

```
import numpy as np
import pandas as pd
import warnings
import cv2
```

```
warnings.filterwarnings('ignore')
```

```
df=pd.read_csv('u.data',sep="\t")
```

```
df.head()
```

	196	242	3	881250949
0	186	302	3	891717742
1	22	377	1	878887116
2	244	51	2	880606923
3	166	346	1	886397596
4	298	474	4	884182806

4 298 474 4 884182806

```
columns_name=['user_id','item_id','rating','timestamp']  
df=pd.read_csv('u.data',sep="\t",names=columns_name)
```

```
df.head()
```

	user_id	item_id	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596

4 166 346 1 886397596

```
movies_title=pd.read_csv('u.item',sep="\\",header=None,encoding="ISO-8859-1")
```

```
movies_title.head()
```

[illegible]

	0	1	2	3		4	5	6	7	8	9	...	14	15	16	17	18	19	20	21	22	23
1	2	GoldenEye (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?GoldenEye%20(...	0	1	1	0	0	...	0	0	0	0	0	0	0	0	1	0	0
2	3	Four Rooms (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Four%20Rooms%...	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0	0
3	4	Get Shorty (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Get%20Shorty%...	0	1	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
4	5	Copycat (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Copycat%20(1995)	0	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0	0

5 rows × 24 columns



In [15]:

```
movies_titles=movies_title[[0,1]]
movies_titles.columns=["item_id","title"]
movies_titles.head()
```

Out[15]:

	item_id	title
0	1	Toy Story (1995)
1	2	GoldenEye (1995)
2	3	Four Rooms (1995)
3	4	Get Shorty (1995)
4	5	Copycat (1995)

In [17]:

```
df=pd.merge(df,movies_titles,on="item_id")
```

In [18]:

df

Out[18]:

	user_id	item_id	rating	timestamp	title
0	196	242	3	881250949	Kolya (1996)
1	63	242	3	875747190	Kolya (1996)
2	226	242	5	883888671	Kolya (1996)
3	154	242	3	879138235	Kolya (1996)
4	306	242	5	876503793	Kolya (1996)
...
99995	840	1674	4	891211682	Mamma Roma (1962)
99996	655	1640	3	888474646	Eighth Day, The (1996)
99997	655	1637	3	888984255	Girls Town (1996)
99998	655	1630	3	887428735	Silence of the Palace, The (Saimt el Qusur) (1...
99999	655	1641	3	887427810	Dadetown (1995)

100000 rows × 5 columns

In [19]:

```
df.tail()
```

Out[19]:

	user_id	item_id	rating	timestamp	title
99995	840	1674	4	891211682	Mamma Roma (1962)
99996	655	1640	3	888474646	Eighth Day, The (1996)
99997	655	1637	3	888984255	Girls Town (1996)
99998	655	1630	3	887428735	Silence of the Palace, The (Saimt el Qusur) (1...
99999	655	1641	3	887427810	Dadetown (1995)

In [26]:

```
ratings=pd.DataFrame(df.groupby('title').mean()['rating'])
```

In [28]:

```
ratings.head()
```

Out[28]:

	rating
title	
'Til There Was You (1997)	2.333333
1-900 (1994)	2.600000
101 Dalmatians (1996)	2.908257
12 Angry Men (1957)	4.344000
187 (1997)	3.024390

In [33]:

```
ratings['num of ratings']=pd.DataFrame(df.groupby('title').count()['rating'])
```

In [34]:

```
df.head()
```

Out[34]:

	user_id	item_id	rating	timestamp	title
0	196	242	3	881250949	Kolya (1996)
1	63	242	3	875747190	Kolya (1996)
2	226	242	5	883888671	Kolya (1996)
3	154	242	3	879138235	Kolya (1996)
4	306	242	5	876503793	Kolya (1996)

In [75]:

```
moviemat=df.pivot_table(index="user_id",columns="title",values="rating")
```

In [37]:

```
moviemat.head()
```

Out[37]:

	'Til There Was You (1997)	1-900 (1994)	101 Dalmatians (1996)	12 Angry Men (1957)	187 (1997)	2 Days in the Valley (1996)	20,000 Leagues Under the Sea (1954)	2001: A Space Odyssey (1968)	3 Ninjas: High Noon At Mega Mountain (1998)	39 Steps, The (1935)	...	Yankee Zulu (1994)	Y of t Hol (199
user_id													
1	NaN	NaN	2.0	5.0	NaN	NaN	3.0	4.0	NaN	NaN	...	NaN	N
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	N
3	NaN	NaN	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	...	NaN	N
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	N
5	NaN	NaN	2.0	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	N

5 rows × 1664 columns



In [87]:

```
starwars_user_ratings=moviemat['Star Wars (1977)']
```

```

-----
KeyError                                Traceback (most recent call last)
c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\core\indexes\base.
py in get_loc(self, key, method, tolerance)
    3079         try:
-> 3080             return self._engine.get_loc(casted_key)
    3081         except KeyError as err:

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_item()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'Star Wars (1977)'

```

The above exception was the direct cause of the following exception:

```

KeyError                                Traceback (most recent call last)
<ipython-input-87-9e9c29266b9a> in <module>
----> 1 starwars_user_ratings=moviemat['Star Wars (1977)']

c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\core\frame.py in __
getitem__(self, key)
    3022         if self.columns.nlevels > 1:
    3023             return self._getitem_multilevel(key)
-> 3024         indexer = self.columns.get_loc(key)
    3025         if is_integer(indexer):
    3026             indexer = [indexer]

c:\users\sony\appdata\local\programs\python\python38\lib\site-packages\pandas\core\indexes\base.
py in get_loc(self, key, method, tolerance)
    3080             return self._engine.get_loc(casted_key)
    3081         except KeyError as err:

```

```
-> 3082         raise KeyError(key) from err
3083
3084     if tolerance is not None:
```

KeyError: 'Star Wars (1977)'

In []:

In []: