```
import pandas as pd
nba=pd.read_csv("nbaallelo.csv")
print(type(nba))
nba.shape
pd.set_option("display.max.columns",None)
pd.set_option("display.precision",2)
nba.tail(10)
nba.info()
```

```
<class 'pandas.core.frame.DataFrame'>
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 126314 entries, 0 to 126313
Data columns (total 23 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   gameorder      126314 non-null  int64
 1   game_id        126314 non-null  object
 2   lg_id          126314 non-null  object
 3   _iscopy        126314 non-null  int64
 4   year_id        126314 non-null  int64
 5   date_game      126314 non-null  object
 6   seasongame     126314 non-null  int64
 7   is_playoffs    126314 non-null  int64
 8   team_id        126314 non-null  object
 9   fran_id        126314 non-null  object
 10  pts            126314 non-null  int64
 11  elo_i          126314 non-null  float64
 12  elo_n          126314 non-null  float64
 13  win_equiv      126314 non-null  float64
 14  opp_id         126314 non-null  object
 15  opp_fran       126314 non-null  object
 16  opp_pts        126314 non-null  int64
 17  opp_elo_i      126314 non-null  float64
 18  opp_elo_n      126314 non-null  float64
 19  game_location  126314 non-null  object
 20  game_result    126314 non-null  object
 21  forecast       126314 non-null  float64
 22  notes          5424 non-null    object
dtypes: float64(6), int64(7), object(10)
memory usage: 22.2+ MB
```

```
nba.describe()
```

	gameorder	_iscopy	year_id	seasongame	is_playoffs	pts	elo_i	
count	126314.00	126314.0	126314.00	126314.00	126314.00	126314.00	126314.00	12
mean	31579.00	0.5	1988.20	43.53	0.06	102.73	1495.24	
std	18231.93	0.5	17.58	25.38	0.24	14.81	112.14	
min	1.00	0.0	1947.00	1.00	0.00	0.00	1091.64	
25%	15790.00	0.0	1975.00	22.00	0.00	93.00	1417.24	
50%	31579.00	0.5	1990.00	43.00	0.00	103.00	1500.95	
75%	47368.00	1.0	2003.00	65.00	0.00	112.00	1576.06	
max	63157.00	1.0	2015.00	108.00	1.00	186.00	1853.10	

```
import numpy as np
nba.describe(include=np.object)
```

```
<ipython-input-29-86bbc4828ab1>:2: DeprecationWarning: `np.object` is a deprecat
Deprecated in NumPy 1.20; for more details and guidance: https://numpy.org/devdo
  nba.describe(include=np.object)
```

	game_id	lg_id	date_game	team_id	fran_id	opp_id	opp_fran	game_loc
count	126314	126314	126314	126314	126314	126314	126314	
unique	63157	2	12426	104	53	104	53	
top	194611010TRH	NBA	4/13/2011	BOS	Lakers	BOS	Lakers	

```
s=pd.Series([1,3,5,6,8])
type(s)
s
```

```
     0    1
     1    3
     2    5
     3    6
     4    8
     dtype: int64
```

```
type(s.values)
s.index
s.values
```

```
     array([1, 3, 5, 6, 8])
```

```
games=pd.Series([5997,5078,5769],index=["BOS","LAL","NYK"])
games
```

```
     BOS    5997
     LAL    5078
     NYK    5769
     dtype: int64
```

```
points=pd.Series({"BOS":104.47,"LAL":108.14})
points.index
points.values
"BOS" in points
points
```

```
     BOS    104.47
     LAL    108.14
     dtype: float64
```

```
data=pd.DataFrame({
"num_games":games,
"average_points":points
})
data
```

	num_games	average_points
BOS	5997	104.47
LAL	5078	108.14
NYK	5769	NaN

```
data.index
data.values
```

```
     array([[5997.  ,  104.47],
            [5078.  ,  108.14],
            [5769.  ,     nan]])
```

```
data.axes
```

```
     [Index(['BOS', 'LAL', 'NYK'], dtype='object'),
      Index(['num_games', 'average_points'], dtype='object')]
```

```
data.keys()
```

```
     Index(['num_games', 'average_points'], dtype='object')
```

```
"num_games" in data
```

```
     True
```

```
nba.index
```

```
     RangeIndex(start=0, stop=126314, step=1)
```

```
nba.axes
```

```
[RangeIndex(start=0, stop=126314, step=1),
 Index(['gameorder', 'game_id', 'lg_id', '_iscopy', 'year_id', 'date_game',
        'seasongame', 'is_playoffs', 'team_id', 'fran_id', 'pts', 'elo_i',
        'elo_n', 'win_equiv', 'opp_id', 'opp_fran', 'opp_pts', 'opp_elo_i',
        'opp_elo_n', 'game_location', 'game_result', 'forecast', 'notes'],
       dtype='object')]
```

```
nba.keys()
```

```
Index(['gameorder', 'game_id', 'lg_id', '_iscopy', 'year_id', 'date_game',
       'seasongame', 'is_playoffs', 'team_id', 'fran_id', 'pts', 'elo_i',
       'elo_n', 'win_equiv', 'opp_id', 'opp_fran', 'opp_pts', 'opp_elo_i',
       'opp_elo_n', 'game_location', 'game_result', 'forecast', 'notes'],
      dtype='object')
```

```
games["BOS"]
```

```
5997
```

```
data
```

	num_games	average_points
BOS	5997	104.47
LAL	5078	108.14
NYK	5769	NaN

```
data.at["LAL","average_points"]
```

➡ 108.14

```
data["num_games"]
```

```
BOS    5997
LAL    5078
NYK    5769
Name: num_games, dtype: int64
```

```
data.num_games
```

```
BOS    5997
LAL    5078
NYK    5769
Name: num_games, dtype: int64
```

```
data.loc["BOS"]
```

```
num_games        5997.00
average_points    104.47
Name: BOS, dtype: float64
```

```
data.loc["LAL":"NYK",["num_games","average_points"]]
```

	num_games	average_points
LAL	5078	108.14
NYK	5769	NaN

```
nba.loc[10001:10010,["team_id","pts","opp_id","opp_pts"]]
```

	team_id	pts	opp_id	opp_pts
10001	STL	89	DET	112
10002	LAL	111	NYK	95
10003	NYK	95	LAL	111
10004	PHW	128	CIN	124
10005	CIN	124	PHW	128
10006	CIN	136	PHW	130
10007	PHW	130	CIN	136
10008	DET	121	SYR	112
10009	SYR	112	DET	121
10010	BOS	120	NYK	106

```
nba["year_id"]>=2010
```

```
0          False
1          False
2          False
3          False
4          False
           ...
126309     True
126310     True
126311     True
126312     True
126313     True
Name: year_id, Length: 126314, dtype: bool
```

```
nba[nba["year_id"]>=2010]
```

	gameorder	game_id	lg_id	_iscopy	year_id	date_game	seasongame	is_playoffs	team_id	fran_id	pts	elo_i	e
111032	55517	200910270CLE	NBA	1	2010	10/27/2009	1	0	BOS	Celtics	95	1616.20	16
111033	55517	200910270CLE	NBA	0	2010	10/27/2009	1	0	CLE	Cavaliers	89	1682.45	16
111034	55518	200910270DAL	NBA	1	2010	10/27/2009	1	0	WAS	Wizards	102	1361.42	13
111035	55518	200910270DAL	NBA	0	2010	10/27/2009	1	0	DAL	Mavericks	91	1586.43	15
111036	55519	200910270LAL	NBA	1	2010	10/27/2009	1	0	LAC	Clippers	92	1317.89	13
...	...	...	...	...	...	...	...	...	...	...	...	...	
126309	63155	201506110CLE	NBA	0	2015	6/11/2015	100	1	CLE	Cavaliers	82	1723.41	17
126310	63156	201506140GSW	NBA	0	2015	6/14/2015	102	1	GSW	Warriors	104	1809.98	18
126311	63156	201506140GSW	NBA	1	2015	6/14/2015	101	1	CLE	Cavaliers	91	1704.39	17
126312	63157	201506170CLE	NBA	0	2015	6/16/2015	102	1	CLE	Cavaliers	97	1700.74	16
126313	63157	201506170CLE	NBA	1	2015	6/16/2015	103	1	GSW	Warriors	105	1813.63	18

15282 rows × 23 columns

```
nba[nba["notes"].notna()]
```

	gameorder	game_id	lg_id	_iscopy	year_id	date_game	seasongame	is_playoffs	team_id	fran_id	pts	elo_i
930	466	194801300BLB	NBA	1	1948	1/30/1948	28	0	WSC	Capitols	71	1469.35
931	466	194801300BLB	NBA	0	1948	1/30/1948	31	0	BLB	Baltimore	95	1434.05
1084	543	194803230CHS	NBA	1	1948	3/23/1948	49	1	WSC	Capitols	70	1465.90
1085	543	194803230CHS	NBA	0	1948	3/23/1948	49	1	CHS	Stags	74	1440.72
1088	545	194803250CHS	NBA	1	1948	3/25/1948	49	1	BLB	Baltimore	75	1477.26
...	...	...	...	...	...	...	...	...	...	...	...	...
122207	61104	201401160ATL	NBA	0	2014	1/16/2014	39	0	ATL	Hawks	110	1505.05
123916	61959	201411120MIN	NBA	1	2015	11/12/2014	8	0	HOU	Rockets	113	1635.77
123917	61959	201411120MIN	NBA	0	2015	11/12/2014	7	0	MIN	Timberwolves	101	1502.92
124862	62432	201501150MIL	NBA	0	2015	1/15/2015	40	0	MIL	Bucks	95	1499.90
124863	62432	201501150MIL	NBA	1	2015	1/15/2015	41	0	NYK	Knicks	79	1304.46

5424 rows × 23 columns

```
nba[(nba["fran_id"].str.endswith("ers"))&(nba["year_id"]>=2010)]
```

	gameorder	game_id	lg_id	_iscopy	year_id	date_game	seasongame	is_playoffs	team_id	fran_id	pts	elo_i
111033	55517	200910270CLE	NBA	0	2010	10/27/2009	1	0	CLE	Cavaliers	89	1682.45
111036	55519	200910270LAL	NBA	1	2010	10/27/2009	1	0	LAC	Clippers	92	1317.89
111037	55519	200910270LAL	NBA	0	2010	10/27/2009	1	0	LAL	Lakers	99	1718.75
111038	55520	200910270POR	NBA	0	2010	10/27/2009	1	0	POR	Trailblazers	96	1600.85
111040	55521	200910280ATL	NBA	1	2010	10/28/2009	1	0	IND	Pacers	109	1495.63
...	...	...	...	...	...	...	...	...	...	...	...	...
126304	63153	201506070GSW	NBA	1	2015	6/7/2015	98	1	CLE	Cavaliers	95	1708.17
126307	63154	201506090CLE	NBA	0	2015	6/9/2015	99	1	CLE	Cavaliers	96	1716.87
126309	63155	201506110CLE	NBA	0	2015	6/11/2015	100	1	CLE	Cavaliers	82	1723.41
126311	63156	201506140GSW	NBA	1	2015	6/14/2015	101	1	CLE	Cavaliers	91	1704.39
126312	63157	201506170CLE	NBA	0	2015	6/16/2015	102	1	CLE	Cavaliers	97	1700.74

3079 rows × 23 columns

```
points=nba["pts"]
points.count()
points.sum()
points.max()
points.min()
points.mean()
points.mode()
points.median()
points.var(ddof=0)
points.std(ddof=0)
points.quantile(0.95)
points.describe()
```

```
    count    126314.00
    mean        102.73
    std          14.81
    min           0.00
    25%          93.00
    50%         103.00
    75%         112.00
    max         186.00
    Name: pts, dtype: float64
```

```
points.corr(nba["opp_pts"])
```

```
    0.5924908479146879
```

```
nba.groupby(["team_id","opp_id"])["pts"].mean()
```

```
    team_id  opp_id
    ANA      DLC       114.70
             DNR       100.70
             HSM       116.00
             INA       121.17
             KEN       106.33
                        ...
    WSC      STB        74.04
             SYR        75.75
             TRH        77.00
             TRI        72.75
             WAT        85.50
    Name: pts, Length: 2650, dtype: float64
```

```
nba[nba["team_id"]=="BOS"].groupby(["year_id","game_result"])["game_id"].count()
```

```
    year_id  game_result
    1947     L              38
             W              22
    1948     L              30
             W              21
    1949     L              35
                            ..
    2013     W              43
    2014     L              57
             W              25
    2015     L              46
             W              40
    Name: game_id, Length: 138, dtype: int64
```

```
nba[nba["team_id"]=="BOS"].groupby("year_id")["game_result"].value_counts()
```

```
    year_id  game_result
    1947     L              38
             W              22
    1948     L              30
             W              21
    1949     L              35
                            ..
    2013     W              43
    2014     L              57
             W              25
    2015     L              46
             W              40
    Name: game_result, Length: 138, dtype: int64
```

```
nba[nba["team_id"]=="BOS"].groupby(nba["year_id"]//10*10)["game_result"].value_counts()
```

```
       year_id  game_result
       1940     L             103
                W              68
       1950     W             439
                L             326
       1960     W             658
                L             276
       1970     W             524
                L             376
       1980     W             683
                L             287
       1990     L             430
                W             392
       2000     W             469
                L             428
       2010     W             284
                L             254
       Name: game_result, dtype: int64
```

```
nba["team_id"]
```

```
       0          TRH
       1          NYK
       2          CHS
       3          NYK
       4          DTF
                 ...
       126309     CLE
       126310     GSW
       126311     CLE
       126312     CLE
       126313     GSW
       Name: team_id, Length: 126314, dtype: object
```

```
nba["team_id"].value_counts()
```

```
       BOS     5997
       NYK     5769
       LAL     5078
       DET     4985
       PHI     4533
              ...
       INJ       60
       PIT       60
       DTF       60
       TRH       60
       SDS       11
       Name: team_id, Length: 104, dtype: int64
```

```
nba["fran_id"].value_counts()
```

```
       Lakers          6024
       Celtics         5997
       Knicks          5769
       Warriors        5657
       Pistons         5650
       Sixers          5644
       Hawks           5572
       Kings           5475
       Wizards         4582
       Spurs           4309
       Bulls           4307
       Pacers          4227
       Thunder         4178
       Rockets         4154
       Nuggets         4120
       Nets            4106
       Suns            4080
       Bucks           4034
       Trailblazers    3870
       Cavaliers       3810
       Clippers        3733
       Jazz            3555
       Mavericks       3013
       Heat            2371
       Pelicans        2254
       Magic           2207
       Timberwolves    2131
       Grizzlies       1657
       Raptors         1634
       Hornets          894
```

```
            Colonels         846
            Squires          799
            Spirits          777
            Stars            756
            Sounds           697
            Baltimore        467
            Floridians       440
            Condors          430
            Capitols         291
            Olympians        282
            Sails            274
            Stags            260
            Bombers          249
            Steamrollers     168
            Packers           72
            Redskins          65
            Rebels            63
            Waterloo          62
            Denver            62
            Falcons           60
            Jets              60
            Ironmen           60
            Huskies           60
            Name: fran_id, dtype: int64
```

```
nba.loc[nba["fran_id"]=="Lakers","team_id"].value_counts()
```

```
    LAL    5078
    MNL     946
    Name: team_id, dtype: int64
```

```
nba.loc[nba["team_id"]=="MNL","date_game"]
```

```
    1136      11/4/1948
    1152      11/6/1948
    1159      11/9/1948
    1170     11/11/1948
    1183     11/13/1948
               ...
    9701      3/19/1960
    9702      3/20/1960
    9709      3/22/1960
    9711      3/24/1960
    9715      3/26/1960
    Name: date_game, Length: 946, dtype: object
```

```
nba.loc[nba["team_id"]=="LAL","date_game"]
```

```
    9730       10/19/1960
    9739       10/22/1960
    9744       10/24/1960
    9747       10/25/1960
    9756       10/28/1960
                 ...
    126016       4/8/2015
    126052      4/10/2015
    126086      4/12/2015
    126115      4/13/2015
    126137      4/15/2015
    Name: date_game, Length: 5078, dtype: object
```

```
df=nba.copy()
df.shape
```

```
    (126314, 23)
```

```
df["lead"]=df.pts-df.opp_pts
df.shape
df
```

	gameorder	game_id	lg_id	_iscopy	year_id	date_game	seasongame	is_playoffs	team_id	fran_id	pts	elo_i	el
0	1	194611010TRH	NBA	0	1947	11/1/1946	1	0	TRH	Huskies	66	1300.00	129
1	1	194611010TRH	NBA	1	1947	11/1/1946	1	0	NYK	Knicks	68	1300.00	130
2	2	194611020CHS	NBA	0	1947	11/2/1946	1	0	CHS	Stags	63	1300.00	130
3	2	194611020CHS	NBA	1	1947	11/2/1946	2	0	NYK	Knicks	47	1306.72	129
4	3	194611020DTF	NBA	0	1947	11/2/1946	1	0	DTF	Falcons	33	1300.00	127
...	...	...	...	...	...	...	...	...	...	...	...	...	...
126309	63155	201506110CLE	NBA	0	2015	6/11/2015	100	1	CLE	Cavaliers	82	1723.41	170
126310	63156	201506140GSW	NBA	0	2015	6/14/2015	102	1	GSW	Warriors	104	1809.98	181
126311	63156	201506140GSW	NBA	1	2015	6/14/2015	101	1	CLE	Cavaliers	91	1704.39	170
126312	63157	201506170CLE	NBA	0	2015	6/16/2015	102	1	CLE	Cavaliers	97	1700.74	169
126313	63157	201506170CLE	NBA	1	2015	6/16/2015	103	1	GSW	Warriors	105	1813.63	182

126314 rows × 24 columns

```python
df["lead"].max()
```

    68

```python
df.rename(
    columns={"game_result":"result","game_location":"location"},
    inplace=True)
df.info()
```

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 126314 entries, 0 to 126313
    Data columns (total 20 columns):
     #   Column   Non-Null Count   Dtype
    --- ------   --------------   -----
     0   gameorder   126314 non-null int64
     1   game_id   126314 non-null object
     2   lg_id   126314 non-null object
     3   _iscopy   126314 non-null int64
     4   year_id   126314 non-null int64
     5   date_game   126314 non-null datetime64[ns]
     6   seasongame   126314 non-null int64
     7   is_playoffs 126314 non-null int64
     8   team_id   126314 non-null object
     9   fran_id   126314 non-null object
     10 pts   126314 non-null int64
     11 win_equiv   126314 non-null float64
     12 opp_id   126314 non-null object
     13 opp_fran   126314 non-null object
     14 opp_pts   126314 non-null int64
     15 location   126314 non-null object
     16 result   126314 non-null object
     17 forecast   126314 non-null float64
     18 notes   5424 non-null   object
     19 lead   126314 non-null int64
    dtypes: datetime64[ns](1), float64(2), int64(8), object(9)
    memory usage: 19.3+ MB

```python
df["date_game"]=pd.to_datetime(df["date_game"])
df.info()
```

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 126314 entries, 0 to 126313
    Data columns (total 20 columns):
     #   Column   Non-Null Count   Dtype
    --- ------   --------------   -----
     0   gameorder   126314 non-null int64
     1   game_id   126314 non-null object
     2   lg_id   126314 non-null object
     3   _iscopy   126314 non-null int64
     4   year_id   126314 non-null int64
     5   date_game   126314 non-null datetime64[ns]
     6   seasongame   126314 non-null int64
     7   is_playoffs 126314 non-null int64

```
 8   team_id     126314 non-null  object
 9   fran_id     126314 non-null  object
10   pts         126314 non-null  int64
11   win_equiv   126314 non-null  float64
12   opp_id      126314 non-null  object
13   opp_fran    126314 non-null  object
14   opp_pts     126314 non-null  int64
15   location    126314 non-null  object
16   result      126314 non-null  object
17   forecast    126314 non-null  float64
18   notes       5424 non-null    object
19   lead        126314 non-null  int64
dtypes: datetime64[ns](1), float64(2), int64(8), object(9)
memory usage: 19.3+ MB
```

```
df["location"].unique()
```

```
array(['H', 'A', 'N'], dtype=object)
```

```
pd.Categorical(df["location"])
```

```
['H', 'A', 'H', 'A', 'H', ..., 'H', 'H', 'A', 'H', 'A']
Length: 126314
Categories (3, object): ['A', 'H', 'N']
```

```
df["location"]=pd.Categorical(df["location"])
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 126314 entries, 0 to 126313
Data columns (total 20 columns):
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   gameorder   126314 non-null  int64
 1   game_id     126314 non-null  object
 2   lg_id       126314 non-null  object
 3   _iscopy     126314 non-null  int64
 4   year_id     126314 non-null  int64
 5   date_game   126314 non-null  datetime64[ns]
 6   seasongame  126314 non-null  int64
 7   is_playoffs 126314 non-null  int64
 8   team_id     126314 non-null  object
 9   fran_id     126314 non-null  object
10   pts         126314 non-null  int64
11   win_equiv   126314 non-null  float64
12   opp_id      126314 non-null  object
```