Data Analysis with Pandas

- **Indexing**
  - **The loc attribute allows indexing and slicing that always references the explicit index**
    - **Label Indexing**
  - **The iloc attribute allows indexing and slicing that always references the implicit index**
    - **Integer Indexing**
- **Grouping**
- **Features**
- **Filtering**
- **Sorting**
- **Statistical**
- **Plotting**
- **Saving**

In [35]:

```
import pandas as pd
birds = pd.read_csv("https://raw.githubusercontent.com/APSSDC-Data-Analysis/DataAnalysis-7/main/Datasets/birds.csv")
birds.head()
```

Out[35]:

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

In [3]:

```
ind = pd.Index([34,56,768,89])
type(ind)
```

Out[3]:

```
pandas.core.indexes.numeric.Int64Index
```

In [4]:

```
ind.tolist()
```

Out[4]:

```
[34, 56, 768, 89]
```

In [5]:

```
birds[["huml","type"]]
```

Out[5]:

	huml	type
0	80.78	SW
1	88.91	SW
2	79.97	SW
3	77.65	SW

	huml	type
4	62.80	SW
...	...	...
415	17.96	SO
416	19.21	SO
417	18.79	SO
418	20.38	SO
419	17.89	SO

420 rows × 2 columns

In [6]:

```
birds["humw"][0]
```

Out[6]:

```
6.68
```

In [8]:

```
birds[0:1]["humw"]
```

Out[8]:

```
0    6.68
Name: humw, dtype: float64
```

Special index attributes

Loc -- used Label index / explicit index / location and label

Iloc -- used integer index / implict indexing / location indexing

In [41]:

```
birds[0]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, key, method, t
olerance)
   2645             try:
-> 2646                 return self._engine.get_loc(key)
   2647             except KeyError:

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_i
tem()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_i
tem()

KeyError: 0

During handling of the above exception, another exception occurred:

KeyError                                  Traceback (most recent call last)
<ipython-input-41-305f12d01a1f> in <module>
----> 1 birds[0]

~\anaconda3\lib\site-packages\pandas\core\frame.py in __getitem__(self, key)
   2798             if self.columns.nlevels > 1:
```

```
   2799                     return self._getitem_multilevel(key)
-> 2800                 indexer = self.columns.get_loc(key)
   2801                 if is_integer(indexer):
   2802                     indexer = [indexer]

~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, key, method, t
olerance)
   2646                     return self._engine.get_loc(key)
   2647                 except KeyError:
-> 2648                     return self._engine.get_loc(self._maybe_cast_indexer(key))
   2649             indexer = self.get_indexer([key], method=method, tolerance=tolerance)
   2650             if indexer.ndim > 1 or indexer.size > 1:

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_i
tem()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_i
tem()

KeyError: 0
```

In [ ]:

```

```

In [42]:

```
birds.iloc[0]
```

Out[42]:

```
huml     80.78
humw      6.68
ulnal    72.01
ulnaw     4.88
feml     41.81
femw       3.7
tibl       5.5
tibw      4.03
tarl      38.7
tarw      3.84
type        SW
Name: 0, dtype: object
```

In [13]:

```
birds.iloc[10:100:20]
```

Out[13]:

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
10	10	165.00	11.45	156.00	8.68	80.25	7.85	143.00	8.25	86.61	6.63	SW
30	30	75.05	12.55	44.05	8.62	75.31	6.76	112.60	7.22	34.31	14.09	SW
50	50	171.00	8.39	187.00	6.68	58.67	6.70	87.57	6.34	50.44	8.41	SW
70	70	26.63	1.63	27.19	1.72	12.21	1.02	31.56	1.07	21.62	0.83	SW
90	90	75.94	5.25	80.66	4.26	30.94	2.85	68.23	2.97	37.38	2.47	SW

In [19]:

```
birds.loc[0,"type"]
```

Out[19]:

```
'SW'
```

In [16]:

```
birds.head()
```

Out[16]:

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

In [20]:

```
birds.loc[10:20,"type"]
```

Out[20]:

```
10    SW
11    SW
12    SW
13    SW
14    SW
15    SW
16    SW
17    SW
18    SW
19    SW
20    SW
Name: type, dtype: object
```

In [21]:

```
birds.loc[1:20,["type","huml"]]
```

Out[21]:

	type	huml
1	SW	88.91
2	SW	79.97
3	SW	77.65
4	SW	62.80
5	SW	61.92
6	SW	79.73
7	SW	86.98
8	SW	118.20
9	SW	145.00
10	SW	165.00
11	SW	186.00
12	SW	172.00
13	SW	148.91
14	SW	149.19
15	SW	140.59
16	SW	135.23
17	SW	124.53
18	SW	127.03

	type	huml
19	SW	106.02
20	SW	113.84

In [45]:

```
birds.loc[[1,100,50,200],["type","huml"]] # fancy slicing using loc
```

Out[45]:

	type	huml
id		
1	SW	88.91
100	SW	67.06
50	SW	171.00
200	T	34.24

In [25]:

```
birds.at[200,"type"]  # at (where -- location -- accessing single value)
```

Out[25]:

```
'T'
```

In [28]:

```
birds.head()
```

Out[28]:

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

In [27]:

```
birds.shape
```

Out[27]:

```
(420, 12)
```

In [36]:

```
birds.set_index('id', inplace = True) # if you want set any label as index
# inplace = True --- ir changes original dataset
```

In [38]:

```
birds
```

Out[38]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
id											
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW
...	...	...	...	...	...	...	...	...	...	...	...
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO

420 rows × 11 columns

In [40]:

```
birds.reset_index().head()
```

Out[40]:

	id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

5. Grouping

1 file IPL

having 1000+ records

in between years 2010 - 2020

CSK Team

MI

RCB

KKR

SRH

RR

wins outs runs overs

In [46]:

```
birds.tail()
```

Out[46]:

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
417	18.79	1.63	19.83	1.33	20.96	1.43	34.45	1.41	22.86	1.21	SO
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO
id											
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO

In [48]:

```
birds.columns
```

Out[48]:

```
Index(['huml', 'humw', 'ulnal', 'ulnaw', 'feml', 'femw', 'tibl', 'tibw',
       'tarl', 'tarw', 'type'],
      dtype='object')
```

In [52]:

```
birds["type"].uncounte()
```

Out[52]:

```
array(['SW', 'W', 'T', 'R', 'P', 'SO'], dtype=object)
```

In [53]:

```
birds["type"].value_counts()
```

Out[53]:

```
SO    128
SW    116
W      65
R      50
P      38
T      23
Name: type, dtype: int64
```

In [54]:

```
bird_grp = birds.groupby("type")    # ["type","huml"]
bird_grp
```

Out[54]:

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001A13623CAC0>
```

In [55]:

```
for name, grp_data in bird_grp:
    print(name, grp_data)
```

```
P        huml   humw   ulnal   ulnaw    feml   femw    tibl   tibw    tarl   tarw  type
id
254   29.26   2.66   27.63    2.23   27.87   2.03   38.90   1.81   25.22   1.62     P
255   30.60   2.45   25.90    1.87   32.36   2.10   50.25   1.88   33.58   1.88     P
256   30.25   2.29   25.14    1.87   31.61   2.18   50.98   2.00   33.50   1.93     P
257   42.52   2.99   36.73    2.91   52.24   3.81   81.57   3.71   61.27   2.71     P
258   45.60   3.35   38.73    3.52   54.45   4.10   83.59   3.72   63.91   2.67     P
259   43.97   3.54   38.38    3.14   52.49   4.25   83.15   4.12   61.08   3.04     P
260   46.38   3.41   39.80    3.48   54.67   4.08   85.88   4.04   63.75   3.09     P
261   45.57   4.64   57.13    2.93   28.32   2.26   36.95   1.96   19.75   2.16     P
262   33.81   2.80   42.82    2.22   18.88   1.45   26.28   1.31   14.45   1.29     P
263   31.72   2.64   40.00    1.99   20.36   1.59   32.21   1.62   17.72   1.52     P
264    9.85   2.88   14.73    1.73   16.27   1.22   20.89   1.25    7.77   1.23     P
265   44.47   4.24   55.77    3.43   38.57   3.06   52.80   2.82   18.57   3.60     P
266   18.46   1.79   21.12    1.47   18.83   1.21   26.33   1.05   11.41   1.56     P
267   18.55   2.12   21.08    1.41   18.19   1.30   25.97   1.17   11.26   1.57     P
268   27.57   2.67   31.60    2.12   18.92   1.52   30.42   1.58   12.46   1.46     P
269   27.07   2.66   31.95    1.87   18.80   1.51   26.38   1.33   12.52   1.16     P
270   26.92   2.73   31.54    2.01   19.10   1.59   26.68   1.35   13.21   1.23     P
271   49.12   3.59   60.95    2.99   26.35   2.00   37.29   1.81   10.39   2.03     P
272   48.80   3.78   59.47    3.17   26.14   2.24   36.01   1.63   11.65   2.00     P
273   47.17   3.96   56.57    2.87   33.60   2.68   50.66   2.49   32.03   2.06     P
274   37.86   3.24   46.72    2.63   27.52   2.26   42.49   1.89   26.88   1.54     P
```

```
274  37.86  3.24  46.72   2.65  27.52  2.26  42.49  1.89  26.88  1.54    P
275  39.98  3.34  47.50   2.64  28.57  2.53  42.96  2.20  27.66  1.67    P
276  44.18  3.88  50.72   3.49  39.58  3.90  50.89  3.61  35.14  2.82    P
277  41.36  3.52  49.34   2.83  30.97  2.46  44.31  2.43  29.97  2.05    P
278  41.66  3.56  48.82   2.80  31.94  2.52  46.19  2.31  30.73  1.95    P
279  31.28  2.83  36.88   2.20  22.30  2.10  32.60  1.68  20.47  1.54    P
280  34.07  3.26  41.34   2.62  24.25  2.30  36.35  2.13  24.69  2.12    P
281  33.79  3.04  42.13   2.56  23.09  2.21  34.74  2.07  22.12  1.78    P
282  34.15  2.98  40.67   2.45  24.72  1.98  37.35  1.93  23.49  1.67    P
283  36.88  3.18  46.28   2.77  26.08  2.19  39.44  2.03  25.60  1.98    P
284  28.55  2.65  33.74   2.38  19.56  1.98  29.21  1.64  18.75  1.53    P
285  30.51  2.69  35.33   2.29  21.90  2.04  32.26  1.65  22.52  1.43    P
286  29.81  2.59  35.05   2.28  21.48  2.11  31.89  1.84  21.67  1.52    P
287  27.72  2.67  32.00   2.21  22.68  2.13  32.50  1.76  19.65  1.64    P
288  29.14  2.75  32.91   2.16  23.13  2.11  33.89  1.80  20.86  1.70    P
289  27.85  2.65  32.62   2.05  22.84  2.14  33.62  1.91  20.82  1.80    P
290  20.06  1.73  23.15   1.49  16.63  1.57  24.71  1.29  15.19  1.24    P
291  41.60  3.74  56.50   3.02  37.00  2.99  62.90  2.82  38.21  2.51    P
R        huml   humw   ulnal  ulnaw   feml  femw    tibl  tibw   tarl  tarw type
id
204   63.76   4.74     NaN    NaN  57.33  4.88   75.67  4.33  60.19  3.82    R
205   57.48   4.74   69.06   4.08  57.56  4.80   80.16  4.25  60.54  3.78    R
206   62.49   4.75   69.66   3.99  57.30  4.60   77.60  4.26  60.31  3.86    R
207   98.08   7.77  113.04   5.76  82.04  7.17  107.47  6.65    NaN   NaN    R
208   54.47   4.37   63.29   3.32  47.83  3.45   64.25  3.15  54.97  2.86    R
209   57.56   4.33   67.15   3.54  55.19  4.21   75.03  3.71  59.33  3.29    R
210   52.88   4.47   62.02   3.76  47.09  3.64   63.42  3.40  54.54  2.90    R
211  110.87   8.26  129.05   6.78  80.80  7.98  112.36  7.18  84.69  7.11    R
212  117.37   9.25  136.52   7.00  87.49  8.92  119.17  7.89  86.61  7.36    R
213  102.35   7.01  118.67   5.72  78.75  6.89  104.01  6.12  80.10  6.10    R
214  108.37   7.49  129.39   6.00  82.29  6.74  113.39  6.38  86.05  6.15    R
215   81.39   5.77   94.83   4.80  65.97  5.52   91.63  5.38  73.29  5.11    R
216   83.13   5.80   95.42   4.40  66.98  5.14   90.60  4.95  75.64  4.56    R
217   89.63   6.44  104.11   5.26  72.24  6.32   97.78  5.81  76.76  5.29    R
218   89.05   6.28  102.09   5.37  70.32  5.90   96.12  5.70  75.05  5.12    R
219   98.94   7.13  119.72   5.60  74.17  6.30  106.50  5.80  83.13  5.86    R
220  121.59   8.17  148.52   6.44  89.67  7.63  116.71  6.87  83.28  6.80    R
221  130.06   9.51  155.00   7.18  94.79  8.64  125.53  7.87  53.83  7.89    R
222  116.41   6.85  131.00   5.80  75.36  6.55  119.87  6.10  99.72  6.06    R
223  107.06   8.62  129.08   6.48  71.03  6.87  100.06  6.10  68.64  5.87    R
224  120.00   8.38  130.00   6.41  89.93  8.46  116.17  7.52  84.18  6.97    R
225   85.45   6.16   99.32   4.81  53.01  4.27   69.15  4.21  37.61  5.47    R
226   83.02   5.95   96.97   4.38  51.87  4.21   66.60  4.06  36.60  5.63    R
227   71.97   6.18   84.68   5.14  63.79  5.43   82.93  4.63  52.71  4.72    R
228   82.58   7.32   97.52   6.19  71.37  6.30   91.66  5.65  56.93  5.99    R
229   45.37   3.72   51.37   3.16  39.02  2.99   53.23  2.68  38.07  2.29    R
230   41.65   3.62   46.01   3.03  38.37  2.85   48.11  2.55  35.49  2.22    R
231   42.46   3.49   47.34   3.27  36.03  2.94   49.17  2.68  34.38  2.25    R
232  145.00  10.37  188.00   7.92  77.09  7.75  125.48  7.90  51.99  9.64    R
233  142.00   9.60   82.00   7.31  77.96  7.03  124.98  7.05  51.82  8.26    R
234  145.00  10.31  188.00   7.85  78.86  7.72  126.54  7.13  53.64  8.88    R
235   57.41   3.82   70.39   3.08  39.70  3.37   66.58  3.33  46.81  2.78    R
236   50.78   3.15   62.89   2.69  36.32  2.84   62.53  2.75  44.08  2.61    R
237  119.12   7.57  137.37   6.06  75.53  6.35  115.02  6.22  60.99  7.91    R
238  117.43   7.13  136.26   5.51  75.54  5.85  115.30  5.45  59.45  8.25    R
239  116.85   7.43  134.80   6.19  73.51  6.10  112.66  5.94  61.25  6.90    R
240  119.58   8.04  138.39   6.25  76.36  6.50  115.90  6.07  59.90  6.92    R
241  126.88   8.01  148.47   6.59  81.42  6.71  124.05  6.58  64.57  7.97    R
242   47.16   2.82   55.41   2.37  34.58  2.47   53.22  2.29  33.77  2.51    R
243   44.91   2.54   48.45   2.14  34.87  2.48   50.87  2.27  29.80  2.70    R
244   61.60   3.42   66.26   2.59  37.04  2.84   63.03  2.87  44.51  2.43    R
245   94.03   5.88  104.09   4.32  56.76  4.98   96.60  4.87  68.65  4.40    R
246   90.13   5.92  104.28   4.40  54.55  4.65   94.99  4.42  67.55  4.14    R
247  100.29   6.33  112.19   4.69  59.71  4.94  102.54  5.33  71.03  4.95    R
248   96.46   5.71  111.85   4.20  59.82  4.65   99.16  4.56  73.95  4.29    R
249   93.37   5.33  102.65   3.90  58.55  4.62  101.24  4.57  71.30  4.38    R
250   99.69   6.09  111.00   4.26  60.84  4.75  104.72  4.92  71.23  4.21    R
251   34.84   2.44   40.68   2.04  38.43  2.54   42.48  2.45  20.07  3.15    R
252   33.12   2.40   39.85   1.80  27.46  2.19   38.99  2.15  19.10  2.89    R
253   35.63   2.40   42.67   2.02  28.51  2.41   42.64  2.46  21.05  3.12    R
SO       huml  humw  ulnal  ulnaw   feml  femw   tibl  tibw   tarl  tarw type
id
292  29.61  2.41  33.93   2.02  30.79  2.19  51.64  2.01  37.33  1.78    SO
```

```
292  29.61  2.41  33.93   2.02  30.79  2.19  51.64  2.01  37.33  1.78   SO
293  30.28  2.89  35.37   2.15  31.09  2.51  53.36  2.31  39.52  1.69   SO
294  28.10  2.44  32.50   2.19  29.68  2.33  50.47  2.21  37.98  1.79   SO
295  28.66  2.48  33.24   2.02  29.33  2.26  50.64  2.05  35.99  1.85   SO
296  36.36  3.31  40.51   2.56  37.52  2.85  53.96  2.63  42.71  2.41   SO
..     ...   ...    ...    ...    ...   ...    ...   ...    ...   ...  ...
415  17.96  1.63  19.25   1.33  18.36  1.54  31.25  1.33  21.99  1.15   SO
416  19.21  1.64  20.76   1.49  19.24  1.45  33.21  1.28  23.60  1.15   SO
417  18.79  1.63  19.83   1.53  20.96  1.43  34.45  1.41  22.86  1.21   SO
418  20.38  1.78  22.53   1.50  21.35  1.48  36.09  1.53  25.98  1.24   SO
419  17.89  1.44  19.26   1.10  17.62  1.34  29.81  1.24  21.69  1.05   SO

[128 rows x 11 columns]
SW        huml  humw   ulnal  ulnaw   feml  femw    tibl  tibw   tarl  tarw type
id
0      80.78  6.68   72.01   4.88  41.81  3.70    5.50  4.03  38.70  3.84   SW
1      88.91  6.63   80.53   5.59  47.04  4.30   80.22  4.51  41.50  4.01   SW
2      79.97  6.37   69.26   5.28  43.07  3.90   75.35  4.04  38.31  3.34   SW
3      77.65  5.70   65.76   4.77  40.04  3.52   69.17  3.40  35.78  3.41   SW
4      62.80  4.84   52.09   3.73  33.95  2.72   56.27  2.96  31.88  3.13   SW
..       ...   ...     ...    ...    ...   ...     ...   ...    ...   ...  ...
111  106.23  6.82  104.02   5.07  40.06  4.96  105.34  4.99  57.27  3.20   SW
112  105.19  7.10   96.85   5.43  37.61  4.43  102.04  4.90  54.78  2.83   SW
113   76.49  5.36   69.06   4.02  30.07  3.03   72.41  3.71  43.11  2.08   SW
114   65.66  3.79   61.06   3.35  24.95  2.44   63.76  2.72  40.19  1.89   SW
115   98.73  6.48   91.69   5.31  35.34  4.02   92.36  4.46  50.53  2.69   SW

[116 rows x 11 columns]
T        huml   humw   ulnal  ulnaw    feml   femw    tibl   tibw    tarl  tarw  \
id
181   43.97   4.64   51.07   3.80   38.76   3.18   55.67   2.88   30.41  2.83
182   44.58   5.76   52.76   4.84   40.47   3.76   56.42   3.31   31.32  3.34
183   20.25   2.35   25.14   1.76   20.17   1.37   27.67   1.41   15.68  1.55
184   48.70   5.46   57.39   3.73   42.56   3.07   52.42   2.85   27.23  3.03
185   35.74   3.91   39.20   2.93   33.01   2.65   42.80   2.18   24.85  2.34
186   37.69   4.18   43.30   3.38   35.00   2.87   49.15   2.37   26.33  2.51
187   31.26   3.60   35.24   2.91   28.62   2.42   39.04   2.32   21.88  2.27
188   30.97   3.65   36.47   2.96   28.03   2.21   37.41   1.99   21.11  2.25
189   30.31   3.36   34.83   2.49   26.36   2.17   34.16   1.88   18.72  2.03
190   32.02   2.99   27.70   2.38   38.27   2.64   53.33   2.71   31.00  2.45
191   32.34   3.17   29.05   2.36   39.64   2.70   55.36   2.51   37.08  2.22
192   30.88   3.32   26.93   2.41   38.19   2.82   53.59   2.62   30.02  2.39
193   32.11   3.47   28.48   2.62   39.55   3.31   56.16   3.00   31.74  2.49
194   32.61   2.98   28.37   2.49   37.47   2.61   48.92   2.39   28.46  2.34
195   34.63   3.10   26.87   2.35   40.87   3.05   53.98   2.79   31.59  2.63
196   32.28   3.34   28.79   2.45   38.46   2.82   52.93   2.50   30.97  2.48
197   92.82   9.60  100.58   7.57   82.94   7.47  118.24   6.61   69.16  5.66
198   55.62   4.79   52.09   4.26   64.80   5.74   86.17   4.76   49.37  4.25
199   32.68   3.26   29.02   2.36   39.38   2.72   56.45   2.50   32.56  2.49
200   34.24   3.39   30.65   2.63   40.41   2.92   53.77   2.72   32.24  2.67
201  118.75  11.90  104.68   4.77  108.85  10.29  183.00   9.69  134.00  7.96
202  127.00  13.79  123.27   9.42  117.07  11.64  189.00  10.03  128.50  8.19
203   39.64   4.07   37.32   2.98   45.50   3.42   64.45   3.37   37.77  2.86

    type
id
181    T
182    T
183    T
184    T
185    T
186    T
187    T
188    T
189    T
190    T
191    T
192    T
193    T
194    T
195    T
196    T
197    T
```

```
197     T
198     T
199     T
200     T
201     T
202     T
203     T
W        huml   humw   ulnal  ulnaw    feml  femw    tibl  tibw    tarl  tarw  \
id
116  157.00   9.00  192.00   7.02   85.60  6.44  227.00  6.84  175.00  6.77
117  190.00  11.92  225.00   8.55  101.75  7.75  240.00  7.71  175.00  7.00
118   70.01   3.71   76.90   2.88   49.42  3.08   79.56  3.35   52.74  3.02
119   92.93   5.28  110.94   3.87   55.11  4.16  137.00  4.29   96.37  3.79
120  118.52   6.87  134.91   5.06   78.20  5.24  132.00  5.38   85.28  4.74
..      ...    ...     ...    ...     ...   ...     ...   ...     ...   ...
176   33.20   2.11   34.31   2.25   20.36  1.57   37.95  1.42   23.67  1.26
177   32.31   2.16   32.74   1.98   19.05  1.40   33.37  1.14   21.20  1.00
178   28.51   1.80   29.35   1.78   17.60  1.21   36.04  1.14   20.81  0.83
179   34.12   1.90   35.57   1.93   21.22  1.39   46.24  1.27   31.39  1.05
180   87.01   5.18   94.34   4.45   35.73  2.80   67.12  2.93   43.43  2.34

    type
id
116    W
117    W
118    W
119    W
120    W
..   ...
176    W
177    W
178    W
179    W
180    W

[65 rows x 11 columns]
```

In [57]:

```
birds.head(1)
```

Out[57]:

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
0	80.78	6.68	72.01	4.88	41.81	3.7	5.5	4.03	38.7	3.84	SW

In [56]:

```
bird_grp.first()   # FIRST record of very group
```

Out[56]:

type	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw
P	29.26	2.66	27.63	2.23	27.87	2.03	38.90	1.81	25.22	1.62
R	63.76	4.74	69.06	4.08	57.33	4.88	75.67	4.33	60.19	3.82
SO	29.61	2.41	33.93	2.02	30.79	2.19	51.64	2.01	37.33	1.78
SW	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84
T	43.97	4.64	51.07	3.80	38.76	3.18	55.67	2.88	30.41	2.83
W	157.00	9.00	192.00	7.02	85.60	6.44	227.00	6.84	175.00	6.77

In [58]:

```
bird_grp.last()
```

Out[58]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw
type										
P	41.60	3.74	56.50	3.02	37.00	2.99	62.90	2.82	38.21	2.51
R	35.63	2.40	42.67	2.02	28.51	2.41	42.64	2.46	21.05	3.12
SO	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05
SW	98.73	6.48	91.69	5.31	35.34	4.02	92.36	4.46	50.53	2.69
T	39.64	4.07	37.32	2.98	45.50	3.42	64.45	3.37	37.77	2.86
W	87.01	5.18	94.34	4.45	35.73	2.80	67.12	2.93	43.43	2.34

In [59]:

```
bird_grp.get_group("W")   # accessing W bird data/ records
```

Out[59]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw
id										
116	157.00	9.00	192.00	7.02	85.60	6.44	227.00	6.84	175.00	6.77
117	190.00	11.92	225.00	8.55	101.75	7.75	240.00	7.71	175.00	7.00
118	70.01	3.71	76.90	2.88	49.42	3.08	79.56	3.35	52.74	3.02
119	92.93	5.28	110.94	3.87	55.11	4.16	137.00	4.29	96.37	3.79
120	118.52	6.87	134.91	5.06	78.20	5.24	132.00	5.38	85.28	4.74
...	...	...	...	...	...	...	...	...	...	...
176	33.20	2.11	34.31	2.25	20.36	1.57	37.95	1.42	23.67	1.26
177	32.31	2.16	32.74	1.98	19.05	1.40	33.37	1.14	21.20	1.00
178	28.51	1.80	29.35	1.78	17.60	1.21	36.04	1.14	20.81	0.83
179	34.12	1.90	35.57	1.93	21.22	1.39	46.24	1.27	31.39	1.05
180	87.01	5.18	94.34	4.45	35.73	2.80	67.12	2.93	43.43	2.34

65 rows × 10 columns

In [60]:

```
bird_grp.get_group("W").mean()
```

Out[60]:

```
huml      73.133077
humw       4.607077
ulnal     78.101385
ulnaw      4.102615
feml      40.119531
femw       3.118594
tibl      76.150000
tibw       3.179231
tarl      47.543385
tarw       2.760308
dtype: float64
```

In [61]:

```
bird_grp.get_group("W").sum()
```

Out[61]:

```
huml      4753.65
humw       299.46
```

```
ulnal     5076.59
ulnaw      266.67
feml      2567.65
femw       199.59
tibl      4949.75
tibw       206.65
tarl      3090.32
tarw       179.42
dtype: float64
```

6. Features

labels/ columns / features

1 file Name roll number internal marks External mark dfb 456 56 NaN

In [62]:

```
birds.head()
```

Out[62]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type
id											
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW

In [63]:

```
# add one more feature to already existing data
birds["len_type"]   =  birds["type"].str.len()
birds
```

Out[63]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type
id												
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2
...	...	...	...	...	...	...	...	...	...	...	...	...
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2

420 rows × 12 columns

In [70]:

```
birds[birds.columns[:-2]].mean(axis = 1)
```

```
birds[birds.columns[:-2]].mean(axis = 1)
```

Out[70]:

```
id
0      26.193
1      36.324
2      32.889
3      30.920
4      25.437
        ...
415    11.579
416    12.303
417    12.410
418    13.386
419    11.244
Length: 420, dtype: float64
```

In [71]:

```
birds["Mean_Features"] = birds[birds.columns[:-2]].mean(axis = 1)
```

In [72]:

```
birds
```

Out[72]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features
id													
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437
...	...	...	...	...	...	...	...	...	...	...	...	...	...
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2	11.579
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2	12.303
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2	12.410
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2	13.386
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2	11.244

420 rows × 13 columns

In [73]:

```
# add "huml humw"
birds["huml_and_humw"] = birds["huml"]+birds["humw"]
birds
```

Out[73]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2	11.579	19.59
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2	12.303	20.85
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2	12.410	20.42
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2	13.386	22.16
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2	11.244	19.33

420 rows × 14 columns

In [75]:

```
birds[["huml","humw"]].sum(axis = 1)
```

Out[75]:

```
id
0      87.46
1      95.54
2      86.34
3      83.35
4      67.64
       ...  
415    19.59
416    20.85
417    20.42
418    22.16
419    19.33
Length: 420, dtype: float64
```

7. Filtering

In [ ]:

```
# cleaning the data
# Extracting wanted data
```

In [76]:

```
birds.head()
```

Out[76]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64

In [77]:

```
birds.count()
```

Out[77]:

```
huml             419
humw             419
ulnal            417
ulnaw            418
feml             418
femw             419
tibl             418
tibw             419
```

```
tarl             419
tarw             419
type             420
len_type         420
Mean_Features    420
huml_and_humw    419
dtype: int64
```

In [78]:

```
birds["humw"].max()
```

Out[78]:

```
17.84
```

In [79]:

```
birds["humw"].min()
```

Out[79]:

```
1.14
```

In [ ]:

```
humw -- value -- 10 - 15
```

In [80]:

```
birds["humw"]>10   # bool array
```

Out[80]:

```
id
0      False
1      False
2      False
3      False
4      False
       ...
415    False
416    False
417    False
418    False
419    False
Name: humw, Length: 420, dtype: bool
```

In [81]:

```
birds[birds["humw"]>10]
```

Out[81]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_hum
id														
9	145.00	10.42	144.00	7.05	70.96	7.44	120.00	7.31	78.67	6.34	SW	2	59.719	155.
10	165.00	11.45	156.00	8.68	80.25	7.85	143.00	8.25	86.61	6.63	SW	2	67.372	176.
29	72.50	14.78	45.83	10.33	77.02	7.76	128.75	6.49	32.48	13.82	SW	2	40.976	87.
30	75.05	12.55	44.05	8.62	75.31	6.76	112.60	7.22	34.31	14.09	SW	2	39.056	87.
31	175.00	10.69	228.00	8.65	47.13	6.16	64.53	4.41	18.42	3.65	SW	2	56.664	185.
32	188.00	11.35	250.00	9.64	52.49	7.60	69.06	5.13	19.52	6.51	SW	2	61.930	199.
33	210.00	13.03	278.00	10.74	56.87	8.03	76.66	5.20	22.54	7.16	SW	2	68.823	223.
34	185.00	10.91	235.00	8.02	51.00	6.71	68.67	4.24	20.14	6.30	SW	2	59.599	195.
35	179.00	11.16	222.00	8.57	47.51	6.39	65.18	4.51	19.67	6.64	SW	2	57.063	190.
36	183.00	11.37	280.00	9.22	52.34	7.24	68.09	4.61	20.08	6.41	SW	2	64.236	194.

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_hum
37	185.00	12.47	242.00	9.80	52.39	7.50	68.01	4.66	19.44	7.21	SW	2	60.848	197.
38	272.00	14.86	320.00	10.42	91.60	9.71	132.00	10.23	81.77	8.91	SW	2	95.150	286.
39	270.00	14.25	310.00	10.90	86.20	9.96	125.76	9.63	79.18	10.05	SW	2	92.593	284.
56	310.00	14.40	315.00	9.51	88.77	8.10	180.00	9.45	96.13	7.69	SW	2	103.905	324.
57	250.00	11.91	252.00	8.31	73.04	7.37	160.00	8.47	82.46	7.04	SW	2	86.060	261.
58	420.00	17.84	422.00	11.72	110.54	9.99	237.00	11.03	128.35	8.93	SW	2	137.740	437.
59	250.00	11.28	247.50	7.50	69.04	6.20	156.00	7.19	83.36	6.13	SW	2	84.420	261.
60	300.00	12.48	300.00	8.65	84.05	8.53	178.00	9.61	99.01	7.55	SW	2	100.788	312.
117	190.00	11.92	225.00	8.55	101.75	7.75	240.00	7.71	175.00	7.00	W	1	97.468	201.
201	118.75	11.90	104.68	4.77	108.85	10.29	183.00	9.69	134.00	7.96	T	1	69.389	130.
202	127.00	13.79	123.27	9.42	117.07	11.64	189.00	10.03	128.50	8.19	T	1	73.791	140.
232	145.00	10.37	188.00	7.92	77.09	7.75	125.48	7.90	51.99	9.64	R	1	63.114	155.
234	145.00	10.31	188.00	7.85	78.86	7.72	126.54	7.13	53.64	8.88	R	1	63.393	155.

In [84]:

```
birds[(birds["humw"]>10)  |  (birds["humw"]<15)]
# Logic gates   -- bitwise and operator
# | or
# & and
```

Out[84]:

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2	11.579	19.59
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2	12.303	20.85
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2	12.410	20.42
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2	13.386	22.16
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2	11.244	19.33

419 rows × 14 columns

In [85]:

```
birds.head()
```

Out[85]:

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64

id

In [86]:

```
# P bird data
birds[birds["type"]  == "P"]
```

Out[86]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
254	29.26	2.66	27.63	2.23	27.87	2.03	38.90	1.81	25.22	1.62	P	1	15.923	31.92
255	30.60	2.45	25.90	1.87	32.36	2.10	50.25	1.88	33.58	1.88	P	1	18.287	33.05
256	30.25	2.29	25.14	1.87	31.61	2.18	50.98	2.00	33.50	1.93	P	1	18.175	32.54
257	42.52	2.99	36.73	2.91	52.24	3.81	81.57	3.71	61.27	2.71	P	1	29.046	45.51
258	45.60	3.35	38.73	3.52	54.45	4.10	83.59	3.72	63.91	2.67	P	1	30.364	48.95
259	43.97	3.54	38.38	3.14	52.49	4.25	83.15	4.12	61.08	3.04	P	1	29.716	47.51
260	46.38	3.41	39.80	3.48	54.67	4.08	85.88	4.04	63.75	3.09	P	1	30.858	49.79
261	45.57	4.64	57.13	2.93	28.32	2.26	36.95	1.96	19.75	2.16	P	1	20.167	50.21
262	33.81	2.80	42.82	2.22	18.88	1.45	26.28	1.31	14.45	1.29	P	1	14.531	36.61
263	31.72	2.64	40.00	1.99	20.36	1.59	32.21	1.62	17.72	1.52	P	1	15.137	34.36
264	9.85	2.88	14.73	1.73	16.27	1.22	20.89	1.25	7.77	1.23	P	1	7.782	12.73
265	44.47	4.24	55.77	3.43	38.57	3.06	52.80	2.82	18.57	3.60	P	1	22.733	48.71
266	18.46	1.79	21.12	1.47	18.83	1.21	26.33	1.05	11.41	1.56	P	1	10.323	20.25
267	18.55	2.12	21.08	1.41	18.19	1.30	25.97	1.17	11.26	1.57	P	1	10.262	20.67
268	27.57	2.67	31.60	2.12	18.92	1.52	30.42	1.58	12.46	1.46	P	1	13.032	30.24
269	27.07	2.66	31.95	1.87	18.80	1.51	26.38	1.33	12.52	1.16	P	1	12.525	29.73
270	26.92	2.73	31.54	2.01	19.10	1.59	26.68	1.35	13.21	1.23	P	1	12.636	29.65
271	49.12	3.59	60.95	2.99	26.35	2.00	37.29	1.81	10.39	2.03	P	1	19.652	52.71
272	48.80	3.78	59.47	3.17	26.14	2.24	36.01	1.63	11.65	2.00	P	1	19.489	52.58
273	47.17	3.96	56.57	2.87	33.60	2.68	50.66	2.49	32.03	2.06	P	1	23.409	51.13
274	37.86	3.24	46.72	2.63	27.52	2.26	42.49	1.89	26.88	1.54	P	1	19.303	41.10
275	39.98	3.34	47.50	2.64	28.57	2.53	42.96	2.20	27.66	1.67	P	1	19.905	43.32
276	44.18	3.88	50.72	3.49	39.58	3.90	50.89	3.61	35.14	2.82	P	1	23.821	48.06
277	41.36	3.52	49.34	2.83	30.97	2.46	44.31	2.43	29.97	2.05	P	1	20.924	44.88
278	41.66	3.56	48.82	2.80	31.94	2.52	46.19	2.31	30.73	1.95	P	1	21.248	45.22
279	31.28	2.83	36.88	2.20	22.30	2.10	32.60	1.68	20.47	1.54	P	1	15.388	34.11
280	34.07	3.26	41.34	2.62	24.25	2.30	36.35	2.13	24.69	2.12	P	1	17.313	37.33
281	33.79	3.04	42.13	2.56	23.09	2.21	34.74	2.07	22.12	1.78	P	1	16.753	36.83
282	34.15	2.98	40.67	2.45	24.72	1.98	37.35	1.93	23.49	1.67	P	1	17.139	37.13
283	36.88	3.18	46.28	2.77	26.08	2.19	39.44	2.03	25.60	1.98	P	1	18.643	40.06
284	28.55	2.65	33.74	2.38	19.56	1.98	29.21	1.64	18.75	1.53	P	1	13.999	31.20
285	30.51	2.69	35.33	2.29	21.90	2.04	32.26	1.65	22.52	1.43	P	1	15.262	33.20
286	29.81	2.59	35.05	2.28	21.48	2.11	31.89	1.84	21.67	1.52	P	1	15.024	32.40
287	27.72	2.67	32.00	2.21	22.68	2.13	32.50	1.76	19.65	1.64	P	1	14.496	30.39
288	29.14	2.75	32.91	2.16	23.13	2.11	33.89	1.80	20.86	1.70	P	1	15.045	31.89
289	27.85	2.65	32.62	2.05	22.84	2.14	33.62	1.91	20.82	1.80	P	1	14.830	30.50
290	20.06	1.73	23.15	1.49	16.63	1.57	24.71	1.29	15.19	1.24	P	1	10.706	21.79

290	20.00	1.75	25.15	1.49	16.05	1.37	24.71	1.29	15.19	1.24	P	1	16.766	21.79
	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
291	41.60	3.74	56.50	3.02	37.00	2.99	62.90	2.82	38.21	2.51	P	1	25.129	45.34
id														

In [88]:

```
birds[(birds["type"]=="SW") | (birds["type"]=="R")]
```

Out[88]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
249	93.37	5.33	102.65	3.90	58.55	4.62	101.24	4.57	71.30	4.38	R	1	44.991	98.70
250	99.69	6.09	111.00	4.26	60.84	4.75	104.72	4.92	71.23	4.21	R	1	47.171	105.78
251	34.84	2.44	40.68	2.04	38.43	2.54	42.48	2.45	20.07	3.15	R	1	18.912	37.28
252	33.12	2.40	39.85	1.80	27.46	2.19	38.99	2.15	19.10	2.89	R	1	16.995	35.52
253	35.63	2.40	42.67	2.02	28.51	2.41	42.64	2.46	21.05	3.12	R	1	18.291	38.03

166 rows × 14 columns

8. Sorting

In [95]:

```
# Index sort
birds.sort_index(axis = 1, ascending=True)
# A-Z
# a-z
```

Out[95]:

	Mean_Features	feml	femw	huml	huml_and_humw	humw	len_type	tarl	tarw	tibl	tibw	type	ulnal	ulnaw
id														
0	26.193	41.81	3.70	80.78	87.46	6.68	2	38.70	3.84	5.50	4.03	SW	72.01	4.88
1	36.324	47.04	4.30	88.91	95.54	6.63	2	41.50	4.01	80.22	4.51	SW	80.53	5.59
2	32.889	43.07	3.90	79.97	86.34	6.37	2	38.31	3.34	75.35	4.04	SW	69.26	5.28
3	30.920	40.04	3.52	77.65	83.35	5.70	2	35.78	3.41	69.17	3.40	SW	65.76	4.77
4	25.437	33.95	2.72	62.80	67.64	4.84	2	31.88	3.13	56.27	2.96	SW	52.09	3.73
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
415	11.579	18.36	1.54	17.96	19.59	1.63	2	21.99	1.15	31.25	1.33	SO	19.25	1.33
416	12.303	19.24	1.45	19.21	20.85	1.64	2	23.60	1.15	33.21	1.28	SO	20.76	1.49
417	12.410	20.96	1.43	18.79	20.42	1.63	2	22.86	1.21	34.45	1.41	SO	19.83	1.53
418	13.386	21.35	1.48	20.38	22.16	1.78	2	25.98	1.24	36.09	1.53	SO	22.53	1.50
419	11.244	17.62	1.34	17.89	19.33	1.44	2	21.69	1.05	29.81	1.24	SO	19.26	1.10

420 rows × 14 columns

In [93]:

```
birds.sort_index(axis = 0, ascending=False)
```

Out[93]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2	11.244	19.33
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2	13.386	22.16
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2	12.410	20.42
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2	12.303	20.85
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2	11.579	19.59
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46

420 rows × 14 columns

In [94]:

```
birds.sort_index(axis = 0, ascending=True)
```

Out[94]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_humw
id														
0	80.78	6.68	72.01	4.88	41.81	3.70	5.50	4.03	38.70	3.84	SW	2	26.193	87.46
1	88.91	6.63	80.53	5.59	47.04	4.30	80.22	4.51	41.50	4.01	SW	2	36.324	95.54
2	79.97	6.37	69.26	5.28	43.07	3.90	75.35	4.04	38.31	3.34	SW	2	32.889	86.34
3	77.65	5.70	65.76	4.77	40.04	3.52	69.17	3.40	35.78	3.41	SW	2	30.920	83.35
4	62.80	4.84	52.09	3.73	33.95	2.72	56.27	2.96	31.88	3.13	SW	2	25.437	67.64
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
415	17.96	1.63	19.25	1.33	18.36	1.54	31.25	1.33	21.99	1.15	SO	2	11.579	19.59
416	19.21	1.64	20.76	1.49	19.24	1.45	33.21	1.28	23.60	1.15	SO	2	12.303	20.85
417	18.79	1.63	19.83	1.53	20.96	1.43	34.45	1.41	22.86	1.21	SO	2	12.410	20.42
418	20.38	1.78	22.53	1.50	21.35	1.48	36.09	1.53	25.98	1.24	SO	2	13.386	22.16
419	17.89	1.44	19.26	1.10	17.62	1.34	29.81	1.24	21.69	1.05	SO	2	11.244	19.33

420 rows × 14 columns

In [98]:

```
birds.sort_values(by = "huml", ascending= False)
```

Out[98]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_hum
id														
58	420.00	17.84	422.00	11.72	110.54	9.99	237.00	11.03	128.35	8.93	SW	2	137.740	437.
56	310.00	14.40	315.00	9.51	88.77	8.10	180.00	9.45	96.13	7.69	SW	2	103.905	324.
60	300.00	12.48	300.00	8.65	84.05	8.53	178.00	9.61	99.01	7.55	SW	2	100.788	312.

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_hum
60	300.00	12.48	300.00	8.65	84.05	8.53	178.00	9.61	99.01	7.55	SW	2	100.788	312.
38	272.00	14.86	320.00	10.42	91.60	9.71	132.00	10.23	81.77	8.91	SW	2	95.150	286.
39	270.00	14.25	310.00	10.90	86.20	9.96	125.76	9.63	79.18	10.05	SW	2	92.593	284.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
367	13.31	1.17	16.47	1.06	12.32	0.93	22.47	0.95	15.97	0.75	SO	2	8.540	14.
413	12.95	1.16	14.09	1.03	13.03	1.03	22.13	0.96	15.19	1.02	SO	2	8.259	14.
377	12.69	1.69	17.44	1.16	11.83	1.03	23.80	0.87	16.97	0.66	SO	2	8.814	14.
264	9.85	2.88	14.73	1.73	16.27	1.22	20.89	1.25	7.77	1.23	P	1	7.782	12.
342	NaN	NaN	NaN	NaN	32.54	2.65	55.06	2.81	38.94	2.25	SO	2	22.375	Na

420 rows × 14 columns

In [99]:

```
# sort multiple columns at a time
birds.sort_values(by = ["huml", "type"], ascending= False)
```

Out[99]:

id	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	tarw	type	len_type	Mean_Features	huml_and_hum
58	420.00	17.84	422.00	11.72	110.54	9.99	237.00	11.03	128.35	8.93	SW	2	137.740	437.
56	310.00	14.40	315.00	9.51	88.77	8.10	180.00	9.45	96.13	7.69	SW	2	103.905	324.
60	300.00	12.48	300.00	8.65	84.05	8.53	178.00	9.61	99.01	7.55	SW	2	100.788	312.
38	272.00	14.86	320.00	10.42	91.60	9.71	132.00	10.23	81.77	8.91	SW	2	95.150	286.
39	270.00	14.25	310.00	10.90	86.20	9.96	125.76	9.63	79.18	10.05	SW	2	92.593	284.
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
367	13.31	1.17	16.47	1.06	12.32	0.93	22.47	0.95	15.97	0.75	SO	2	8.540	14.
413	12.95	1.16	14.09	1.03	13.03	1.03	22.13	0.96	15.19	1.02	SO	2	8.259	14.
377	12.69	1.69	17.44	1.16	11.83	1.03	23.80	0.87	16.97	0.66	SO	2	8.814	14.
264	9.85	2.88	14.73	1.73	16.27	1.22	20.89	1.25	7.77	1.23	P	1	7.782	12.
342	NaN	NaN	NaN	NaN	32.54	2.65	55.06	2.81	38.94	2.25	SO	2	22.375	Na

420 rows × 14 columns

In [100]:

```
help(birds.sort_index)
```

```
Help on method sort_index in module pandas.core.frame:

sort_index(axis=0, level=None, ascending=True, inplace=False, kind='quicksort', na_positi
on='last', sort_remaining=True, ignore_index: bool = False) method of pandas.core.frame.D
ataFrame instance
    Sort object by labels (along an axis).

    Parameters
    ----------
    axis : {0 or 'index', 1 or 'columns'}, default 0
        The axis along which to sort.  The value 0 identifies the rows,
        and 1 identifies the columns.
    level : int or level name or list of ints or list of level names
        If not None, sort on values in specified index level(s).
    ascending : bool, default True
        Sort ascending vs. descending.
    inplace : bool, default False
        If True, perform operation in-place.
    kind : {'quicksort', 'mergesort', 'heapsort'}, default 'quicksort'
    kind : {'quicksort', 'mergesort', 'heapsort'}, default 'quicksort'
        Choice of sorting algorithm. See also ndarray.np.sort for more
        information.  `mergesort` is the only stable algorithm. For
        DataFrames, this option is only applied when sorting on a single
        column or label.
    na_position : {'first', 'last'}, default 'last'
        Puts NaNs at the beginning if `first`; `last` puts NaNs at the end.
        Not implemented for MultiIndex.
    sort_remaining : bool, default True
        If True and sorting by level and index is multilevel, sort by other
        levels too (in order) after sorting by specified level.
    ignore_index : bool, default False
        If True, the resulting axis will be labeled 0, 1, …, n - 1.

        .. versionadded:: 1.0.0

    Returns
    -------
    sorted_obj : DataFrame or None
        DataFrame with sorted index if inplace=False, None otherwise.
```

9. Statistical

max value min mean median varience std dev sum count mode

In [101]:

```
birds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 420 entries, 0 to 419
Data columns (total 14 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   huml           419 non-null    float64
 1   humw           419 non-null    float64
 2   ulnal          417 non-null    float64
 3   ulnaw          418 non-null    float64
 4   feml           418 non-null    float64
 5   femw           419 non-null    float64
 6   tibl           418 non-null    float64
 7   tibw           419 non-null    float64
 8   tarl           419 non-null    float64
 9   tarw           419 non-null    float64
 10  type           420 non-null    object 
 11  len_type       420 non-null    int64  
 12  Mean_Features  420 non-null    float64
 13  huml_and_humw  419 non-null    float64
dtypes: float64(12), int64(1), object(1)
memory usage: 69.2+ KB
```

In [102]:

```
birds.describe()
```

Out[102]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	ta
count	419.000000	419.000000	417.000000	418.000000	418.000000	419.000000	418.000000	419.000000	419.000000	419.0000
mean	64.650501	4.370573	69.115372	3.597249	36.872416	3.220883	64.662823	3.182339	39.229976	2.9300
std	53.834549	2.854617	58.784775	2.186747	19.979082	2.023581	37.838145	2.080827	23.184313	2.1856
min	9.850000	1.140000	14.090000	1.000000	11.830000	0.930000	5.500000	0.870000	7.770000	0.6600
25%	25.170000	2.190000	28.050000	1.870000	21.297500	1.715000	36.417500	1.565000	23.035000	1.4250
50%	44.180000	3.500000	43.710000	2.945000	31.130000	2.520000	52.120000	2.490000	31.740000	2.2300

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	ta
75%	90.310000	5.810000	97.520000	4.770000	47.120000	4.135000	82.870000	4.255000	50.250000	3.5000
max	420.000000	17.840000	422.000000	12.000000	117.070000	11.640000	240.000000	11.030000	175.000000	14.0900

In [103]:

```
birds["ulnal"].median()
```

Out[103]:

```
43.71
```

In [104]:

```
birds["ulnal"].mean()
```

Out[104]:

```
69.11537170263787
```

In [106]:

```
birds["type"].mode()   # frequency count of entire column
```

Out[106]:

```
0    SO
dtype: object
```

In [113]:

```
birds["type"].value_counts()
```

Out[113]:

```
SO    128
SW    116
W      65
R      50
P      38
T      23
Name: type, dtype: int64
```

In [114]:

```
birds["ulnal"].std()
```

Out[114]:

```
58.7847749488504
```

In [117]:

```
import numpy as np
birds[["ulnal"]].agg([np.min, np.max, np.mean])
```

Out[117]:

	ulnal
amin	14.090000
amax	422.000000
mean	69.115372

In [121]:

```
birds[["ulnal"]].max()
```

Out[121]:

```
ulnal    422.0
dtype: float64
```

```
dtype: float64
```

In [122]:

```
birds["ulnal"].idxmax()   # find max value index position
```

Out[122]:

```
58
```

In [123]:

```
birds.iloc[58]
```

Out[123]:

```
huml                  420
humw                17.84
ulnal                 422
ulnaw               11.72
feml               110.54
femw                 9.99
tibl                  237
tibw                11.03
tarl               128.35
tarw                 8.93
type                   SW
len_type                2
Mean_Features      137.74
huml_and_humw      437.84
Name: 58, dtype: object
```

In [125]:

```
birds["ulnal"].idxmin()
```

Out[125]:

```
413
```

10. Plotting

11. Saving

In [ ]:

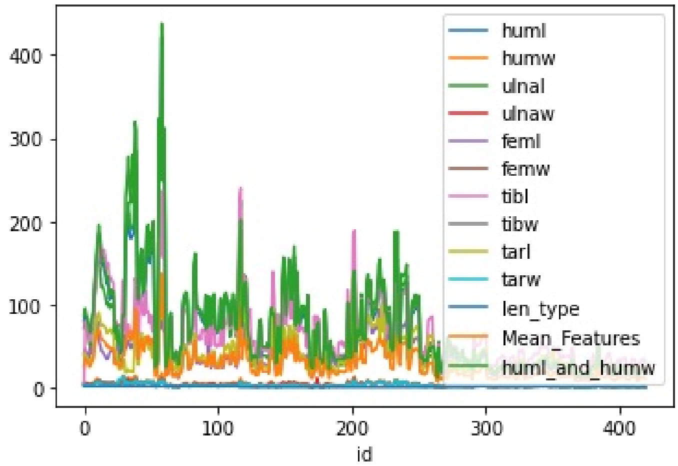
```
import matplotlib.pyplot as plt
```

In [126]:

```
birds.plot()
```

Out[126]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a13636f5b0>
```

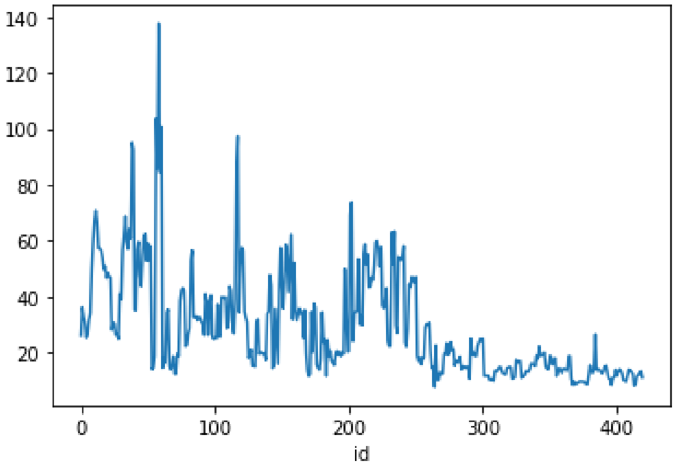


In [130]:

```
birds["Mean_Features"].plot()
```

Out[130]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a13852cfd0>
```



In [129]:

```
birds.columns
```

Out[129]:

```
Index(['huml', 'humw', 'ulnal', 'ulnaw', 'feml', 'femw', 'tibl', 'tibw',
       'tarl', 'tarw', 'type', 'len_type', 'Mean_Features', 'huml_and_humw'],
      dtype='object')
```

In [131]:

```
birds.describe()
```

Out[131]:

	huml	humw	ulnal	ulnaw	feml	femw	tibl	tibw	tarl	ta
count	419.000000	419.000000	417.000000	418.000000	418.000000	419.000000	418.000000	419.000000	419.000000	419.0000
mean	64.650501	4.370573	69.115372	3.597249	36.872416	3.220883	64.662823	3.182339	39.229976	2.9300
std	53.834549	2.854617	58.784775	2.186747	19.979082	2.023581	37.838145	2.080827	23.184313	2.1856
min	9.850000	1.140000	14.090000	1.000000	11.830000	0.930000	5.500000	0.870000	7.770000	0.6600
25%	25.170000	2.190000	28.050000	1.870000	21.297500	1.715000	36.417500	1.565000	23.035000	1.4250
50%	44.180000	3.500000	43.710000	2.945000	31.130000	2.520000	52.120000	2.490000	31.740000	2.2300
75%	90.310000	5.810000	97.520000	4.770000	47.120000	4.135000	82.870000	4.255000	50.250000	3.5000
max	420.000000	17.840000	422.000000	12.000000	117.070000	11.640000	240.000000	11.030000	175.000000	14.0900

In [133]:

```
birds.describe().to_csv("birds.csv")
```

In [134]:

```
birds.describe().to_excel("birds_describe.xlsx")
```

In [ ]:

```

```

In [ ]:

In [ ]:

In [ ]:

Cleaning data in Python

- **NaN : not a number -- special floating-point value**
- **Working with duplicates and missing values**
  - **isnull()**
  - **notnull()**
  - **dropna()**
  - **fillna()**
  - **replace()**
- **Which values should be replaced with missing values based on data identifying and eliminating outliers**
- **Dropping duplicate data**

2 file

first file 100 records -- 5 persons responded useless file -- unwanted second file 100 records -- 95 persons responded -- will take -- filling missing with meaning fulldata

In [135]:

```
emp = pd.read_csv("https://raw.githubusercontent.com/APSSDC-Data-Analysis/DataAnalysis-7/main/Datasets/employe.csv")
emp.head()
```

Out[135]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services

In [136]:

```
emp.shape   # 1000 employe details
```

Out[136]:

```
(1000, 8)
```

In [138]:

```
emp.isnull().sum()  # finding missing data
```

Out[138]:

```
First Name             67
Gender                145
Start Date              0
Last Login Time         0
Salary                  0
Bonus %                 0
Senior Management      67
```

```
Team                  43
dtype: int64
```

In [139]:

```
# notnull()
emp.notnull().sum()   # non null value count
```

Out[139]:

```
First Name             933
Gender                 855
Start Date            1000
Last Login Time       1000
Salary                1000
Bonus %               1000
Senior Management      933
Team                   957
dtype: int64
```

In [140]:

```
emp.describe()   # statastical
```

Out[140]:

	Salary	Bonus %
count	1000.000000	1000.000000
mean	90662.181000	10.207555
std	32923.693342	5.528481
min	35013.000000	1.015000
25%	62613.000000	5.401750
50%	90428.000000	9.838500
75%	118740.250000	14.838000
max	149908.000000	19.944000

In [142]:

```
emp.dropna().shape
```

Out[142]:

```
(764, 8)
```

In [143]:

```
1000 - 764   # 236 are entired details with missing data
```

Out[143]:

```
236
```

In [144]:

```
emp.dropna()  # it drops entire record having min one missing value
```

Out[144]:

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	8/6/1993	12:42 PM	97308	6.945	True	Marketing
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	3/4/2005	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
5	Dennis	Male	4/18/1987	1:35 AM	115163	10.125	False	Legal

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
...	...	...	...	...	...	...	...	...
994	George	Male	6/21/2013	5:47 PM	98874	4.479	True	Marketing
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

764 rows × 8 columns

In [146]:

```
emp["First Name"].isnull().sum()
```

Out[146]:

```
67
```

In [148]:

```
# fillna() -- replacing null values with given data

emp["Gender"].fillna("No Gender")   # filling with constant
```

Out[148]:

```
0            Male
1            Male
2          Female
3            Male
4            Male
         ...
995     No Gender
996          Male
997          Male
998          Male
999          Male
Name: Gender, Length: 1000, dtype: object
```

In [149]:

```
emp["Gender"].fillna(0)
```

Out[149]:

```
0         Male
1         Male
2       Female
3         Male
4         Male
        ...
995          0
996       Male
997       Male
998       Male
999       Male
Name: Gender, Length: 1000, dtype: object
```

In [150]:

```
# bfill   -- backword one -- filling next one detils
# and pad --  forword one -- filling missing value with previoue one
emp["Gender"].fillna(method = "pad")
```

Out[150]:

```
0         Male
1         Male
2       Female
3         Male
4         Male
```

```
4         Male
        ...
995       Male
996       Male
997       Male
998       Male
999       Male
Name: Gender, Length: 1000, dtype: object
```

In [151]:

```
emp["Gender"].fillna(method = "bfill")
```

Out[151]:

```
0         Male
1         Male
2       Female
3         Male
4         Male
        ...
995       Male
996       Male
997       Male
998       Male
999       Male
Name: Gender, Length: 1000, dtype: object
```

In [ ]: