
                                                                                     Bond",
                                        sends him
                                                           "nam...
                                                                      "name...
                                                                                       "cr...
                                               0...
                                                                    [{"id": 849,
                                                                                 [{"cast id":
                                                         [{"id": 28,
                                         Following
                                                                       "name":
                             The Dark
                                                          "name":
                                                                                                                 [{"cr
                                         the death
                                                                           "dc
                                                                                "character":
                   49026
            3
                                Knight
                                         of District
                                                         "Action"},
                                                                                               "52fe4781c3a36847f81
                                                                                     "Bruce
                                                                     comics"},
                                 Rises
                                          Attorney
                                                         {"id": 80,
                                                                         {"id":
                                                                                   Wayne /
                                           Harve...
                                                           "nam...
                                                                        853,...
                                                                                       Ba...
                                              John
                                                                                 [{"cast_id":
                                                                    [{"id": 818,
                                                         [{"id": 28,
                                        Carter is a
                                                                                         5,
                                                          "name":
                                                                      "name":
                                                                                                                 [{"cr
                                              war-
                                 John
                                                                                "character":
                   49529
                                            weary,
                                                         "Action"},
                                                                    "based on
                                                                                               "52fe479ac3a36847f81
                                                                                      "John
                                Carter
                                                         {"id": 12,
                                            former
                                                                      novel"},
                                                                                    Carter",
                                           military
                                                           "nam...
                                                                       {"id":...
                                                                                        "c...
                                              ca...
```

```
In [7]: movies.isnull().sum()
 Out[7]: movie_id
         title
                     0
         overview
                     3
                     0
         genres
         keywords
                     0
         cast
                     0
         crew
                     0
         dtype: int64
In [8]: #remove's all null value rows
         movies.dropna(inplace=True)
In [9]: |movies.shape
Out[9]: (4806, 7)
In [10]: movies.duplicated().sum()
Out[10]: 0
In [11]: movies.iloc[0].genres
Out[11]: '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 1
         4, "name": "Fantasy"}, {"id": 878, "name": "Science Fiction"}]'
In [12]: import ast
In [13]: ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Advent
Out[13]: [{'id': 28, 'name': 'Action'},
          {'id': 12, 'name': 'Adventure'},
          {'id': 14, 'name': 'Fantasy'},
          {'id': 878, 'name': 'Science Fiction'}]
In [14]: | def convert(text):
             L = []
             for i in ast.literal_eval(text):
                 L.append(i['name'])
             return L
```

```
In [15]: movies['genres']=movies['genres'].apply(convert)
movies.head()
```

Out[15]:	mov	ie_id	title	overview	genres	keywords	cast		
	0 1	9995	Avatar	In the 22nd century, a paraplegic Marine is di	[Action, Adventure, Fantasy, Science Fiction]	[{"id": 1463, "name": "culture clash"}, {"id":	[{"cast_id": 242, "character": "Jake Sully", "	[{"cre "52fe48009251416c750a	
	1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[Adventure, Fantasy, Action]	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[{"cast_id": 4, "character": "Captain Jack Spa	[{"cre "52fe4232c3a36847f800	
	2 20	6647	Spectre	A cryptic message from Bond's past sends him o	[Action, Adventure, Crime]	[{"id": 470, "name": "spy"}, {"id": 818, "name	[{"cast_id": 1, "character": "James Bond", "cr	[{"cre "54805967c3a36829b500	
	3 4	9026	The Dark Knight Rises	Following the death of District Attorney Harve	[Action, Crime, Drama, Thriller]	[{"id": 849, "name": "dc comics"}, {"id": 853,	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"cre "52fe4781c3a36847f813	
	4 4	9529	John Carter	John Carter is a war- weary, former military ca	[Action, Adventure, Science Fiction]	[{"id": 818, "name": "based on novel"}, {"id":	[{"cast_id": 5, "character": "John Carter", "c	[{"cre "52fe479ac3a36847f813	
	4)	
In [16]:	<pre>movies['keywords']=movies['keywords'].apply(convert)</pre>								
In [17]:	<pre>def convert3(text): L = [] counter = 0 for i in ast.literal_eval(text): if counter < 3: L.append(i['name']) counter+=1 return L</pre>								
In [18]:	<pre>movies['cast']=movies['cast'].apply(convert3)</pre>								

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

Out[19]:	ı	movie_id	title	overview	genres	keywords	cast		
	0	19995	Avatar	In the 22nd century, a paraplegic Marine is di	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon	[Sam Worthington, Zoe Saldana, Sigourney Weaver]	[{"crı "52fe48009251416c750	
	1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india	[Johnny Depp, Orlando Bloom, Keira Knightley]	[{"crı "52fe4232c3a36847f80	
	2	206647	Spectre	A cryptic message from Bond's past sends him o	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi	[Daniel Craig, Christoph Waltz, Léa Seydoux]	[{"cre "54805967c3a36829b50	
	3	49026	The Dark Knight Rises	Following the death of District Attorney Harve	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i	[Christian Bale, Michael Caine, Gary Oldman]	[{"cro "52fe4781c3a36847f81	
	4	49529	John Carter	John Carter is a war- weary, former military ca	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel	[Taylor Kitsch, Lynn Collins, Samantha Morton]	[{"crı "52fe479ac3a36847f81	
	4)	
In [20]:	<pre>def fetch_director(text): L = [] for i in ast.literal_eval(text): if i['job'] == 'Director': L.append(i['name']) return L</pre>								
In [21]:	<pre>movies['crew']=movies['crew'].apply(fetch_director)</pre>								

In [22]: movies.head()

Out[22]:

```
movie id
                                    title
                                            overview
                                                            genres
                                                                       keywords
                                                                                             cast
                                                                                                          crew
                                                                          [culture
                                          In the 22nd
                                                            [Action,
                                                                                            [Sam
                                                                           clash,
                                                         Adventure,
                                                                                     Worthington,
                                            century, a
                                                                           future,
                                                                                                        [James
             0
                    19995
                                                           Fantasy,
                                 Avatar
                                           paraplegic
                                                                                    Zoe Saldana,
                                                                       space war,
                                                                                                     Cameron]
                                                            Science
                                            Marine is
                                                                                       Sigourney
                                                                           space
                                                  di...
                                                            Fiction]
                                                                                         Weaver]
                                                                          colon...
                                              Captain
                              Pirates of
                                                                          [ocean,
                                           Barbossa,
                                                                                   [Johnny Depp,
                                    the
                                                         [Adventure,
                                                                      drug abuse,
                                                 long
                                                                                         Orlando
                                                                                                         [Gore
             1
                      285
                             Caribbean:
                                                           Fantasy,
                                                                           exotic
                                           believed to
                                                                                     Bloom, Keira
                                                                                                      Verbinski]
                              At World's
                                                             Action]
                                                                      island, east
                                             be dead,
                                                                                       Knightley]
                                    Fnd
                                                                          india ...
                                                 ha...
                                             A cryptic
                                                                      [spy, based
                                                                                    [Daniel Craig,
                                             message
                                                            [Action,
                                                                        on novel,
                                                                                        Christoph
                                                                                                          [Sam
             2
                  206647
                                Spectre
                                          from Bond's
                                                         Adventure,
                                                                           secret
                                                                                       Waltz, Léa
                                                                                                       Mendes]
                                                             Crime]
                                           past sends
                                                                           agent,
                                                                                        Seydoux]
                                              him o...
                                                                      sequel, mi...
                                            Following
                                                                       [dc comics,
                                                            [Action,
                                                                                        [Christian
                               The Dark
                                          the death of
                                                                            crime
                                                             Crime,
                                                                                    Bale, Michael
                                                                                                   [Christopher
             3
                    49026
                                 Knight
                                               District
                                                                           fighter,
                                                                                      Caine, Gary
                                                            Drama,
                                                                                                         Nolan]
                                  Rises
                                             Attorney
                                                                         terrorist,
                                                            Thriller]
                                                                                         Oldman]
                                             Harve...
                                                                        secret i...
                                          John Carter
                                                                        [based on
                                                                                    [Taylor Kitsch,
                                                            [Action,
                                             is a war-
                                                                      novel, mars,
                                                         Adventure,
                                                                                     Lynn Collins,
                                                                                                       [Andrew
                    49529
                            John Carter
                                               weary,
                                                                       medallion,
                                                            Science
                                                                                       Samantha
                                                                                                       Stanton]
                                               former
                                                                           space
                                                            Fiction]
                                                                                          Morton]
                                          military ca...
                                                                          travel...
In [23]:
            #to remove spaces
            def collapse(L):
                 L1 = []
                 for i in L:
                       L1.append(i.replace(" ",""))
                 return L1
            movies['cast'] = movies['cast'].apply(collapse)
In [24]:
            movies['crew'] = movies['crew'].apply(collapse)
            movies['genres'] = movies['genres'].apply(collapse)
            movies['keywords'] = movies['keywords'].apply(collapse)
```

In [25]: movies.head()

Out[25]:	movie_id		ovie_id title		genres	keywords	cast				
	0	19995	Avatar	In the 22nd century, a paraplegic Marine is di	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, 	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[James(
	1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[Gore\			
	2	206647	Spectre	A cryptic message from Bond's past sends him o	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6,	[DanielCraig, ChristophWaltz, LéaSeydoux]	[Sam			
	3	49026	The Dark Knight Rises	Following the death of District Attorney Harve	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden	[ChristianBale, MichaelCaine, GaryOldman]	[Christoph			
	4	49529	John Carter	John Carter is a war- weary, former military ca	[Action, Adventure, ScienceFiction]	[basedonnovel, mars, medallion, spacetravel, p	[TaylorKitsch, LynnCollins, SamanthaMorton]	[Andrew			
	4							•			
In [26]:	movi	<pre>movies['overview'] = movies['overview'].apply(lambda x:x.split())</pre>									
In [27]:	movi	<pre>movies['tags'] = movies['overview'] + movies['genres'] + movies['keywords']</pre>									
In [28]:	new_	<pre>new_df = movies.drop(columns=['overview','genres','keywords','cast','crew']</pre>									
In [29]:	<pre>new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x)) new_df.head()</pre>										
Out[29]:	movie_id title										
	0 19995			Avatar			In the 22nd century, a paraplegic Marine is di				
	1	285	Pirates o	Pirates of the Caribbean: At World's Captain Barbossa, long believed to End							
	2	206647			Spectre	A cryptic message from Bond's past sends hi					
	3 49026			The Dar	k Knight Rises	Following the death of District Attorney Harve					
	4	49529		John Carter John Carter is a war-weary, former military ca							

```
# ['love','loving','lover']
# ['love','love','love']
In [30]:
           import nltk
In [31]: | from nltk.stem.porter import PorterStemmer
           ps=PorterStemmer()
In [32]:
          def stem(text):
               y=[]
               for i in text.split():
                    y.append(ps.stem(i))
               return " ".join(y)
In [37]: new_df['tags']=new_df['tags'].apply(stem)
          new df.head()
In [38]:
Out[38]:
                                                     title
               movie_id
                                                                                               tags
            0
                  19995
                                                   Avatar
                                                           in the 22nd century, a parapleg marin is dispa...
                           Pirates of the Caribbean: At World's
            1
                    285
                                                           captain barbossa, long believ to be dead, ha c...
                                                             a cryptic messag from bond' past send him on
            2
                206647
                                                  Spectre
                 49026
                                      The Dark Knight Rises
            3
                                                             follow the death of district attorney harvey d...
                  49529
                                               John Carter
                                                             john carter is a war-weary, former militari ca...
          from sklearn.feature_extraction.text import CountVectorizer
           cv=CountVectorizer(max_features=5000,stop_words='english')
          vectors=cv.fit transform(new df['tags']).toarray()
In [43]:
In [46]:
          vectors
Out[46]: array([[0, 0, 0, ..., 0, 0, 0],
                   [0, 0, 0, \ldots, 0, 0, 0],
                   [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

```
In [48]:
         cv.get_feature_names()
Out[48]: ['000',
           '007'
           '10',
           '100',
           '11',
           '12',
           '13',
          '14',
           '15',
           '16',
           '17',
           '17th',
           '18',
           '18th',
          '18thcenturi',
          '19',
           '1910',
           '1920',
           '1930',
In [49]:
         from sklearn.metrics.pairwise import cosine_similarity
In [50]:
         similarity = cosine_similarity(vectors)
In [51]: | similarity
                            , 0.08346223, 0.0860309 , ..., 0.04499213, 0.
Out[51]: array([[1.
                  0.
                                        , 0.06063391, ..., 0.02378257, 0.
                 [0.08346223, 1.
                 0.02615329],
                 [0.0860309 , 0.06063391, 1.
                                                 , ..., 0.02451452, 0.
                 0.
                            ],
                 . . . ,
                 [0.04499213, 0.02378257, 0.02451452, ..., 1.
                                                                     , 0.03962144,
                 0.04229549],
                 [0.
                                                    , ..., 0.03962144, 1.
                                        , 0.
                 0.08714204],
                            , 0.02615329, 0.
                 [0.
                                                    , ..., 0.04229549, 0.08714204,
                 1.
                            ]])
In [53]:
         sorted(list(enumerate(similarity[0])),reverse=True,key = lambda x: x[1])[1:
Out[53]: [(1216, 0.2867696673382022),
          (2409, 0.26901379342448517),
           (3730, 0.2605130246476754),
           (507, 0.255608593705383),
           (539, 0.25038669783359574)]
```

```
In [77]: def recommend(movie):
    index = new_df[new_df['title'] == movie].index[0]
    distances = sorted(list(enumerate(similarity[index])),reverse=True,key
    for i in distances[1:6]:
        print(new_df.iloc[i[0]].title)
In [79]: recommend('Batman Begins')
The Dark Knight
Batman
Batman
The Dark Knight Rises
10th & Wolf

In []:
```