```
import pandas as pd
import numpy as np

data=pd.read_csv('Downloads/Movie.csv')
data
```

	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
...	...	...	...
8987	7087	GoldenEye (1995)	3.0
8988	7088	GoldenEye (1995)	1.0
8989	7105	GoldenEye (1995)	2.0
8990	7113	GoldenEye (1995)	3.0
8991	7117	GoldenEye (1995)	3.0

8992 rows × 3 columns

```
len(data['userId'].unique())
```

```
4081
```

```
len(data['movie'].unique())
```

```
10
```

```
data1=data.pivot(index='userId',columns='movie',values='rating').reset_index(drop=True)
data1
```

movie	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)
0	NaN	NaN	NaN	NaN	3.5	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN
3	NaN	4.0	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
4076	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4077	3.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN
4078	NaN	3.0	4.0	5.0	NaN	3.0	1.0	NaN	4.0	NaN
4079	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	5.0	NaN
4080	NaN	NaN	NaN	NaN	4.0	4.0	NaN	NaN	4.5	NaN

```
data1.index=data['userId'].unique()
data1
```

movie	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)
3	NaN	NaN	NaN	NaN	3.5	NaN	NaN	NaN	NaN	NaN
6	NaN	NaN	4.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN
10	NaN	4.0	NaN	3.0	NaN	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	3.0	NaN	NaN	NaN	NaN	NaN

data1.fillna(0,inplace=True)
data1

movie	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)
3	0.0	0.0	0.0	0.0	3.5	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
10	0.0	4.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
7044	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7070	3.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0
7080	0.0	3.0	4.0	5.0	0.0	3.0	1.0	0.0	4.0	0.0
7087	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0
7105	0.0	0.0	0.0	0.0	4.0	4.0	0.0	0.0	4.5	0.0

4081 rows × 10 columns

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

sim=1-pairwise_distances(data1,metric='cosine')
sim

```
array([[1.        , 0.        , 0.        , ..., 0.        , 0.        ,
        0.55337157],
       [0.        , 1.        , 0.        , ..., 0.45883147, 0.        ,
        0.        ],
       [0.        , 0.        , 1.        , ..., 0.45883147, 1.        ,
        0.62254302],
       ...,
       [0.        , 0.45883147, 0.45883147, ..., 1.        , 0.45883147,
        0.47607054],
       [0.        , 0.        , 1.        , ..., 0.45883147, 1.        ,
        0.62254302],
       [0.55337157, 0.        , 0.62254302, ..., 0.47607054, 0.62254302,
        1.        ]])
```

sim1=pd.DataFrame(sim)
sim1

	0	1	2	3	4	5	6	7	8	9	...	4071	4072	4073	4074	4075
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
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
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
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
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
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
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
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
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
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
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

sim1.iloc[0:5,0:5]

	0	1	2	3	4
0	1.0	0.0	0.0	0.0	1.0
1	0.0	1.0	0.0	0.0	0.0
2	0.0	0.0	1.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0
4	1.0	0.0	0.0	0.0	1.0

data[(data['userId']==6) | (data['userId']==126)]

	userId	movie	rating
1	6	Toy Story (1995)	5.0
40	126	Toy Story (1995)	4.0
3725	6	Grumpier Old Men (1995)	3.0
3739	126	Grumpier Old Men (1995)	4.0
4554	126	Father of the Bride Part II (1995)	4.0
5226	126	Heat (1995)	3.0

user1 = data[(data['userId']==6)]

user2 = data[(data['userId']==126)]

user1.movie

```
1              Toy Story (1995)
3725    Grumpier Old Men (1995)
6464             Sabrina (1995)
Name: movie, dtype: object
```

user2.movie

```
40                          Toy Story (1995)
3739                Grumpier Old Men (1995)
4554    Father of the Bride Part II (1995)
5226                              Heat (1995)
Name: movie, dtype: object
```

pd.merge(user1,user2,on='movie',how='outer')

	userId_x	movie	rating_x	userId_y	rating_y
0	6.0	Toy Story (1995)	5.0	126.0	4.0
1	6.0	Grumpier Old Men (1995)	3.0	126.0	4.0
2	6.0	Sabrina (1995)	5.0	NaN	NaN
3	NaN	Father of the Bride Part II (1995)	NaN	126.0	4.0
4	NaN	Heat (1995)	NaN	126.0	3.0