```
import numpy as np # linear algebra
In [1]:
           import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
           movies=pd.read csv('tmdb 5000 movies.csv')
In [4]:
           credits=pd.read_csv('tmdb_5000_credits.csv')
           movies.head(2)
In [5]:
Out[5]:
                 budget
                                                                                         id keywords original_la
                                 genres
                                                                        homepage
                                                                                                  [{"id":
                               [{"id": 28,
                                                                                                  1463,
                                "name":
                                                                                               "name":
              237000000
                               "Action"},
                                                       http://www.avatarmovie.com/ 19995
                                                                                                "culture
                               {"id": 12,
                                                                                                clash"},
                                 "nam...
                                                                                                 {"id":...
                                                                                                  [{"id":
                               [{"id": 12,
                                                                                                   270,
                                "name":
                                                                                               "name":
              300000000
                                          http://disney.go.com/disneypictures/pirates/
                                                                                        285
                           "Adventure"},
                                                                                               "ocean"},
                            {"id": 14, "...
                                                                                              {"id": 726,
                                                                                                  "na...
           movies.shape
In [6]:
           (4803, 20)
Out[6]:
           credits.head()
In [7]:
                                             title
Out[7]:
              movie_id
                                                                          cast
                                                                                                          crew
                                                                                                   [{"credit_id":
                                                     [{"cast_id": 242, "character":
           0
                 19995
                                                                                  "52fe48009251416c750aca23",
                                           Avatar
                                                                "Jake Sully", "...
                                                                                                          "de...
                                                                                                   [{"credit_id":
                          Pirates of the Caribbean:
                                                       [{"cast_id": 4, "character":
           1
                    285
                                                                                  "52fe4232c3a36847f800b579",
                                   At World's End
                                                            "Captain Jack Spa...
                                                                                                          "de...
                                                                                                   [{"credit_id":
                                                       [{"cast_id": 1, "character":
           2
                206647
                                                                                 "54805967c3a36829b5002c41",
                                          Spectre
                                                             "James Bond", "cr...
                                                                                                   [{"credit_id":
                                                       [{"cast_id": 2, "character":
           3
                            The Dark Knight Rises
                                                                                  "52fe4781c3a36847f81398c3",
                 49026
                                                            "Bruce Wayne / Ba...
                                                                                                          "de...
                                                                                                   [{"credit_id":
                                                       [{"cast_id": 5, "character":
                                      John Carter
                                                                                  "52fe479ac3a36847f813eaa3",
           4
                 49529
                                                              "John Carter", "c...
                                                                                                          "de...
          movies = movies.merge(credits,on='title')
In [8]:
           # to inclede the colomns
In [9]:
           # genres
           # id
           # keywords
           # title
```

```
# overview
             # cast
             # crew
             movies=movies[['movie id','title','genres','overview','keywords','cast','crew']]
In [10]:
             movies.head()
Out[10]:
                movie id
                                  title
                                                         overview
                                                                     keywords
                                               genres
                                                                                        cast
                                                                                                                         CI
                                                             In the
                                                                         [{"id":
                                             [{"id": 28,
                                                                                  [{"cast_id":
                                                             22nd
                                                                          1463,
                                              "name":
                                                                                        242,
                                                                                                                  [{"credit_
                                                         century, a
                                                                       "name":
             0
                                                                                 "character":
                    19995
                                Avatar
                                             "Action"},
                                                                                                "52fe48009251416c750aca2
                                                        paraplegic
                                                                       "culture
                                             {"id": 12,
                                                                                       "Jake
                                                          Marine is
                                                                        clash"},
                                                                                   Sully", "...
                                                "nam...
                                                               di...
                                                                        {"id":...
                                                                         [{"id":
                                                           Captain
                             Pirates of
                                                                                  [{"cast_id":
                                             [{"id": 12,
                                                         Barbossa,
                                                                           270,
                                   the
                                                                                          4,
                                                                                                                  [{"credit
                                              "name":
                                                                       "name":
                                                              long
                                                                                  "character":
                      285 Caribbean:
                                                                                                "52fe4232c3a36847f800b57
                                         "Adventure"},
                                                          believed
                                                                      "ocean"},
                            At World's
                                                                                    "Captain
                                                                     {"id": 726,
                                           {"id": 14, "...
                                                             to be
                                   End
                                                                                  Jack Spa...
                                                         dead, ha...
                                                                          "na...
                                                          A cryptic
                                                                         [{"id":
                                             [{"id": 28,
                                                                                  [{"cast_id":
                                                                          470,
                                                          message
                                              "name":
                                                                                                                  [{"credit_
                                                                                           1.
                                                              from
                                                                       "name":
             2
                  206647
                                                                                 "character":
                                                                                               "54805967c3a36829b5002c4
                               Spectre
                                             "Action"},
                                                            Bond's
                                                                         "spy"},
                                              {"id": 12,
                                                                                      "James
                                                        past sends
                                                                     {"id": 818,
                                                                                 Bond", "cr...
                                               "nam...
                                                           him o...
                                                                       "name...
                                                                          [{"id":
                                                                                  [{"cast_id":
                                             [{"id": 28,
                                                         Following
                                                                           849,
                                                                                           2,
                              The Dark
                                              "name":
                                                         the death
                                                                       "name":
                                                                                                                  [{"credit
                                                                                 "character":
                   49026
             3
                                                                            "dc
                                                                                                "52fe4781c3a36847f81398c
                                Knight
                                             "Action"},
                                                         of District
                                                                                      "Bruce
                                  Rises
                                             {"id": 80,
                                                          Attorney
                                                                      comics"},
                                                                                    Wayne /
                                               "nam...
                                                           Harve...
                                                                          {"id":
                                                                                        Ва...
                                                                         853,...
                                                              John
                                                                         [{"id":
                                             [{"id": 28,
                                                                                  [{"cast_id":
                                                         Carter is a
                                                                          818,
                                              "name":
                                                                                           5,
                                                                                                                  [{"credit_
                                                              war-
                                  John
                                                                       "name":
                                                                                 "character":
                    49529
                                             "Action"},
                                                            weary,
                                                                                                "52fe479ac3a36847f813eaa
                                 Carter
                                                                     "based on
                                              {"id": 12,
                                                            former
                                                                                       "John
                                                                        novel"},
                                                                                 Carter", "c...
                                                "nam...
                                                           military
                                                                        {"id":...
In [11]:
             import ast
In [12]:
             def convert(text):
                  L = []
                  for i in ast.literal eval(text):
                       L.append(i['name'])
                  return L
             movies.dropna(inplace=True)
In [13]:
             movies['genres'] = movies['genres'].apply(convert)
In [14]:
             movies.head()
```

Out[14]:		movie_id	title	genres	overview	keywords	cast	cre
	0	19995	Avatar	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	[{"id": 1463, "name": "culture clash"}, {"id":	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_id "52fe48009251416c750aca23 "de
	1	285	Pirates of the Caribbean: At World's End	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	[{"id": 270, "name": "ocean"}, {"id": 726, "na	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_id "52fe4232c3a36847f800b579 "de
	2	206647	Spectre	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o	[{"id": 470, "name": "spy"}, {"id": 818, "name	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_id "54805967c3a36829b5002c41 "de
	3	49026	The Dark Knight Rises	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve	[{"id": 849, "name": "dc comics"}, {"id": 853,	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_id "52fe4781c3a36847f81398c3 "de
	4	49529	John Carter	[Action, Adventure, Science Fiction]	John Carter is a war- weary, former military ca	[{"id": 818, "name": "based on novel"}, {"id":	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_id "52fe479ac3a36847f813eaa3 "de
4								<b>)</b>
In [15]:	mov	vies[' <mark>ke</mark> y	/words'] =	movies['ke	eywords'].	apply(con	vert)	

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

Out[15]:		movie_id	title	genres	overview	keywords	cast	cre
	0	19995	Avatar	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	[culture clash, future, space war, space colon	[{"cast_id": 242, "character": "Jake Sully", "	[{"credit_ic "52fe48009251416c750aca23 "de
	1	285	Pirates of the Caribbean: At World's End	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	[ocean, drug abuse, exotic island, east india 	[{"cast_id": 4, "character": "Captain Jack Spa	[{"credit_ic "52fe4232c3a36847f800b579 "de
	2	206647	Spectre	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o	[spy, based on novel, secret agent, sequel, mi	[{"cast_id": 1, "character": "James Bond", "cr	[{"credit_ic "54805967c3a36829b5002c41 "de
	3	49026	The Dark Knight Rises	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve	[dc comics, crime fighter, terrorist, secret i	[{"cast_id": 2, "character": "Bruce Wayne / Ba	[{"credit_ic "52fe4781c3a36847f81398c3 "de
	4	49529	John Carter	[Action, Adventure, Science Fiction]	John Carter is a war- weary, former military ca	[based on novel, mars, medallion, space travel	[{"cast_id": 5, "character": "John Carter", "c	[{"credit_ic "52fe479ac3a36847f813eaa3 "de
4								<b>•</b>
In [16]:		oort ast	_eval(' <mark>[{</mark>	"id": 28, '	'name": "A	ction"},	{"id": 12,	"name": "Adventure"},
Out[16]:	{	id': 14,	'name':	Action'}, Adventure' Fantasy'}, 'Science F				
In [17]:	def	L = [] counter for i i if	n = 0 in ast.lite counter < L.append unter+=1	eral_eval(t 3: (i['name'])				
In [18]:		/ies[ˈ <mark>cas</mark> /ies.head		ies['cast']	].apply(co	nvert)		

Out[18]:		movie_id	title	genres	overview	keywords	cast	C
	0	19995	Avatar	[Action, Adventure, Fantasy, Science Fiction]	In the 22nd century, a paraplegic Marine is di	[culture clash, future, space war, space colon	[Sam Worthington, Zoe Saldana, Sigourney Weave	[{"credit_ "52fe48009251416c750aca "،
	1	285	Pirates of the Caribbean: At World's End	[Adventure, Fantasy, Action]	Captain Barbossa, long believed to be dead, ha	[ocean, drug abuse, exotic island, east india 	[Johnny Depp, Orlando Bloom, Keira Knightley,	[{"credit_ "52fe4232c3a36847f800b5 ",
	2	206647	Spectre	[Action, Adventure, Crime]	A cryptic message from Bond's past sends him o	[spy, based on novel, secret agent, sequel, mi	[Daniel Craig, Christoph Waltz, Léa Seydoux, R	[{"credit_ "54805967c3a36829b5002c ",
	3	49026	The Dark Knight Rises	[Action, Crime, Drama, Thriller]	Following the death of District Attorney Harve	[dc comics, crime fighter, terrorist, secret i	[Christian Bale, Michael Caine, Gary Oldman, A	[{"credit_ "52fe4781c3a36847f81398 ",
	4	49529	John Carter	[Action, Adventure, Science Fiction]	John Carter is a war- weary, former military ca	[based on novel, mars, medallion, space travel	[Taylor Kitsch, Lynn Collins, Samantha Morton,	[{"credit_ "52fe479ac3a36847f813ea "،
4								<b>•</b>
In [19]:	mov	ies[ˈcas	t'] = mov	ies['cast']	.apply( <b>la</b>	mbda x:x[0	ð:3])	
In [20]:	def	L = [] for i i	i['job'] = L.append	ext): eral_eval(t == 'Directo (i['name'])	or':			
In [21]:	mov	ies['cre	w'] = mov	ies['crew']	.apply(fe	tch_direct	tor)	
In [22]:		vies['ov ies.samp	_	= movies['c	overview']	.apply(lan	nbda x:x.spl	it())

Out[22]:	n	novie_id	title	genres	overview	keywords	cast	crew
	3064	24034	Teacher's Pet	[Animation, Drama, Family, Music]	Spot (Lane) is a dog who can talk and read. Po	0	[Nathan Lane, Kelsey Grammer, Shaun Fleming]	[Timothy Bjorklund]
	2161	107985	The World's End	[Comedy, Action, Science Fiction]	Five friends who reunite in an attempt to top	[end of the world, apocalypse, trilogy, homage]	[Simon Pegg, Nick Frost, Martin Freeman]	[Edgar Wright]
	3676	11935	Capricorn One	[Drama, Action, Thriller, Science Fiction]	In order to protect the reputation of the Amer	[helicopter, nasa, texas, spacecraft, beguilem	[Elliott Gould, James Brolin, Brenda Vaccaro]	[Peter Hyams]
	2055	17709	Dudley Do-Right	[Comedy, Family, Romance]	Based on the 60's-era cartoon of the same name	[based on cartoon, mountie]	[Brendan Fraser, Sarah Jessica Parker, Alfred	[Hugh Wilson]
	196	38055	Megamind	[Animation, Action, Comedy, Family, Science Fi	Bumbling supervillain Megamind finally defeats	[saving the world, date, prison, secret identi	[Will Ferrell, Brad Pitt, Tina Fey]	[Tom McGrath]
In [23]:	L1 for	llapse(l = [] r i in l L1.app turn L1	- <b>:</b>	ace(" ","")	)			
In [24]:	movies movies	'crew' 'genres	= movies[ s'] = movie	'crew'].app s['genres']	ly(collapse) ly(collapse) .apply(colla ds'].apply(c			
In [25]:	movies	.head()						

	cast	keywords	overview	genres	title	movie_id	
[JamesCan	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[cultureclash, future, spacewar, spacecolony,	In the 22nd century, a paraplegic Marine is di	[Action, Adventure, Fantasy, ScienceFiction]	Avatar	19995	0
[GoreVer	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[ocean, drugabuse, exoticisland, eastindiatrad	Captain Barbossa, long believed to be dead, ha	[Adventure, Fantasy, Action]	Pirates of the Caribbean: At World's End	285	1
[SamM	[DanielCraig, ChristophWaltz, LéaSeydoux]	[spy, basedonnovel, secretagent, sequel, mi6,	A cryptic message from Bond's past sends him o	[Action, Adventure, Crime]	Spectre	206647	2
[Christopher	[ChristianBale, MichaelCaine, GaryOldman]	[dccomics, crimefighter, terrorist, secretiden	Following the death of District Attorney Harve	[Action, Crime, Drama, Thriller]	The Dark Knight Rises	49026	3
	[TaylorKitsch, LynnCollins,	[basedonnovel, mars, medallion,	John Carter is a war- weary,	[Action,	John	40520	
[AndrewSt	SamanthaMorton]	spacetravel, p	former military ca	Adventure, ScienceFiction]	Carter	49529	4
[AndrewSt	•		former military		Carter	49529	4
[AndrewSt	SamanthaMorton]		former military ca	ScienceFiction]			
	SamanthaMorton]	spacetravel, p  oly(lambda x:	former military ca	ScienceFiction] movies['over	erview'] =	ovies[' <mark>ove</mark>	mov
nds']	SamanthaMorton]  x.split())	spacetravel, p  oly(lambda x:: es['genres']	former military ca  view'].app	ScienceFiction]  movies['over ies['overview	erview'] = gs'] = mov es.drop(co	ovies['ove	mov
_	SamanthaMorton]  x.split())  + movies['keywor	spacetravel, p  oly(lambda x:: es['genres'] - res','keyword:	former military ca  view'].app  '] + movid iew','genu	ScienceFiction]  movies['over ies['overview	erview'] = gs'] = mov es.drop(co	ovies['ove ovies['tag ew = movie new.head()	mov new #ne
rds']	SamanthaMorton]  x.split())  + movies['keywor	spacetravel, p  oly(lambda x:: es['genres'] - res','keyword:	former military ca  view'].app  '] + movid iew','genu	ScienceFiction]  movies['over ies['overview lumns=['overv	erview'] = gs'] = mov es.drop(co	<pre>ovies['ove ovies['tag ew = movie new.head() ew['tags']</pre>	mov nev nev
rds']	SamanthaMorton]  x.split())  + movies['keywor	spacetravel, p  oly(lambda x:: es['genres'] - res','keyword: " ".join(x))	former military ca  view'].app '] + movio iew','gent ambda x:	ScienceFiction]  movies['over ies['overview lumns=['overv	erview'] = gs'] = mov es.drop(co	<pre>ovies['ove ovies['tag ovies['tag ew = movie new.head() ew['tags'] ew.head() movie_id</pre>	mov nev nev
tags arine is di	SamanthaMorton]  x.split())  + movies['keywors','cast','crew'	spacetravel, p  oly(lambda x:: es['genres']  res','keyword: " ".join(x))  In the 22nd cer	former military ca  view'].app '] + movio iew','gent ambda x:  title  Avatar	ScienceFiction]  movies['over ies['overview lumns=['overv	erview'] = gs'] = mov es.drop(co ) ] = new['t	ovies['ove ovies['tag ew = movie new.head() ew['tags'] ew.head() movie_id	mov nev nev
tags arine is di	SamanthaMorton]  x.split())  + movies['keywor' s','cast','crew' ntury, a paraplegic Marketing of the state of	spacetravel, p  oly(lambda x:: es['genres']  res','keyword: " ".join(x))  In the 22nd cer  Captain Barbossa	former military ca  view'].app  '] + movid  iew', 'gent  ambda x:  title  Avatar  At World's	ScienceFiction]  movies['overview lumns=['overview ags'].apply(1	erview'] = gs'] = mov es.drop(co ) ] = new['t	ovies['ove ovies['tag ew = movie new.head() ew['tags'] ew.head() movie_id 19995 285	mov nev nev
tags arine is di e dead, ha t sends him o	SamanthaMorton]  x.split())  movies['keywors','crew'  atury, a paraplegic Mana, long believed to be	spacetravel, p  oly(lambda x:: es['genres']  res','keyword: " ".join(x))  In the 22nd cer  Captain Barbossa  A cryptic messa	former military ca  view'].app  '] + movi  iew', 'gen  ambda x:  title  Avatar  At World's End  Spectre	ScienceFiction]  movies['overview lumns=['overview ags'].apply(1	erview'] = gs'] = mov es.drop(co ) ] = new['t	ovies['ove ovies['tag ew = movie fnew.head() ew['tags'] ew.head() movie_id 19995 285 206647	move #new new o
tags arine is di e dead, ha t sends him o ney Harve	SamanthaMorton]  **.split())  * movies['keywor' s','cast','crew'  ntury, a paraplegic Manage from Bond's past	spacetravel, p  oly(lambda x:: es['genres']  res','keyword: " ".join(x))  In the 22nd cer  Captain Barbossa A cryptic messa	former military ca  view'].app  '] + movi  iew', 'gen  ambda x:  title  Avatar  At World's End  Spectre	ScienceFiction]  movies['overview lumns=['overv ags'].apply(1	erview'] = gs'] = mov es.drop(co ) ] = new['t	ovies['ove ovies['tag ew = movie inew.head() ew['tags'] ew.head() movie_id 19995 285 206647 49026	movement mov

```
cv = CountVectorizer(max_features=5000,stop_words='english')
         vector = cv.fit transform(new['tags']).toarray()
In [32]:
         vector.shape
In [33]:
         (4806, 5000)
Out[33]:
         from sklearn.metrics.pairwise import cosine similarity
In [34]:
         similarity = cosine_similarity(vector)
In [35]:
         similarity
In [36]:
                           , 0.09583148, 0.06277648, ..., 0.02465568, 0.0270666 ,
         array([[1.
Out[36]:
                           ],
                [0.09583148, 1.
                                       , 0.07018624, ..., 0.02756589, 0.
                 0.
                           ],
                [0.06277648, 0.07018624, 1. , ..., 0.02708645, 0.
                0.
                           ],
                [0.02465568, 0.02756589, 0.02708645, ..., 1. , 0.07007128,
                 0.04732485],
                                       , 0. , ..., 0.07007128, 1.
                [0.0270666, 0.
                 0.05195243],
                          , 0.
                                       , 0. , ..., 0.04732485, 0.05195243,
                [0.
                 1.
                           ]])
         new[new['title'] == 'The Lego Movie'].index[0]
In [37]:
         744
Out[37]:
         def recommend(movie):
In [38]:
             index = new[new['title'] == movie].index[0]
             distances = sorted(list(enumerate(similarity[index])),reverse=True,key = lambda
             for i in distances[1:6]:
                 print(new.iloc[i[0]].title)
In [39]:
         recommend('Gandhi')
         Gandhi, My Father
         The Wind That Shakes the Barley
         A Passage to India
         Guiana 1838
         Ramanujan
In [40]: import pickle
         pickle.dump(new,open('movie_list.pkl','wb'))
In [41]:
         pickle.dump(similarity,open('similarity.pkl','wb'))
In [ ]:
```