Pandas

- pandas is one of the libary that is avialble in python.it is used for data maipulation and data analysis

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

In [4]:
```
a = pd.Series(("a","b","c"))
a
```

Out[4]:
```
0    a
1    b
2    c
dtype: object
```

In [5]:
```
b = pd.Series(["a","b",23,4,9.5])
b
```

Out[5]:
```
0      a
1      b
2     23
3      4
4    9.5
dtype: object
```

In [8]:
```
pd.Series({"name":["a","b","c"],"des":["ass","lect","trainer"]})
```

Out[8]:
```
name                [a, b, c]
des     [ass, lect, trainer]
dtype: object
```

In [9]:
```
pd.DataFrame({"name":["a","b","c"],"des":["ass","lect","trainer"]})
```

Out[9]:

	name	des
0	a	ass
1	b	lect
2	c	trainer

In [10]:
```
pd.DataFrame({"name":["a","b","c"],"des":["ass","lect","trainer"]},index=["vignai
```

Out[10]:

	name	des
vignan	a	ass
layola	b	lect
vrsec	c	trainer

In [11]:
```
df=pd.DataFrame([[1,2,3,4],[5,6,7,8]],columns=['a','b','c','d'],index=[1,2])
df
```

Out[11]:

	a	b	c	d
1	1	2	3	4
2	5	6	7	8

In [13]:
```
df = pd.read_csv("movie_metadata.csv")
df
```

Out[13]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_fa
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	
4	NaN	Doug Walker	NaN	NaN	131.0	
5	Color	Andrew Stanton	462.0	132.0	475.0	
6	Color	Sam Raimi	392.0	156.0	0.0	
7	Color	Nathan Greno	324.0	100.0	15.0	
8	Color	Joss Whedon	635.0	141.0	0.0	

In [14]:
```
df.shape
```

Out[14]:
```
(5043, 28)
```

In [15]:
```
df.columns
```

Out[15]:
```
Index(['color', 'director_name', 'num_critic_for_reviews', 'duration',
       'director_facebook_likes', 'actor_3_facebook_likes', 'actor_2_name',
       'actor_1_facebook_likes', 'gross', 'genres', 'actor_1_name',
       'movie_title', 'num_voted_users', 'cast_total_facebook_likes',
       'actor_3_name', 'facenumber_in_poster', 'plot_keywords',
       'movie_imdb_link', 'num_user_for_reviews', 'language', 'country',
       'content_rating', 'budget', 'title_year', 'actor_2_facebook_likes',
       'imdb_score', 'aspect_ratio', 'movie_facebook_likes'],
      dtype='object')
```

In [17]:
```
df.head(2)
```

Out[17]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	

2 rows × 28 columns

In [18]:
```
df.tail(4)
```

Out[18]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
5039	Color	NaN	43.0	43.0	NaN	
5040	Color	Benjamin Roberds	13.0	76.0	0.0	
5041	Color	Daniel Hsia	14.0	100.0	0.0	
5042	Color	Jon Gunn	43.0	90.0	16.0	

4 rows × 28 columns

In [19]:
```
df.sample(3)
```

Out[19]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
4831	Color	Niall Johnson	4.0	114.0	7.0	
2394	Color	Larry Charles	343.0	82.0	119.0	
4521	Color	Regardt van den Bergh	5.0	116.0	12.0	

3 rows × 28 columns

In [21]:
```
df["color"].head()
```

Out[21]:
```
0    Color
1    Color
2    Color
3    Color
4      NaN
Name: color, dtype: object
```

In [23]:
```
#loc - get rows or columns with particular lables from the index
#iloc - get rows or columns at particular position based on the index(it takes o
df.loc[5,"color"]
```

Out[23]: 'Color'

In [25]:
```
df.head(7)
```

Out[25]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	2
4	NaN	Doug Walker	NaN	NaN	131.0	
5	Color	Andrew Stanton	462.0	132.0	475.0	
6	Color	Sam Raimi	392.0	156.0	0.0	

7 rows × 28 columns

In [28]:
```
df.iloc[5,2]
```

Out[28]: 462.0

In [29]:
```
df.isnull()
```

Out[29]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
0	False	False	False	False	False	
1	False	False	False	False	False	
2	False	False	False	False	False	
3	False	False	False	False	False	
4	True	False	True	True	False	
5	False	False	False	False	False	
6	False	False	False	False	False	
7	False	False	False	False	False	
8	False	False	False	False	False	
9	False	False	False	False	False	
10	False	False	False	False	False	
11	False	False	False	False	False	
12	False	False	False	False	False	
13	False	False	False	False	False	
14	False	False	False	False	False	
15	False	False	False	False	False	
16	False	False	False	False	False	
17	False	False	False	False	False	
18	False	False	False	False	False	
19	False	False	False	False	False	
20	False	False	False	False	False	
21	False	False	False	False	False	
22	False	False	False	False	False	
23	False	False	False	False	False	
24	False	False	False	False	False	
25	False	False	False	False	False	
26	False	False	False	False	False	
27	False	False	False	False	False	
28	False	False	False	False	False	
29	False	False	False	False	False	
...	...	...	...	...	...	
5013	False	False	False	False	False	
5014	False	False	False	False	False	
5015	False	False	False	False	False	

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
5016	False	False	True	False	False	
5017	False	False	False	False	False	
5018	False	False	False	False	False	
5019	False	False	False	False	False	
5020	True	False	True	False	False	
5021	False	False	False	False	False	
5022	False	False	False	False	False	
5023	False	False	False	False	False	
5024	False	False	False	False	False	
5025	False	False	False	False	False	
5026	False	False	False	False	False	
5027	False	False	False	False	False	
5028	False	False	False	False	False	
5029	False	False	False	False	False	
5030	False	False	True	False	False	
5031	False	False	False	False	False	
5032	False	False	False	False	False	
5033	False	False	False	False	False	
5034	False	False	False	False	False	
5035	False	False	False	False	False	
5036	False	False	True	False	False	
5037	False	False	False	False	False	
5038	False	False	False	False	False	
5039	False	True	False	False	True	
5040	False	False	False	False	False	
5041	False	False	False	False	False	
5042	False	False	False	False	False	

5043 rows × 28 columns

In [30]:
```
df.isnull().sum()
```

Out[30]:
```
color                        19
director_name               104
num_critic_for_reviews       50
duration                     15
director_facebook_likes     104
actor_3_facebook_likes       23
actor_2_name                 13
actor_1_facebook_likes        7
gross                       884
genres                        0
actor_1_name                  7
movie_title                   0
num_voted_users               0
cast_total_facebook_likes     0
actor_3_name                 23
facenumber_in_poster         13
plot_keywords               153
movie_imdb_link               0
num_user_for_reviews         21
language                     12
country                       5
content_rating              303
budget                      492
title_year                  108
actor_2_facebook_likes       13
imdb_score                    0
aspect_ratio                329
movie_facebook_likes          0
dtype: int64
```

In [31]:
```
df.isnull().sum().sum()
```

Out[31]:
```
2698
```

In [32]:
```
x = df.iloc[:,:-1].values
x
```

Out[32]:
```
array([['Color', 'James Cameron', 723.0, ..., 936.0, 7.9, 1.78],
       ['Color', 'Gore Verbinski', 302.0, ..., 5000.0, 7.1, 2.35],
       ['Color', 'Sam Mendes', 602.0, ..., 393.0, 6.8, 2.35],
       ...,
       ['Color', 'Benjamin Roberds', 13.0, ..., 0.0, 6.3, nan],
       ['Color', 'Daniel Hsia', 14.0, ..., 719.0, 6.3, 2.35],
       ['Color', 'Jon Gunn', 43.0, ..., 23.0, 6.6, 1.85]], dtype=object)
```

In [36]:
```
df1 = df.dropna()
df1
```

Out[36]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_fa
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	
4	0	Doug Walker	0.0	0.0	131.0	
5	Color	Andrew Stanton	462.0	132.0	475.0	
6	Color	Sam Raimi	392.0	156.0	0.0	
7	Color	Nathan Greno	324.0	100.0	15.0	
8	Color	Joss Whedon	635.0	141.0	0.0	

In [34]:
```
df.fillna(value=0,inplace=True)
```

In [35]:
```
df.head()
```

Out[35]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	2
4	0	Doug Walker	0.0	0.0	131.0	

5 rows × 28 columns

In [39]:
```
df.drop(["color"],axis=1)
```

Out[39]:

	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_
0	James Cameron	723.0	178.0	0.0	
1	Gore Verbinski	302.0	169.0	563.0	1
2	Sam Mendes	602.0	148.0	0.0	
3	Christopher Nolan	813.0	164.0	22000.0	23
4	Doug Walker	0.0	0.0	131.0	
5	Andrew Stanton	462.0	132.0	475.0	
6	Sam Raimi	392.0	156.0	0.0	4
7	Nathan Greno	324.0	100.0	15.0	

In [46]:
```
df[10:13]
```

Out[46]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceboc
10	Color	Zack Snyder	673.0	183.0	0.0	
11	Color	Bryan Singer	434.0	169.0	0.0	
12	Color	Marc Forster	403.0	106.0	395.0	

3 rows × 28 columns

In [45]:
```
df.head(2)
```

Out[45]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	

2 rows × 28 columns

In [48]:
```
# masking
df["duration"]<169.0
```

Out[48]:
```
0       False
1       False
2        True
3        True
4        True
5        True
6        True
7        True
8        True
9        True
10      False
11      False
12       True
13       True
14       True
15       True
16       True
17      False
18       True
19       True
20       True
21       True
22       True
23      False
24       True
25      False
26      False
27       True
28       True
29       True
        ...
5013     True
5014     True
5015     True
5016     True
5017     True
5018     True
5019     True
5020     True
5021     True
5022     True
5023     True
5024     True
5025     True
5026     True
5027     True
5028     True
5029     True
5030     True
5031     True
5032     True
5033     True
5034     True
5035     True
```

```
5036     True
5037     True
5038     True
5039     True
5040     True
5041     True
5042     True
Name: duration, Length: 5043, dtype: bool
```

In [49]:
```
df[df["duration"]<169.0]
```

Out[49]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_fa
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	
4	0	Doug Walker	0.0	0.0	131.0	
5	Color	Andrew Stanton	462.0	132.0	475.0	
6	Color	Sam Raimi	392.0	156.0	0.0	
7	Color	Nathan Greno	324.0	100.0	15.0	
8	Color	Joss Whedon	635.0	141.0	0.0	
9	Color	David Yates	375.0	153.0	282.0	

In [51]:
```
df[df.duplicated()]
```

Out[51]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
137	Color	David Yates	248.0	110.0	282.0	
187	Color	Bill Condon	322.0	115.0	386.0	
204	Color	Hideaki Anno	1.0	120.0	28.0	
303	Color	Joe Wright	256.0	111.0	456.0	
389	Color	Josh Trank	369.0	100.0	128.0	
395	Color	Rob Cohen	187.0	106.0	357.0	
590	Color	Brett Ratner	245.0	101.0	420.0	
656	Color	Paul Verhoeven	196.0	113.0	719.0	
794	Color	Joss Whedon	703.0	173.0	0.0	
1220	Color	Angelina Jolie Pitt	322.0	137.0	11000.0	
1305	Color	Paul McGuigan	159.0	110.0	118.0	
1449	Color	Albert Hughes	208.0	122.0	117.0	
2169	Color	Paul McGuigan	98.0	114.0	118.0	
2292	Color	Frank Oz	168.0	87.0	0.0	
2472	Color	Jon Cassar	45.0	90.0	78.0	
2493	Black and White	Yimou Zhang	283.0	80.0	611.0	
2533	Color	Neil Burger	236.0	110.0	168.0	
2562	Color	Jon Lucas	81.0	100.0	24.0	
2568	Color	Vic Armstrong	169.0	110.0	179.0	
2619	Color	John Carpenter	318.0	101.0	0.0	
2771	Color	Ole Bornedal	264.0	92.0	30.0	
2777	Color	Stephen Frears	51.0	119.0	350.0	
2798	Color	Shawn Levy	69.0	88.0	189.0	
2971	Color	John Lee Hancock	106.0	137.0	102.0	
3117	Color	Guy Ritchie	151.0	104.0	0.0	

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceb
3345	Color	Herbert Ross	60.0	107.0	71.0	
3452	Color	Paul Haggis	287.0	115.0	549.0	
3480	Color	Michael Winterbottom	71.0	115.0	187.0	
3729	Color	Tim Blake Nelson	92.0	95.0	596.0	
3900	Color	0	9.0	60.0	0.0	
3915	Color	Wes Craven	160.0	107.0	0.0	
4182	Color	Rob Zombie	220.0	119.0	0.0	
4226	Color	William Friedkin	138.0	104.0	607.0	
4282	Color	Kenneth Branagh	85.0	150.0	0.0	
4313	Color	Bruce McCulloch	52.0	85.0	54.0	
4408	Color	Yimou Zhang	101.0	95.0	611.0	
4565	Color	Peter Cattaneo	122.0	91.0	11.0	
4573	Color	Mel Brooks	48.0	92.0	0.0	
4631	Color	Danny Boyle	393.0	101.0	0.0	
4769	Color	Tamra Davis	111.0	93.0	33.0	
4882	Color	Dan Curtis	0.0	99.0	45.0	
4927	Color	Jason Stone	48.0	108.0	14.0	
4942	Color	Paul Schrader	130.0	93.0	261.0	
4950	Color	David Hewlett	8.0	88.0	686.0	
4951	Black and White	George A. Romero	284.0	96.0	0.0	

45 rows × 28 columns

In [54]:
```
df.drop_duplicates().shape
```

Out[54]:
```
(4998, 28)
```

In [55]:
```
df["color"].unique()
```

Out[55]:
```
array(['Color', 0, ' Black and White'], dtype=object)
```

In [56]:
```
df["color"].value_counts()
```

Out[56]:
```
Color               4815
 Black and White     209
0                     19
Name: color, dtype: int64
```

In [57]:
```
df.describe()
```

Out[57]:

	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_1
count	5043.000000	5043.000000	5043.000000	5043.000000	
mean	138.804283	106.882213	672.351576	642.068015	
std	121.792053	25.828463	2785.871819	1661.808199	
min	0.000000	0.000000	0.000000	0.000000	
25%	48.000000	93.000000	6.000000	130.000000	
50%	109.000000	103.000000	45.000000	367.000000	
75%	194.000000	118.000000	189.000000	635.000000	
max	813.000000	511.000000	23000.000000	23000.000000	

In [58]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5043 entries, 0 to 5042
Data columns (total 28 columns):
color                      5043 non-null object
director_name              5043 non-null object
num_critic_for_reviews     5043 non-null float64
duration                   5043 non-null float64
director_facebook_likes    5043 non-null float64
actor_3_facebook_likes     5043 non-null float64
actor_2_name               5043 non-null object
actor_1_facebook_likes     5043 non-null float64
gross                      5043 non-null float64
genres                     5043 non-null object
actor_1_name               5043 non-null object
movie_title                5043 non-null object
num_voted_users            5043 non-null int64
cast_total_facebook_likes  5043 non-null int64
actor_3_name               5043 non-null object
facenumber_in_poster       5043 non-null float64
plot_keywords              5043 non-null object
movie_imdb_link            5043 non-null object
num_user_for_reviews       5043 non-null float64
language                   5043 non-null object
country                    5043 non-null object
content_rating             5043 non-null object
budget                     5043 non-null float64
title_year                 5043 non-null float64
actor_2_facebook_likes     5043 non-null float64
imdb_score                 5043 non-null float64
aspect_ratio               5043 non-null float64
movie_facebook_likes       5043 non-null int64
dtypes: float64(13), int64(3), object(12)
memory usage: 1.1+ MB
```

In [68]:
```
import numpy as np
s = df.sort_values(by=["num_critic_for_reviews"])
s
```

Out[68]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_fa
2351	Color	Jonathan Jakubowicz	0.0	105.0	23.0	
4444	Color	Frank Lotito	0.0	102.0	5.0	
4767	Color	Patrick Gilles	0.0	90.0	0.0	
4989	Color	Daniel Mellitz	0.0	0.0	0.0	
4711	Color	Gene Teigland	0.0	103.0	0.0	
4763	Color	Daston Kalili	0.0	127.0	2.0	
4622	Color	Michael Taliferro	0.0	138.0	105.0	
4882	Color	Dan Curtis	0.0	99.0	45.0	

In [70]:
```
s.reindex(np.arange(0,5043))
```

Out[70]:

	color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_fa
0	Color	James Cameron	723.0	178.0	0.0	
1	Color	Gore Verbinski	302.0	169.0	563.0	
2	Color	Sam Mendes	602.0	148.0	0.0	
3	Color	Christopher Nolan	813.0	164.0	22000.0	
4	0	Doug Walker	0.0	0.0	131.0	
5	Color	Andrew Stanton	462.0	132.0	475.0	
6	Color	Sam Raimi	392.0	156.0	0.0	
7	Color	Nathan Greno	324.0	100.0	15.0	
8	Color	Joss Whedon	635.0	141.0	0.0	

In [ ]: