```
import pandas as pd
In [1]:
          import numpy as np
          import warnings; warnings.simplefilter('ignore')
          movies = pd.read_csv('tmdb_5000_movies.csv')
In [2]:
          credits = pd.read_csv('tmdb_5000_credits.csv')
In [3]:
          movies.head()
Out[3]:
                budget
                              genres
                                                                      homepage
                                                                                      id keywords original_language origi
                                                                                              [{"id":
                             [{"id": 28,
                                                                                              1463,
                              "name":
                                                                                            "name":
             237000000
                            "Action"},
                                                     http://www.avatarmovie.com/
                                                                                   19995
                                                                                                                    en
                                                                                            "culture
                             {"id": 12,
                                                                                             clash"},
                               "nam...
                                                                                             {"id":...
                                                                                              [{"id":
                             [{"id": 12,
                                                                                               270,
                                                                                                                        Pirat
                              "name":
                                                                                            "name":
                                                                                                                          Ca
             300000000
                                                                                     285
                                        http://disney.go.com/disneypictures/pirates/
                                                                                                                    en
                         "Adventure"},
                                                                                           "ocean"},
                                                                                                                           A<sub>1</sub>
                           {"id": 14, "...
                                                                                           {"id": 726,
                                                                                               "na...
                                                                                              [{"id":
                             [{"id": 28,
                                                                                               470,
                              "name":
                                                                                            "name":
          2 245000000
                            "Action"},
                                      http://www.sonypictures.com/movies/spectre/ 206647
                                                                                                                    en
                                                                                              "spy"},
                             {"id": 12,
                                                                                           {"id": 818,
                               "nam...
                                                                                            "name...
                                                                                              [{"id":
                             [{"id": 28,
                                                                                               849,
                                                                                            "name":
                              "name":
          3 250000000
                            "Action"},
                                                http://www.thedarkknightrises.com/
                                                                                   49026
                                                                                                "dc
                                                                                                                    en
                                                                                                                         Knic
                             {"id": 80,
                                                                                           comics"},
                               "nam...
                                                                                              {"id":
                                                                                              853,...
                                                                                              [{"id":
                             [{"id": 28,
                                                                                               818,
                              "name":
                                                                                            "name":
             260000000
                            "Action"},
                                               http://movies.disney.com/john-carter
                                                                                   49529
                                                                                                                         Joh
                                                                                                                    en
                                                                                          "based on
                             {"id": 12,
                                                                                            novel"},
                               "nam...
                                                                                             {"id":...
          movies.columns
In [4]:
          Index(['budget', 'genres', 'homepage', 'id', 'keywords', 'original_language',
Out[4]:
                  'original_title', 'overview', 'popularity', 'production_companies',
                  'production_countries', 'release_date', 'revenue', 'runtime',
                  'spoken_languages', 'status', 'tagline', 'title', 'vote_average',
                  'vote_count'],
                 dtype='object')
          movies=movies[['genres', 'id', 'keywords', 'original_language',
In [5]:
                   'original_title', 'overview', 'popularity', 'production_companies', 'release_date', 'runt
                   'spoken_languages', 'tagline', 'title', 'vote_average',
                   'vote_count']]
          movies.head()
In [6]:
```

Out[6]:		genres	id	keywords	original_language	original_title	overview	popularity	production_companies
	0	[{"id": 28, "name": "Action"}, {"id": 12, "nam	19995	[{"id": 1463, "name": "culture clash"}, {"id":	en	Avatar	In the 22nd century, a paraplegic Marine is di	150.437577	[{"name": "Ingenious Film Partners", "id": 289
	1	[{"id": 12, "name": "Adventure"}, {"id": 14, "	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"
	2	[{"id": 28, "name": "Action"}, {"id": 12, "nam	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name	en	Spectre	A cryptic message from Bond's past sends him o	107.376788	[{"name": "Columbia Pictures", "id": 5}, {"nam
	3	[{"id": 28, "name": "Action"}, {"id": 80, "nam	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,	en	The Dark Knight Rises	Following the death of District Attorney Harve	112.312950	[{"name": "Legendary Pictures", "id": 923}, {"
	4	[{"id": 28, "name": "Action"}, {"id": 12, "nam	49529	[{"id": 818, "name": "based on novel"}, {"id":	en	John Carter	John Carter is a war- weary, former military ca	43.926995	[{"name": "Walt Disney Pictures", "id": 2}]
									•

Out[7]:	movie_id		title	cast	crew					
	0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id": "52fe48009251416c750aca23", "de					
	1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id": "52fe4232c3a36847f800b579", "de					
	2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id": "54805967c3a36829b5002c41", "de					
	3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id": "52fe4781c3a36847f81398c3", "de					
	4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id": "52fe479ac3a36847f813eaa3", "de					
n [8]:	credits.columns									
ut[8]:	<pre>Index(['movie_id', 'title', 'cast', 'crew'], dtype='object')</pre>									
n [9]:	<pre>credits=credits[['title', 'cast', 'crew']]</pre>									
[10]:	<pre>df=pd.merge(movies,credits,on='title') df.head()</pre>									

In [7]: credits.head()

Out[10]:		genres	id	keywords	original_language	original_title	overview	popularity	production_companies
	0	[{"id": 28, "name": "Action"}, {"id": 12, "nam	19995	[{"id": 1463, "name": "culture clash"}, {"id":	en	Avatar	In the 22nd century, a paraplegic Marine is di	150.437577	[{"name": "Ingenious Film Partners", "id": 289
	1	[{"id": 12, "name": "Adventure"}, {"id": 14, "	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha	139.082615	[{"name": "Walt Disney Pictures", "id": 2}, {"
	2	[{"id": 28, "name": "Action"}, {"id": 12, "nam	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name	en	Spectre	A cryptic message from Bond's past sends him o	107.376788	[{"name": "Columbia Pictures", "id": 5}, {"nam
	3	[{"id": 28, "name": "Action"}, {"id": 80, "nam	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,	en	The Dark Knight Rises	Following the death of District Attorney Harve	112.312950	[{"name": "Legendary Pictures", "id": 923}, {"
	4	[{"id": 28, "name": "Action"}, {"id": 12, "nam	49529	[{"id": 818, "name": "based on novel"}, {"id":	en	John Carter	John Carter is a war- weary, former military ca	43.926995	[{"name": "Walt Disney Pictures", "id": 2}]
4									<b>&gt;</b>
In [11]:	df	.shape							
Out[11]:	(4	(4809, 17)							
In [12]:	df	.isna().sum	()						
Out[12]:	id ke or or over poper record to the control of the	keywords original_language original_title overview 3 popularity production_companies orelease_date runtime 2 spoken_languages tagline title vote_average vote_count cast crew dtype: int64							
In [13]:		=df.dropna( .shape	subset=	overview'	)				

```
(4806, 17)
Out[13]:
            df.isna().sum()
In [14]:
            genres
                                               0
Out[14]:
            id
                                               0
                                               0
            keywords
            original_language
                                               0
            original_title
                                               0
            overview
                                               0
                                               0
            popularity
            production_companies
                                               0
            release_date
                                               1
                                               0
            runtime
            spoken_languages
                                               0
            tagline
                                            841
            title
                                               0
                                               0
            vote_average
                                               0
            vote count
            cast
                                               0
            crew
                                               0
            dtype: int64
            df.duplicated().sum()
In [15]:
Out[15]:
            df.head()
In [16]:
Out[16]:
                                    id keywords original_language original_title
                                                                                          overview
                                                                                                      popularity production_companies
                      genres
                                             [{"id":
                                                                                              In the
                    [{"id": 28,
                                              1463,
                                                                                               22nd
                     "name":
                                                                                                                       [{"name": "Ingenious
                                           "name":
                                                                                          century, a
            0
                    "Action"},
                                 19995
                                                                                                      150.437577
                                                                                                                         Film Partners", "id":
                                                                                 Avatar
                                                                     en
                                           "culture
                                                                                          paraplegic
                     {"id": 12,
                                                                                                                                       289...
                                                                                           Marine is
                                            clash"},
                       "nam...
                                            {"id":...
                                                                                                di...
                                             [{"id":
                                                                                            Captain
                    [{"id": 12,
                                               270,
                                                                          Pirates of the
                                                                                           Barbossa,
                                                                             Caribbean:
                                                                                                                      [{"name": "Walt Disney
                     "name":
                                           "name":
                                                                                               long
                                   285
                                                                                                      139.082615
                                                                     en
                "Adventure"},
                                          "ocean"},
                                                                             At World's
                                                                                            believed
                                                                                                                       Pictures", "id": 2}, {"...
                  {"id": 14, "...
                                         {"id": 726,
                                                                                    End
                                                                                               to be
                                                                                          dead, ha...
                                              "na...
                                                                                           A cryptic
                                             [{"id":
                    [{"id": 28,
                                                                                           message
                                               470,
                     "name":
                                                                                               from
                                                                                                                        [{"name": "Columbia
                                           "name":
            2
                    "Action"},
                               206647
                                                                                Spectre
                                                                                             Bond's
                                                                                                      107.376788
                                                                                                                           Pictures", "id": 5},
                                                                     en
                                             "spy"},
                     {"id": 12,
                                                                                                                                    {"nam...
                                                                                                past
                                         {"id": 818,
                       "nam...
                                                                                          sends him
                                           "name...
                                                                                                 0...
                                             [{"id":
                    [{"id": 28,
                                               849,
                                                                                          Following
                     "name":
                                           "name":
                                                                                          the death
                                                                                                                       [{"name": "Legendary
                                                                               The Dark
            3
                                49026
                                               "dc
                                                                                          of District
                                                                                                      112.312950
                    "Action"},
                                                                     en
                                                                           Knight Rises
                                                                                                                     Pictures", "id": 923}, {"...
                     {"id": 80,
                                          comics"},
                                                                                           Attorney
                       "nam...
                                              {"id":
                                                                                             Harve...
                                             853,...
                                                                                               John
                                             [{"id":
                    [{"id": 28,
                                                                                          Carter is a
                                               818,
                     "name":
                                                                                               war-
                                           "name":
                                                                                                                      [{"name": "Walt Disney
                    "Action"},
                                49529
                                                                            John Carter
                                                                                              weary,
                                                                                                       43.926995
                                         "based on
                                                                                                                           Pictures", "id": 2}]
                     {"id": 12,
                                                                                             former
                                           novel"},
                       "nam...
                                                                                             military
                                            {"id":...
                                                                                                ca...
```

```
In [40]:
           #data=df['id']
            #data.head()
                   19995
           0
Out[40]:
           1
                     285
           2
                  206647
           3
                   49026
           4
                   49529
           Name: id, dtype: int64
In [41]:
           #data.to_csv('ids.csv',index=False)
            df.columns
In [18]:
           Index(['genres', 'id', 'keywords', 'original_language', 'original_title',
Out[18]:
                     'overview', 'popularity', 'production_companies', 'release_date',
                    'runtime', 'spoken_languages', 'tagline', 'title', 'vote_average',
                    'vote_count', 'cast', 'crew'],
                   dtype='object')
             newDf=df[['genres', 'id', 'keywords',
In [19]:
                     'overview', 'title', 'cast', 'crew']]
            newDf.head()
In [20]:
Out[20]:
                                    id
                                          keywords
                                                         overview
                                                                           title
                                                                                          cast
                       genres
                                                                                                                          crew
                                                                                     [{"cast_id":
                                         [{"id": 1463,
                                                       In the 22nd
                     [{"id": 28,
                                            "name":
                                                         century, a
                                                                                          242,
                                                                                                                   [{"credit_id":
                      "name":
                                19995
                                                                                   "character":
                                                                                                  "52fe48009251416c750aca23",
                                            "culture
                                                        paraplegic
                                                                          Avatar
                "Action"}, {"id":
                                             clash"},
                                                         Marine is
                                                                                   "Jake Sully",
                                                                                                                          "de...
                    12, "nam...
                                              {"id":...
                                                              di...
                                                           Captain
                                          [{"id": 270,
                                                                       Pirates of
                     [{"id": 12,
                                                         Barbossa,
                                                                                   [{"cast_id": 4,
                                            "name":
                                                                            the
                                                                                                                   [{"credit_id":
                      "name":
                                                                                   "character":
                                                             long
                                  285
                                           "ocean"},
                                                                      Caribbean:
                                                                                                  "52fe4232c3a36847f800b579",
                                                       believed to
                                                                                  "Captain Jack
                 "Adventure"},
                                                                      At World's
                                           {"id": 726,
                                                                                                                          "de...
                  {"id": 14, "...
                                                          be dead,
                                                                                         Spa...
                                               "na...
                                                                            End
                                                              ha...
                                          [{"id": 470,
                                                          A cryptic
                     [{"id": 28,
                                                                                   [{"cast_id": 1,
                                            "name":
                                                                                                                   [{"credit_id":
                                                          message
                      "name":
                                                                                    "character":
                               206647
                                                       from Bond's
                                                                                                 "54805967c3a36829b5002c41",
                                         "spy"}, {"id":
                                                                        Spectre
                "Action"}, {"id":
                                                                                        "James
                                                818,
                                                        past sends
                                                                                                                          "de...
                    12, "nam...
                                                                                   Bond", "cr...
                                            "name...
                                                           him o...
                                                         Following
                     [{"id": 28,
                                          [{"id": 849,
                                                                                   [{"cast_id": 2,
                                                       the death of
                                                                                                                   [{"credit_id":
                                         "name": "dc
                      "name":
                                                                       The Dark
                                                                                    "character":
                                49026
                                                                                                   "52fe4781c3a36847f81398c3",
                                                           District
                                           comics"},
                "Action"}, {"id":
                                                                    Knight Rises
                                                                                        "Bruce
                                                          Attorney
                                                                                                                          "de...
                   80, "nam...
                                         {"id": 853,...
                                                                                  Wayne / Ba...
                                                           Harve...
                                          [{"id": 818,
                                                       John Carter
                     [{"id": 28,
                                                                                  [{"cast_id": 5,
                                                                                                                   [{"credit_id":
                                            "name":
                                                          is a war-
                      "name":
                                                                                   "character":
                                49529
                                                                                                   "52fe479ac3a36847f813eaa3",
                                          "based on
                                                                     John Carter
                                                           weary,
                                                                                  "John Carter",
                "Action"}, {"id":
                                                                                                                          "de...
                                            novel"},
                                                           former
                                                                                           "c...
                    12, "nam...
                                              {"id":...
                                                       military ca...
            newDf['genres']=newDf['genres'].apply(lambda x: [i['name'] for i in eval(x)])
In [21]:
            newDf['keywords']=newDf['keywords'].apply(lambda x: [i['name'] for i in eval(x)])
            newDf['cast']=newDf['cast'].apply(lambda x: [i['name'] for i in eval(x)[:3]])
            def get dir(crew list):
In [22]:
                 for crew in eval(crew_list):
                      if crew['job'] == 'Director':
                           return crew['name']
                 return None
```

```
newDf['crew'] = newDf['crew'].apply(get_dir)
In [23]:
In [24]:
            newDf.head()
                                                                                          title
Out[24]:
                                       id
                                                 keywords
                         genres
                                                                    overview
                                                                                                                cast
                                                                                                                             crew
                                              [culture clash,
                                                                  In the 22nd
                         [Action,
                                                                                                  [Sam Worthington,
                     Adventure,
                                               future, space
                                                                    century, a
                                                                                                                            James
                                   19995
                                                                                        Avatar
                                                                                                       Zoe Saldana.
                Fantasy, Science
                                                 war, space
                                                                   paraplegic
                                                                                                                         Cameron
                                                                                                 Sigourney Weaver]
                                                                 Marine is di...
                         Fiction]
                                                    colon...
                                               [ocean, drug
                                                                      Captain
                                                                                  Pirates of the
                                                                                                      [Johnny Depp,
                     [Adventure,
                                               abuse, exotic
                                                                Barbossa, long
                                                                                                                             Gore
                                     285
                                                                                  Caribbean: At
                                                                                                     Orlando Bloom,
                 Fantasy, Action]
                                           island, east india
                                                                believed to be
                                                                                                                         Verbinski
                                                                                   World's End
                                                                                                     Keira Knightley]
                                                                    dead, ha...
                                             [spy, based on
                                                                     A cryptic
                         [Action,
                                                                                                       [Daniel Craig,
                                               novel, secret
                                                                message from
                                                                                                                              Sam
            2
                                  206647
                     Adventure,
                                                                                       Spectre
                                                                                                    Christoph Waltz,
                                              agent, sequel,
                                                                  Bond's past
                                                                                                                          Mendes
                         Crime]
                                                                                                       Léa Seydoux]
                                                                 sends him o...
                                                       mi...
                                                [dc comics,
                                                                 Following the
                                                                                                     [Christian Bale,
                  [Action, Crime,
                                              crime fighter,
                                                                                      The Dark
                                                                                                                       Christopher
            3
                                   49026
                                                              death of District
                                                                                                 Michael Caine, Gary
                 Drama, Thriller]
                                             terrorist, secret
                                                                                   Knight Rises
                                                                                                                            Nolan
                                                              Attorney Harve...
                                                                                                           Oldman]
                                                               John Carter is a
                                           [based on novel,
                                                                                                  [Taylor Kitsch, Lynn
                         [Action,
                                                                                                                           Andrew
                                                                   war-weary,
                                   49529
                     Adventure,
                                           mars, medallion,
                                                                                   John Carter
                                                                                                   Collins, Samantha
                                                               former military
                                                                                                                           Stanton
                 Science Fiction]
                                              space travel...
                                                                                                            Morton]
            newDf['genres']=newDf['genres'].apply(lambda x :' '.join([i.replace(" ",'') for i in x]))
In [25]:
            newDf['keywords']=newDf['keywords'].apply(lambda x :' '.join([i.replace(" ",'') for i in x]))
            newDf['cast']=newDf['cast'].apply(lambda x :' '.join([i.replace(" ",'') for i in x]))
            newDf.head()
Out[25]:
                         genres
                                      id
                                                      keywords
                                                                      overview
                                                                                          title
                                                                                                                cast
                                                                                                                             crew
                         Action
                                                                    In the 22nd
                                              cultureclash future
                                                                                                    SamWorthington
                     Adventure
                                                                                                                            James
                                                                      century, a
            0
                                   19995
                                                                                                         ZoeSaldana
                                                      spacewar
                                                                                        Avatar
                        Fantasy
                                                                     paraplegic
                                                                                                                         Cameron
                                             spacecolony socie...
                                                                                                   SigourneyWeaver
                  ScienceFiction
                                                                   Marine is di...
                                                                        Captain
                                               ocean drugabuse
                                                                      Barbossa,
                                                                                  Pirates of the
                                                                                                        JohnnyDepp
                     Adventure
                                                                                                                             Gore
            1
                                     285
                                                    exoticisland
                                                                  long believed
                                                                                  Caribbean: At
                                                                                                      OrlandoBloom
                  Fantasy Action
                                                                                                                         Verbinski
                                                                                   World's End
                                              eastindiatradingc...
                                                                    to be dead,
                                                                                                      KeiraKnightley
                                                                           ha...
                                                                       A cryptic
                         Action
                                              spy basedonnovel
                                                                                                         DanielCraig
                                                                  message from
                                                                                                                              Sam
            2
                     Adventure
                                  206647
                                              secretagent sequel
                                                                                       Spectre
                                                                                                     ChristophWaltz
                                                                    Bond's past
                                                                                                                          Mendes
                                                     mi6 britis...
                          Crime
                                                                                                         LéaSeydoux
                                                                  sends him o...
                                                                  Following the
                                                       dccomics
                                                                       death of
                                                                                                        ChristianBale
                   Action Crime
                                                                                      The Dark
                                                                                                                       Christopher
            3
                                   49026
                                            crimefighter terrorist
                                                                                                       MichaelCaine
                                                                        District
                  Drama Thriller
                                                                                   Knight Rises
                                                                                                                            Nolan
                                                                                                        GaryOldman
                                                 secretidentity...
                                                                       Attorney
                                                                        Harve...
                                                                  John Carter is
                         Action
                                             basedonnovel mars
                                                                                                         TaylorKitsch
                                                                                                                           Andrew
                                                                   a war-weary,
                     Adventure
                                   49529 medallion spacetravel
                                                                                   John Carter
                                                                                                         LynnCollins
                                                                         former
                                                                                                                          Stanton
                  ScienceFiction
                                                        prince...
                                                                                                   SamanthaMorton
                                                                    military ca...
            newDf['tags']=newDf['genres'] + ' ' + newDf['keywords'] + ' ' + newDf['overview'] + ' ' + newDf
In [26]:
            newDf.head()
```

Action   Adventure   Page	Out[26]:		gen	res	id	keywords	overview	title	cast	crew	tags	
Action Action  Action		0	Advent Fant	ure tasy	19995	spacewar spacecolony	22nd century, a paraplegic Marine is	Avatar	ZoeSaldana		Adventure Fantasy ScienceFiction	
Action Adventure 206647 sequel mi6 brits seretagent sequel mi6 brits seretagent sequel mi6 brits sends him. O  3 Action Crime Drama Thriller 49026 crimefighter terrorist secretidentity basedonnovel action and the sequel mi6 brits seretidentity basedonnovel mars medallion spacetravel prince basedonnovel mars marktime p		1	Fant	asy	285	exoticisland	Barbossa, long believed to be	the Caribbean: At World's	OrlandoBloom		Fantasy Action ocean drugabuse	
Action Crime Drama Thriller  Action Adventure ScienceFiction Spacetravel prince  In [27]: finalDf=newDf[['id', 'title', 'tags']] finalDf-nead()  Out[27]: id title  O 19995 Avatar Action Adventure Fantasy ScienceFiction cultur  1 285 Pirates of the Caribbean: At World's End Adventure Fantasy Action ocean drugabuse exoti  2 206647 Spectre Action Adventure Fantasy Action ocean drugabuse exoti  3 49026 The Dark Knight Rises Action Crime Drama Thriller dccomics crimefigh  Action Adventure Fantasy ScienceFiction basedonnovel secret  Action Adventure Fantasy ScienceFiction cultur Action Adventure Fantasy ScienceFiction basedonnovel secret  Action Adventure Fantasy ScienceFiction cultur  Action Adventure Fantasy Action ocean drugabuse exoti  Action Adventure Fantasy Action Drama Thriller Action Adventure ScienceFiction basedonnovel m  In [28]: import nltk as nl from nltk.stem.porter import PorterStemmer  ps = PorterStemmer()  In [29]:  def stem(text): y_append(ps.stem(i)) return " ".join(y)  In [30]: finalDf['tags'] = finalDf['tags'].astype(str)  In [31]: finalDf['tags'] = finalDf['tags'].aspply(stem)		2	Advent	ure	206647	secretagent	message from Bond's past sends him	Spectre	ChristophWaltz		Adventure Crime spy basedonnovel	
Action Adventure ScienceFiction 49529		3			49026	crimefighter terrorist	the death of District Attorney	Knight	MichaelCaine	•	Drama Thriller dccomics	
<pre>finalDf.head()  Out[27]:</pre>		4	Advent	ure	49529	mars medallion spacetravel	Carter is a war- weary, former military	_	LynnCollins		Adventure ScienceFiction basedonnovel	
<pre>0 19995</pre>	In [27]:											
1 285 Pirates of the Caribbean: At World's End Adventure Fantasy Action ocean drugabuse exoti 2 206647	Out[27]:	id title								tags		
<pre>2 206647</pre>		0	<b>0</b> 19995			Avatar Action Adventure Fantasy ScienceFiction				tion cultur		
<pre>3  49026</pre>		1	285 P	irate	s of the C	aribbean: At World's	End Adver	, 5				
John Carter Action Adventure ScienceFiction basedonnovel m  In [28]: import nltk as nl from nltk.stem.porter import PorterStemmer ps = PorterStemmer()  In [29]: def stem(text):						·						
<pre>In [28]: import nltk as nl     from nltk.stem.porter import PorterStemmer     ps = PorterStemmer()  In [29]: def stem(text):</pre>												
<pre>from nltk.stem.porter import PorterStemmer ps = PorterStemmer()  In [29]: def stem(text):     y=[]     for i in text.split():         y.append(ps.stem(i))     return " ".join(y)  In [30]: finalDf['tags'] = finalDf['tags'].astype(str)  In [31]: finalDf['tags']=finalDf['tags'].apply(stem)</pre>		<b>4</b> 49529 John Carter Action Adventure ScienceFiction basedonnovel m										
<pre>y=[]     for i in text.split():</pre>	In [28]:	<pre>from nltk.stem.porter import PorterStemmer</pre>										
<pre>In [31]: finalDf['tags']=finalDf['tags'].apply(stem)</pre>	In [29]:	<pre>y=[] for i in text.split():     y.append(ps.stem(i))</pre>										
	In [30]:	fi	<pre>finalDf['tags'] = finalDf['tags'].astype(str)</pre>									
<pre>In [32]: finalDf.head()</pre>	In [31]:	<pre>finalDf['tags']=finalDf['tags'].apply(stem)</pre>										
	In [32]:	fi	.nalDf.hea	ad()								

```
Out[32]:
                 id
                                                  title
                                                                                            tags
             19995
          0
                                                           action adventur fantasi sciencefict culturecla...
                                                Avatar
                285
                     Pirates of the Caribbean: At World's End
                                                        adventur fantasi action ocean drugabus exotici...
          2 206647
                                               Spectre
                                                        action adventur crime spi basedonnovel secreta...
              49026
                                    The Dark Knight Rises
                                                         action crime drama thriller dccomic crimefight...
              49529
                                            John Carter action adventur sciencefict basedonnovel mar m...
          from sklearn.feature_extraction.text import CountVectorizer
In [33]:
          cv = CountVectorizer(max_features=5000, stop_words='english')
          vectors=cv.fit_transform(finalDf['tags']).toarray()
 In [ ]:
          vectors[0]
In [34]:
          array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
Out[34]:
          from sklearn.metrics.pairwise import cosine_similarity
In [35]:
          similarity = cosine_similarity(vectors)
In [36]:
          def recommend(movie):
               movie_index=finalDf[finalDf['title']==movie].index[0]
               distance=similarity[movie_index]
               movie_list = sorted(list(enumerate(distance)), reverse=True, key=lambda x:x[1])[1:10]
               for i in movie_list:
                   print(finalDf.iloc[i[0]].title)
          recommend('Batman Begins')
In [37]:
          The Dark Knight
          Batman
          Batman
          The Dark Knight Rises
          Batman Returns
          Batman v Superman: Dawn of Justice
          Batman & Robin
          10th & Wolf
          Rockaway
In [38]:
          recommend('Iron Man')
          Iron Man 2
          Iron Man 3
          Avengers: Age of Ultron
          Captain America: Civil War
          The Avengers
          Ant-Man
          Guardians of the Galaxy
          X-Men
```

X-Men Origins: Wolverine