Importing dataset

```
In [2]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
```

```
In [3]: df = pd.read_csv('https://raw.githubusercontent.com/AshishJangra27/Da
        df.head()
```

Out[3]:

	name	year	duration	genre	rating	director	type
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movie
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie
4	12 Angry Men	1957	96	Crime,Drama	9.0	Sidney Lumet	Movie

```
In [193]: ## preprocessing the genre column
          data = []

          for i in df.values:
              if (i[3] != 'Nan'):
                  data.append([i[0],i[1],i[2],i[3],i[4],i[5],i[6]])

          df = pd.DataFrame(data,columns = ['name','year','duration','genre','r
          df
```

Out[193]:

	name	year	duration	genre	rating	director	typ
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movi
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movi
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movi
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movi
4	12 Angry Men	1957	96	Crime,Drama	9.0	Sidney Lumet	Movi
...	...	...	...	...	...	...	.
1221	Luther	2010	60	Crime,Drama,Mystery	8.4	Idris Elba,Dermot Crowley,Michael Smiley	Serie
1222	Wentworth	2013	45	Crime,Drama	8.4	Kate Atkinson,Katrina Milosevic,Robbie Magasiva	Serie
1223	Vikings	2013	44	Action,Adventure,Drama	8.4	Katheryn Winnick,Gustaf Skarsgård,Alexander Lu...	Serie
1224	Sseulsseulhago Chalranhashin: Dokkaebi	2016	120	Comedy,Drama,Fantasy	8.4	Gong Yoo,Kim Go-eun,Lee Dong-Wook	Serie
1225	Normal People	2020	339	Drama,Romance	8.4	Daisy Edgar-Jones,Paul Mescal,Desmond Eastwood	Serie

1226 rows × 7 columns

```
In [194]: genres = []
          for i in df['genre']:
              genres += (i.split(','))
          genres = list(set(genres))
```

```
In [195]: one_hot = []
          for i in df['genre']:
              i = [j.strip() for j in i.split(',')]
              enc = [0 for encode in range(len(genres))]
              for genre in i:
                  enc[genres.index(genre)] = 1
              one_hot.append(enc)
```

```
In [196]: one_hot = np.array(one_hot)
```

```
In [197]: for i in range(len(genres)):
              df[genres[i]] = one_hot[:,i]


          df.head()
```

Out[197]:

	name	year	duration	genre	rating	director	type	Music	Adventu
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movie	0	
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie	0	
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie	0	
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie	0	
4	12 Angry Men	1957	96	Crime,Drama	9.0	Sidney Lumet	Movie	0	

5 rows × 34 columns

```
In [198]: for i in range(len(genres)):
              df[genres[i]] = one_hot[:,i]

          df.head()
```

Out[198]:

	name	year	duration	genre	rating	director	type	Music	Adventu
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movie	0	
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie	0	
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie	0	
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie	0	
4	12 Angry Men	1957	96	Crime,Drama	9.0	Sidney Lumet	Movie	0	

5 rows × 34 columns

```
In [201]: len(df.columns) - len(genres)
```

Out[201]: 7

1. Ask from the user to select a movie

```
In [9]: movie_selection = input('select your movie :')
        movie_selection = movie_selection.strip()
        user_enc = df[df['name'] == movie_selection].values[0][7:]
        count = 1
        for i in df.values:
            if (list(user_enc - i[7:]).count(0) == len(user_enc)):
                print(count ,'|',i[0],'|',i[3])
                count += 1
```

```
select you movie :The Godfather
1 | The Godfather | Crime,Drama
2 | The Godfather: Part II | Crime,Drama
3 | 12 Angry Men | Crime,Drama
4 | Pulp Fiction | Crime,Drama
5 | Cidade de Deus | Crime,Drama
6 | Once Upon a Time in America | Crime,Drama
7 | Scarface | Crime,Drama
8 | Taxi Driver | Crime,Drama
9 | Casino | Crime,Drama
10 | To Kill a Mockingbird | Crime,Drama
11 | La haine | Crime,Drama
12 | Cool Hand Luke | Crime,Drama
13 | Les quatre cents coups | Crime,Drama
14 | Dogville | Crime,Drama
15 | Bound by Honor | Crime,Drama
16 | Manbiki kazoku | Crime,Drama
17 | Bir Zamanlar Anadolu'da | Crime,Drama
18 | Un prophète | Crime,Drama
19 | Once Were Warriors | Crime,Drama
20 | Drive | Crime,Drama
21 | Lilja 4-ever | Crime,Drama
22 | Boyz n the Hood | Crime,Drama
23 | À bout de souffle | Crime,Drama
24 | This Is England | Crime,Drama
25 | Leviafan | Crime,Drama
26 | The Godfather: Part III | Crime,Drama
27 | The Sopranos | Crime,Drama
28 | Better Call Saul | Crime,Drama
29 | Peaky Blinders | Crime,Drama
30 | Mr Inbetween | Crime,Drama
31 | Delhi Crime | Crime,Drama
32 | Wentworth | Crime,Drama
```

2. Find the directors of the movie that user has selected

```
In [10]: ## single line solution to find the name of the director the input mc
         df[df['name'] == movie_selection]['director']
```

```
Out[10]: 1    Francis Ford Coppola
         Name: director, dtype: object
```

```
In [11]:  # using data structure
          for i in df.values:
              if (i[0] == movie_selection ):
                  print('Name of the movie is  :',movie_selection)
                  print('Director of the movie :',i[5])

          Name of the movie is  : The Godfather
          Director of the movie : Francis Ford Coppola
```

3. Recommend them top 10 movies based on the directors of the movie that user has selected

```
In [13]:  data = []
          for i in df.values:
              if (i[0] == movie_selection) and (i[6] == 'Movie'):
                  for j in range(20):
                      try:
                          data.append(df[df['director'] == i[5]].values[j][:7])
                      except:
                          pass

          df_ = pd.DataFrame(data,columns= ['name','year','duration','genre','
          df_
```

Out[13]:

	name	year	duration	genre	rating	director	type
0	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie
1	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie
2	Apocalypse Now	1979	147	Drama,Mystery,War	8.4	Francis Ford Coppola	Movie
3	The Conversation	1974	113	Drama,Mystery,Thriller	7.8	Francis Ford Coppola	Movie
4	The Godfather: Part III	1990	162	Crime,Drama	7.6	Francis Ford Coppola	Movie

4. Also recommend the movies based on the name of the movie (you need to perform one hot encoding on name column of the dataset and finding the similarity) for example The Godfather and The Godfather - II

```
In [241]:  movies = []
           for i in df['name']:
               movies += i.split(' ')
           movies = (list(set(movies)))
```

```
In [242]: hot_encode = []

          for i in df['name']:
              i = ([words.strip() for words in i.split(' ')])
              enco = [0] * len(movies)
              for movie in i:
                  enco[movies.index(movie)] = 1
              hot_encode.append(enco)

          hot_encode = np.array(hot_encode)
```

```
In [243]: for i in range(len(movies)):
              df[movies[i]] = hot_encode[:,i]

          df.head()
```

Out[243]:

	name	year	duration	genre	rating	director	type	Music	Adventu...
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Movie	0	
1	The Godfather	1972	175	Crime,Drama	9.2	Francis Ford Coppola	Movie	0	
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie	0	

	name	year	duration	genre	rating	director	type	Music	Adventu...
3	The Godfather: Part II	1974	202	Crime,Drama	9.0	Francis Ford Coppola	Movie	0	

In [244]: `len(df.columns)-len(movies)`

Out[244]: 24

In [245]:
```python
def mod(lst):
    s = 0
    for i in lst:
        if i < 0:
            s += -i
        else:
            s += i
    return s
```

In [246]: `# top 10 movies`

```
In [251]: movie_name = input('Enter your movie name : ').strip()
          user = df[df['name'] == movie_name].values[0][23:]

          rel = []

          for i in df.values:
              rel.append(mod(i[23:] - user))

          df['rel'] = rel
          df.sort_values(by = 'rel',ascending= True).head(10)
```

Enter your movie name : The Dark Knight

Out[251]:

	name	year	duration	genre	rating	director	type	M
2	The Dark Knight	2008	152	Action,Crime,Drama	9.0	Christopher Nolan	Movie	
63	The Dark Knight Rises	2012	164	Action,Adventure	8.4	Christopher Nolan	Movie	
271	The Thing	1982	109	Horror,Mystery,Sci-Fi	8.1	John Carpenter	Movie	
347	The Help	2011	146	Drama	8.0	Tate Taylor	Movie	
36	The Prestige	2006	130	Drama,Mystery,Sci-Fi	8.5	Christopher Nolan	Movie	
1186	The Office	2001	30	Comedy,Drama	8.5	Ricky Gervais,Martin Freeman,Mackenzie Crook	Series	
357	The Avengers	2012	143	Action,Adventure,Sci-Fi	8.0	Joss Whedon	Movie	
481	The Artist	2011	100	Comedy,Drama,Romance	7.9	Michel Hazanavicius	Movie	
1183	The Expanse	2015	60	Drama,Mystery,Sci-Fi	8.5	Wes Chatham,Shohreh Aghdashloo,Steven Strait	Series	
838	The Verdict	1982	129	Drama	7.7	Sidney Lumet	Movie	

10 rows × 2074 columns