

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

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

```
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 [5]: data.head(10)
```

```
Out[5]:
```

	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
5	6	One Magic Christmas	1985	3.8	5333.0
6	7	Muriel's Wedding	1994	3.5	6323.0
7	8	Mother's Boys	1994	3.4	5733.0
8	9	Nosferatu: Original Version	1929	3.5	5651.0
9	10	Nick of Time	1995	3.4	5333.0

```
In [6]: data.isna().sum()
```

```
Out[6]: srno      0
movie      0
year      0
rating    38776
time      3754
dtype: int64
```

```
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
```

above line to check type of columns

In [12]: data2=data.groupby(['year']).count()

```
In [11]: data2
```

```
Out[11]:
```

	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 [14]: data2.to_csv('movies1.txt')
```

In [15]: data2

Out[15]:

	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 []: