In [1]:

```
import pandas as pd
import numpy as np

df_movielens = pd.read_csv('movielens100Nones.csv',  encoding = 'ISO-8859-1')
df_movielens.shape
```

Out[1]:

```
(100000, 29)
```

In [2]:

```
df_movielens.index
```

Out[2]:

```
RangeIndex(start=0, stop=100000, step=1)
```

In [53]:

```
import random

df_movielens = df_movielens.sample(100000)
```

In [3]:

```
df_movielens.shape
```

Out[3]:

```
(100000, 29)
```

In [4]:

```
df_movielens.columns
```

Out[4]:

```
Index(['userID', 'movieID', 'gender', 'age', 'occupation', 'zip_code
', 'state',
       'ratings', 'timestamp', 'year', 'title', 'Animation', 'Childr
en's',
       'Comedy', 'Adventure', 'Fantasy', 'Romance', 'Drama', 'Action
', 'Crime',
       'Thriller', 'Horror', 'Sci-Fi', 'Documentary', 'War', 'Musica
l',
       'Mystery', 'Film-Noir', 'Western'],
      dtype='object')
```

In [5]:

```
df_movielens.loc[df_movielens['gender'] == 'M', 'gender'] = 0
df_movielens.loc[df_movielens['gender'] == 'F', 'gender'] = 1
```

In [7]:

```
print(df_movielens.iloc[:10,2])
```

```
0    0
1    0
2    0
3    0
4    0
5    1
6    0
7    0
8    0
9    1
Name: gender, dtype: object
```

In [8]:

```
df_movielens.to_csv('Table100.csv', index = False)
```