```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
movies=pd.read_csv('/content/Movie_Recommendar/dataset/tmdb_5000_movies.csv')
credits=pd.read_csv('/content/Movie_Recommendar/dataset/tmdb_5000_credits.csv')
movies.head(2)
             budget
                                                                    homepage
                                                                                   id keywords origina
                            genres
                                                                                            [{"id":
                          [{"id": 28,
                                                                                            1463,
                            "name":
                                                                                          "name":
      0 237000000
                           "Action"},
                                                  http://www.avatarmovie.com/ 19995
                                                                                          "culture
                           {"id": 12,
                                                                                          clash"},
                             "nam...
                                                                                          {"id":...
                                                                                            [{"id":
                          [{"id": 12,
                                                                                             270,
                            "name":
                                                                                          "name":
      1 300000000 "Adventure"},
                                     http://disney.go.com/disneypictures/pirates/
                                                                                        "ocean"},
                        {"id": 14, "...
                                                                                        {"id": 726,
                                                                                            "na...
                Generate code with movies
                                                View recommended plots
 Next steps:
credits.head(2)
          movie_id
                                  title
                                                         cast
                                                                                              crew
                                                                                                       \blacksquare
                                               [{"cast id": 242,
                                                                                       [{"credit_id":
                                                                     "52fe48009251416c750aca23",
       0
             19995
                                  Avatar
                                              "character": "Jake
                                                     Sully", "...
                                                                                              "de...
                Generate code with credits
 Next steps:
                                                 View recommended plots
movies.shape
      (4803, 20)
credits.shape
      (4803, 4)
#merging datset
movies=movies.merge(credits,on="title")
movies.head(2)
                                                                                       keywords origina
             budget
                            genres
                                                                    homepage
                                                                                           [{"id":
                          [{"id": 28,
                                                                                            1463.
                            "name":
                                                                                          "name":
       0 237000000
                           "Action"},
                                                  http://www.avatarmovie.com/ 19995
                                                                                          "culture
                           {"id": 12,
                                                                                          clash"},
                                                                                          {"id":...
                                                                                           [{"id":
                           [{"id": 12,
                                                                                            270,
                            "name":
                                                                                          "name":
       1 300000000
                                     http://disney.go.com/disneypictures/pirates/
                                                                                  285
                       "Adventure"},
                                                                                        "ocean"},
                        {"id": 14, "...
                                                                                        {"id": 726,
                                                                                            "na...
      2 rows × 23 columns
```

movies.shape

```
(4809, 23)
# node need of budget,homepage,original language,popularity,
movies.columns
     'production_countries', 'release_date', 'revenue', 'runtime', 'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
            'vote_count', 'movie_id', 'cast', 'crew'],
dtype='object')
movies=movies[['movie_id','title','overview','genres',"keywords","cast","crew"]];
movies.head(2)
         movie id
                         title overview
                                                 genres keywords
                                                                           cast
                                     In the
                                                              [{"id":
                                                [{"id": 28,
                                                                      [{"cast_id":
                                     22nd
                                                              1463,
                                                 "name":
                                                                           242,
                                                                                                   [{"cr
                                 century, a
                                                            "name":
             19995
                                                "Action"},
                                                                                  "52fe48009251416c750
                        Avatar
                                                                      "character":
                                 paraplegic
                                                            "culture
                                                {"id": 12,
                                                                           "Jake
                                                            clash"},
                                  Marine is
                                                  "nam...
                                                                       Sully", "...
                                       di...
                                                             {"id":...
 Next steps:
               Generate code with movies
                                              View recommended plots
movies.shape
     (4809, 7)
movies.isnull().sum()
     movie id
     title
                  0
     overview
                  3
     genres
                  0
     kevwords
     cast
                  a
     dtype: int64
movies.dropna(inplace=True)
movies.isnull().sum()
     movie_id
                  0
     title
                  0
     overview
                  0
     genres
                   0
     keywords
                  0
     cast
     crew
     dtype: int64
movies.shape
     (4806, 7)
movies.duplicated().sum()
movies.iloc[0]['genres']
     '[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasv"}. {"id": 878. "name": "Science Fiction"}]'
```

```
import ast
ast.literal_eval('[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}, {"id": 878, "name": "Sc.
      [{'id': 28, 'name': 'Action'},
       {'id': 12, 'name': 'Adventure'},
       {'id': 14, 'name': 'Fantasy'},
       {'id': 878, 'name': 'Science Fiction'}]
import ast
def convert(text):
  1=[]
  for i in ast.literal_eval(text):
    1.append(i['name'])
  return 1
movies['genres']=movies['genres'].apply(convert)
movies.head(2)
          movie id
                          title overview
                                                 genres keywords
                                                                            cast
                                                               [{"id":
                                      In the
                                                 [Action,
                                                                       [{"cast_id":
                                       22nd
                                                               1463,
                                                                             242,
                                                                                                     [{"cred
                                              Adventure.
                                                             "name":
                                  century, a
                                                                                   "52fe48009251416c750a
             19995
                         Avatar
                                                Fantasy,
                                                                      "character":
                                                             "culture
                                  paraplegic
                                                 Science
                                                                            "Jake
                                   Marine is
                                                             clash"},
                                                  Fiction1
                                                                        Sully", "...
                                        di...
                                                             {"id":...
     4
 Next steps:
                Generate code with movies
                                                View recommended plots
movies.iloc[0]['keywords']
      '[{"id": 1463, "name": "culture clash"}, {"id": 2964, "name": "future"}, {"id": 3386,
      "name": "space war"}, {"id": 3388, "name": "space colony"}, {"id": 3679, "name": "soci
      ety"}, {"id": 3801, "name": "space travel"}, {"id": 9685, "name": "futuristic"}, {"i
      d": 9840, "name": "romance"}, {"id": 9882, "name": "space"}, {"id": 9951, "name": "ali
      en"}, {"id": 10148, "name": "tribe"}, {"id": 10158, "name": "alien planet"}, {"id": 10
     987, "name": "cgi"}, {"id": 11399, "name": "marine"}, {"id": 13065, "name": "soldie r"}. {"id": 14643. "name": "battle"}. {"id": 14720. "name": "love affair"}. {"id": 165
movies['keywords']=movies['keywords'].apply(convert)
movies.head(2)
          movie id
                          title overview
                                                 genres keywords
                                                                            cast
                                                             [culture
                                      In the
                                                 [Action,
                                                              clash,
                                                                       [{"cast_id":
                                       22nd
                                                                             242.
                                                                                                     [{"cred
                                              Adventure,
                                                              future,
                                  century, a
                                                                                   "52fe48009251416c750a
              19995
                                                                       'character":
                         Avatar
                                                Fantasy.
                                                              space
                                  paraplegic
                                                 Science
                                                                war.
                                                                            "Jake
                                   Marine is
                                                                        Sully", "...
                                                  Fiction]
                                                              space
                                        di
                                                             colon
     4
 Next steps:
                Generate code with movies
                                                View recommended plots
movies.iloc[0]['cast']
      '[{"cast_id": 242, "character": "Jake Sully", "credit_id": "5602a8a7c3a3685532001c9a",
      "gender": 2, "id": 65731, "name": "Sam Worthington", "order": 0}, {"cast id": 3, "char
      acter": "Neytiri", "credit_id": "52fe48009251416c750ac9cb", "gender": 1, "id": 8691,
      "name": "Zoe Saldana", "order": 1}, {"cast_id": 25, "character": "Dr. Grace Augustin
      e", "credit_id": "52fe48009251416c750aca39", "gender": 1, "id": 10205, "name": "Sigour
     ney Weaver", "order": 2}, {"cast_id": 4, "character": "Col. Quaritch", "credit_id": "5 2fe48009251416c750ac9cf", "gender": 2, "id": 32747, "name": "Stephen Lang", "order":
      3}, {"cast_id": 5, "character": "Trudy Chacon", "credit_id": "52fe48009251416c750ac9d 3". "gender": 1. "id": 17647. "name": "Michelle Rodriguez". "order": 4}. {"cast_id":
movies['cast'].head()
           [{"cast_id": 242, "character": "Jake Sully", "...
           [{"cast_id": 4, "character": "Captain Jack Spa...
```

```
[{"cast_id": 1, "character": "James Bond", "cr...
            [{"cast_id": 2, "character": "Bruce Wayne / Ba...
      3
            [{"cast_id": 5, "character": "John Carter", "c...
      4
      Name: cast, dtype: object
# movies['cast']=movies['cast'].apply(convert)
import ast
def convert_cast(text):
  1 = []
  counter=0:
  for i in ast.literal eval(text):
    if counter < 3:
       1.append(i['name'])
     counter +=1
  return 1
movies['cast'] = movies['cast'].apply(convert cast)
movies.head(2)
          movie id
                           title overview
                                                    genres keywords
                                                                                 cast
                                                                [culture
                                        In the
                                                                                 [Sam
                                                    [Action,
                                                                 clash,
                                                                         Worthington,
                                         22nd
                                                 Adventure,
                                                                 future,
                                    century, a
                                                                                  Zoe
              19995
                           Avatar
                                                                                        "52fe48009251416c750
                                                   Fantasv.
                                                                 space
                                    paraplegic
                                                                             Saldana,
                                                   Science
                                                                   war,
                                     Marine is
                                                                            Sigourney
                                                    Fiction]
                                                                 space
                                          di
                                                                             Weave
     4
 Next steps:
                 Generate code with movies
                                                  View recommended plots
movies.iloc[0]['crew']
      '[{"credit_id": "52fe48009251416c750aca23", "department": "Editing", "gender": 0, "i
      d": 1721, "job": "Editor", "name": "Stephen E. Rivkin"}, {"credit_id": "539c47ecc3a368
      10e3001f87", "department": "Art", "gender": 2, "id": 496, "job": "Production Design", "name": "Rick Carter"}, {"credit_id": "54491c89c3a3680fb4001cf7", "department": "Soun
                                                                                     "department": "Soun
      d", "gender": 0, "id": 900, "job": "Sound Designer", "name": "Christopher Boyes"}, {"credit_id": "54491cb70e0a267480001bd0", "department": "Sound", "gender": 0, "id": 900,
      "job": "Supervising Sound Editor", "name": "Christopher Boyes"}, {"credit_id": "53944a
4cc3a36810c9002101". "department": "Production". "gender": 1. "id": 1262. "ioh": "Cast
import ast
def fetch_director(text):
  1=[]
  for i in ast.literal_eval(text):
    if i['job'] == "Director":
       1.append(i['name'])
       break;
  return 1
movies['crew']=movies['crew'].apply(fetch_director)
movies.head(2)
                                                                                                           \blacksquare
          movie_id
                            title
                                      overview
                                                       genres
                                                                  keywords
                                                                                      cast
                                                                                                   crew
                                                                    [culture
                                                                                                            ıl.
                                     In the 22nd
                                                       [Action.
                                                                                      [Sam
                                                                     clash,
                                      century, a
                                                    Adventure,
                                                                               Worthington,
                                                                     future,
                                                                                                [James
              19995
                            Avatar
                                      paraplegic
                                                      Fantasy,
                                                                              Zoe Saldana,
                                                                                              Cameron]
                                                                 space war.
                                       Marine is
                                                      Science
                                                                                 Sigourney
                                                                     space
                                            di
                                                       Fiction]
                                                                                  Weave...
                                                                    colon...
                                                                    [ocean.
                Generate code with movies
                                                  View recommended plots
 Next steps:
movies.iloc[0]['overview']
```

'In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting an alien civilization '

 $\verb"movies['overview'] = \verb"movies['overview'].apply(lambda x:x.split())$

movies.head()

	mov	ie_id	title	overview	genres	keywords	cast	crew
	0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon	[San Worthington Zoe Saldana Sigourne Weaver	[James Cameron]
	1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india	[Johnny Depp Orlandd Bloom, Keira Knightley	[Gore Verbinski]
				[A, cryptic, message,	[Action_	[spy, based on	[Danie	
ext	steps:	Generate code with movies			View recommended plots		plots	

movies.iloc[0]['overview']

```
['In',
'the',
       '22nd<sup>'</sup>,
       'century,',
       'a',
       'paraplegic',
       'Marine',
       'is',
       'dispatched',
      'to',
'the',
       'moon',
       'Pandora',
       'on',
       'a',
       'unique',
       'mission,',
       'but',
       'becomes',
       'torn',
       'between',
       'following',
       'orders',
       'and',
       'protecting',
      'an',
'alien',
       'civilization.']
def removespace(word):
  1 = []
  for i in word:
    1.append(i.replace(" ",""))
  return 1;
movies['cast']=movies['cast'].apply(removespace)
movies['crew']=movies['crew'].apply(removespace)
movies['genres']=movies['genres'].apply(removespace)
movies['keywords']=movies['keywords'].apply(removespace)
# movies['cast']=movies['cast'].apply(removespace)
movies.head()
```

```
movie_id
                          title
                                   overview
                                                       genres
                                                                     keywords
                                                                                             cast
                                      [In, the,
                                                                  [cultureclash,
                                                      [Action,
                                       22nd,
                                                                        future.
                                                                                 [SamWorthington,
                                                   Adventure,
       0
              19995
                          Avatar
                                                                                      ZoeSaldana.
                                  century,, a,
                                                                     spacewar.
                                                                                                     [James
                                                     Fantasy,
                                                                                SigourneyWeaver]
                                  paraplegic,
                                                                  spacecolony,
                                               ScienceFiction]
                                      Marin...
                       Pirates of
                                    [Captain,
                                                                       [ocean,
                                  Barbossa..
                                                  [Adventure,
                                                                                     [JohnnyDepp,
                             the
                                                                   drugabuse,
       1
                285 Caribbean:
                                                                                                       [Gore
                                        long,
                                                     Fantasy,
                                                                                    OrlandoBloom,
                                                                  exoticisland,
                       At World's
                                    believed,
                                                       Action]
                                                                                    KeiraKnightley]
                                                                eastindiatrad...
                            End
                                   to, be, d...
                                   [A, cryptic,
                                   message,
                                                                          [spy,
                                                      [Action,
                                                                                      [DanielCraig,
                                        from,
                                                                basedonnovel,
       2
            206647
                         Spectre
                                                                                   ChristophWaltz.
                                                                                                        [Sai
                                                   Adventure,
                                                                  secretagent,
                                      Bond's,
                                                       Crime]
                                                                                     LéaSeydoux]
                                        past,
                                                                sequel, mi6, ...
     4
 Next steps:
                Generate code with movies
                                                 View recommended plots
movies['tags']=movies['overview']+movies['genres']+movies['keywords']+movies['cast']+movies['crew']
movies.head()
          movie_id
                          title
                                   overview
                                                       genres
                                                                     keywords
                                                                                             cast
                                                                  [cultureclash,
                                      [In, the,
                                                      [Action,
                                       22nd,
                                                                        future,
                                                                                 [SamWorthington,
                                                   Adventure,
       0
              19995
                                  century,, a,
                                                                                      ZoeSaldana.
                          Avatar
                                                                     spacewar.
                                                                                                     [James
                                                     Fantasy,
                                                                                SigourneyWeaver]
                                  paraplegic,
                                                                  spacecolony,
                                               ScienceFiction]
                                      Marin...
                       Pirates of
                                    [Captain,
                                                                       [ocean,
                                   Barbossa..
                                                  [Adventure,
                                                                                     [JohnnyDepp,
                             the
                                                                   drugabuse,
       1
                285
                     Caribbean:
                                                                                    OrlandoBloom.
                                                                                                       [Gore
                                        long,
                                                     Fantasy,
                                                                  exoticisland,
                      At World's
                                    believed,
                                                       Action]
                                                                                    KeiraKnightley]
                                                                eastindiatrad...
                            End
                                   to, be, d...
                                   [A, cryptic,
                                   message,
                                                                          [spy,
                                                      [Action,
                                                                                      [DanielCraig,
                                        from,
                                                                basedonnovel,
       2
            206647
                         Spectre
                                                                                   ChristophWaltz,
                                                                                                        [Sai
                                                   Adventure.
                                      Bond's,
                                                                  secretagent,
                                                       Crime]
                                                                                     LéaSeydoux]
                                                                sequel, mi6, ...
                                        past,
                                      send..
                                  [Following,
                                                                    Idccomics.
                       The Dark
                                                [Action, Crime,
                                                                                    [ChristianBale,
                                   the, death,
                                                                  crimefighter,
       3
             49026
                                                                                    MichaelCaine, [Christor
                          Knight
                                                      Drama.
                                   of District
                                                                       tarroriet
 Next steps:
                Generate code with movies
                                                 View recommended plots
movies.iloc[0]['tags']
      ['In',
        the'
       '22nd',
       'century,',
       'paraplegic',
       'Marine',
       'is',
       'dispatched',
       'to',
       'the'
       'moon',
       'Pandora',
       'on',
       'a',
       'unique',
       'mission,',
       'but',
       'becomes',
       'torn',
       'between'.
       'following',
       'orders',
       'and',
```

'protecting',
'an',
'alien',
'civilization.',
'Action',
'Adventure',

```
'Fantasy',
       'ScienceFiction',
       'cultureclash',
       'future',
       'spacewar',
       'spacecolony',
       'society',
       'spacetravel',
       'futuristic',
       'romance',
       'space',
       'alien',
       'tribe'
       'alienplanet',
       'cgi',
       'marine'
       'soldier',
       'battle',
       'loveaffair',
       'antiwar',
       'powerrelations',
       'mindandsoul',
       '3d',
       'SamWorthington',
       'ZoeSaldana',
       'SigourneyWeaver',
       'JamesCameron']
new_df=movies[["movie_id","title","tags"]]
new_df.head()
          movie_id
                                                  title
                                                                                                      翩
                                                                                             tags
                                                               [In, the, 22nd, century,, a, paraplegic,
       0
             19995
                                                 Avatar
                      Pirates of the Caribbean: At World's
                                                          [Captain, Barbossa,, long, believed, to, be,
               285
                                                   End
                                                            [A, cryptic, message, from, Bond's, past,
            206647
                                                Spectre
                                                               ___[Following, the, death, of, District,
                                                View recommended plots
 Next steps:
                Generate code with new df
new_df['tags']=new_df['tags'].apply(lambda x: " ".join(x))
      <ipython-input-221-6204029b9a1d>: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: \frac{https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html\#returning-a-view-versus-a-c\\ new_df['tags']=new_df['tags'].apply(lambda x: " ".join(x))
     4
new_df.head()
                                                 title
          movie_id
                                                                                             tags
                                                          In the 22nd century, a paraplegic Marine is
             19995
                                                 Avatar
                     Pirates of the Caribbean: At World's
                                                         Captain Barbossa, long believed to be dead,
               285
       1
                                                          A cryptic message from Bond's past sends
            206647
                                               Spectre
                                                           Following the death of District Attorney
               Generate code with new_df
                                               View recommended plots
 Next steps:
```

```
\label{lem:new_df['tags']=new_df['tags'].apply(lambda x: x.lower())} % \begin{center} \begin{c
           <ipython-input-223-d51011ffdf14>: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-c
               new_df['tags']=new_df['tags'].apply(lambda x: x.lower())
new df.head()
                                                                                          title
                                                                                                                                                                                            扁
                  movie_id
                                                                                                                                                                            tags
                                                                                                            in the 22nd century, a paraplegic marine is
             0
                         19995
                                                                                          Avatar
                                        Pirates of the Caribbean: At World's
                                                                                                          captain barbossa, long believed to be dead,
             1
                             285
                                                                                                             a cryptic message from bond's past sends
                      206647
             2
                                                                                        Spectre
                                                                                                                                                                       him o...
                                                                                                                    following the death of district attorney
                             Generate code with new df
                                                                                        View recommended plots
   Next steps:
import nltk
from nltk.stem import PorterStemmer
ps =PorterStemmer()
def stems(text):
    1=[]
    for i in text.split():
        1.append(ps.stem(i))
    return " ".join(1)
new_df['tags']=new_df['tags'].apply(stems)
           <ipython-input-231-2fbf24761000>: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: <a href="https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c">https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-c</a>
               new_df['tags']=new_df['tags'].apply(stems)
new_df.iloc[0]['tags']
           'in the 22nd century, a parapleg marin is dispatch to the moon pandora on a uniqu miss
           ion, but becom torn between follow order and protect an alien civilization. action adv
           entur fantasi sciencefict cultureclash futur spacewar spacecoloni societi spacetravel
           futurist romand space alien tribe alienplanet ogi marin soldier battl loveaffair antiw
from sklearn.feature_extraction.text import CountVectorizer
cv=CountVectorizer(max_features=5000,stop_words='english')
vector=cv.fit_transform(new_df['tags']).toarray()
vector
           array([[0, 0, 0, ..., 0, 0, 0],
                         [0, 0, 0, ..., 0, 0, 0],
                         [0, 0, 0, ..., 0, 0, 0],
                         [0, 0, 0, \ldots, 0, 0, 0],
                         [0, 0, 0, ..., 0, 0, 0],
[0, 0, 0, ..., 0, 0, 0]])
vector.shape
           (4806, 5000)
```

from sklearn.metrics.pairwise import cosine_similarity

```
similarity=cosine_similarity(vector)
similarity
    array([[1.
                      , 0.08346223, 0.0860309 , ..., 0.04499213, 0.
           [0.08346223, 1.
                                 , 0.06063391, ..., 0.02378257, 0.
            0.02615329],
           [0.0860309 , 0.06063391, 1.
                                            , ..., 0.02451452, 0.
                     ],
           [0.04499213, 0.02378257, 0.02451452, ..., 1.
                                                            , 0.03962144,
            0.04229549],
           [0.
                     , 0.
                                , 0.
                                           , ..., 0.03962144, 1.
            0.08714204],
                     , 0.02615329, 0.
           [0.
                                            , ..., 0.04229549, 0.08714204,
                     ]])
similarity.shape
    (4806, 4806)
new_df[new_df['title'] == 'Spider-Man'].index[0]
    159
def recommend(movie):
   index = new_df[new_df['title'] == movie].index[0]
   for i in distances[1:10]:
       print(new_df.iloc[i[0]].title)
recommend('Spider-Man')
    Spider-Man 3
    Spider-Man 2
    The Amazing Spider-Man 2
    Arachnophobia
    Kick-Ass
    The Amazing Spider-Man
    21 Jump Street
    X-Men
    Eight Legged Freaks
recommend("The Dark Knight Rises")
    The Dark Knight
    Batman Returns
    Batman
    Batman Forever
    Batman Begins
    Batman
    Batman & Robin
    Nighthawks
    Slow Burn
import pickle
pickle.dump(new\_df, open('\_/content/Movie\_Recommendar/artificats/movie\_list.pkl' \ , \ "wb"))
pickle.dump(new_df, open('/content/Movie_Recommendar/artificats/similarity.pkl' , "wb"))
Start coding or generate with AI.
Start coding or generate with AI.
```

