

```
In [9]: #import dataset and libraries
import pandas as pd
import numpy as np
import ast
from sklearn.feature_extraction.text import CountVectorizer
import nltk
from nltk.stem.porter import PorterStemmer
ps=PorterStemmer()
from sklearn.metrics.pairwise import cosine_similarity

movies=pd.read_csv('tmdb_5000_movies.csv')
credits=pd.read_csv('tmdb_5000_credits.csv')
```

```
In [2]: movies.head()
```

| Out[2]: | budget | genres | homepage | id | keywords | original_language | original_title | overview |
|---------|-----------|--|--|--------|--|-------------------|--|--|
| 0 | 237000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://www.avatarmovie.com/ | 19995 | [{"id": 1463, "name": "culture clash"}, {"id": 1463, "name": "culture clash"}] | en | Avatar | In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes torn between following orders and protecting those who have become his family. |
| 1 | 300000000 | [{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}] | http://disney.go.com/disneypictures/pirates/ | 285 | [{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}] | en | Pirates of the Caribbean: At World's End | Captain Barbossa, long believed to be dead, has returned to the Caribbean Sea. |
| 2 | 245000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://www.sonypictures.com/movies/spectre/ | 206647 | [{"id": 470, "name": "spy"}, {"id": 818, "name": "thriller"}] | en | Spectre | A cryptic message from Bond's past sends him on a quest to find a new villain who can challenge him on his own terms. |
| 3 | 250000000 | [{"id": 28, "name": "Action"}, {"id": 80, "name": "Crime"}] | http://www.thedarkknightrises.com/ | 49026 | [{"id": 849, "name": "dc comics"}, {"id": 853, "name": "superhero"}] | en | The Dark Knight Rises | Following the death of District Attorney Harvey Dent, Batman deduces that the only person who can stop the gangster Hugo Weasel is the man who has become his enemy: the Dark Knight. |
| 4 | 260000000 | [{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}] | http://movies.disney.com/john-carter | 49529 | [{"id": 818, "name": "based on novel"}, {"id": 818, "name": "based on novel"}] | en | John Carter | John Carter is a war-weary, former military man, lately unable to find a peaceful life in the only place he feels at home in America: the desert. |

```
In [9]: credits.head()
```

| Out[9]: | movie_id | title | cast | crew |
|---------|----------|--|---|--|
| 0 | 19995 | Avatar | [{"cast_id": 242, "character": "Jake Sully", "credit_id": "52fe48009251416c750aca23", "de..."}] | [{"credit_id": "52fe48009251416c750aca23", "de..."}] |
| 1 | 285 | Pirates of the Caribbean: At World's End | [{"cast_id": 4, "character": "Captain Jack Sparrow", "credit_id": "52fe4232c3a36847f800b579", "de..."}] | [{"credit_id": "52fe4232c3a36847f800b579", "de..."}] |
| 2 | 206647 | Spectre | [{"cast_id": 1, "character": "James Bond", "credit_id": "54805967c3a36829b5002c41", "de..."}] | [{"credit_id": "54805967c3a36829b5002c41", "de..."}] |
| 3 | 49026 | The Dark Knight Rises | [{"cast_id": 2, "character": "Bruce Wayne / Batman", "credit_id": "52fe4781c3a36847f81398c3", "de..."}] | [{"credit_id": "52fe4781c3a36847f81398c3", "de..."}] |
| 4 | 49529 | John Carter | [{"cast_id": 5, "character": "John Carter", "credit_id": "52fe479ac3a36847f81398c3", "de..."}] | [{"credit_id": "52fe479ac3a36847f81398c3", "de..."}] |

```
In [11]: # Data Cleaning and processesing
```

```
In [13]: movies=movies.merge(credits,on='title')
```

```
In [15]: movies['title'] = movies['title'].apply(lambda x: x.lower())
```

```
In [17]: movies=movies[['title','genres','movie_id','keywords','overview','cast','crew']]
```

```
In [19]: movies.isnull().sum()
```

Out[19]: title 0
genres 0
movie_id 0
keywords 0
overview 3
cast 0
crew 0
dtype: int64

In [21]: movies.dropna(inplace=True)

In [23]: movies.duplicated().sum()

Out[23]: 0

In [25]: movies.drop_duplicates()

Out[25]:

| | title | genres | movie_id | keywords | overview | cast | crew |
|------|--|---|----------|--|---|---|---|
| 0 | avatar | [{"id": 28, "name": "Action"}, {"id": 12, "nam... | 19995 | [{"id": 1463, "name": "culture clash"}, {"id": ... | In the 22nd century, a paraplegic Marine is di... | [{"cast_id": 242, "character": "Jake Sully", "... | [{"credit_id": "52fe48009251416c750aca23", "de... |
| 1 | pirates of the caribbean: at world's end | [{"id": 12, "name": "Adventure"}, {"id": 14, "... | 285 | [{"id": 270, "name": "ocean"}, {"id": 726, "na... | Captain Barbossa, long believed to be dead, ha... | [{"cast_id": 4, "character": "Captain Jack Spa... | [{"credit_id": "52fe4232c3a36847f800b579", "de... |
| 2 | spectre | [{"id": 28, "name": "Action"}, {"id": 12, "nam... | 206647 | [{"id": 470, "name": "spy"}, {"id": 818, "name... | A cryptic message from Bond's past sends him o... | [{"cast_id": 1, "character": "James Bond", "cr... | [{"credit_id": "54805967c3a36829b5002c41", "de... |
| 3 | the dark knight rises | [{"id": 28, "name": "Action"}, {"id": 80, "nam... | 49026 | [{"id": 849, "name": "dc comics"}, {"id": 853, ... | Following the death of District Attorney Harve... | [{"cast_id": 2, "character": "Bruce Wayne / Ba... | [{"credit_id": "52fe4781c3a36847f81398c3", "de... |
| 4 | john carter | [{"id": 28, "name": "Action"}, {"id": 12, "nam... | 49529 | [{"id": 818, "name": "based on novel"}, {"id": ... | John Carter is a war-weary, former military ca... | [{"cast_id": 5, "character": "John Carter", "c... | [{"credit_id": "52fe479ac3a36847f81398c3", "de... |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 4804 | el mariachi | [{"id": 28, "name": "Action"}, {"id": 80, "nam... | 9367 | [{"id": 5616, "name": "united states\u2013mexi... | El Mariachi just wants to play his guitar and ... | [{"cast_id": 1, "character": "El Mariachi", "c... | [{"credit_id": "52fe44eec3a36847f80b280b", "de... |
| 4805 | newlyweds | [{"id": 35, "name": "Comedy"}, {"id": 10749, "... | 72766 | [] | A newlywed couple's honeymoon is upended by th... | [{"cast_id": 1, "character": "Buzzy", "credit_... | [{"credit_id": "52fe487dc3a368484e0fb013", "de... |
| 4806 | signed, sealed, delivered | [{"id": 35, "name": "Comedy"}, {"id": 18, "nam... | 231617 | [{"id": 248, "name": "date"}, {"id": 699, "nam... | "Signed, Sealed, Delivered" introduces a dedic... | [{"cast_id": 8, "character": "Oliver Olu2019To... | [{"credit_id": "52fe4df3c3a36847f8275ecf", "de... |
| 4807 | shanghai calling | [] | 126186 | [] | When ambitious New York attorney Sam is sent t... | [{"cast_id": 3, "character": "Sam", "credit_id... | [{"credit_id": "52fe4ad9c3a368484e16a36b", "de... |
| 4808 | my date with drew | [{"id": 99, "name": "Documentary"}] | 25975 | [{"id": 1523, "name": "obsession"}, {"id": 224... | Ever since the second grade when he first saw ... | [{"cast_id": 3, "character": "Herself", "credi... | [{"credit_id": "58ce021b9251415a390165d9", "de... |

4806 rows × 7 columns

In [27]: def convert_1(obj):
L=[]
for i in ast.literal_eval(obj):
L.append(i['name'])
return L

In [29]: movies['genres']=movies['genres'].apply(convert_1)

In [31]: movies['keywords']=movies['keywords'].apply(convert_1)

In [33]: def convert_2(obj):
L=[]
counter=0
for i in ast.literal_eval(obj):
if counter !=5:

```

        L.append(i['name'])
        counter+=1
    else:
        break
return L

```

```
In [35]: movies['cast']=movies['cast'].apply(convert_2)
```

```
In [36]: def director(obj):
        L=[]
        for i in ast.literal_eval(obj):
            if i['job']=='Director':
                L.append(i['name'])
                break
        return L

```

```
In [37]: movies['crew']=movies['crew'].apply(director)
```

```
In [38]: movies['overview']=movies['overview'].apply(lambda x:x.split())
```

```
In [39]: movies['genres']=movies['genres'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [40]: movies['keywords']=movies['keywords'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [41]: movies['cast']=movies['cast'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [44]: movies['crew']=movies['crew'].apply(lambda x:[i.replace(" ","") for i in x])
```

```
In [45]: movies['tags']=movies['crew']+movies['cast']+movies['keywords']+movies['genres']+movies['overview']
```

```
In [47]: new_df=movies[['movie_id','title','tags']]
```

```
In [50]: new_df
```

```
Out[50]:
```

| | movie_id | title | tags |
|------|----------|--|---|
| 0 | 19995 | avatar | [JamesCameron, SamWorthington, ZoeSaldana, Sig... |
| 1 | 285 | pirates of the caribbean: at world's end | [GoreVerbinski, JohnnyDepp, OrlandoBloom, Keir... |
| 2 | 206647 | spectre | [SamMendes, DanielCraig, ChristophWaltz, LéaSe... |
| 3 | 49026 | the dark knight rises | [ChristopherNolan, ChristianBale, MichaelCaine... |
| 4 | 49529 | john carter | [AndrewStanton, TaylorKitsch, LynnCollins, Sam... |
| ... | ... | ... | ... |
| 4804 | 9367 | el mariachi | [RobertRodriguez, CarlosGallardo, JaimeHoyos... |
| 4805 | 72766 | newlyweds | [EdwardBurns, EdwardBurns, KerryBishé, MarshaD... |
| 4806 | 231617 | signed, sealed, delivered | [ScottSmith, EricMabius, KristinBooth, Crystal... |
| 4807 | 126186 | shanghai calling | [DanielHsia, DanielHenney, ElizaCoupe, BillPax... |
| 4808 | 25975 | my date with drew | [BrianHerzlinger, DrewBarrymore, BrianHerzling... |

4806 rows × 3 columns

```
In [58]: new_df = new_df.copy()
        new_df['tags'] = new_df['tags'].apply(lambda x: " ".join(x))
```

```
In [60]: new_df
```

| Out[60]: | movie_id | | title | | tags |
|----------|----------|--------|--|---|------|
| | 0 | 19995 | avatar | JamesCameron SamWorthington ZoeSaldana Sigourn... | |
| | 1 | 285 | pirates of the caribbean: at world's end | GoreVerbinski JohnnyDepp OrlandoBloom KeiraKni... | |
| | 2 | 206647 | spectre | SamMendes DanielCraig ChristophWaltz LéaSeydou... | |
| | 3 | 49026 | the dark knight rises | ChristopherNolan ChristianBale MichaelCaine Ga... | |
| | 4 | 49529 | john carter | AndrewStanton TaylorKitsch LynnCollins Samanth... | |
| | ... | ... | ... | | ... |
| | 4804 | 9367 | el mariachi | RobertRodriguez CarlosGallardo JaimeHoyos Pe... | |
| | 4805 | 72766 | newlyweds | EdwardBurns EdwardBurns KerryBishé MarshaDietl... | |
| | 4806 | 231617 | signed, sealed, delivered | ScottSmith EricMabius KristinBooth CrystalLowe... | |
| | 4807 | 126186 | shanghai calling | DanielHsia DanielHenney ElizaCoupe BillPaxton ... | |
| | 4808 | 25975 | my date with drew | BrianHerzlinger DrewBarrymore BrianHerzlinger ... | |

4806 rows × 3 columns

```
In [62]: new_df['tags'] = new_df['tags'].apply(lambda x:x.lower())
```

```
In [64]: new_df['tags'][0]
```

```
Out[64]: 'jamescameron samworthington zoesaldana sigourneyweaver stephenlang michellerodriguez cultureclash future space
war spacecolony society spacetravel futuristic romance space alien tribe alienplanet cgi marine soldier battle
loveaffair antiwar powerrelations mindandsoul 3d action adventure fantasy sciencefiction in the 22nd century, a
paraplegic marine is dispatched to the moon pandora on a unique mission, but becomes torn between following ord
ers and protecting an alien civilization.'
```

```
In [66]: # Vectorization
```

```
In [68]: def stem(text):
    y = []
    for word in nltk.word_tokenize(text):
        y.append(ps.stem(word))
    return " ".join(y)
```

```
In [70]: new_df['tags']=new_df['tags'].apply(stem)
```

```
In [71]: cv=CountVectorizer(max_features=5000,stop_words='english')
```

```
In [72]: vectors=cv.fit_transform(new_df['tags']).toarray()
```

```
In [73]: feature_names = cv.get_feature_names_out()
print(feature_names)

['000' '007' '10' ... 'zone' 'zoo' 'zoeydeschanel']
```

```
In [78]: new_df['tags'][1]
```

```
Out[78]: "goreverbinski johnnydepp orlandobloom keiraknightley stellanskarsgård chowyun-fat ocean drugabus exoticisland
eastindiatradingcompani loveofone'slif traitor shipwreck strongwoman ship allianc calypso afterlif fighter pira
t swashbuckl aftercreditssting adventur fantasi action captain barbossa , long believ to be dead , ha come back
to life and is head to the edg of the earth with will turner and elizabeth swann . but noth is quit as it seem
."
```

```
In [80]: similarity=cosine_similarity(vectors)
```

```
In [81]: sorted(list(enumerate(similarity[0])),reverse=True,key=lambda x:x[1])[1:11]
```

```
Out[81]: [(1216, 0.2555993162216239),
(2409, 0.24795914953354392),
(507, 0.23995201439520164),
(3730, 0.23855936180684895),
(539, 0.2342606428329091),
(582, 0.2295655645606626),
(61, 0.22315184957216838),
(778, 0.2186870084203477),
(1204, 0.2186870084203477),
(151, 0.2186432666440485)]
```

```
In [84]: # Main method
```

```
In [86]: def to_title_case(s):
    return ' '.join(word.capitalize() for word in s.split())
def recommend(movie):
    # Convert the input movie to lowercase
```

```

movie = movie.lower()

# Convert the titles in the dataframe to lowercase for comparison
new_df['title_lower'] = new_df['title'].str.lower()

if movie in new_df['title_lower'].values:
    movie_index = new_df[new_df['title_lower'] == movie].index[0]
    distances = similarity[movie_index]
    movies_list = sorted(list(enumerate(distances)), reverse=True, key=lambda x: x[1])[1:11]

    print(f"The Top 10 Recommendations based on '{movie}' are:")
    for i in movies_list:
        output = to_title_case(new_df.iloc[i[0]].title)
        print(output)

    # for i in movies_list:
    #     output = new_df.iloc[i[0]].title.title()
    #     print(output)

else:
    print('Movie not found')

```

In [88]: # Output

In [90]: # Output 1
recommend('batman')

The Top 10 Recommendations based on 'batman' are:
 Batman
 Batman & Robin
 Batman Begins
 Batman Returns
 The R.m.
 The Dark Knight Rises
 Rockaway
 Micmacs
 Code Of Honor
 Batman Forever

In [92]: # Output 2
recommend('IRON MAN')

The Top 10 Recommendations based on 'iron man' are:
 Iron Man 3
 Iron Man 2
 Avengers: Age Of Ultron
 The Avengers
 Captain America: Civil War
 Thor: The Dark World
 Ant-man
 Guardians Of The Galaxy
 X-men Origins: Wolverine
 X-men

In [94]: # Output 3
recommend('HitMaN')

The Top 10 Recommendations based on 'hitman' are:
 Silent Trigger
 From Russia With Love
 Mr. & Mrs. Smith
 The Bourne Legacy
 Patriot Games
 The Assassin
 Amidst The Devil's Wings
 The Sentinel
 Ninja Assassin
 Bangkok Dangerous

In []: # Done by Piyush Verma as Data Science intern at Growintern
 # For contact pverma20269968@gmail.com