```python
In [26]: import warnings
         warnings.filterwarnings("ignore")
         import pandas as pd
```

```python
In [ ]:
```

```python
In [2]: data=pd.read_csv("/home/placement/Downloads/movies.csv")
```

```python
In [3]: data.describe()
```

Out[3]:

	srno	year	rating	time
count	49590.000000	49590.000000	10814.000000	45836.000000
mean	24795.500000	2002.303428	3.451248	2628.445436
std	14315.544261	12.534555	0.495601	1604.646265
min	1.000000	1913.000000	1.400000	52.000000
25%	12398.250000	1999.000000	3.100000	1356.000000
50%	24795.500000	2007.000000	3.500000	2563.000000
75%	37192.750000	2010.000000	3.800000	2877.000000
max	49590.000000	2014.000000	4.500000	28813.000000

```python
In [4]: data.head()
```

Out[4]:

	srno	movie	year	rating	time
0	1	The Nightmare Before	1993	3.9	4568.0
1	2	The Mummy	1932	3.5	4388.0
2	3	Orphans of the Storm	1921	3.2	9062.0
3	4	The Object of Beauty	1991	2.8	6150.0
4	5	Night Tide	1963	2.8	5126.0

In [5]: `data.isna().sum()`

Out[5]:
```
srno           0
movie          0
year           0
rating     38776
time        3754
dtype: int64
```

In [7]:
```
data1=data.fillna(data.median())
data1
```

/tmp/ipykernel_11660/4065881331.py:1: FutureWarning: The default value of numeric_only in DataFrame.median is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=None' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
  data1=data.fillna(data.median())

Out[7]:

	srno	movie	year	rating	time
0	1	The Nightmare Before	1993	3.9	4568.0
1	2	The Mummy	1932	3.5	4388.0
2	3	Orphans of the Storm	1921	3.2	9062.0
3	4	The Object of Beauty	1991	2.8	6150.0
4	5	Night Tide	1963	2.8	5126.0
...	...	...	...	...	...
49585	49586	Winter Wonderland	2013	2.8	1812.0
49586	49587	Top Gear: Series 19: Africa Special	2013	3.5	6822.0
49587	49588	Fireplace For Your Home: Crackling Fireplace w...	2010	3.5	3610.0
49588	49589	Kate Plus Ei8ht	2010	2.7	2563.0
49589	49590	Kate Plus Ei8ht: Season 1	2010	2.7	2563.0

49590 rows × 5 columns

In [8]: `data.shape`

Out[8]: (49590, 5)

In [9]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49590 entries, 0 to 49589
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   srno    49590 non-null  int64
 1   movie   49590 non-null  object
 2   year    49590 non-null  int64
 3   rating  10814 non-null  float64
 4   time    45836 non-null  float64
dtypes: float64(2), int64(2), object(1)
memory usage: 1.9+ MB
```

In [ ]: `show lines in check type of columns`

In [20]: `data1=data.groupby(['year']).count()`

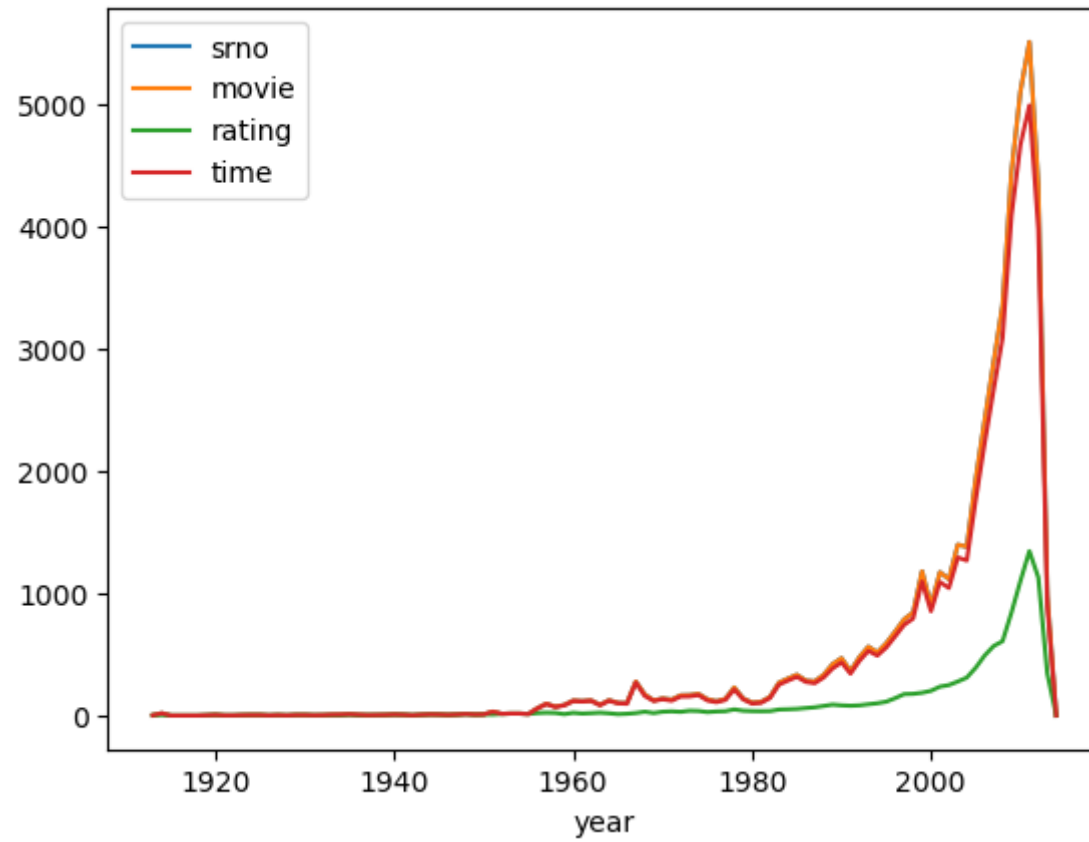In [21]: `data1`

Out[21]:

	srno	movie	rating	time
year				
1913	3	3	3	3
1914	20	20	5	18
1915	1	1	1	1
1916	1	1	1	1
1918	1	1	1	1
...	...	...	...	...
2010	5107	5107	1102	4671
2011	5511	5511	1346	4992
2012	4339	4339	1130	3978
2013	981	981	345	901
2014	1	1	1	1

101 rows × 4 columns

In [22]: `data1.to_csv('movie.csv')`

In [23]: `data1.plot()`

Out[23]: `<Axes: xlabel='year'>`

In [ ]:

In [ ]: