In [2]:
```
import pandas as pd
```

In [3]:
```
df=pd.read_csv("g:/dataset/analysis/movies.csv")
df
```

Out[3]:

	star_rating	title	content_rating	category	duration	actors_list
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
1	9.2	The Godfather	R	Crime	175	[u'Marlon Brando', u'Al Pacino', u'James Caan']
2	9.1	The Godfather: Part II	R	Crime	200	[u'Al Pacino', u'Robert De Niro', u'Robert Duv...
3	9.0	The Dark Knight	PG-13	Action	152	[u'Christian Bale', u'Heath Ledger', u'Aaron E...
4	8.9	Pulp Fiction	R	Crime	154	[u'John Travolta', u'Uma Thurman', u'Samuel L....
...	...	...	...	...	...	...
974	7.4	Tootsie	PG	Comedy	116	[u'Dustin Hoffman', u'Jessica Lange', u'Teri G...
975	7.4	Back to the Future Part III	PG	Adventure	118	[u'Michael J. Fox', u'Christopher Lloyd', u'Ma...
976	7.4	Master and Commander: The Far Side of the World	PG-13	Action	138	[u'Russell Crowe', u'Paul Bettany', u'Billy Bo...
977	7.4	Poltergeist	PG	Horror	114	[u'JoBeth Williams', u"Heather O'Rourke", u'Cr...
978	7.4	Wall Street	R	Crime	126	[u'Charlie Sheen', u'Michael Douglas', u'Tamar...

979 rows × 6 columns

In [5]:
```
df[df.actors_list=='Robbins']
```

Out[5]:

	star_rating	title	content_rating	category	duration	actors_list

In [6]:
```
df[df.actors_list.str.contains('Robbins')]
```

Out[6]:

	star_rating	title	content_rating	category	duration	actors_list
0	9.3	The Shawshank Redemption	R	Crime	142	[u'Tim Robbins', u'Morgan Freeman', u'Bob Gunt...
365	8.0	Mystic River	R	Crime	138	[u'Sean Penn', u'Tim Robbins', u'Kevin Bacon']
611	7.7	Short Cuts	R	Comedy	187	[u'Andie MacDowell', u'Julianne Moore', u'Tim ...
693	7.7	The Player	R	Comedy	124	[u'Tim Robbins', u'Greta Scacchi', u'Fred Ward']
819	7.6	Jacob's Ladder	R	Drama	113	[u'Tim Robbins', u'Elizabeth Pe\xf1a', u'Danny...

In [7]:
```
df[df.actors_list.str.contains('amir')]
```

Out[7]:

	star_rating	title	content_rating	category	duration	actors_list
47	8.5	Taare Zameen Par	PG	Drama	165	[u'Darsheel Safary', u'Aamir Khan', u'Tanay Ch...
60	8.5	3 Idiots	PG-13	Comedy	170	[u'Aamir Khan', u'Madhavan', u'Mona Singh']
72	8.4	Rang De Basanti	NOT RATED	Drama	157	[u'Aamir Khan', u'Soha Ali Khan', u'Siddharth']
109	8.3	Dil Chahta Hai	NOT RATED	Comedy	183	[u'Aamir Khan', u'Saif Ali Khan', u'Akshaye Kh...
142	8.3	Lagaan: Once Upon a Time in India	PG	Adventure	224	[u'Aamir Khan', u'Gracy Singh', u'Rachel Shell...
239	8.1	Elite Squad: The Enemy Within	UNRATED	Action	115	[u'Wagner Moura', u'Irandhir Santos', u'Andr\x...
261	8.1	Elite Squad	R	Action	115	[u'Wagner Moura', u'Andr\xe9 Ramiro', u'Caio J...

In [8]:
```
df[df.actors_list.str.contains('amir Khan',case=False)]
```

Out[8]:

	star_rating	title	content_rating	category	duration	actors_list
47	8.5	Taare Zameen Par	PG	Drama	165	[u'Darsheel Safary', u'Aamir Khan', u'Tanay Ch...
60	8.5	3 Idiots	PG-13	Comedy	170	[u'Aamir Khan', u'Madhavan', u'Mona Singh']
72	8.4	Rang De Basanti	NOT RATED	Drama	157	[u'Aamir Khan', u'Soha Ali Khan', u'Siddharth']
109	8.3	Dil Chahta Hai	NOT RATED	Comedy	183	[u'Aamir Khan', u'Saif Ali Khan', u'Akshaye Kh...
142	8.3	Lagaan: Once Upon a Time in India	PG	Adventure	224	[u'Aamir Khan', u'Gracy Singh', u'Rachel Shell...

In [13]:
```
df[df.actors_list.str.contains('Akshay',case=False)]
```

Out[13]:

	star_rating	title	content_rating	category	duration	actors_list
109	8.3	Dil Chahta Hai	NOT RATED	Comedy	183	[u'Aamir Khan', u'Saif Ali Khan', u'Akshaye Kh...

In [14]:
```
df[df.category.isin(['Comedy','Drama'])]
```

Out[14]:

	star_rating	title	content_rating	category	duration	actors_list
5	8.9	12 Angry Men	NOT RATED	Drama	96	[u'Henry Fonda', u'Lee J. Cobb', u'Martin Bals...
9	8.9	Fight Club	R	Drama	139	[u'Brad Pitt', u'Edward Norton', u'Helena Bonh...
13	8.8	Forrest Gump	PG-13	Drama	142	[u'Tom Hanks', u'Robin Wright', u'Gary Sinise']
16	8.7	One Flew Over the Cuckoo's Nest	R	Drama	133	[u'Jack Nicholson', u'Louise Fletcher', u'Mich...
17	8.7	Seven Samurai	UNRATED	Drama	207	[u'Toshir\xf4 Mifune', u'Takashi Shimura', u'K...
...	...	...	...	...	...	...
970	7.4	Wonder Boys	R	Drama	107	[u'Michael Douglas', u'Tobey Maguire', u'Franc...
971	7.4	Death at a Funeral	R	Comedy	90	[u'Matthew Macfadyen', u'Peter Dinklage', u'Ew...
972	7.4	Blue Valentine	NC-17	Drama	112	[u'Ryan Gosling', u'Michelle Williams', u'John...
973	7.4	The Cider House Rules	PG-13	Drama	126	[u'Tobey Maguire', u'Charlize Theron', u'Micha...
974	7.4	Tootsie	PG	Comedy	116	[u'Dustin Hoffman', u'Jessica Lange', u'Teri G...

434 rows × 6 columns

In [15]:
```
df[(df.category=='Drama')|(df.category=='Comedy')]
```

Out[15]:

	star_rating	title	content_rating	category	duration	actors_list
5	8.9	12 Angry Men	NOT RATED	Drama	96	[u'Henry Fonda', u'Lee J. Cobb', u'Martin Bals...
9	8.9	Fight Club	R	Drama	139	[u'Brad Pitt', u'Edward Norton', u'Helena Bonh...
13	8.8	Forrest Gump	PG-13	Drama	142	[u'Tom Hanks', u'Robin Wright', u'Gary Sinise']
16	8.7	One Flew Over the Cuckoo's Nest	R	Drama	133	[u'Jack Nicholson', u'Louise Fletcher', u'Mich...
17	8.7	Seven Samurai	UNRATED	Drama	207	[u'Toshir\xf4 Mifune', u'Takashi Shimura', u'K...
...	...	...	...	...	...	...
970	7.4	Wonder Boys	R	Drama	107	[u'Michael Douglas', u'Tobey Maguire', u'Franc...
971	7.4	Death at a Funeral	R	Comedy	90	[u'Matthew Macfadyen', u'Peter Dinklage', u'Ew...
972	7.4	Blue Valentine	NC-17	Drama	112	[u'Ryan Gosling', u'Michelle Williams', u'John...
973	7.4	The Cider House Rules	PG-13	Drama	126	[u'Tobey Maguire', u'Charlize Theron', u'Micha...
974	7.4	Tootsie	PG	Comedy	116	[u'Dustin Hoffman', u'Jessica Lange', u'Teri G...

434 rows × 6 columns

In [16]:
```
df=pd.read_csv("g:/dataset/analysis/ufo.csv")
df
```

Out[16]:

	City	Colors Reported	Shape Reported	State	Time
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:00
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:00
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:00
3	Abilene	NaN	DISK	KS	6/1/1931 13:00
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:00
...	...	...	...	...	...
18236	Grant Park	NaN	TRIANGLE	IL	12/31/2000 23:00
18237	Spirit Lake	NaN	DISK	IA	12/31/2000 23:00
18238	Eagle River	NaN	NaN	WI	12/31/2000 23:45
18239	Eagle River	RED	LIGHT	WI	12/31/2000 23:45
18240	Ybor	NaN	OVAL	FL	12/31/2000 23:59

18241 rows × 5 columns

In [17]:
```
df.dtypes
```

Out[17]:
```
City               object
Colors Reported    object
Shape Reported     object
State              object
Time               object
dtype: object
```

In [19]:
```
df['date_time']=pd.to_datetime(df.Time)
```

In [20]:
```
df
```

Out[20]:

	City	Colors Reported	Shape Reported	State	Time	date_time
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:00	1930-06-01 22:00:00
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:00	1930-06-30 20:00:00
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:00	1931-02-15 14:00:00
3	Abilene	NaN	DISK	KS	6/1/1931 13:00	1931-06-01 13:00:00
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:00	1933-04-18 19:00:00
...	...	...	...	...	...	...
18236	Grant Park	NaN	TRIANGLE	IL	12/31/2000 23:00	2000-12-31 23:00:00
18237	Spirit Lake	NaN	DISK	IA	12/31/2000 23:00	2000-12-31 23:00:00
18238	Eagle River	NaN	NaN	WI	12/31/2000 23:45	2000-12-31 23:45:00
18239	Eagle River	RED	LIGHT	WI	12/31/2000 23:45	2000-12-31 23:45:00
18240	Ybor	NaN	OVAL	FL	12/31/2000 23:59	2000-12-31 23:59:00

18241 rows × 6 columns

In [25]:
```
df['year']=df.date_time.dt.year
df['month']=df.date_time.dt.month_name()
df['day']=df.date_time.dt.day_name()
```

In [38]:
```
df
```

Out[38]:

	City	Colors Reported	Shape Reported	State	Time	date_time	year	month	day
0	Ithaca	NaN	TRIANGLE	NY	6/1/1930 22:00	1930-06-01 22:00:00	1930	June	Sunday
1	Willingboro	NaN	OTHER	NJ	6/30/1930 20:00	1930-06-30 20:00:00	1930	June	Monday
2	Holyoke	NaN	OVAL	CO	2/15/1931 14:00	1931-02-15 14:00:00	1931	February	Sunday
3	Abilene	NaN	DISK	KS	6/1/1931 13:00	1931-06-01 13:00:00	1931	June	Monday
4	New York Worlds Fair	NaN	LIGHT	NY	4/18/1933 19:00	1933-04-18 19:00:00	1933	April	Tuesday
...	...	...	...	...	...	...	...	...	...
18236	Grant Park	NaN	TRIANGLE	IL	12/31/2000 23:00	2000-12-31 23:00:00	2000	December	Sunday
18237	Spirit Lake	NaN	DISK	IA	12/31/2000 23:00	2000-12-31 23:00:00	2000	December	Sunday
18238	Eagle River	NaN	NaN	WI	12/31/2000 23:45	2000-12-31 23:45:00	2000	December	Sunday
18239	Eagle River	RED	LIGHT	WI	12/31/2000 23:45	2000-12-31 23:45:00	2000	December	Sunday
18240	Ybor	NaN	OVAL	FL	12/31/2000 23:59	2000-12-31 23:59:00	2000	December	Sunday

18241 rows × 9 columns

In [39]:
```
df.State.value_counts()
```

Out[39]:
```
CA    2529
WA    1322
TX    1027
NY     914
FL     837
AZ     738
OH     667
IL     613
PA     598
MI     591
OR     534
MO     448
NJ     370
CO     367
WI     357
NC     356
IN     326
GA     325
MA     322
VA     299
TN     286
NV     284
MN     254
KY     244
NM     241
CT     225
MD     215
AR     206
UT     193
OK     193
AL     193
ME     181
KS     176
LA     174
SC     166
IA     162
MT     144
MS     139
WV     132
ID     130
NH     125
AK     116
NE     101
HI      85
WY      69
RI      67
SD      57
ND      51
VT      44
DE      43
Fl       4
Ca       1
Name: State, dtype: int64
```

In [27]:
```
df.year.value_counts()
```

Out[27]:
```
1999    2774
2000    2635
1998    1743
1995    1344
1997    1237
        ...
1936       2
1930       2
1935       1
1934       1
1933       1
Name: year, Length: 68, dtype: int64
```

In [41]:
```
df.day.value_counts()
```

Out[41]:
```
Tuesday      2822
Sunday       2689
Saturday     2687
Friday       2669
Thursday     2598
Wednesday    2476
Monday       2300
Name: day, dtype: int64
```

In [42]:
```
df.City.value_counts()
```

Out[42]:
```
Seattle                        187
New York City                  161
Phoenix                        137
Houston                        108
Las Vegas                      105
                              ...
Neb.-Mo. Line                    1
Carlsberg                        1
Yukon                            1
Dolly Sods Wilderness Area       1
Ybor                             1
Name: City, Length: 6476, dtype: int64
```

In [43]:
```
df.month.value_counts()
```

Out[43]:
```
June         3059
July         2345
August       1948
October      1723
September    1635
November     1509
May          1168
March        1096
April        1045
December     1034
January       862
February      817
Name: month, dtype: int64
```

In [44]:
```
df[(df.year==1999)&(df.month=='June')].shape
```

Out[44]:
```
(236, 9)
```

In [45]:
```
df[(df.year==1999)&(df.month=='June')&(df.day=='Tuesday')].shape
```

Out[45]:
```
(80, 9)
```

In [46]:
```
df=pd.read_csv('g:/dataset/analysis/restaurant.csv')
df
```

Out[46]:

	total_bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

In [47]:
```
df.corr()
```

```
C:\Users\panka\AppData\Local\Temp\ipykernel_4536\1134722465.py:1: FutureWarning: T
he default value of numeric_only in DataFrame.corr is deprecated. In a future vers
ion, it will default to False. Select only valid columns or specify the value of n
umeric_only to silence this warning.
  df.corr()
```

Out[47]:

	total_bill	tip	size
total_bill	1.000000	0.675734	0.598315
tip	0.675734	1.000000	0.489299
size	0.598315	0.489299	1.000000

In [ ]:
```
```