

```
In [1]: import pandas as pd

In [2]: df=pd.read_csv("f:/dataset/recommend/movies_content.csv")
df
```

Out[2]:	movie_id	description	language	released	rating	writer	director	cast	genre
0	tt5286444	Neerja is the story of the courageous Neerja B...	["Hindi"]	2016-02-19T00:00:00.000Z	7.9	["Saiwyn Quadras", "Sanyukta Shaikh Chawla"]	["Ram Madhvani"]	["Sonam Kapoor", "Shabana Azmi", "Yogendra Ti..."]	"Biogra... "Dr... "Thri..."
1	tt4434004	A story that revolves around drug abuse in the...	["Hindi", "Panjabi"]	2016-06-17T00:00:00.000Z	7.9	["Sudip Sharma", "Abhishek Chaubey"]	["Abhishek Chaubey"]	["Alia Bhatt", "Shahid Kapoor", "Diljit Dosan..."]	["Ci... "Dr... "Thri..."
2	tt0248126	Yashvardhan Raichand lives a very wealthy life...	["Hindi", "English", "Urdu"]	2001-12-14T00:00:00.000Z	7.5	["Karan Johar", "Sheena Parikh"]	["Karan Johar"]	["Amitabh Bachchan", "Jaya Bhaduri", "Shah Ru..."]	["Dr... "Mu... "Roma..."
3	tt0347304	Naina, an introverted, perpetually depressed g...	["Hindi", "Urdu", "Gujarati", "Panjabi"]	2003-11-28T00:00:00.000Z	8.0	["Niranjan Iyengar", "Karan Johar"]	["Nikkhil Advani"]	["Shah Rukh Khan", "Preity Zinta", "Saif Ali ..."]	["Corr... "Dr... "Roma..."
4	tt3043252	'Parched' is a story about women set in the he...	["Hindi", "English"]	2016-06-17T00:00:00.000Z	7.6	["Supratik Sen", "Leena Yadav"]	["Leena Yadav"]	["Tannishtha Chatterjee", "Radhika Apte", "Le..."]	["Dra..."
...
2845	tt0308793	NaN	["Kannada"]	NaN	7.7	[]	["V. Ravichandran"]	["V. Ravichandran", "Kushboo", "Anant Nag", "..."]	["Cri..."
2846	tt3006508	NaN	["Kannada"]	2013-02-22	7.7	[]	["Nagashekar"]	["Chethan Kumar", "Nithya Menon", "Sarath Kum..."]	["Rom..."
2847	tt5348692	The plot is simple. Ranganna (Rangayana Raghu)...	["Kannada"]	2013-11-29	7.7	["Pc Shekhar"]	["Pc Shekhar"]	["Rangayana Raghu", "Sadhu Kokila", "Avinash"]	["Corr... "Dra..."
2848	tt3171754	NaN	["Kannada"]	2013-12-27	7.6	[]	["Manju Swaraj"]	["Ganesh", "Amulya", "Anant Nag", "Tara", "Av..."]	["Corr... "Roma..."
2849	tt0231455	NaN	["Kannada"]	NaN	7.5	[]	["S.R. Puttana Kanagal"]	[]	[]

2850 rows × 10 columns

```
In [3]: df=df[['name','description']]
df
```

```
Out[3]:
```

	name	description
0	Neerja	Neerja is the story of the courageous Neerja B...
1	Udta Punjab	A story that revolves around drug abuse in the...
2	Kabhi Khushi Kabhie Gham...	Yashvardhan Raichand lives a very wealthy life...
3	Kal Ho Naa Ho	Naina, an introverted, perpetually depressed g...
4	Parched	'Parched' is a story about women set in the he...
...
2845	Ranadheera	NaN
2846	Myna	NaN
2847	Chaddi Dosth	The plot is simple. Ranganna (Rangayana Raghu)...
2848	Shravani Subramanya	NaN
2849	Dharmasere	NaN

2850 rows × 2 columns

```
In [4]: df.drop_duplicates(inplace=True)
```

```
C:\Users\Ducat\AppData\Local\Temp\ipykernel_10804\3006716147.py:1: SettingWithCopyWarnin
g:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
guide/indexing.html#returning-a-view-versus-a-copy
df.drop_duplicates(inplace=True)
```

```
In [5]: df
```

```
Out[5]:
```

	name	description
0	Neerja	Neerja is the story of the courageous Neerja B...
1	Udta Punjab	A story that revolves around drug abuse in the...
2	Kabhi Khushi Kabhie Gham...	Yashvardhan Raichand lives a very wealthy life...
3	Kal Ho Naa Ho	Naina, an introverted, perpetually depressed g...
4	Parched	'Parched' is a story about women set in the he...
...
2845	Ranadheera	NaN
2846	Myna	NaN
2847	Chaddi Dosth	The plot is simple. Ranganna (Rangayana Raghu)...
2848	Shravani Subramanya	NaN
2849	Dharmasere	NaN

2837 rows × 2 columns

```
In [6]: df.dropna(inplace=True)
```

C:\Users\Ducat\AppData\Local\Temp\ipykernel_10804\1379821321.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df.dropna(inplace=True)

```
In [7]: df
```

```
Out[7]:
```

	name	description
0	Neerja	Neerja is the story of the courageous Neerja B...
1	Udta Punjab	A story that revolves around drug abuse in the...
2	Kabhi Khushi Kabhie Gham...	Yashvardhan Raichand lives a very wealthy life...
3	Kal Ho Naa Ho	Naina, an introverted, perpetually depressed g...
4	Parched	'Parched' is a story about women set in the he...
...
2838	Run Antony	Antony is depressed and wants to commit suicid...
2841	Jolly Days	A bubbly tale, the story is about four pairs o...
2842	The Great Story of Sodabuddi	The film tells the story of Sodabuddi, who spe...
2844	Vamshi	Vamshi, under the eye of his mother is living ...
2847	Chaddi Dosth	The plot is simple. Ranganna (Rangayana Raghu)...

1751 rows × 2 columns

```
In [8]: s1="python is a Programming language"  
s2="india lost world cup 2023"  
s3="java was trending programming Language"
```

```
In [15]: from sklearn.feature_extraction.text import TfidfVectorizer, ENGLISH_STOP_WORDS
```

```
In [16]: tf=TfidfVectorizer(lowercase=True, stop_words=list(ENGLISH_STOP_WORDS))  
X=tf.fit_transform([s1,s2,s3]).toarray()
```

```
In [17]: X
```

```
Out[17]: array([[0.         , 0.         , 0.         , 0.         , 0.51785612,  
          0.         , 0.51785612, 0.68091856, 0.         , 0.         ],  
        [0.4472136 , 0.4472136 , 0.4472136 , 0.         , 0.         ,  
          0.4472136 , 0.         , 0.         , 0.         , 0.4472136 ],  
        [0.         , 0.         , 0.         , 0.5628291 , 0.42804604,  
          0.         , 0.42804604, 0.         , 0.5628291 , 0.         ]])
```

```
In [ ]: #how to find similarty b/w 2 docs  
>obtain vector representation of each doc using any vectorizer technique.  
>use euclidean or cosine metric to find similiarty scores(cosine is prefered)
```

```
In [23]: from sklearn.metrics.pairwise import cosine_similarity, cosine_distances
```

In [19]: cosine_similarity(X)

Out[19]: array([[1. , 0. , 0.44333251],
 [0. , 1. , 0.],
 [0.44333251, 0. , 1.]])

In [20]: s1

Out[20]: 'python is a programming language'

In [21]: s2

Out[21]: 'india lost world cup 2023'

In [22]: s3

Out[22]: 'java was trending programming language'

In [24]: cosine_distances(X)

Out[24]: array([[0. , 1. , 0.55666749],
 [1. , 0. , 1.],
 [0.55666749, 1. , 0.]])

In [25]: df

Out[25]:

	name	description
0	Neerja	Neerja is the story of the courageous Neerja B...
1	Udta Punjab	A story that revolves around drug abuse in the...
2	Kabhi Khushi Kabhie Gham...	Yashvardhan Raichand lives a very wealthy life...
3	Kal Ho Naa Ho	Naina, an introverted, perpetually depressed g...
4	Parched	'Parched' is a story about women set in the he...
...
2838	Run Antony	Antony is depressed and wants to commit suicid...
2841	Jolly Days	A bubbly tale, the story is about four pairs o...
2842	The Great Story of Sodabuddi	The film tells the story of Sodabuddi, who spe...
2844	Vamshi	Vamshi, under the eye of his mother is living ...
2847	Chaddi Dosth	The plot is simple. Ranganna (Rangayana Raghu)...

1751 rows × 2 columns

In [26]: df=df.reset_index(drop=True)

In [27]: df

Out[27]:

	name	description
0	Neerja	Neerja is the story of the courageous Neerja B...
1	Udta Punjab	A story that revolves around drug abuse in the...
2	Kabhi Khushi Kabhie Gham...	Yashvardhan Raichand lives a very wealthy life...
3	Kal Ho Naa Ho	Naina, an introverted, perpetually depressed g...

4	Parched	'Parched' is a story about women set in the he...
...
1746	Run Antony	Antony is depressed and wants to commit suicid...
1747	Jolly Days	A bubbly tale, the story is about four pairs o...
1748	The Great Story of Sodabuddi	The film tells the story of Sodabuddi, who spe...
1749	Vamshi	Vamshi, under the eye of his mother is living ...
1750	Chaddi Dosth	The plot is simple. Ranganna (Rangayana Raghu)...

1751 rows × 2 columns

```
In [28]: tf=TfidfVectorizer(lowercase=True,stop_words='english')
X=tf.fit_transform(df.description).toarray()
```

```
In [30]: len(tf.get_feature_names_out())
```

```
Out[30]: 7702
```

```
In [32]: from sklearn.neighbors import NearestNeighbors
```

```
In [33]: model=NearestNeighbors(metric='cosine')
model.fit(X)
```

```
Out[33]: ▼ NearestNeighbors
NearestNeighbors(metric='cosine')
```

```
In [34]: name="Neerja".lower()
X[0]
```

```
Out[34]: array([0., 0., 0., ..., 0., 0., 0.])
```

```
In [36]: model.kneighbors([X[0]],n_neighbors=5)
```

```
Out[36]: (array([[0.          , 0.85030796, 0.87983038, 0.92082846, 0.93608081]]),
array([[ 0, 1327, 107, 1563, 560]], dtype=int64))
```

```
In [41]: df.iloc[560]
```

```
Out[41]: name                                Rikki-Tikki-Tavi
description  Rikki is a young mongoose who is adopted by a ...
Name: 560, dtype: object
```

```
In [42]: model.kneighbors([X[3]],n_neighbors=5)
```

```
Out[42]: (array([[0.          , 0.83292146, 0.83760027, 0.84226349, 0.87019565]]),
array([[ 3, 105, 993, 146, 1286]], dtype=int64))
```

```
In [44]: df.iloc[993]
```

```
Out[44]: name                                Thenmavin Kombath
description  Best friends fall in love with the same girl.
Name: 993, dtype: object
```

```
In [71]: name=input("enter movie name:")
if name in df.name.values:
    index=df[df.name==name].index[0]
    dis,idx=model.kneighbors([X[index]],n_neighbors=6)
    for i in idx[0][1:]:
```

```
        print(df.iloc[i].values[0])  
else:  
    print("movie not found")
```

Sikander
Gelo
Tiger
Maachis
Aa Dinagalu

In []: