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

In [2]:
```
data=pd.read_csv("Movie.csv")
data.head()
```

Out[2]:

	userId	movie	rating
0	3	Toy Story (1995)	4.0
1	6	Toy Story (1995)	5.0
2	8	Toy Story (1995)	4.0
3	10	Toy Story (1995)	4.0
4	11	Toy Story (1995)	4.5

In [3]:
```
df=pd.pivot_table(data, values="rating", index="userId" ,columns="movie")
df
```

Out[3]:

movie userId	Father of the Bride Part II (1995)	GoldenEye (1995)	Grumpier Old Men (1995)	Heat (1995)	Jumanji (1995)	Sabrina (1995)	Sudden Death (1995)	Tom and Huck (1995)	Toy Story (1995)	Waiting to Exhale (1995)
1	NaN	NaN	NaN	NaN	3.5	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN
4	NaN	4.0	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
7115	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7116	3.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN
7117	NaN	3.0	4.0	5.0	NaN	3.0	1.0	NaN	4.0	NaN
7119	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0	NaN
7120	NaN	NaN	NaN	NaN	4.0	4.0	NaN	NaN	4.5	NaN

4081 rows × 10 columns

In [4]:
```
df=df.fillna(0)
df
```

Out[4]:

movie userId	Father of the Bride Part II (1995)	GoldenEye (1995)	Grumpier Old Men (1995)	Heat (1995)	Jumanji (1995)	Sabrina (1995)	Sudden Death (1995)	Tom and Huck (1995)	Toy Story (1995)	Waiting to Exhale (1995)
1	0.0	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
4	0.0	4.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
7115	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7116	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
7117	0.0	3.0	4.0	5.0	0.0	3.0	1.0	0.0	4.0	0.0
7119	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
7120	0.0	0.0	0.0	0.0	4.0	4.0	0.0	0.0	4.5	0.0

4081 rows × 10 columns

In [5]:
```
#similarity score

from sklearn.metrics import pairwise_distances
from scipy.spatial.distance import correlation,cosine
```

In [6]:
```
user_sim=1-pairwise_distances(df.values,metric="cosine")
user_sim_df=pd.DataFrame(user_sim)
user_sim_df
```

Out[6]:

	0	1	2	3	4	5	6	7	8	9	...	4071	4072	4073	4074	4075	4076	4077	4078	
0	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	1.000000	0.707107	0.000000	0.000000	0.000000	0.000000	0.00
1	0.000000	1.000000	0.000000	0.000000	0.000000	0.390567	0.707107	0.615457	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.458831	0.00
2	0.000000	0.000000	1.000000	0.000000	0.000000	0.650945	0.000000	0.492366	1.000000	0.874157	...	0.000000	1.000000	0.000000	0.707107	0.000000	0.000000	0.752577	0.458831	1.00
3	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.615457	0.000000	0.388514	...	0.800000	0.000000	0.000000	0.000000	0.989949	0.000000	0.000000	0.619422	0.00
4	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	1.000000	0.707107	0.000000	0.000000	0.000000	0.000000	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4076	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.658505	0.000000	0.00
4077	0.000000	0.000000	0.752577	0.000000	0.000000	0.489886	0.000000	0.370543	0.752577	0.657870	...	0.000000	0.752577	0.000000	0.532152	0.000000	0.658505	1.000000	0.345306	0.75
4078	0.000000	0.458831	0.458831	0.619422	0.000000	0.701884	0.567775	0.889532	0.458831	0.568212	...	0.344124	0.458831	0.000000	0.324443	0.648886	0.000000	0.345306	1.000000	0.45
4079	0.000000	0.000000	1.000000	0.000000	0.000000	0.650945	0.000000	0.492366	1.000000	0.874157	...	0.000000	1.000000	0.000000	0.707107	0.000000	0.000000	0.752577	0.458831	1.00
4080	0.553372	0.000000	0.622543	0.000000	0.553372	0.765455	0.391293	0.306519	0.622543	0.544201	...	0.000000	0.622543	0.553372	0.831497	0.000000	0.000000	0.468511	0.476071	0.62

4081 rows × 4081 columns

In [7]:
```
import numpy as np
```

In [8]:
```
np.fill_diagonal(user_sim,0)
user_sim_df.loc[0:5,0:5]
```

Out[8]:

	0	1	2	3	4	5
0	0.0	0.000000	0.000000	0.0	1.0	0.000000
1	0.0	0.000000	0.000000	0.0	0.0	0.390567
2	0.0	0.000000	0.000000	0.0	0.0	0.650945
3	0.0	0.000000	0.000000	0.0	0.0	0.000000
4	1.0	0.000000	0.000000	0.0	0.0	0.000000
5	0.0	0.390567	0.650945	0.0	0.0	0.000000

In [9]:
```
user_sim_df.index=list(df.index)
user_sim_df.columns=list(df.index)
```

In [10]:
```
user_sim_df.sort_values([8],ascending=False).index
```

Out[10]:
```
Int64Index([4831,  614, 2274, 5043, 5843, 4758, 1672, 6739, 3593, 6784,
            ...
            4259, 4454, 4292, 4325, 4330, 4359, 4397, 4409, 4453, 3546],
           dtype='int64', length=4081)
```

In [11]:
```
data[(data["userId"]==8) |(data["userId"]==4831)]
```

Out[11]:

	userId	movie	rating
2	8	Toy Story (1995)	4.0
1766	4831	Toy Story (1995)	3.5
3727	8	Grumpier Old Men (1995)	5.0
4200	4831	Grumpier Old Men (1995)	3.5
5205	8	Heat (1995)	3.0
6066	4831	Heat (1995)	2.5
7445	8	GoldenEye (1995)	4.0
8526	4831	GoldenEye (1995)	3.0