

In [29]: `import numpy as np
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import linear_kernel`

In [30]: `movies_df=pd.read_csv('tmdb_5000_movies.csv')
credit_df=pd.read_csv('tmdb_5000_credits.csv')`

In [31]: `movies_df.head()`

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_companies	production_
0	237000000	{["id": 28, "name": "Action"], ["id": 12, "nam...	http://www.avatarmovie.com/	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...	{["iso_3166": "name": "Uni
1	300000000	{["id": 12, "name": "Adventure"], {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	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], {"...	{["iso_3166": "name": "Uni
2	245000000	{["id": 28, "name": "Action"], {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	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...	{["iso_3166": "nam": "K
3	250000000	{["id": 28, "name": "Action"], {"id": 80, "nam...	http://www.thedarkknightrises.com/	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], {"...	{["iso_3166": "name": "Uni
4	260000000	{["id": 28, "name": "Action"], {"id": 12, "nam...	http://movies.disney.com/john-carter	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]]	{["iso_3166": "name": "Uni

In [32]: `credit_df.head()`

	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...

In [33]: `movies_df[ "overview" ]`

0 In the 22nd century, a paraplegic Marine is di...
1 Captain Barbossa, long believed to be dead, ha...
2 A cryptic message from Bond's past sends him o...
3 Following the death of District Attorney Harve...
4 John Carter is a war-weary, former military ca...

...
4798 El Mariachi just wants to play his guitar and ...
4799 A newlywed couple's honeymoon is upended by th...
4800 "Signed, Sealed, Delivered" introduces a dedic...
4801 When ambitious New York attorney Sam is sent t...
4802 Ever since the second grade when he first saw ...
Name: overview, Length: 4803, dtype: object

In [34]: `tfidf=TfidfVectorizer(stop_words="english")
movies_df['overview']= movies_df[ 'overview' ].fillna("")
tfidf_matrix=tfidf.fit_transform(movies_df[ "overview" ])`

In [35]: `cosine_sim= linear_kernel(tfidf_matrix,tfidf_matrix)`

In [36]: `indices= pd.Series(movies_df.index, index=movies_df['original_title']).drop_duplicates()
indices`

original\_title
Avatar 0
Pirates of the Caribbean: At World's End 1
Spectre 2
The Dark Knight Rises 3
John Carter 4

...
El Mariachi 4798
Newlyweds 4799
Signed, Sealed, Delivered 4800
Shanghai Calling 4801
My Date with Drew 4802
Length: 4803, dtype: int64

In [37]: `def get_recomendation(title,cosine_sim=cosine_sim):
print(title)

get_recomendation('The Dark Knight Rises')`

The Dark Knight Rises

In [38]: `def get_recomendation(title,cosine_sim = cosine_sim):
idx = indices[title]
sim_scores = enumerate(cosine_sim[idx])
sim_scores = sorted(sim_scores,key=lambda x: x[1], reverse=True)
sim_scores = sim_scores[1:11]

sim_index = [i[0]for i in sim_scores]
print(movies_df[ "original_title" ].iloc[sim_index])

get_recomendation('The Dark Knight Rises')`

65 The Dark Knight
299 Batman Forever
428 Batman Returns
1359 Batman
3854 Batman: The Dark Knight Returns, Part 2
119 Batman Begins
2507 Slow Burn
9 Batman v Superman: Dawn of Justice
1181 JFK
210 Batman & Robin
Name: original\_title, dtype: object

In [39]: `get_recomendation('The Dark Knight Rises')`

65 The Dark Knight
299 Batman Forever
428 Batman Returns
1359 Batman
3854 Batman: The Dark Knight Returns, Part 2
119 Batman Begins
2507 Slow Burn
9 Batman v Superman: Dawn of Justice
1181 JFK
210 Batman & Robin
Name: original\_title, dtype: object

In [40]: `get_recomendation('The Matrix')`

1281 Hackers
2996 Commando
2088 Pulse
1341 Obitaemyy Ostrov
333 Transcendence
0 Avatar
261 Live Free or Die Hard
775 Supernova
125 The Matrix Reloaded
2614 The Love Letter
Name: original\_title, dtype: object

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: