In [48]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

In [49]:
```
data=pd.read_csv("/home/placement/Downloads/movies.csv")
```

reading a file

In [50]:
```
data.head(10)
```

Out[50]:

	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

Describe data

In [51]:
```
data.describe()
```

Out[51]:

	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 [52]:
```
data1=data.loc[(data.time>5000)]
data1
```

Out[52]:

	srno	movie	year	rating	time
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
...	...	...	...	...	...
49564	49565	American Addict	2013	3.5	5377.0
49579	49580	Underground: The Julian Assange Story	2012	3.7	5665.0
49583	49584	Sunset Strip	2012	3.0	5770.0
49584	49585	Silver Bells	2013	3.5	5287.0
49586	49587	Top Gear: Series 19: Africa Special	2013	NaN	6822.0

5897 rows × 5 columns

In [53]:
```
data.head(100)
```

Out[53]:

	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
...	...	...	...	...	...
95	96	The Hunted	1995	3.4	6605.0
96	97	The Great Waldo Pepper	1975	3.5	6467.0
97	98	Godzilla: King of the Monsters	1956	3.5	4828.0
98	99	Highlander 2: Renegade Version	1991	3.1	6585.0
99	100	High Noon	1952	3.9	5087.0

100 rows × 5 columns

In [54]:
```
data.tail()
```

Out[54]:

	srno	movie	year	rating	time
49585	49586	Winter Wonderland	2013	2.8	1812.0
49586	49587	Top Gear: Series 19: Africa Special	2013	NaN	6822.0
49587	49588	Fireplace For Your Home: Crackling Fireplace w...	2010	NaN	3610.0
49588	49589	Kate Plus Ei8ht	2010	2.7	NaN
49589	49590	Kate Plus Ei8ht: Season 1	2010	2.7	NaN

In [55]:
```
data.tail(10)
```

Out[55]:

	srno	movie	year	rating	time
49580	49581	Curious George: A Very Monkey Christmas	2009	3.8	3438.0
49581	49582	Mumfie's White Christmas	1996	2.4	1350.0
49582	49583	Lady Gaga & The Muppets' Holiday Spectacular	2013	3.1	3496.0
49583	49584	Sunset Strip	2012	3.0	5770.0
49584	49585	Silver Bells	2013	3.5	5287.0
49585	49586	Winter Wonderland	2013	2.8	1812.0
49586	49587	Top Gear: Series 19: Africa Special	2013	NaN	6822.0
49587	49588	Fireplace For Your Home: Crackling Fireplace w...	2010	NaN	3610.0
49588	49589	Kate Plus Ei8ht	2010	2.7	NaN
49589	49590	Kate Plus Ei8ht: Season 1	2010	2.7	NaN

In [56]:
```
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
```

Unique Values

In [57]:
```
data['movie'].unique()
```

Out[57]:
```
array(['The Nightmare Before', 'The Mummy', 'Orphans of the Storm', ...,
       'Fireplace For Your Home: Crackling Fireplace with Music',
       'Kate Plus Ei8ht', 'Kate Plus Ei8ht: Season 1'], dtype=object)
```

loc keyword using

In [58]:
```
data1=data.loc[(data.rating>4)&(data.year>2000)&(data.year<=2010)]
```

In [59]:
```
data1
```

Out[59]:

	srno	movie	year	rating	time
984	985	Paid in Full	2002	4.1	5883.0
1061	1062	The Pianist	2002	4.1	8945.0
1285	1286	Pirates of the Caribbean: The Curse of the Bla...	2003	4.1	8586.0
1305	1306	Ken Burns: Jazz	2001	4.1	NaN
1363	1364	The Life of Mammals	2002	4.2	NaN
...	...	...	...	...	...
47765	47766	Brew Masters: Season 1	2010	4.1	NaN
48212	48213	Cake Boss: Next Great Baker: Season 1	2010	4.1	NaN
48854	48855	Pit Bulls & Parolees	2009	4.3	NaN
48874	48875	Brew Masters	2010	4.1	NaN
49025	49026	Cake Boss: Next Great Baker	2010	4.1	NaN

434 rows × 5 columns

In [60]:
```
data.shape
```

Out[60]:
```
(49590, 5)
```

In [61]:
```
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
```

12345

In [62]:
```
data1=data.groupby(['year']).count()
```

In [63]:
```
data1
```

Out[63]:

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 [64]:
```
data1.to_csv('movies.csv')
```

In [65]:
```
data.head()
```

Out[65]:

	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 [66]:
```
data.count()
```

Out[66]:
```
srno      49590
movie     49590
year      49590
rating    10814
time      45836
dtype: int64
```

In [67]:
```
import warnings
warnings.filterwarnings('ignore')
```

In [68]:
```
data=data.corr()
data
```

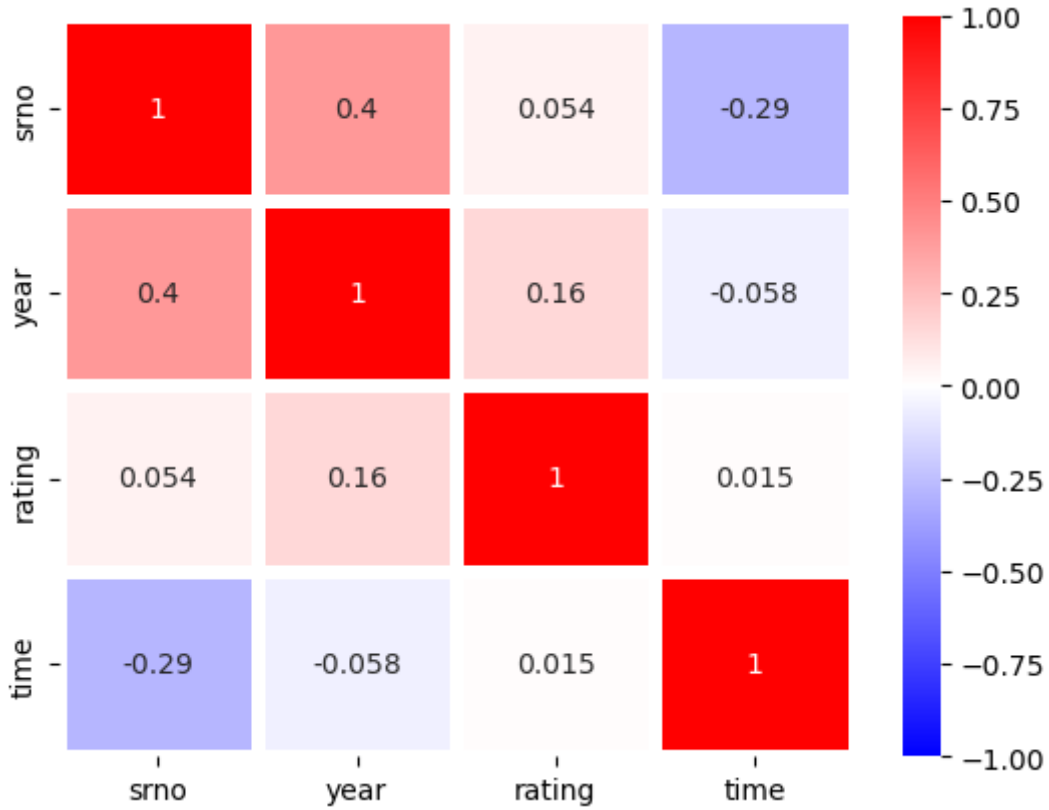
Out[68]:

	srno	year	rating	time
srno	1.000000	0.401153	0.054288	-0.286048
year	0.401153	1.000000	0.156210	-0.058444
rating	0.054288	0.156210	1.000000	0.015285
time	-0.286048	-0.058444	0.015285	1.000000

Correlation

In [69]:
```
sns.heatmap(data,vmax=1,vmin=-1,annot=True,linewidth=5,cmap='bwr')
```

Out[69]: <Axes: >



In [ ]: