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

```python
In [2]: import numpy as np
```

```python
In [8]: from sklearn.feature_extraction.text import CountVectorizer
```

```python
In [9]: from sklearn.metrics.pairwise import cosine_similarity
```

```python
In [10]: df = pd.read_csv("movie_dataset.csv")
```

```python
In [11]: df
```

Out[11]:

	index	budget	genres	homepage	id	keywords	original_langua
0	0	237000000	Action Adventure Fantasy Science Fiction	http://www.avatarmovie.com/	19995	culture clash future space war space colony so...	
1	1	300000000	Adventure Fantasy Action	http://disney.go.com/disneypictures/pirates/	285	ocean drug abuse exotic island east india trad...	
2	2	245000000	Action Adventure Crime	http://www.sonypictures.com/movies/spectre/	206647	spy based on novel secret agent sequel mi6	
3	3	250000000	Action Crime Drama Thriller	http://www.thedarkknightrises.com/	49026	dc comics crime fighter terrorist secret ident...	
4	4	260000000	Action Adventure Science Fiction	http://movies.disney.com/john-carter	49529	based on novel mars medallion space travel pri...	
...	...	...	...	...	...	...	
4798	4798	220000	Action Crime Thriller	NaN	9367	united states\u2013mexico barrier legs arms pa...	
4799	4799	9000	Comedy Romance	NaN	72766	NaN	
4800	4800	0	Comedy Drama Romance TV Movie	http://www.hallmarkchannel.com/signedsealeddel...	231617	date love at first sight narration investigati...	
4801	4801	0	NaN	http://shanghaicalling.com/	126186	NaN	
4802	4802	0	Documentary	NaN	25975	obsession camcorder crush dream girl	

4803 rows × 24 columns

```python
In [12]: features = ['keywords','cast','genres','director']
```

```python
In [13]: features
```

Out[13]: ['keywords', 'cast', 'genres', 'director']

```python
In [14]: def combine_features(row):
             return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']
```

```python
In [15]: def combine_features(row):
             return row['keywords']+" "+row['cast']+" "+row['genres']+" "+row['director']
```

```python
In [16]: for feature in features:
             df[feature] = df[feature].fillna('') #filling all NaNs with blank string
```

```python
In [20]: df["combined_features"] = df.apply(combine_features,axis=1) #applying combined_features() method over each rows of dataframe and storing the combined string in "combined_features" column
         df.iloc[0].combined_features
```

Out[20]: 'culture clash future space war space colony society Sam Worthington Zoe Saldana Sigourney Weaver Stephen Lang Michelle Rodriguez Action Adventure Fantasy Science Fiction James Cameron'

```python
In [ ]:
```

```python
In [26]: cv = CountVectorizer() #creating new CountVectorizer() object
         count_matrix = cv.fit_transform(df["combined_features"]) #feeding combined strings(movie contents) to CountVectorizer() object
```

```python
In [27]: cosine_sim = cosine_similarity(count_matrix)
```

```python
In [29]: def get_title_from_index(index):
             return df[df.index == index]["title"].values[0]
         def get_index_from_title(title):
             return df[df.title == title]["index"].values[0]
```

```python
In [31]: movie_user_likes = "Avatar"
         movie_index = get_index_from_title(movie_user_likes)
         similar_movies = list(enumerate(cosine_sim[movie_index]))
```

```python
In [32]: sorted_similar_movies = sorted(similar_movies,key=lambda x:x[1],reverse=True)[1:]
```

```python
In [36]: i=0
         print("Top 5 similar movies to "+movie_user_likes+" are:\n")
         for element in sorted_similar_movies:
             print(get_title_from_index(element[0]))
             i=i+1
             if i>50:
                 break
```

```
Top 5 similar movies to Avatar are:

Guardians of the Galaxy
Aliens
Star Wars: Clone Wars: Volume 1
Star Trek Into Darkness
Star Trek Beyond
Alien
Lockout
Jason X
The Helix... Loaded
Moonraker
Planet of the Apes
Galaxy Quest
Gravity
Alien³
Jupiter Ascending
The Wolverine
Silent Running
Zathura: A Space Adventure
Trekkies
Cargo
Wing Commander
Star Trek
Lost in Space
Babylon A.D.
The Fifth Element
Oblivion
Titan A.E.
AVP: Alien vs. Predator
The Empire Strikes Back
Dragonball Evolution
Superman Returns
Divergent
John Carter
The Black Hole
The Ice Pirates
Memoirs of an Invisible Man
Starship Troopers
The Astronaut's Wife
Machete Kills
Soldier
The Abyss
Damnation Alley
Men in Black
Space Cowboys
Space Dogs
The Time Machine
Sheena
Captain America: Civil War
Star Trek: Insurrection
Oz: The Great and Powerful
The One
```

```python
In [ ]:
```