In [1]:
```
import pandas as pd
```

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

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

In [4]:
```
data.isna().sum()
```

Out[4]:
```
srno          0
movie         0
year          0
rating    38776
time       3754
dtype: int64
```

In [6]:
```
data1=data.fillna(data.median())
```

```
/tmp/ipykernel_15197/3060338577.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' i
s deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
  data1=data.fillna(data.median())
```

In [7]:
```
data.shape
```

Out[7]:
```
(49590, 5)
```

In [10]:
```
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 [9]:
```

```

```
<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 [18]:
```
data2=data.groupby(['year']).count()
```

In [19]:
```
data2
```

Out[19]:

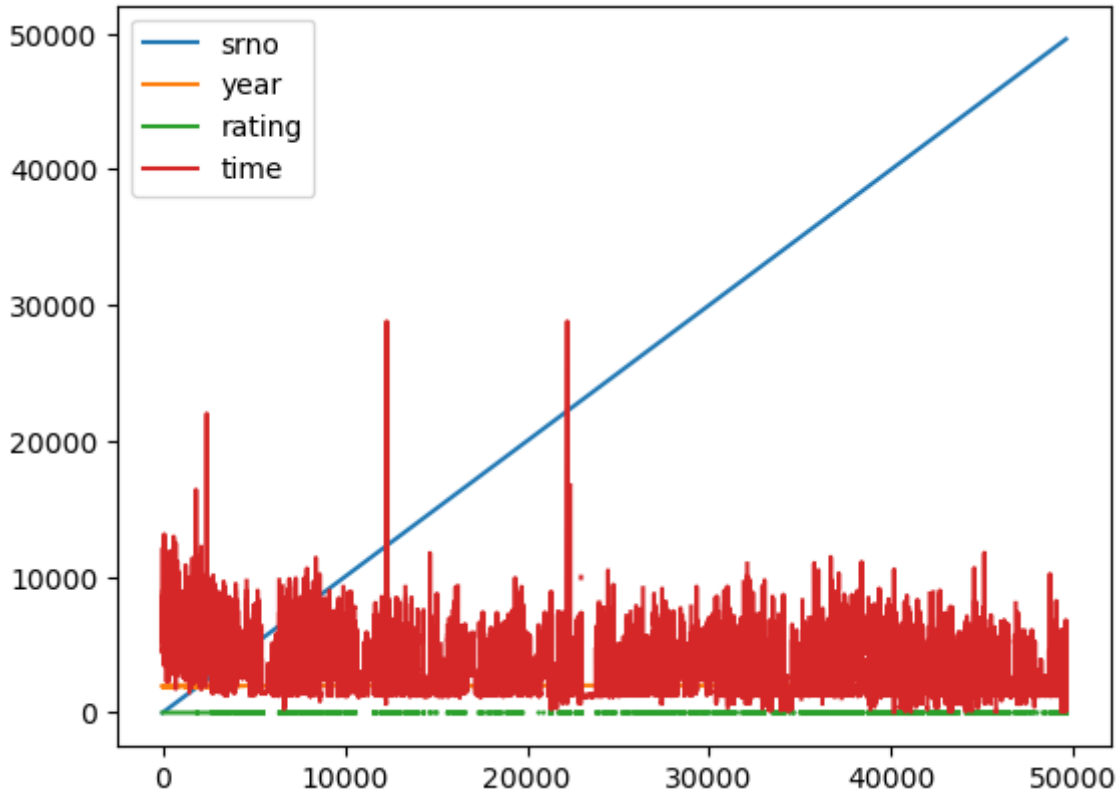
year	srno	movie	rating	time
1913	3	3	3	3
1914	20	20	5	18
1915	1	1	1	1
1916	1	1	1	1
1918	1	1	1	1
...	...	...	...	...
2010	5107	5107	1102	4671
2011	5511	5511	1346	4992
2012	4339	4339	1130	3978
2013	981	981	345	901
2014	1	1	1	1

101 rows × 4 columns

In [23]:
```
data2.to_csv('movies2.csv')
```

In [24]:
```
data.plot()
```

Out[24]: <Axes: >



In [ ]: