

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

```
In [2]: credits = pd.read_csv('data/tmdb_5000_credits.csv')
movies = pd.read_csv('data/tmdb_5000_movies.csv')
```

```
In [3]: movies.head(2)
```

```
Out[3]:
```

	budget	genres	homepage	id	keywords
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	1995	[{"id": 1463, "name": "culture clash"}, {"id": 1464, "name": "culture clash"}]
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Fantasy"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "ocean"}]

```
In [4]: credits.head(2)
```

```
Out[4]:
```

	movie_id	title	cast	crew
0	1995	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..."}]

```
In [5]: movies.shape
```

```
Out[5]: (4803, 20)
```

```
In [6]: credits.shape
```

```
Out[6]: (4803, 4)
```

```
In [7]: movies = movies.merge(credits,on='title')
```

```
In [8]: movies.head(2)
```

Out[8]:

	budget	genres	homepage	id	keywords
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}]	http://disney.go.com/disneypictures/pirates/	285	[{"id": 1463, "name": "culture clash"}, {"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]

2 rows × 23 columns

In [9]: `movies = movies[['movie_id', 'title', 'overview', 'genres', 'keywords', 'cast', 'crew']]`

In [10]: `movies.head(2)`

Out[10]:

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[{"id": 28, "name": "Action"}, {"id": 12, "name": "Adventure"}]	[{"id": 1463, "name": "culture clash"}, {"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]	[{"cast_id": 242, "character": "Jake Sully", "..."}]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[{"id": 12, "name": "Adventure"}, {"id": 14, "name": "Action"}]	[{"id": 270, "name": "ocean"}, {"id": 726, "name": "pirates"}]	[{"cast_id": 4, "character": "Captain Jack Spa..."}]

In [11]: `movies.isna().sum()`

Out[11]:

```

movie_id    0
title       0
overview    3
genres      0
keywords    0
cast        0
crew        0
dtype: int64

```

In [12]: `movies = movies.dropna()`

In [13]: `movies.isna().sum()`

```
Out[13]: movie_id    0
         title      0
         overview   0
         genres     0
         keywords   0
         cast       0
         crew       0
         dtype: int64
```

```
In [14]: movies.shape
```

```
Out[14]: (4806, 7)
```

```
In [15]: import ast
         def convert(text):
             L = []
             for i in ast.literal_eval(text):
                 L.append(i['name'])
             return L
```

```
In [16]: movies['genres'] = movies['genres'].apply(convert)
         movies.head()
```

Out[16]:

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

In [17]: `movies['keywords'] = movies['keywords'].apply(convert)`
`movies.head()`

Out[17]:

	movie_id	title	overview	genres	keywords	cast
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	{{"cast_id": 242, "character": "Jake Sully", "...
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	{{"cast_id": 4, "character": "Captain Jack Spa...
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	{{"cast_id": 1, "character": "James Bond", "cr...
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	{{"cast_id": 2, "character": "Bruce Wayne / Ba...
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	{{"cast_id": 5, "character": "John Carter", "c...

```
In [18]: movies['cast'] = movies['cast'].apply(convert)
movies.head()
```

Out[18]:

	movie_id	title	overview	genres	keywords	cast	
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	"52fe480092514
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	"52fe4232c3a36
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	"54805967c3a36
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman, A...	"52fe4781c3a36
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton,...	"52fe479ac3a36

```
In [19]: def fetch_director(text):
L = []
for i in ast.literal_eval(text):
    if i['job'] == 'Director':
        L.append(i['name'])
return L
```

```
In [20]: movies['crew'] = movies['crew'].apply(fetch_director)
```

```
In [21]: movies.head(5)
```

Out[21]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman, A...	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton,...	[Andrew Stanton]

In [22]: `movies.head()`

Out[22]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	[Action, Adventure, Fantasy, Science Fiction]	[culture clash, future, space war, space colon...	[Sam Worthington, Zoe Saldana, Sigourney Weave...	[James Cameron]
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	[Adventure, Fantasy, Action]	[ocean, drug abuse, exotic island, east india ...	[Johnny Depp, Orlando Bloom, Keira Knightley, ...	[Gore Verbinski]
2	206647	Spectre	A cryptic message from Bond's past sends him o...	[Action, Adventure, Crime]	[spy, based on novel, secret agent, sequel, mi...	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...	[Sam Mendes]
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	[Action, Crime, Drama, Thriller]	[dc comics, crime fighter, terrorist, secret i...	[Christian Bale, Michael Caine, Gary Oldman, A...	[Christopher Nolan]
4	49529	John Carter	John Carter is a war-weary, former military ca...	[Action, Adventure, Science Fiction]	[based on novel, mars, medallion, space travel...	[Taylor Kitsch, Lynn Collins, Samantha Morton,...	[Andrew Stanton]

```
In [23]: movies['genres'] = movies['genres'].apply(lambda x: ' '.join(x) if isinstance(x, list) else x)
movies['cast'] = movies['cast'].apply(lambda x: ' '.join(x) if isinstance(x, list) else x)
movies['crew'] = movies['crew'].apply(lambda x: ' '.join(x) if isinstance(x, list) else x)
movies['keywords'] = movies['keywords'].apply(lambda x: ' '.join(x) if isinstance(x, list) else x)
```

```
In [24]: movies.head(5)
```


Out[24]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	Action Adventure Fantasy Science Fiction	culture clash future space war space colony so...	Sam Worthington Zoe Saldana Sigourney Weaver S...	James Cameron
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	Adventure Fantasy Action	ocean drug abuse exotic island east india trad...	Johnny Depp Orlando Bloom Keira Knightley Stel...	Gore Verbinski
2	206647	Spectre	A cryptic message from Bond's past sends him o...	Action Adventure Crime	spy based on novel secret agent sequel mi6 bri...	Daniel Craig Christoph Waltz Léa Seydoux Ralph...	Sam Mendes
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	Action Crime Drama Thriller	dc comics crime fighter terrorist secret ident...	Christian Bale Michael Caine Gary Oldman Anne ...	Christopher Nolan
4	49529	John Carter	John Carter is a war-weary, former military ca...	Action Adventure Science Fiction	based on novel mars medallion space travel pri...	Taylor Kitsch Lynn Collins Samantha Morton Wil...	Andrew Stanton

```
In [126... movies['overview'] = movies['overview'].apply(lambda x:x.split())
```

```
In [33]: movies['overview'] = movies['overview'].apply(lambda x: ' '.join(x) if isinstance(x, list) else x)
```

```
In [34]: movies.head()
```

Out[34]:

	movie_id	title	overview	genres	keywords	cast	crew
0	19995	Avatar	In the 22nd century, a paraplegic Marine is di...	Action Adventure Fantasy Science Fiction	culture clash future space war space colony so...	Sam Worthington Zoe Saldana Sigourney Weaver S...	James Cameron
1	285	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	Adventure Fantasy Action	ocean drug abuse exotic island east india trad...	Johnny Depp Orlando Bloom Keira Knightley Stel...	Gore Verbinski
2	206647	Spectre	A cryptic message from Bond's past sends him o...	Action Adventure Crime	spy based on novel secret agent sequel mi6 bri...	Daniel Craig Christoph Waltz Léa Seydoux Ralph...	Sam Mendes
3	49026	The Dark Knight Rises	Following the death of District Attorney Harve...	Action Crime Drama Thriller	dc comics crime fighter terrorist secret ident...	Christian Bale Michael Caine Gary Oldman Anne ...	Christopher Nolan
4	49529	John Carter	John Carter is a war-weary, former military ca...	Action Adventure Science Fiction	based on novel mars medallion space travel pri...	Taylor Kitsch Lynn Collins Samantha Morton Wil...	Andrew Stanton

```
In [36]: movies['tags'] = movies['overview'] + " " + movies['genres'] + " " + movies['key
```

```
In [37]: df = movies.drop(columns=['overview', 'genres', 'keywords', 'cast', 'crew'])
```

```
In [38]: df.head()
```

Out[38]:

	movie_id		title	tags
0	19995		Avatar	In the 22nd century, a paraplegic Marine is di...
1	285	Pirates of the Caribbean: At World's End		Captain Barbossa, long believed to be dead, ha...
2	206647		Spectre	A cryptic message from Bond's past sends him o...
3	49026	The Dark Knight Rises		Following the death of District Attorney Harve...
4	49529		John Carter	John Carter is a war-weary, former military ca...

In [134...]: `df.to_csv('data/Movies.csv')`

In []: `df['tags'] = df['tags'].str.lower()`

In [39]: `from sklearn.feature_extraction.text import CountVectorizer`
`cv = CountVectorizer(max_features=5000, stop_words='english')`

In [40]: `vector = cv.fit_transform(df['tags']).toarray()`

In [41]: `vector.shape`

Out[41]: (4806, 5000)

In [42]: `from sklearn.metrics.pairwise import cosine_similarity`

In [43]: `similarity = cosine_similarity(vector)`

In [44]: `similarity`

Out[44]: array([[1. , 0.10082339, 0.07024632, ..., 0.02909332, 0.03261062,
0.02351585],
[0.10082339, 1. , 0.04657262, ..., 0.04243489, 0.01585505,
0.01143324],
[0.07024632, 0.04657262, 1. , ..., 0.02687774, 0.06025436,
0.],
...,
[0.02909332, 0.04243489, 0.02687774, ..., 1. , 0.06862635,
0.04948717],
[0.03261062, 0.01585505, 0.06025436, ..., 0.06862635, 1. ,
0.03698001],
[0.02351585, 0.01143324, 0. , ..., 0.04948717, 0.03698001,
1.]])

In [45]: `df[df['title'] == 'The Lego Movie'].index[0]`

Out[45]: 744

In [48]: `def recommend(movie):`
`index = df[df['title'] == movie].index[0]`
`distances = sorted(list(enumerate(similarity[index])), reverse=True, key = lam`

```
for i in distances[1:6]:
    print(df.iloc[i[0]].title)
```

In [58]: `recommend('Batman')`

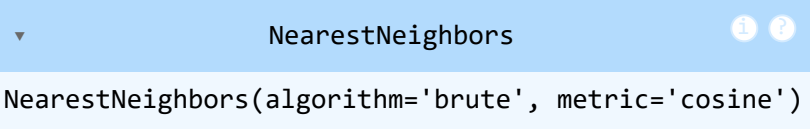
```
Batman
Batman
Batman & Robin
Batman Returns
The Dark Knight
```

In [60]: `from sklearn.model_selection import train_test_split`

In [61]: `y=df['title']`
`X_train, X_test, y_train, y_test = train_test_split(vector, y, test_size=0.2, ra`

In [63]: `from sklearn.neighbors import NearestNeighbors`

In [64]: `knn = NearestNeighbors(metric='cosine', algorithm='brute')`
`knn.fit(X_train)`

Out[64]:  `NearestNeighbors(algorithm='brute', metric='cosine')`

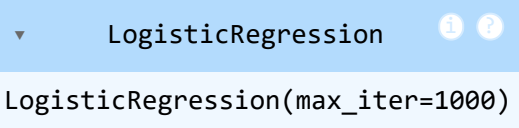
In [118... `def knn_recommendation():`
`distances, indices = knn.kneighbors(X_test[:1], n_neighbors=5)`
`recommendations = []`
`for i in range(len(indices)):`
 `recommendations.append([y_train.iloc[j] for j in indices[i]])`
`return recommendations`

In [119... `print(knn_recommendation())`

```
[['The Savages', 'Last Orders', 'Ramona and Beezus', "Things to Do in Denver When
You're Dead", "Your Sister's Sister"]]
```

In [82]: `from sklearn.linear_model import LogisticRegression`

In [86]: `lr = LogisticRegression(max_iter=1000)`
`lr.fit(X_train, y_train)`

Out[86]:  `LogisticRegression(max_iter=1000)`

In [115... `def lr_accuracy():`
`y_pred = lr.predict(X_test)`
`return accuracy_score(y_test, y_pred)`

In [116... `print(lr_accuracy())`

```
0.002079002079002079
```

In [98]: `from sklearn.naive_bayes import MultinomialNB`

```
In [99]: nb = MultinomialNB()
         nb.fit(X_train, y_train)
```

```
Out[99]: ▼ MultinomialNB ⓘ ?
         MultinomialNB()
```

```
In [108... from sklearn.metrics import accuracy_score, mean_squared_error
```

```
In [124... def nb_accuracy():
            y_pred = nb.predict(X_test)
            return accuracy_score(y_test, y_pred)
```

```
In [127... print(nb_accuracy())
```

```
0.0031185031185031187
```

```
In [ ]: cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
def get_recommendations(title, cosine_sim=cosine_sim):
    idx = movies[movies['title'].str.contains(title, case=False)].index[0]
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    sim_scores = sim_scores[1:11]
    movie_indices = [i[0] for i in sim_scores]
    return movies['title'].iloc[movie_indices]
print("Recommended Movies:")
recommendations = get_recommendations('Avatar')
for movie in recommendations:
    print(movie)
```

```
In [129... cosine_sim = cosine_similarity(vector, vector)
```

```
In [130... def get_recommendations(title, cosine_sim=cosine_sim):
            idx = movies[movies['title'].str.contains(title, case=False)].index[0]
            sim_scores = list(enumerate(cosine_sim[idx]))
            sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
            sim_scores = sim_scores[1:11]
            movie_indices = [i[0] for i in sim_scores]
            return movies['title'].iloc[movie_indices]
```

```
In [133... print("Recommended Movies:")
            recommendations = get_recommendations('Batman')
            for movie in recommendations:
                print(movie)
```

```
Recommended Movies:
The Dark Knight
Man of Steel
The Dark Knight Rises
Real Steel
Superman
Now You See Me 2
Batman Begins
Suicide Squad
Ted 2
Rock of Ages
```