

```
In [1]: ▶ 1 import pandas as pd
          2 import numpy as np
          3 from sklearn.feature_extraction.text import TfidfVectorizer #for each
          4 from sklearn.metrics.pairwise import cosine_similarity
          5
```

In [2]: ▶

```
1 df=pd.read_csv("movies1.csv")
2 df
3
```

Out[2]:

	index	budget		genres	homepage	id
<b>0</b>	0	237000000		Action Adventure Fantasy Science Fiction	<a href="http://www.avatarmovie.com/">http://www.avatarmovie.com/</a>	19995
<b>1</b>	1	300000000		Adventure Fantasy Action	<a href="http://disney.go.com/disneypictures/pirates/">http://disney.go.com/disneypictures/pirates/</a>	285
<b>2</b>	2	245000000		Action Adventure Crime	<a href="http://www.sonypictures.com/movies/spectre/">http://www.sonypictures.com/movies/spectre/</a>	206647
<b>3</b>	3	250000000		Action Crime Drama Thriller	<a href="http://www.thedarkknighttrises.com/">http://www.thedarkknighttrises.com/</a>	49026
<b>4</b>	4	260000000		Action Adventure Science Fiction	<a href="http://movies.disney.com/john-carter">http://movies.disney.com/john-carter</a>	49529
...	...	...		...	...	...
<b>4798</b>	4798	220000		Action Crime Thriller	NaN	9367
<b>4799</b>	4799	9000		Comedy Romance	NaN	72766
<b>4800</b>	4800	0		Comedy Drama Romance TV Movie	<a href="http://www.hallmarkchannel.com/signedsealeddel...">http://www.hallmarkchannel.com/signedsealeddel...</a>	231617
<b>4801</b>	4801	0		NaN	<a href="http://shanghaicalling.com/">http://shanghaicalling.com/</a>	126186

	index	budget	genres	homepage	id
	4802	4802	0 Documentary	NaN	25975

4803 rows × 24 columns

In [3]: 1 df.columns

Out[3]: Index(['index', 'budget', 'genres', 'homepage', 'id', 'keywords', 'original\_language', 'original\_title', 'overview', 'popularity', 'production\_companies', 'production\_countries', 'release\_date', 'revenue', 'runtime', 'spoken\_languages', 'status', 'tagline', 'title', 'vote\_average', 'vote\_count', 'cast', 'crew', 'director'], dtype='object')

In [4]: 1 df.shape

Out[4]: (4803, 24)

In [5]: 1 features=['genres','keywords','original\_language','title','cast','director']  
2 for feature in features:  
3 df[feature]=df[feature].fillna('') #filling missing values with space  
4

In [6]: 1 def combined\_features(row):  
2 return row['title']+", "+row['genres']+", "+row['keywords']+",  
3 "+row['original\_language']+", "+row['cast']+", "+row['director']  
4 df['combined\_features'] = df.apply(combined\_features,axis=1)  
5 #column wise axis=1 dataframe same columns merge  
6 df['combined\_features'] #one more column with all those combined features  
7

Out[6]: 0 Avatar,Action Adventure Fantasy Science Fiction  
1 Pirates of the Caribbean: At World's End,Adventure  
2 Spectre,Action Adventure Crime,spy based on novel  
3 The Dark Knight Rises,Action Crime Drama Thriller  
4 John Carter,Action Adventure Science Fiction,b...  
...  
4798 El Mariachi,Action Crime Thriller,united state...  
4799 Newlyweds,Comedy Romance,,en,Edward Burns Kerr...  
4800 Signed, Sealed, Delivered,Comedy Drama Romance...  
4801 Shanghai Calling,,,en,Daniel Henney Eliza Coup...  
4802 My Date with Drew,Documentary,obsession camcor...  
Name: combined\_features, Length: 4803, dtype: object

```
In [7]: 1 tfidf=TfidfVectorizer()
2         tfidv=tfidf.fit_transform(df['combined_features']) #text word vector
3         tfidv.toarray()
4
```

```
Out[7]: array([[0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               ...,
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.],
               [0., 0., 0., ..., 0., 0., 0.]])
```

```
In [8]: 1 tfidv.shape
```

```
Out[8]: (4803, 17502)
```

```
In [9]: 1 cosine_sim=cosine_similarity(tfidv)
2         cosine_sim #compare the vectors
3
```

```
Out[9]: array([[1.00000000e+00, 2.48272960e-02, 4.22228416e-02, ...,
                1.12200797e-03, 1.34747833e-03, 1.11917020e-03],
               [2.48272960e-02, 1.00000000e+00, 1.42630716e-02, ...,
                4.07129084e-02, 1.11032424e-03, 9.22197982e-04],
               [4.22228416e-02, 1.42630716e-02, 1.00000000e+00, ...,
                1.15436886e-03, 5.90455012e-02, 1.15144924e-03],
               ...,
               [1.12200797e-03, 4.07129084e-02, 1.15436886e-03, ...,
                1.00000000e+00, 1.17574481e-03, 5.83911241e-02],
               [1.34747833e-03, 1.11032424e-03, 5.90455012e-02, ...,
                1.17574481e-03, 1.00000000e+00, 1.17277112e-03],
               [1.11917020e-03, 9.22197982e-04, 1.15144924e-03, ...,
                5.83911241e-02, 1.17277112e-03, 1.00000000e+00]])
```

```
In [10]: 1 cosine_sim.shape
```

```
Out[10]: (4803, 4803)
```

```
In [11]: 1 movie=input("Enter movie name: ")
2         def get_index(mn):
3             return df[df.title==mn].index[0]
4         mi=get_index(movie)
5         mi
6
```

Enter movie name: Avatar

```
Out[11]: 0
```

In [12]:

```
1 sm=list(enumerate(cosine_sim[mi])) #0 se aage badhenge
2 print(sm)
3
```

```
[(0, 1.0000000000000002), (1, 0.024827295986620684), (2, 0.04222284156
9151265), (3, 0.008160607303690774), (4, 0.10625865464875475), (5, 0.0
6782484284827672), (6, 0.0013395252359036881), (7, 0.0369512312714523
5), (8, 0.022342422412007637), (9, 0.027496115325422923), (10, 0.06727
359460847546), (11, 0.015687532413989805), (12, 0.0268928002342309),
(13, 0.026225825804090122), (14, 0.05361238394260148), (15, 0.02013212
146436357), (16, 0.03874126084538814), (17, 0.027494314158529494), (1
8, 0.06237344854351263), (19, 0.027289742758032448), (20, 0.0285754600
84002833), (21, 0.015171695489496225), (22, 0.02126049131775769), (23,
0.04444890691031617), (24, 0.015793538029016164), (25, 0.0464680778643
0685), (26, 0.08523979691334363), (27, 0.03667482299340698), (28, 0.03
904258280160722), (29, 0.042527070873668735), (30, 0.0676870099620024
8), (31, 0.05857713702365093), (32, 0.03158621013232506), (33, 0.03771
431221443131), (34, 0.0013210362060361365), (35, 0.03829621557186802),
(36, 0.03440671321126578), (37, 0.09007853975155278), (38, 0.028762470
70167063), (39, 0.03543766832556243), (40, 0.010817383893253206), (41,
0.04043087219409347), (42, 0.001323843755365448), (43, 0.1102056299480
7107), (44, 0.09848714935864478), (45, 0.056294011657772404), (46, 0.0
9876911686079831), (47, 0.19866721509827104), (48, 0.02044222023930852
5), (49, 0.0013395252359036881), (50, 0.03695123127145235), (51, 0.02
2342422412007637), (52, 0.027496115325422923), (53, 0.06727359460847546),
(54, 0.015687532413989805), (55, 0.0268928002342309), (56, 0.026225825804090122),
(57, 0.05361238394260148), (58, 0.02013212146436357), (59, 0.03874126084538814),
(60, 0.027494314158529494), (61, 0.06237344854351263), (62, 0.027289742758032448),
(63, 0.028575460084002833), (64, 0.015171695489496225), (65, 0.02126049131775769),
(66, 0.04444890691031617), (67, 0.015793538029016164), (68, 0.04646807786430685),
(69, 0.08523979691334363), (70, 0.03667482299340698), (71, 0.03904258280160722),
(72, 0.042527070873668735), (73, 0.06768700996200248), (74, 0.05857713702365093),
(75, 0.03158621013232506), (76, 0.03771431221443131), (77, 0.0013210362060361365),
(78, 0.03829621557186802), (79, 0.03440671321126578), (80, 0.09007853975155278),
(81, 0.02876247070167063), (82, 0.03543766832556243), (83, 0.010817383893253206),
(84, 0.04043087219409347), (85, 0.001323843755365448), (86, 0.11020562994807107),
(87, 0.09848714935864478), (88, 0.056294011657772404), (89, 0.09876911686079831),
(90, 0.19866721509827104), (91, 0.020442220239308525), (92, 0.0013395252359036881),
(93, 0.03695123127145235), (94, 0.022342422412007637), (95, 0.027496115325422923),
(96, 0.06727359460847546), (97, 0.015687532413989805), (98, 0.0268928002342309),
(99, 0.026225825804090122), (100, 0.05361238394260148), (101, 0.02013212146436357),
(102, 0.03874126084538814), (103, 0.027494314158529494), (104, 0.06237344854351263),
(105, 0.027289742758032448), (106, 0.028575460084002833), (107, 0.015171695489496225),
(108, 0.02126049131775769), (109, 0.04444890691031617), (110, 0.015793538029016164),
(111, 0.04646807786430685), (112, 0.08523979691334363), (113, 0.03667482299340698),
(114, 0.03904258280160722), (115, 0.042527070873668735), (116, 0.06768700996200248),
(117, 0.05857713702365093), (118, 0.03158621013232506), (119, 0.03771431221443131),
(120, 0.0013210362060361365), (121, 0.03829621557186802), (122, 0.03440671321126578),
(123, 0.09007853975155278), (124, 0.02876247070167063), (125, 0.03543766832556243),
(126, 0.010817383893253206), (127, 0.04043087219409347), (128, 0.001323843755365448),
(129, 0.11020562994807107), (130, 0.09848714935864478), (131, 0.056294011657772404),
(132, 0.09876911686079831), (133, 0.19866721509827104), (134, 0.020442220239308525),
(135, 0.0013395252359036881), (136, 0.03695123127145235), (137, 0.022342422412007637),
(138, 0.027496115325422923), (139, 0.06727359460847546), (140, 0.015687532413989805),
(141, 0.0268928002342309), (142, 0.026225825804090122), (143, 0.05361238394260148),
(144, 0.02013212146436357), (145, 0.03874126084538814), (146, 0.027494314158529494),
(147, 0.06237344854351263), (148, 0.027289742758032448), (149, 0.028575460084002833),
(150, 0.015171695489496225), (151, 0.02126049131775769), (152, 0.04444890691031617),
(153, 0.015793538029016164), (154, 0.04646807786430685), (155, 0.08523979691334363),
(156, 0.03667482299340698), (157, 0.03904258280160722), (158, 0.042527070873668735),
(159, 0.06768700996200248), (160, 0.05857713702365093), (161, 0.03158621013232506),
(162, 0.03771431221443131), (163, 0.0013210362060361365), (164, 0.03829621557186802),
(165, 0.03440671321126578), (166, 0.09007853975155278), (167, 0.02876247070167063),
(168, 0.03543766832556243), (169, 0.010817383893253206), (170, 0.04043087219409347),
(171, 0.001323843755365448), (172, 0.11020562994807107), (173, 0.09848714935864478),
(174, 0.056294011657772404), (175, 0.09876911686079831), (176, 0.19866721509827104),
(177, 0.020442220239308525), (178, 0.0013395252359036881), (179, 0.03695123127145235),
(180, 0.022342422412007637), (181, 0.027496115325422923), (182, 0.06727359460847546),
(183, 0.015687532413989805), (184, 0.0268928002342309), (185, 0.026225825804090122),
(186, 0.05361238394260148), (187, 0.02013212146436357), (188, 0.03874126084538814),
(189, 0.027494314158529494), (190, 0.06237344854351263), (191, 0.027289742758032448),
(192, 0.028575460084002833), (193, 0.015171695489496225), (194, 0.02126049131775769),
(195, 0.04444890691031617), (196, 0.015793538029016164), (197, 0.04646807786430685),
(198, 0.08523979691334363), (199, 0.03667482299340698), (200, 0.03904258280160722),
(201, 0.042527070873668735), (202, 0.06768700996200248), (203, 0.05857713702365093),
(204, 0.03158621013232506), (205, 0.03771431221443131), (206, 0.0013210362060361365),
(207, 0.03829621557186802), (208, 0.03440671321126578), (209, 0.09007853975155278),
(210, 0.02876247070167063), (211, 0.03543766832556243), (212, 0.010817383893253206),
(213, 0.04043087219409347), (214, 0.001323843755365448), (215, 0.11020562994807107),
(216, 0.09848714935864478), (217, 0.056294011657772404), (218, 0.09876911686079831),
(219, 0.19866721509827104), (220, 0.020442220239308525), (221, 0.0013395252359036881),
(222, 0.03695123127145235), (223, 0.022342422412007637), (224, 0.027496115325422923),
(225, 0.06727359460847546), (226, 0.015687532413989805), (227, 0.0268928002342309),
(228, 0.026225825804090122), (229, 0.05361238394260148), (230, 0.02013212146436357),
(231, 0.03874126084538814), (232, 0.027494314158529494), (233, 0.06237344854351263),
(234, 0.027289742758032448), (235, 0.028575460084002833), (236, 0.015171695489496225),
(237, 0.02126049131775769), (238, 0.04444890691031617), (239, 0.015793538029016164),
(240, 0.04646807786430685), (241, 0.08523979691334363), (242, 0.03667482299340698),
(243, 0.03904258280160722), (244, 0.042527070873668735), (245, 0.06768700996200248),
(246, 0.05857713702365093), (247, 0.03158621013232506), (248, 0.03771431221443131),
(249, 0.0013210362060361365), (250, 0.03829621557186802), (251, 0.03440671321126578),
(252, 0.09007853975155278), (253, 0.02876247070167063), (254, 0.03543766832556243),
(255, 0.010817383893253206), (256, 0.04043087219409347), (257, 0.001323843755365448),
(258, 0.11020562994807107), (259, 0.09848714935864478), (260, 0.056294011657772404),
(261, 0.09876911686079831), (262, 0.19866721509827104), (263, 0.020442220239308525),
(264, 0.0013395252359036881), (265, 0.03695123127145235), (266, 0.022342422412007637),
(267, 0.027496115325422923), (268, 0.06727359460847546), (269, 0.015687532413989805),
(270, 0.0268928002342309), (271, 0.026225825804090122), (272, 0.05361238394260148),
(273, 0.02013212146436357), (274, 0.03874126084538814), (275, 0.027494314158529494),
(276, 0.06237344854351263), (277, 0.027289742758032448), (278, 0.028575460084002833),
(279, 0.015171695489496225), (280, 0.02126049131775769), (281, 0.04444890691031617),
(282, 0.015793538029016164), (283, 0.04646807786430685), (284, 0.08523979691334363),
(285, 0.03667482299340698), (286, 0.03904258280160722), (287, 0.042527070873668735),
(288, 0.06768700996200248), (289, 0.05857713702365093), (290, 0.03158621013232506),
(291, 0.03771431221443131), (292, 0.0013210362060361365), (293, 0.03829621557186802),
(294, 0.03440671321126578), (295, 0.09007853975155278), (296, 0.02876247070167063),
(297, 0.03543766832556243), (298, 0.010817383893253206), (299, 0.04043087219409347),
(300, 0.001323843755365448), (301, 0.11020562994807107), (302, 0.09848714935864478),
(303, 0.056294011657772404), (304, 0.09876911686079831), (305, 0.19866721509827104),
(306, 0.020442220239308525), (307, 0.0013395252359036881), (308, 0.03695123127145235),
(309, 0.022342422412007637), (310, 0.027496115325422923), (311, 0.06727359460847546),
(312, 0.015687532413989805), (313, 0.0268928002342309), (314, 0.026225825804090122),
(315, 0.05361238394260148), (316, 0.02013212146436357), (317, 0.03874126084538814),
(318, 0.027494314158529494), (319, 0.06237344854351263), (320, 0.027289742758032448),
(321, 0.028575460084002833), (322, 0.015171695489496225), (323, 0.02126049131775769),
(324, 0.04444890691031617), (325, 0.015793538029016164), (326, 0.04646807786430685),
(327, 0.08523979691334363), (328, 0.03667482299340698), (329, 0.03904258280160722),
(330, 0.042527070873668735), (331, 0.06768700996200248), (332, 0.05857713702365093),
(333, 0.03158621013232506), (334, 0.03771431221443131), (335, 0.0013210362060361365),
(336, 0.03829621557186802), (337, 0.03440671321126578), (338, 0.09007853975155278),
(339, 0.02876247070167063), (340, 0.03543766832556243), (341, 0.010817383893253206),
(342, 0.04043087219409347), (343, 0.001323843755365448), (344, 0.11020562994807107),
(345, 0.09848714935864478), (346, 0.056294011657772404), (347, 0.09876911686079831),
(348, 0.19866721509827104), (349, 0.020442220239308525), (350, 0.0013395252359036881),
(351, 0.03695123127145235), (352, 0.022342422412007637), (353, 0.027496115325422923),
(354, 0.06727359460847546), (355, 0.015687532413989805), (356, 0.0268928002342309),
(357, 0.026225825804090122), (358, 0.05361238394260148), (359, 0.02013212146436357),
(360, 0.03874126084538814), (361, 0.027494314158529494), (362, 0.06237344854351263),
(363, 0.027289742758032448), (364, 0.028575460084002833), (365, 0.015171695489496225),
(366, 0.02126049131775769), (367, 0.04444890691031617), (368, 0.015793538029016164),
(369, 0.04646807786430685), (370, 0.08523979691334363), (371, 0.03667482299340698),
(372, 0.03904258280160722), (373, 0.042527070873668735), (374, 0.06768700996200248),
(375, 0.05857713702365093), (376, 0.03158621013232506), (377, 0.03771431221443131),
(378, 0.0013210362060361365), (379, 0.03829621557186802), (380, 0.03440671321126578),
(381, 0.09007853975155278), (382, 0.02876247070167063), (383, 0.03543766832556243),
(384, 0.010817383893253206), (385, 0.04043087219409347), (386, 0.001323843755365448),
(387, 0.11020562994807107), (388, 0.09848714935864478), (389, 0.056294011657772404),
(390, 0.09876911686079831), (391, 0.19866721509827104), (392, 0.020442220239308525),
(393, 0.0013395252359036881), (394, 0.03695123127145235), (395, 0.022342422412007637),
(396, 0.027496115325422923), (397, 0.06727359460847546), (398, 0.015687532413989805),
(399, 0.0268928002342309), (400, 0.026225825804090122), (401, 0.05361238394260148),
(402, 0.02013212146436357), (403, 0.03874126084538814), (404, 0.027494314158529494),
(405, 0.06237344854351263), (406, 0.027289742758032448), (407, 0.028575460084002833),
(408, 0.015171695489496225), (409, 0.02126049131775769), (410, 0.04444890691031617),
(411, 0.015793538029016164), (412, 0.04646807786430685), (413, 0.08523979691334363),
(414, 0.03667482299340698), (415, 0.03904258280160722), (416, 0.042527070873668735),
(417, 0.06768700996200248), (418, 0.05857713702365093), (419, 0.03158621013232506),
(420, 0.03771431221443131), (421, 0.0013210362060361365), (422, 0.03829621557186802),
(423, 0.03440671321126578), (424, 0.09007853975155278), (425, 0.02876247070167063),
(426, 0.03543766832556243), (427, 0.010817383893253206), (428, 0.04043087219409347),
(429, 0.001323843755365448), (430, 0.11020562994807107), (431, 0.09848714935864478),
(432, 0.056294011657772404), (433, 0.09876911686079831), (434, 0.19866721509827104),
(435, 0.020442220239308525), (436, 0.0013395252359036881), (437, 0.03695123127145235),
(438, 0.022342422412007637), (439, 0.027496115325422923), (440, 0.06727359460847546),
(441, 0.015687532413989805), (442, 0.0268928002342309), (443, 0.026225825804090122),
(444, 0.05361238394260148), (445, 0.02013212146436357), (446, 0.03874126084538814),
(447, 0.027494314158529494), (448, 0.06237344854351263), (449, 0.027289742758032448),
(450, 0.028575460084002833), (451, 0.015171695489496225), (452, 0.02126049131775769),
(453, 0.04444890691031617), (454, 0.015793538029016164), (455, 0.04646807786430685),
(456, 0.08523979691334363), (457, 0.03667482299340698), (458, 0.03904258280160722),
(459, 0.0425270708736
```

```

In [14]: ▶ 1 def get_info(index):
2         return df[df.index==index]['title'].values[0]+":
3         "+ df[df.index==index]['cast'].values[0]
4         i=0
5         for movie in sorted_sm:
6             print(get_info(movie[0]))
7             i=i+1 #FOR NEXT RECOMMENDATION
8             if i>10:
9                 break
10

```

Avatar: Sam Worthington Zoe Saldana Sigourney Weaver Stephen Lang Michelle Rodriguez  
Guardians of the Galaxy: Chris Pratt Zoe Saldana Dave Bautista Vin Diesel Bradley Cooper  
Aliens: Sigourney Weaver Michael Biehn James Remar Paul Reiser Lance Henriksen  
Alien: Tom Skerritt Sigourney Weaver Veronica Cartwright Harry Dean Stanton John Hurt  
Galaxy Quest: Tim Allen Sigourney Weaver Alan Rickman Tony Shalhoub Sam Rockwell  
Star Trek Into Darkness: Chris Pine Zachary Quinto Zoe Saldana Karl Urban Simon Pegg  
Star Trek Beyond: Chris Pine Zachary Quinto Karl Urban Simon Pegg Zoe Saldana  
Jason X: Kane Hodder Lexa Doig Chuck Campbell Lisa Ryder David Cronenberg  
Space Dogs: Anna Bolshova Evgeny Mironov Sergey Garmash Aleksandr Bashirov Elena Yakovleva  
Alien³: Sigourney Weaver Charles S. Dutton Charles Dance Pete Postlethwaite Ralph Brown  
Gravity: Sandra Bullock George Clooney Ed Harris Orto Ignatiussen Phaldu t Sharma