```
import pandas as pd
import numpy as np
```

```
movies = pd.read_csv('/content/imdb-top-1000.csv')
```

```
movies.head()
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	Star1	No_of_Votes	Gross	Metascore
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Tim Robbins	2343110	28341469.0	80.0
1	The Godfather	1972	175	Crime	9.2	Francis Ford Coppola	Marlon Brando	1620367	134966411.0	100.0
2	The Dark Knight	2008	152	Action	9.0	Christopher Nolan	Christian Bale	2303232	534858444.0	84.0
	The Godfather: Part					Francis Ford				

```
genres = movies.groupby('Genre')
```

```
# Applying builtin aggregation fuctions on groupby objects
genres.std()
```

	Runtime	IMDB_Rating	No_of_Votes	Gross	Metascore
Genre					
Action	28.500706	0.304258	432946.814748	2.256724e+08	12.421252
Adventure	33.317320	0.229781	301188.347642	1.697543e+08	12.345393
Animation	14.530471	0.253221	262173.231571	2.091840e+08	8.813646
Biography	25.514466	0.267140	271284.191372	1.363251e+08	11.028187
Comedy	22.946213	0.228771	188653.570564	1.946513e+08	11.829160
Crime	27.689231	0.335477	373999.730656	1.571191e+08	13.099102
Drama	27.740490	0.267229	305554.162841	2.201164e+08	12.744687
Family	10.606602	0.000000	137008.302816	3.048412e+08	16.970563
Fantasy	12.727922	0.141421	22179.111299	7.606861e+07	NaN
Film-Noir	4.000000	0.152753	54649.083277	7.048472e+07	1.527525
Horror	13.604812	0.311302	234883.508691	9.965017e+07	15.362291
Mystery	14.475423	0.310791	404621.915297	1.567524e+08	18.604435
Thriller	NaN	NaN	NaN	NaN	NaN
Western	17.153717	0.420317	263489.554280	1.230626e+07	9.032349

```
# find the top 3 genres by total earning
movies.groupby('Genre').sum()['Gross'].sort_values(ascending=False).head(3)
```

```
Genre
Drama     3.540997e+10
Action    3.263226e+10
Comedy    1.566387e+10
Name: Gross, dtype: float64
```

```
movies.groupby('Genre')['Gross'].sum().sort_values(ascending=False).head(3)
```

```
Genre
Drama     3.540997e+10
Action    3.263226e+10
Comedy    1.566387e+10
Name: Gross, dtype: float64
```

```
# find the genre with highest avg IMDB rating
movies.groupby('Genre')['IMDB_Rating'].mean().sort_values(ascending=False).head(1)
```

```
Genre
Western    8.35
Name: IMDB_Rating, dtype: float64
```

```
# find director with most popularity
movies.groupby('Director')['No_of_Votes'].sum().sort_values(ascending=False).head(1)
```

```
Director
Christopher Nolan    11578345
Name: No_of_Votes, dtype: int64
```

```
# find the highest rated movie of each genre
# movies.groupby('Genre')['IMDB_Rating'].max()
```

```
Genre
Action       9.0
Adventure    8.6
Animation    8.6
Biography    8.9
Comedy       8.6
Crime        9.2
Drama        9.3
Family       7.8
Fantasy      8.1
Film-Noir    8.1
Horror       8.5
Mystery      8.4
Thriller     7.8
Western      8.8
Name: IMDB_Rating, dtype: float64
```

```
# find number of movies done by each actor
# movies['Star1'].value_counts()

movies.groupby('Star1')['Series_Title'].count().sort_values(ascending=False)
```

```
Star1
Tom Hanks             12
Robert De Niro        11
Clint Eastwood        10
Al Pacino             10
Leonardo DiCaprio      9
                      ..
Glen Hansard           1
Giuseppe Battiston     1
Giulietta Masina       1
Gerardo Taracena       1
Ömer Faruk Sorak       1
Name: Series_Title, Length: 660, dtype: int64
```

```
# GroupBy Attributes and Methods
# find total number of groups -> len
# find items in each group -> size
# first()/last() -> nth item
# get_group -> vs filtering
# groups
# describe
# sample
# nunique
```

```
len(movies.groupby('Genre'))
```

```
14
```

```
movies['Genre'].nunique()
```

```
14
```

```
movies.groupby('Genre').size()
```

```
Genre
Action       172
Adventure     72
Animation     82
Biography     88
Comedy       155
Crime        107
Drama        289
Family         2
Fantasy        2
Film-Noir      3
Horror        11
Mystery       12
Thriller       1
Western        4
dtype: int64
```

```
genres = movies.groupby('Genre')
# genres.first()
# genres.last()
genres.nth(6)
```

Genre	Series_Title	Released_Year	Runtime	IMDB_Rating	Director	Star1
Action	Star Wars: Episode V - The Empire Strikes Back	1980	124	8.7	Irvin Kershner	Mark Hamil
Adventure	North by Northwest	1959	136	8.3	Alfred Hitchcock	Cary Grant
Animation	WALL·E	2008	98	8.4	Andrew Stanton	Ben Burtt
Biography	Braveheart	1995	178	8.3	Mel Gibson	Mel Gibson
Comedy	The Great Dictator	1940	125	8.4	Charles Chaplin	Charles Chaplin

```
movies['Genre'].value_counts()
```

```
Drama        289
Action       172
Comedy       155
Crime        107
Biography     88
Animation     82
Adventure     72
Mystery       12
Horror        11
Western        4
Film-Noir      3
Fantasy        2
Family         2
Thriller       1
Name: Genre, dtype: int64
```

```
genres.get_group('Fantasy')

movies[movies['Genre'] == 'Fantasy']
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	Star1	No_of_Votes	Gross	Metascore
321	Das Cabinet des Dr. Caligari	1920	76	Fantasy	8.1	Robert Wiene	Werner Krauss	57428	337574718.0	NaN

```
genres.groups
```

```
{'Action': [2, 5, 8, 10, 13, 14, 16, 29, 30, 31, 39, 42, 44, 55, 57, 59, 60, 63, 68, 72, 106, 109, 129, 130, 134, 140,
142, 144, 152, 155, 160, 161, 166, 168, 171, 172, 177, 181, 194, 201, 202, 216, 217, 223, 224, 236, 241, 262, 275, 294,
308, 320, 325, 326, 331, 337, 339, 340, 343, 345, 348, 351, 353, 356, 357, 362, 368, 369, 375, 376, 390, 410, 431, 436,
473, 477, 479, 482, 488, 493, 496, 502, 507, 511, 532, 535, 540, 543, 564, 569, 570, 573, 577, 582, 583, 602, 605, 608,
615, 623, ...], 'Adventure': [21, 47, 93, 110, 114, 116, 118, 137, 178, 179, 191, 193, 209, 226, 231, 247, 267, 273,
281, 300, 301, 304, 306, 323, 329, 361, 366, 377, 402, 406, 415, 426, 458, 470, 497, 498, 506, 513, 514, 537, 549, 552,
553, 566, 576, 604, 609, 618, 638, 647, 675, 681, 686, 692, 711, 713, 739, 755, 781, 797, 798, 851, 873, 884, 912, 919,
947, 957, 964, 966, 984, 991], 'Animation': [23, 43, 46, 56, 58, 61, 66, 70, 101, 135, 146, 151, 158, 170, 197, 205,
211, 213, 219, 229, 230, 242, 245, 246, 270, 330, 332, 358, 367, 378, 386, 389, 394, 395, 399, 401, 405, 409, 469, 499,
510, 516, 518, 522, 578, 586, 592, 595, 596, 599, 633, 640, 643, 651, 665, 672, 694, 728, 740, 741, 744, 756, 758, 761,
771, 783, 796, 799, 822, 828, 843, 875, 891, 892, 902, 906, 920, 956, 971, 976, 986, 992], 'Biography': [7, 15, 18, 35,
38, 54, 102, 107, 131, 139, 147, 157, 159, 173, 176, 212, 215, 218, 228, 235, 243, 263, 276, 282, 290, 298, 317, 328,
338, 342, 346, 359, 360, 365, 372, 373, 385, 411, 416, 418, 424, 429, 484, 525, 536, 542, 545, 575, 579, 587, 600, 606,
614, 622, 632, 635, 644, 649, 650, 657, 671, 673, 684, 729, 748, 753, 757, 759, 766, 770, 779, 809, 810, 815, 820, 831,
849, 858, 877, 882, 897, 910, 915, 923, 940, 949, 952, 987], 'Comedy': [19, 26, 51, 52, 64, 78, 83, 95, 96, 112, 117,
120, 127, 128, 132, 153, 169, 183, 192, 204, 207, 208, 214, 221, 233, 238, 240, 250, 251, 252, 256, 261, 266, 277, 284,
311, 313, 316, 318, 322, 327, 374, 379, 381, 392, 396, 403, 413, 414, 417, 427, 435, 445, 446, 449, 455, 459, 460, 463,
464, 466, 471, 472, 475, 481, 490, 494, 500, 503, 509, 526, 528, 530, 531, 533, 538, 539, 541, 547, 557, 558, 562, 563,
565, 574, 591, 593, 594, 598, 613, 626, 630, 660, 662, 667, 679, 680, 683, 687, 701, ...], 'Crime': [1, 3, 4, 6, 22,
25, 27, 28, 33, 37, 41, 71, 77, 79, 86, 87, 103, 108, 111, 113, 123, 125, 133, 136, 162, 163, 164, 165, 180, 186, 187,
189, 198, 222, 232, 239, 255, 257, 287, 288, 299, 305, 335, 363, 364, 380, 384, 397, 437, 438, 441, 442, 444, 450, 451,
465, 474, 480, 485, 487, 505, 512, 519, 520, 523, 527, 546, 556, 560, 584, 597, 603, 607, 611, 621, 639, 653, 664, 669,
```

```
676, 695, 708, 723, 762, 763, 767, 775, 791, 795, 802, 811, 823, 827, 833, 885, 895, 921, 922, 926, 938, ...], 'Drama':
[0, 9, 11, 17, 20, 24, 32, 34, 36, 40, 45, 50, 53, 62, 65, 67, 73, 74, 76, 80, 82, 84, 85, 88, 89, 90, 91, 92, 94, 97,
98, 99, 100, 104, 105, 121, 122, 124, 126, 138, 141, 143, 148, 149, 150, 154, 156, 167, 174, 175, 182, 184, 185, 188,
190, 195, 196, 199, 200, 203, 206, 210, 225, 227, 234, 237, 244, 248, 249, 253, 254, 258, 259, 260, 264, 265, 268, 269,
272, 274, 278, 279, 280, 283, 285, 286, 289, 291, 292, 293, 295, 296, 297, 302, 303, 307, 310, 312, 314, 315, ...],
'Family': [688, 698], 'Fantasy': [321, 568], 'Film-Noir': [309, 456, 712], 'Horror': [49, 75, 271, 419, 544, 707, 724,
844, 876, 932, 948], 'Mystery': [69, 81, 119, 145, 220, 393, 420, 714, 829, 899, 959, 961], 'Thriller': [700],
'Western': [12, 48, 115, 691]}
```

```
genres.describe()
```

	Runtime								IMDB_Rati	
	count	mean	std	min	25%	50%	75%	max	count	mea
Genre										
Action	172.0	129.046512	28.500706	45.0	110.75	127.5	143.25	321.0	172.0	7.9
Adventure	72.0	134.111111	33.317320	88.0	109.00	127.0	149.00	228.0	72.0	7.9
Animation	82.0	99.585366	14.530471	71.0	90.00	99.5	106.75	137.0	82.0	7.9
Biography	88.0	136.022727	25.514466	93.0	120.00	129.0	146.25	209.0	88.0	7.9
Comedy	155.0	112.129032	22.946213	68.0	96.00	106.0	124.50	188.0	155.0	7.9
Crime	107.0	126.392523	27.689231	80.0	106.50	122.0	141.50	229.0	107.0	8.0
Drama	289.0	124.737024	27.740490	64.0	105.00	121.0	137.00	242.0	289.0	7.9
Family	2.0	107.500000	10.606602	100.0	103.75	107.5	111.25	115.0	2.0	7.8
Fantasy	2.0	85.000000	12.727922	76.0	80.50	85.0	89.50	94.0	2.0	8.0
Film-Noir	3.0	104.000000	4.000000	100.0	102.00	104.0	106.00	108.0	3.0	7.9
Horror	11.0	102.090909	13.604812	71.0	98.00	103.0	109.00	122.0	11.0	7.9
Mystery	12.0	119.083333	14.475423	96.0	110.75	117.5	130.25	138.0	12.0	7.9
Thriller	1.0	108.000000	NaN	108.0	108.00	108.0	108.00	108.0	1.0	7.8
Western	4.0	148.250000	17.153717	132.0	134.25	148.0	162.00	165.0	4.0	8.3

14 rows × 40 columns

```
genres.sample(2,replace=True)
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	
944	Batoru rowaiaru	2000	114	Action	7.6	Kinji Fukasaku	T Fu
625	Apocalypto	2006	139	Action	7.8	Mel Gibson	G Ta
991	Kelly's Heroes	1970	144	Adventure	7.6	Brian G. Hutton	Eas
300	Ben-Hur	1959	212	Adventure	8.1	William Wyler	Cl I
891	Incredibles 2	2018	118	Animation	7.6	Brad Bird	C I
389	The Iron Giant	1999	86	Animation	8.0	Brad Bird	Mar
536	All the President's Men	1976	138	Biography	7.9	Alan J. Pakula	He
635	Walk the Line	2005	136	Biography	7.8	James Mangold	J P
826	Barton Fink	1991	116	Comedy	7.7	Joel Coen	
732	Me and Earl and the Dying Girl	2015	105	Comedy	7.7	Alfonso Gomez-Rejon	T
438	Touch of Evil	1958	95	Crime	8.0	Orson Welles	Cl I
222	Prisoners	2013	153	Crime	8.1	Denis Villeneuve	Ja
555	High Noon	1952	85	Drama	7.9	Fred Zinnemann	C
314	Gone with the Wind	1939	238	Drama	8.1	Victor Fleming	C
698	Willy Wonka & the Chocolate Factory	1971	100	Family	7.8	Mel Stuart	
698	Willy Wonka & the Chocolate Factory	1971	100	Family	7.8	Mel Stuart	
321	Das Cabinet des Dr. Caligari	1920	76	Fantasy	8.1	Robert Wiene	V I

```
genres.nunique()
```

	Series_Title	Released_Year	Runtime	IMDB_Rating	Director	Star1	No_of_Votes	Gross	Metascore
Genre									
Action	172	61	78	15	123	121	172	172	50
Adventure	72	49	58	10	59	59	72	72	33
Animation	82	35	41	11	51	77	82	82	29
Biography	88	44	56	13	76	72	88	88	40
Comedy	155	72	70	11	113	133	155	155	44
Crime	106	56	65	14	86	85	107	107	39
Drama	289	83	95	14	211	250	288	287	52
Family	2	2	2	1	2	2	2	2	2
Fantasy	2	2	2	2	2	2	2	2	0
Film-Noir	3	3	3	3	3	3	3	3	3
Horror	11	11	10	8	10	11	11	11	9
Mystery	12	11	10	8	10	11	12	12	7
Thriller	1	1	1	1	1	1	1	1	1
Western	4	4	4	4	2	2	4	4	4

```
# agg method
# passing dict
genres.agg(
    {
        'Runtime':'mean',
        'IMDB_Rating':'mean',
        'No_of_Votes':'sum',
        'Gross':'sum',
        'Metascore':'min'
    }
)
```

	Runtime	IMDB_Rating	No_of_Votes	Gross	Metascore
Genre					
Action	129.046512	7.949419	72282412	3.263226e+10	33.0
Adventure	134.111111	7.937500	22576163	9.496922e+09	41.0
Animation	99.585366	7.930488	21978630	1.463147e+10	61.0
Biography	136.022727	7.938636	24006844	8.276358e+09	48.0
Comedy	112.129032	7.901290	27620327	1.566387e+10	45.0
Crime	126.392523	8.016822	33533615	8.452632e+09	47.0
Drama	124.737024	7.957439	61367304	3.540997e+10	28.0
Family	107.500000	7.800000	551221	4.391106e+08	67.0
Fantasy	85.000000	8.000000	146222	7.827267e+08	NaN
Film-Noir	104.000000	7.966667	367215	1.259105e+08	94.0
Horror	102.090909	7.909091	3742556	1.034649e+09	46.0
Mystery	119.083333	7.975000	4203004	1.256417e+09	52.0
Thriller	108.000000	7.800000	27733	1.755074e+07	81.0
Western	148.250000	8.350000	1289665	5.822151e+07	69.0

```
# passing list
genres.agg(['min','max','mean','sum'])
```

	Runtime				IMDB_Rating				No_of_Votes	
	min	max	mean	sum	min	max	mean	sum	min	max
Genre										
Action	45	321	129.046512	22196	7.6	9.0	7.949419	1367.3	25312	2303232
Adventure	88	228	134.111111	9656	7.6	8.6	7.937500	571.5	29999	1512360
Animation	71	137	99.585366	8166	7.6	8.6	7.930488	650.3	25229	999790
Biography	93	209	136.022727	11970	7.6	8.9	7.938636	698.6	27254	1213505
Comedy	68	188	112.129032	17380	7.6	8.6	7.901290	1224.7	26337	939631
Crime	80	229	126.392523	13524	7.6	9.2	8.016822	857.8	27712	1826188
Drama	64	242	124.737024	36049	7.6	9.3	7.957439	2299.7	25088	2343110
Family	100	115	107.500000	215	7.8	7.8	7.800000	15.6	178731	372490
Fantasy	76	94	85.000000	170	7.9	8.1	8.000000	16.0	57428	88794
Film-Noir	100	108	104.000000	312	7.8	8.1	7.966667	23.9	59556	158731
Horror	71	122	102.090909	1123	7.6	8.5	7.909091	87.0	27007	787806
Mystery	96	138	119.083333	1429	7.6	8.4	7.975000	95.7	33982	1129894
Thriller	108	108	108.000000	108	7.8	7.8	7.800000	7.8	27733	27733
Western	132	165	148.250000	593	7.8	8.8	8.350000	33.4	65659	688390

```
# Adding both the syntax
genres.agg(
    {
        'Runtime':['min','mean'],
        'IMDB_Rating':'mean',
        'No_of_Votes':['sum','max'],
        'Gross':'sum',
        'Metascore':'min'
    }
)
```

	Runtime		IMDB_Rating	No_of_Votes		Gross	Metascore
	min	mean	mean	sum	max	sum	min
Genre							
Action	45	129.046512	7.949419	72282412	2303232	3.263226e+10	33.0
Adventure	88	134.111111	7.937500	22576163	1512360	9.496922e+09	41.0
Animation	71	99.585366	7.930488	21978630	999790	1.463147e+10	61.0
Biography	93	136.022727	7.938636	24006844	1213505	8.276358e+09	48.0
Comedy	68	112.129032	7.901290	27620327	939631	1.566387e+10	45.0
Crime	80	126.392523	8.016822	33533615	1826188	8.452632e+09	47.0
Drama	64	124.737024	7.957439	61367304	2343110	3.540997e+10	28.0
Family	100	107.500000	7.800000	551221	372490	4.391106e+08	67.0
Fantasy	76	85.000000	8.000000	146222	88794	7.827267e+08	NaN
Film-Noir	100	104.000000	7.966667	367215	158731	1.259105e+08	94.0
Horror	71	102.090909	7.909091	3742556	787806	1.034649e+09	46.0
Mystery	96	119.083333	7.975000	4203004	1129894	1.256417e+09	52.0
Thriller	108	108.000000	7.800000	27733	27733	1.755074e+07	81.0
Western	132	148.250000	8.350000	1289665	688390	5.822151e+07	69.0

```
# looping on groups
df = pd.DataFrame(columns=movies.columns)
for group,data in genres:
  df = df.append(data[data['IMDB_Rating'] == data['IMDB_Rating'].max()])

df
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	
2	The Dark Knight	2008	152	Action	9.0	Christopher Nolan	Chr
21	Interstellar	2014	169	Adventure	8.6	Christopher Nolan	McC
23	Sen to Chihiro no kamikakushi	2001	125	Animation	8.6	Hayao Miyazaki	
7	Schindler's List	1993	195	Biography	8.9	Steven Spielberg	Lia
19	Gisaengchung	2019	132	Comedy	8.6	Bong Joon Ho	Kan
26	La vita è bella	1997	116	Comedy	8.6	Roberto Benigni	
1	The Godfather	1972	175	Crime	9.2	Francis Ford Coppola	
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Ti
688	E.T. the Extra-Terrestrial	1982	115	Family	7.8	Steven Spielberg	
698	Willy Wonka & the Chocolate Factory	1971	100	Family	7.8	Mel Stuart	G
[illegible]	Das Cabinet	[illegible]	[illegible]	[illegible]	[illegible]	Robert	

```
# split (apply) combine
# apply -> builtin function

genres.apply(min)
```

Genre	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director
Action	300	1924	45	Action	7.6	Abhishek Chaubey
Adventure	2001: A Space Odyssey	1925	88	Adventure	7.6	Akira Kurosawa
Animation	Akira	1940	71	Animation	7.6	Adam Elliot
Biography	12 Years a Slave	1928	93	Biography	7.6	Adam McKay
Comedy	(500) Days of Summer	1921	68	Comedy	7.6	Alejandro G. Iñárritu
Crime	12 Angry Men	1931	80	Crime	7.6	Akira Kurosawa
Drama	1917	1925	64	Drama	7.6	Aamir Khan
Family	E.T. the Extra-Terrestrial	1971	100	Family	7.8	Mel Stuart
Fantasy	Das Cabinet des Dr. Caligari	1920	76	Fantasy	7.9	F.W. Murnau

```
# find number of movies starting with A for each group

def foo(group):
  return group['Series_Title'].str.startswith('A').sum()
```

```
genres.apply(foo)
```

```
Genre
Action       10
Adventure     2
Animation     2
Biography     9
Comedy       14
Crime         4
Drama        21
Family        0
Fantasy       0
Film-Noir     0
Horror        1
Mystery       0
Thriller      0
Western       0
dtype: int64
```

```
# find ranking of each movie in the group according to IMDB score

def rank_movie(group):
  group['genre_rank'] = group['IMDB_Rating'].rank(ascending=False)
  return group


genres.apply(rank_movie)
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	St
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Rob
1	The Godfather	1972	175	Crime	9.2	Francis Ford Coppola	Ma Bra
2	The Dark Knight	2008	152	Action	9.0	Christopher Nolan	Chris
3	The Godfather: Part II	1974	202	Crime	9.0	Francis Ford Coppola	Al Pa
4	12 Angry Men	1957	96	Crime	9.0	Sidney Lumet	H Fc
...	...	...	...	...	...	...	
995	Breakfast at Tiffany's	1961	115	Comedy	7.6	Blake Edwards	Au Hep
996	Giant	1956	201	Drama	7.6	George	Eliza

```
# find normalized IMDB rating group wise

def normal(group):
  group['norm_rating'] = (group['IMDB_Rating'] - group['IMDB_Rating'].min())/(group['IMDB_Rating'].max() - group['IMDB_Ratir
  return group

genres.apply(normal)
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	St
0	The Shawshank Redemption	1994	142	Drama	9.3	Frank Darabont	Rob
1	The Godfather	1972	175	Crime	9.2	Francis Ford Coppola	Ma Bra
2	The Dark Knight	2008	152	Action	9.0	Christopher Nolan	Chris
3	The Godfather: Part II	1974	202	Crime	9.0	Francis Ford Coppola	Al Pa
4	12 Angry Men	1957	96	Crime	9.0	Sidney Lumet	H Fc
...	...	...	...	...	...	...	
995	Breakfast at Tiffany's	1961	115	Comedy	7.6	Blake Edwards	Au Hep
996	Giant	1956	201	Drama	7.6	George	Eliza

```
# groupby on multiple cols
duo = movies.groupby(['Director','Star1'])
duo
# size
duo.size()
# get_group
duo.get_group(('Aamir Khan','Amole Gupte'))
```

	Series_Title	Released_Year	Runtime	Genre	IMDB_Rating	Director	Star1	No_of_Votes	Gross	Metascore
65	Taare Zameen Par	2007	165	Drama	8.4	Aamir Khan	Amole Gupte	168895	1223869.0	NaN

```
# find the most earning actor->director combo
duo['Gross'].sum().sort_values(ascending=False).head(1)
```

```
Director        Star1
Akira Kurosawa  Toshirô Mifune    2.999877e+09
Name: Gross, dtype: float64
```

```
# find the best(in-terms of metascore(avg)) actor->genre combo
movies.groupby(['Star1','Genre'])['Metascore'].mean().reset_index().sort_values('Metascore',ascending=False).head(1)
```

	Star1	Genre	Metascore
230	Ellar Coltrane	Drama	100.0

```
# agg on multiple groupby
duo.agg(['min','max','mean'])
```

		Runtime			IMDB_Rating			No_of_Votes			Gro
		min	max	mean	min	max	mean	min	max	mean	min
Director	**Star1**										
Aamir Khan	**Amole Gupte**	165	165	165.0	8.4	8.4	8.4	168895	168895	168895.0	1
Aaron Sorkin	**Eddie Redmayne**	129	129	129.0	7.8	7.8	7.8	89896	89896	89896.0	853
Abdellatif Kechiche	**Léa Seydoux**	180	180	180.0	7.7	7.7	7.7	138741	138741	138741.0	2
Abhishek Chaubey	**Shahid Kapoor**	148	148	148.0	7.8	7.8	7.8	27175	27175	27175.0	218
Abhishek Kapoor	**Amit Sadh**	130	130	130.0	7.7	7.7	7.7	32628	32628	32628.0	1
...	**...**	...	...	...	...	...	...	...	...	...	
Zaza Urushadze	**Lembit Ulfsak**	87	87	87.0	8.2	8.2	8.2	40382	40382	40382.0	
Zoya Akhtar	**Hrithik Roshan**	155	155	155.0	8.1	8.1	8.1	67927	67927	67927.0	3

...

Excercise

```
ipl = pd.read_csv('/content/deliveries.csv')
ipl.head()
```

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_strik
0	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	1	DA Warner	S Dhaw
1	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	2	DA Warner	S Dhaw
2	1	1	Sunrisers Hyderabad	Royal Challengers Bangalore	1	3	DA Warner	S Dhaw
			Sunrisers	Royal			DA	
4	1	1	Hyderabad	Challengers	1	5	Warner	S Dhaw

```
ipl.shape
```

```
(179078, 21)
```

```
# find the top 10 batsman in terms of runs
ipl.groupby('batsman')['batsman_runs'].sum().sort_values(ascending=False).head(10)
```

```
batsman
V Kohli          5434
SK Raina         5415
RG Sharma        4914
DA Warner        4741
S Dhawan         4632
CH Gayle         4560
MS Dhoni         4477
RV Uthappa       4446
AB de Villiers   4428
G Gambhir        4223
Name: batsman_runs, dtype: int64
```

```
# find the batsman with max no of sixes
six = ipl[ipl['batsman_runs'] == 6]

six.groupby('batsman')['batsman'].count().sort_values(ascending=False).head(1).index[0]
```

```
'CH Gayle'
```

```
# find batsman with most number of 4's and 6's in last 5 overs
temp_df = ipl[ipl['over'] > 15]
temp_df = temp_df[(temp_df['batsman_runs'] == 4) | (temp_df['batsman_runs'] == 6)]
temp_df.groupby('batsman')['batsman'].count().sort_values(ascending=False).head(1).index[0]
```

```
'MS Dhoni'
```

```
# find V Kohli's record against all teams
temp_df = ipl[ipl['batsman'] == 'V Kohli']

temp_df.groupby('bowling_team')['batsman_runs'].sum().reset_index()
```

	bowling_team	batsman_runs
0	Chennai Super Kings	749
1	Deccan Chargers	306
2	Delhi Capitals	66
3	Delhi Daredevils	763
4	Gujarat Lions	283
5	Kings XI Punjab	636
6	Kochi Tuskers Kerala	50
7	Kolkata Knight Riders	675
8	Mumbai Indians	628
9	Pune Warriors	128