

```
In [58]: import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
data=pd.read_csv("/home/placement/Desktop/analysis/movies.csv")
print(data)
```

```

      Unnamed: 0  srno      movie \
0              0      1  The Nightmare Before
1              1      2          Bhahubali
2              2      3  Orphans of the Storm
3              3      4  The Object of Beauty
4              4      5      Night Tide
...          ...    ...
49585          49585  49586      Winter Wonderland
49586          49586  49587      Top Gear: Series 19: Africa Special
49587          49587  49588  Fireplace For Your Home: Crackling Fireplace w...
49588          49588  49589      Kate Plus Ei8ht
49589          49589  49590  Kate Plus Ei8ht: Season 1

```

```

      year  rating  time
0    1993    3.9  4568.0
1    2017    3.5  4388.0
2    1921    3.2  9062.0
3    1991    2.8  6150.0
4    1963    2.8  5126.0
...    ...    ...
49585  2013    2.8  1812.0
49586  2013    NaN  6822.0
49587  2010    NaN  3610.0
49588  2010    2.7    NaN
49589  2010    2.7    NaN

```

```
[49590 rows x 6 columns]
```

In [40]: data.head(10)

Out[40]:

	Unnamed: 0	srno	movie	year	rating	time
0	0	1	The Nightmare Before	1993	3.9	4568.0
1	1	2	Bhahubali	2017	3.5	4388.0
2	2	3	Orphans of the Storm	1921	3.2	9062.0
3	3	4	The Object of Beauty	1991	2.8	6150.0
4	4	5	Night Tide	1963	2.8	5126.0
5	5	6	One Magic Christmas	1985	3.8	5333.0
6	6	7	Muriel's Wedding	1994	3.5	6323.0
7	7	8	Mother's Boys	1994	3.4	5733.0
8	8	9	Nosferatu: Original Version	1929	3.5	5651.0
9	9	10	Nick of Time	1995	3.4	5333.0

In [41]: data.describe()

Out[41]:

	Unnamed: 0	srno	year	rating	time
count	49590.000000	49590.000000	49590.000000	10814.000000	45836.000000
mean	24794.500000	24795.500000	2002.305142	3.451248	2628.445436
std	14315.544261	14315.544261	12.530752	0.495601	1604.646265
min	0.000000	1.000000	1913.000000	1.400000	52.000000
25%	12397.250000	12398.250000	1999.000000	3.100000	1356.000000
50%	24794.500000	24795.500000	2007.000000	3.500000	2563.000000
75%	37191.750000	37192.750000	2010.000000	3.800000	2877.000000
max	49589.000000	49590.000000	2017.000000	4.500000	28813.000000

to find the missing data we have to use isna()

```
In [42]: data.isna().sum()
```

```
Out[42]: Unnamed: 0      0
         srno          0
         movie          0
         year          0
         rating    38776
         time      3754
         dtype: int64
```

to know the no. of rows and columns

```
In [43]: data.shape
```

```
Out[43]: (49590, 6)
```

```
In [44]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49590 entries, 0 to 49589
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Unnamed: 0  49590 non-null  int64  
 1   srno        49590 non-null  int64  
 2   movie       49590 non-null  object  
 3   year        49590 non-null  int64  
 4   rating      10814 non-null  float64 
 5   time        45836 non-null  float64 
dtypes: float64(2), int64(3), object(1)
memory usage: 2.3+ MB
```

```
In [45]: data.sample()
```

```
Out[45]:
```

	Unnamed: 0	srno	movie	year	rating	time
27606	27606	27607	Frasier: Season 3: Frasier Loves Roz	1995	NaN	1362.0

```
In [46]: data.index
```

```
Out[46]: RangeIndex(start=0, stop=49590, step=1)
```

```
In [47]: data.columns
```

```
Out[47]: Index(['Unnamed: 0', 'srno', 'movie', 'year', 'rating', 'time'], dtype='object')
```

to know the columns

```
In [48]: data.dtypes
```

```
Out[48]: Unnamed: 0      int64
srno      int64
movie      object
year      int64
rating     float64
time      float64
dtype: object
```

```
In [49]: data2=data.groupby(['year']).count() #get the data from "year" column using groupby()
```

In [50]: data2

Out[50]:

	Unnamed: 0	srno	movie	rating	time
year					
1913	3	3	3	3	3
1914	20	20	20	5	18
1915	1	1	1	1	1
1916	1	1	1	1	1
1918	1	1	1	1	1
...
2011	5511	5511	5511	1346	4992
2012	4339	4339	4339	1130	3978
2013	981	981	981	345	901
2014	1	1	1	1	1
2017	1	1	1	1	1

102 rows × 5 columns

In [51]: data

Out[51]:

	Unnamed: 0	srno	movie	year	rating	time
0	0	1	The Nightmare Before	1993	3.9	4568.0
1	1	2	Bhahubali	2017	3.5	4388.0
2	2	3	Orphans of the Storm	1921	3.2	9062.0
3	3	4	The Object of Beauty	1991	2.8	6150.0
4	4	5	Night Tide	1963	2.8	5126.0
...
49585	49585	49586	Winter Wonderland	2013	2.8	1812.0
49586	49586	49587	Top Gear: Series 19: Africa Special	2013	NaN	6822.0
49587	49587	49588	Fireplace For Your Home: Crackling Fireplace w...	2010	NaN	3610.0
49588	49588	49589	Kate Plus Ei8ht	2010	2.7	NaN
49589	49589	49590	Kate Plus Ei8ht: Season 1	2010	2.7	NaN

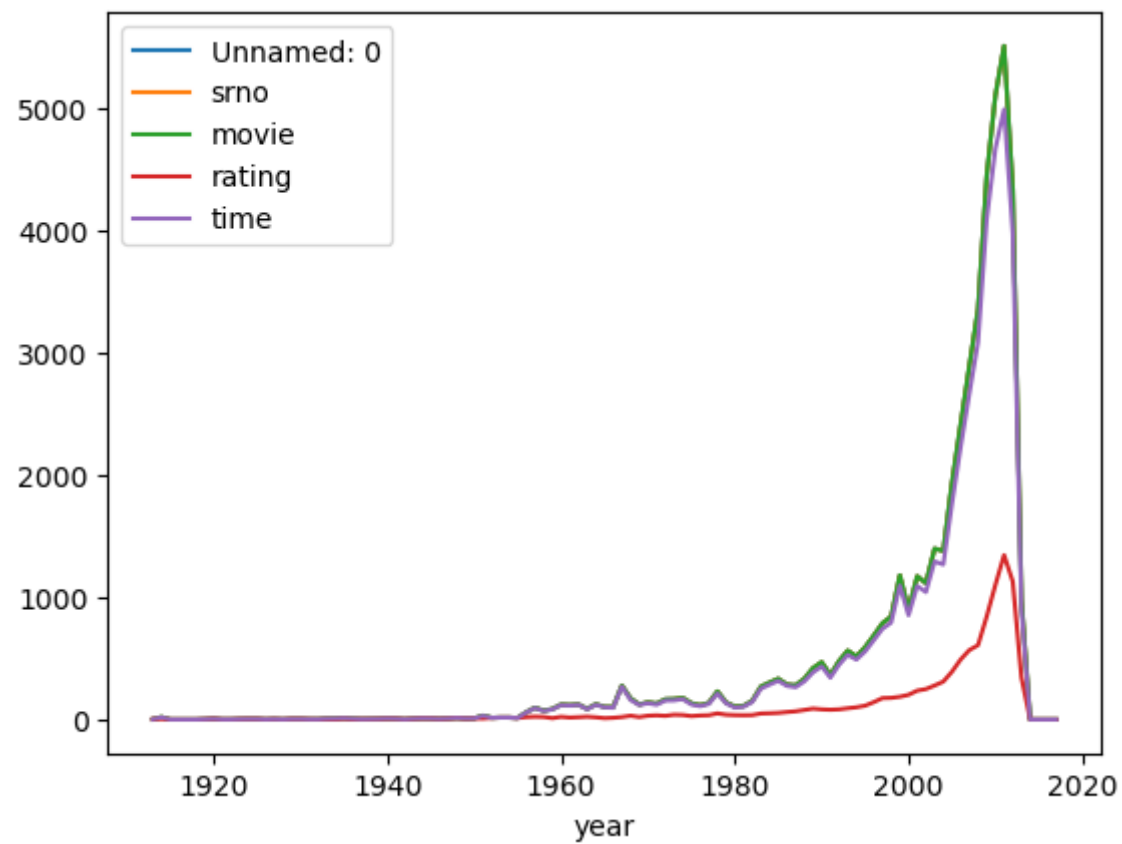
49590 rows × 6 columns

plotting data

In [34]: data.to_csv("movies.csv")

```
In [53]: data2.plot()
```

```
Out[53]: <Axes: xlabel='year'>
```



In []:

In []: