

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_table(r'C:\Users\HP\Desktop\New folder (2)\movies.data')
df
```

	1	Unnamed: 1	Toy Story (1995)	Unnamed: 3
\				
0	2	NaN	Jumanji (1995)	NaN
1	3	NaN	Grumpier Old Men (1995)	NaN
2	4	NaN	Waiting to Exhale (1995)	NaN
3	5	NaN	Father of the Bride Part II (1995)	NaN
4	6	NaN	Heat (1995)	NaN
...	...	...	...	...
3877	3948	NaN	Meet the Parents (2000)	NaN
3878	3949	NaN	Requiem for a Dream (2000)	NaN
3879	3950	NaN	Tigerland (2000)	NaN
3880	3951	NaN	Two Family House (2000)	NaN
3881	3952	NaN	Contender, The (2000)	NaN

	Animation Children's Comedy	Unnamed: 5	Unnamed: 6
0	Adventure Children's Fantasy	NaN	NaN
1	Comedy Romance	NaN	NaN
2	Comedy Drama	NaN	NaN
3	Comedy	NaN	NaN
4	Action Crime Thriller	NaN	NaN
...	...	...	...
3877	Comedy	NaN	NaN
3878	Drama	NaN	NaN
3879	Drama	NaN	NaN
3880	Drama	NaN	NaN
3881	Drama Thriller	NaN	NaN

```
[3882 rows x 7 columns]
```

```
df.head()
```

1	Unnamed: 1	Toy Story (1995)	Unnamed: 3	\
0	2	NaN	Jumanji (1995)	NaN
1	3	NaN	Grumpier Old Men (1995)	NaN
2	4	NaN	Waiting to Exhale (1995)	NaN
3	5	NaN	Father of the Bride Part II (1995)	NaN
4	6	NaN	Heat (1995)	NaN

	Animation Children's Comedy	Unnamed: 5	Unnamed: 6
0	Adventure Children's Fantasy	NaN	NaN
1	Comedy Romance	NaN	NaN
2	Comedy Drama	NaN	NaN
3	Comedy	NaN	NaN
4	Action Crime Thriller	NaN	NaN

```
df1 = pd.read_csv(r"C:\Users\HP\Desktop\ps\movies.csv")
df1
```

	MovieID	Unnamed: 1	Title	Unnamed:
3	\			
0	1	NaN	Toy Story (1995)	
NaN				
1	2	NaN	Jumanji (1995)	
NaN				
2	3	NaN	Grumpier Old Men (1995)	
NaN				
3	4	NaN	Waiting to Exhale (1995)	
NaN				
4	5	NaN	Father of the Bride Part II (1995)	
NaN				
...	...	...	...	.
..				
3878	3948	NaN	Meet the Parents (2000)	
NaN				
3879	3949	NaN	Requiem for a Dream (2000)	
NaN				
3880	3950	NaN	Tigerland (2000)	
NaN				
3881	3951	NaN	Two Family House (2000)	
NaN				
3882	3952	NaN	Contender, The (2000)	
NaN				

	Genres	Unnamed: 5	Unnamed: 6
0	Animation Children's Comedy	NaN	NaN
1	Adventure Children's Fantasy	NaN	NaN
2	Comedy Romance	NaN	NaN
3	Comedy Drama	NaN	NaN
4	Comedy	NaN	NaN
...	...	...	...
3878	Comedy	NaN	NaN

3879		Drama	NaN	NaN
3880		Drama	NaN	NaN
3881		Drama	NaN	NaN
3882		Drama Thriller	NaN	NaN

[3883 rows x 7 columns]

```
df1.drop(['Unnamed: 1','Unnamed: 3','Unnamed: 5','Unnamed: 6'],axis=1,inplace=True)
df1
```

	MovieID	Title \
0	1	Toy Story (1995)
1	2	Jumanji (1995)
2	3	Grumpier Old Men (1995)
3	4	Waiting to Exhale (1995)
4	5	Father of the Bride Part II (1995)
...	...	...
3878	3948	Meet the Parents (2000)
3879	3949	Requiem for a Dream (2000)
3880	3950	Tigerland (2000)
3881	3951	Two Family House (2000)
3882	3952	Contender, The (2000)

	Genres
0	Animation Children's Comedy
1	Adventure Children's Fantasy
2	Comedy Romance
3	Comedy Drama
4	Comedy
...	...
3878	Comedy
3879	Drama
3880	Drama
3881	Drama
3882	Drama Thriller

[3883 rows x 3 columns]

```
df1.head()
```

	MovieID	Title	Genres
0	1	Toy Story (1995)	Animation Children's Comedy
1	2	Jumanji (1995)	Adventure Children's Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama

4            5    Father of the Bride Part II (1995)  
Comedy

```
df2 = pd.read_csv(r"C:\Users\HP\Desktop\ps\rating.csv")  
df2
```

	UserID	Unnamed: 1	MovieID	Unnamed: 3	Rating	Unnamed:
5 \						
0	1	NaN	1193	NaN	5	NaN
1	1	NaN	661	NaN	3	NaN
2	1	NaN	914	NaN	3	NaN
3	1	NaN	3408	NaN	4	NaN
4	1	NaN	2355	NaN	5	NaN
...	...	...	...	...	...	...
1000204	6040	NaN	1091	NaN	1	NaN
1000205	6040	NaN	1094	NaN	5	NaN
1000206	6040	NaN	562	NaN	5	NaN
1000207	6040	NaN	1096	NaN	4	NaN
1000208	6040	NaN	1097	NaN	4	NaN

	Timestamp
0	978300760
1	978302109
2	978301968
3	978300275
4	978824291
...	...
1000204	956716541
1000205	956704887
1000206	956704746
1000207	956715648
1000208	956715569

[1000209 rows x 7 columns]

```
df2.drop(['Unnamed: 1','Unnamed: 3','Unnamed: 5'],axis=1,inplace=True)  
df2
```

	UserID	MovieID	Rating	Timestamp
0	1	1193	5	978300760
1	1	661	3	978302109
2	1	914	3	978301968
3	1	3408	4	978300275
4	1	2355	5	978824291
...	...	...	...	...
1000204	6040	1091	1	956716541
1000205	6040	1094	5	956704887
1000206	6040	562	5	956704746
1000207	6040	1096	4	956715648
1000208	6040	1097	4	956715569

[1000209 rows x 4 columns]

df2.head()

	UserID	MovieID	Rating	Timestamp
0	1	1193	5	978300760
1	1	661	3	978302109
2	1	914	3	978301968
3	1	3408	4	978300275
4	1	2355	5	978824291

df3 = pd.read\_csv(r"C:\Users\HP\Desktop\ps\user.csv")  
df3

	UserID	Occupation \	Unnamed: 1	Gender	Unnamed: 3	Age	Unnamed: 5
0	1		NaN	F	NaN	1	NaN
10							
1	2		NaN	M	NaN	56	NaN
16							
2	3		NaN	M	NaN	25	NaN
15							
3	4		NaN	M	NaN	45	NaN
7							
4	5		NaN	M	NaN	25	NaN
20							
...	...		...	...	...	...	...
...							
6035	6036		NaN	F	NaN	25	NaN
15							
6036	6037		NaN	F	NaN	45	NaN
1							
6037	6038		NaN	F	NaN	56	NaN
1							
6038	6039		NaN	F	NaN	45	NaN
0							
6039	6040		NaN	M	NaN	25	NaN
6							

	Unnamed: 7	Zip-code
0	NaN	48067
1	NaN	70072
2	NaN	55117
3	NaN	2460
4	NaN	55455
...	...	...
6035	NaN	32603
6036	NaN	76006
6037	NaN	14706
6038	NaN	1060
6039	NaN	11106

[6040 rows x 9 columns]

```
df3.drop(['Unnamed: 1','Unnamed: 3','Unnamed: 5','Unnamed: 7'],axis=1,inplace=True)
df3
```

	UserID	Gender	Age	Occupation	Zip-code
0	1	F	1	10	48067
1	2	M	56	16	70072
2	3	M	25	15	55117
3	4	M	45	7	2460
4	5	M	25	20	55455
...	...	...	...	...	...
6035	6036	F	25	15	32603
6036	6037	F	45	1	76006
6037	6038	F	56	1	14706
6038	6039	F	45	0	1060
6039	6040	M	25	6	11106

[6040 rows x 5 columns]

```
df3.head()
```

	UserID	Gender	Age	Occupation	Zip-code
0	1	F	1	10	48067
1	2	M	56	16	70072
2	3	M	25	15	55117
3	4	M	45	7	2460
4	5	M	25	20	55455

```
df4= pd.merge(df1,df2,on='MovieID',how='inner')
df4
```

	MovieID	Title	Genres
UserID \			
0	1	Toy Story (1995)	Animation Children's Comedy
1			

1	1	Toy Story (1995)	Animation Children's Comedy
6			
2	1	Toy Story (1995)	Animation Children's Comedy
8			
3	1	Toy Story (1995)	Animation Children's Comedy
9			
4	1	Toy Story (1995)	Animation Children's Comedy
10			
...	...	...	...
...			
1000204	3952	Contender, The (2000)	Drama Thriller
5812			
1000205	3952	Contender, The (2000)	Drama Thriller
5831			
1000206	3952	Contender, The (2000)	Drama Thriller
5837			
1000207	3952	Contender, The (2000)	Drama Thriller
5927			
1000208	3952	Contender, The (2000)	Drama Thriller
5998			

	Rating	Timestamp
0	5	978824268
1	4	978237008
2	4	978233496
3	5	978225952
4	5	978226474
...	...	...
1000204	4	992072099
1000205	3	986223125
1000206	4	1011902656
1000207	1	979852537
1000208	4	1001781044

[1000209 rows x 6 columns]

```
df4.drop(['Genres', 'Timestamp'], axis=1, inplace=True)
df4
```

	MovieID	Title	UserID	Rating
0	1	Toy Story (1995)	1	5
1	1	Toy Story (1995)	6	4
2	1	Toy Story (1995)	8	4
3	1	Toy Story (1995)	9	5
4	1	Toy Story (1995)	10	5
...	...	...	...	...
1000204	3952	Contender, The (2000)	5812	4
1000205	3952	Contender, The (2000)	5831	3
1000206	3952	Contender, The (2000)	5837	4
1000207	3952	Contender, The (2000)	5927	1

```
1000208      3952  Contender, The (2000)      5998      4
```

```
[1000209 rows x 4 columns]
```

```
df5= pd.merge(df4,df3,on='UserID',how='inner')
df5
```

	MovieID	Title	UserID	Rating	Gender
Age \					
0	1	Toy Story (1995)	1	5	F
1	48	Pocahontas (1995)	1	5	F
1	150	Apollo 13 (1995)	1	5	F
2	260	Star Wars	1	4	F
1	527	Schindler's List (1993)	1	5	F
3	...	...	...	...	...
1	...	...	...	...	...
1000204	3513	Rules of Engagement (2000)	5727	4	M
25	3535	American Psycho (2000)	5727	2	M
1000205	3536	Keeping the Faith (2000)	5727	5	M
25	3555	U-571 (2000)	5727	3	M
1000206	3578	Gladiator (2000)	5727	5	M
25					
	Occupation	Zip-code			
0	10	48067			
1	10	48067			
2	10	48067			
3	10	48067			
4	10	48067			
...	...	...			
1000204	4	92843			
1000205	4	92843			
1000206	4	92843			
1000207	4	92843			
1000208	4	92843			

```
[1000209 rows x 8 columns]
```

```
df5.drop('Zip-code',axis=1,inplace=True)
df5
```



Age \	MovieID	Title	UserID	Rating	Gender
0	1	Toy Story (1995)	1	5	F
1	48	Pocahontas (1995)	1	5	F
1	150	Apollo 13 (1995)	1	5	F
2	260	Star Wars	1	4	F
1	527	Schindler's List (1993)	1	5	F
3					
1					
4					
1					
...	...	...	...	...	...
...					
1000204	3513	Rules of Engagement (2000)	5727	4	M
25					
1000205	3535	American Psycho (2000)	5727	2	M
25					
1000206	3536	Keeping the Faith (2000)	5727	5	M
25					
1000207	3555	U-571 (2000)	5727	3	M
25					
1000208	3578	Gladiator (2000)	5727	5	M
25					

	Occupation
0	10
1	10
2	10
3	10
4	10
...	...
1000204	4
1000205	4
1000206	4
1000207	4
1000208	4

[1000209 rows x 7 columns]

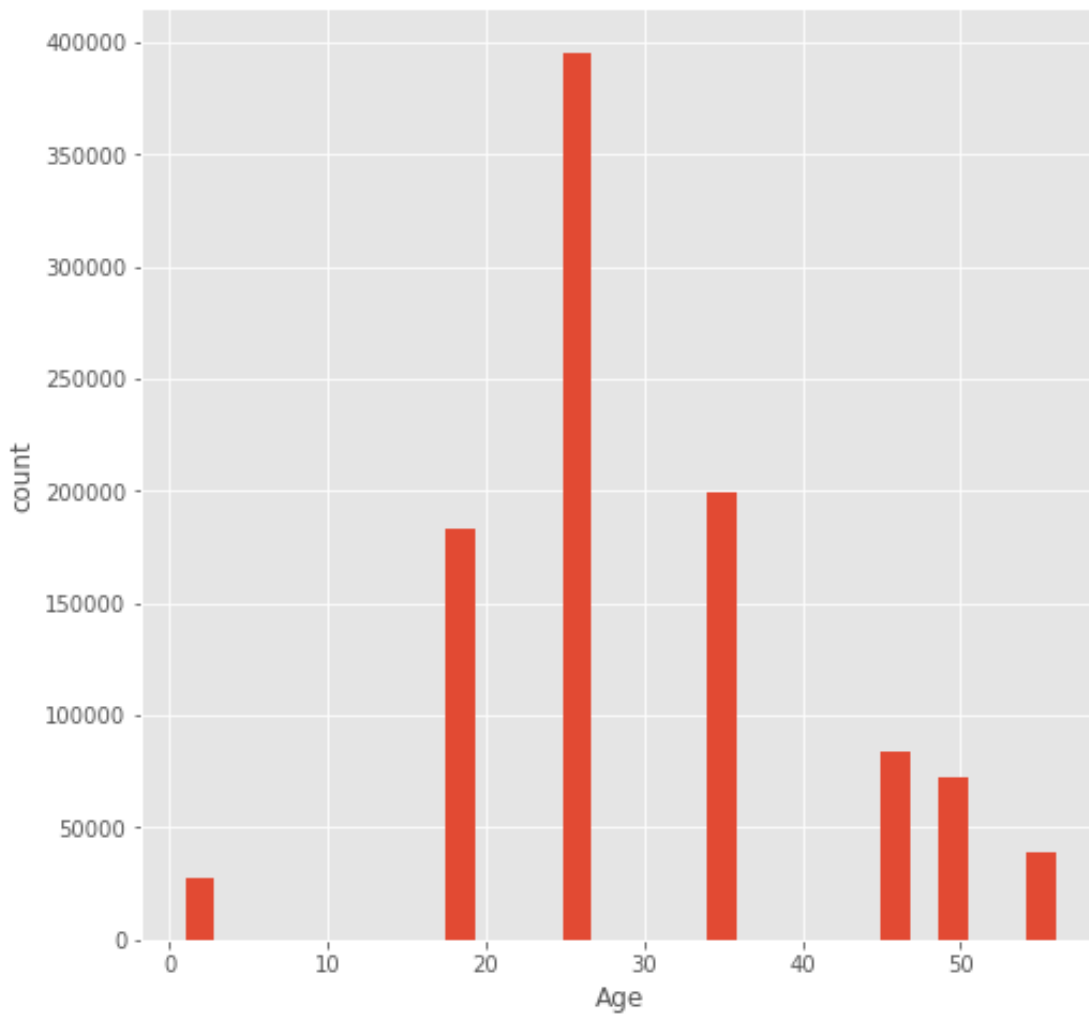
df5.head()

	MovieID	Title	UserID	Rating	Gender	Age
0	1	Toy Story (1995)	1	5	F	1
10						
1	48	Pocahontas (1995)	1	5	F	1
10						
2	150	Apollo 13 (1995)	1	5	F	1
10						

3	260	Star Wars	1	4	F	1
10						
4	527	Schindler's List (1993)	1	5	F	1
10						

```
from matplotlib import style
```

```
style.use('ggplot')
plt.figure(figsize=(8,8))
plt.hist(df5['Age'],bins=30)
plt.xlabel('Age')
plt.ylabel('count')
plt.show()
```



```
df6=df5[['Title','Rating']]
df6
```

	Title	Rating
0	Toy Story (1995)	5
1	Pocahontas (1995)	5

2	Apollo 13 (1995)	5
3	Star Wars	4
4	Schindler's List (1993)	5
...	...	...
1000204	Rules of Engagement (2000)	4
1000205	American Psycho (2000)	2
1000206	Keeping the Faith (2000)	5
1000207	U-571 (2000)	3
1000208	Gladiator (2000)	5

[1000209 rows x 2 columns]

```
df7=pd.DataFrame(df6.iloc[0])
df7
```

	0
Title	Toy Story (1995)
Rating	5

