

Descriptive summary statistics

```
In [1]: import pandas as pd  
df=pd.read_csv('Z:/sem 7/EmployeeAttrition.csv')  
df
```

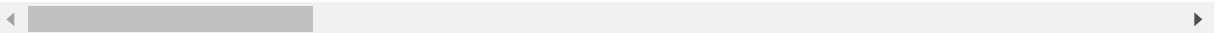
Out[1]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Edu
0	41	Yes	Travel_Rarely	1102	Sales	1	2
1	49	No	Travel_Frequently	279	Research & Development	8	1
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2
3	33	No	Travel_Frequently	1392	Research & Development	3	4
4	27	No	Travel_Rarely	591	Research & Development	2	1
5	32	No	Travel_Frequently	1005	Research & Development	2	2
6	59	No	Travel_Rarely	1324	Research & Development	3	3
7	30	No	Travel_Rarely	1358	Research & Development	24	1
8	38	No	Travel_Frequently	216	Research & Development	23	3
9	36	No	Travel_Rarely	1299	Research & Development	27	3
10	35	No	Travel_Rarely	809	Research & Development	16	3
11	29	No	Travel_Rarely	153	Research & Development	15	2
12	31	No	Travel_Rarely	670	Research & Development	26	1
13	34	No	Travel_Rarely	1346	Research & Development	19	2
14	28	Yes	Travel_Rarely	103	Research & Development	24	3
15	29	No	Travel_Rarely	1389	Research & Development	21	4
16	32	No	Travel_Rarely	334	Research & Development	5	2
17	22	No	Non-Travel	1123	Research & Development	16	2
18	53	No	Travel_Rarely	1219	Sales	2	4

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Edu
19	38	No	Travel_Rarely	371	Research & Development	2	3
20	24	No	Non-Travel	673	Research & Development	11	2
21	36	Yes	Travel_Rarely	1218	Sales	9	4
22	34	No	Travel_Rarely	419	Research & Development	7	4
23	21	No	Travel_Rarely	391	Research & Development	15	2
24	34	Yes	Travel_Rarely	699	Research & Development	6	1
25	53	No	Travel_Rarely	1282	Research & Development	5	3
26	32	Yes	Travel_Frequently	1125	Research & Development	16	1
27	42	No	Travel_Rarely	691	Sales	8	4
28	44	No	Travel_Rarely	477	Research & Development	7	4
29	46	No	Travel_Rarely	705	Sales	2	4
...
1440	36	No	Travel_Frequently	688	Research & Development	4	2
1441	56	No	Non-Travel	667	Research & Development	1	4
1442	29	Yes	Travel_Rarely	1092	Research & Development	1	4
1443	42	No	Travel_Rarely	300	Research & Development	2	3
1444	56	Yes	Travel_Rarely	310	Research & Development	7	2
1445	41	No	Travel_Rarely	582	Research & Development	28	4
1446	34	No	Travel_Rarely	704	Sales	28	3
1447	36	No	Non-Travel	301	Sales	15	4
1448	41	No	Travel_Rarely	930	Sales	3	3
1449	32	No	Travel_Rarely	529	Research & Development	2	3

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Edu
1450	35	No	Travel_Rarely	1146	Human Resources	26	4
1451	38	No	Travel_Rarely	345	Sales	10	2
1452	50	Yes	Travel_Frequently	878	Sales	1	4
1453	36	No	Travel_Rarely	1120	Sales	11	4
1454	45	No	Travel_Rarely	374	Sales	20	3
1455	40	No	Travel_Rarely	1322	Research & Development	2	4
1456	35	No	Travel_Frequently	1199	Research & Development	18	4
1457	40	No	Travel_Rarely	1194	Research & Development	2	4
1458	35	No	Travel_Rarely	287	Research & Development	1	4
1459	29	No	Travel_Rarely	1378	Research & Development	13	2
1460	29	No	Travel_Rarely	468	Research & Development	28	4
1461	50	Yes	Travel_Rarely	410	Sales	28	3
1462	39	No	Travel_Rarely	722	Sales	24	1
1463	31	No	Non-Travel	325	Research & Development	5	3
1464	26	No	Travel_Rarely	1167	Sales	5	3
1465	36	No	Travel_Frequently	884	Research & Development	23	2
1466	39	No	Travel_Rarely	613	Research & Development	6	1
1467	27	No	Travel_Rarely	155	Research & Development	4	3
1468	49	No	Travel_Frequently	1023	Sales	2	3
1469	34	No	Travel_Rarely	628	Research & Development	8	3

1470 rows × 35 columns



In [2]: `df.describe`

```

Out[2]: <bound method NDFrame.describe of
lyRate      Department \
0      41      Yes      Travel_Rarely      1102      Sales
1      49      No      Travel_Frequently      279      Research & Development
2      37      Yes      Travel_Rarely      1373      Research & Development
3      33      No      Travel_Frequently      1392      Research & Development
4      27      No      Travel_Rarely      591      Research & Development
5      32      No      Travel_Frequently      1005      Research & Development
6      59      No      Travel_Rarely      1324      Research & Development
7      30      No      Travel_Rarely      1358      Research & Development
8      38      No      Travel_Frequently      216      Research & Development
9      36      No      Travel_Rarely      1299      Research & Development
10     35      No      Travel_Rarely      809      Research & Development
11     29      No      Travel_Rarely      153      Research & Development
12     31      No      Travel_Rarely      670      Research & Development
13     34      No      Travel_Rarely      1346      Research & Development
14     28      Yes      Travel_Rarely      103      Research & Development
15     29      No      Travel_Rarely      1389      Research & Development
16     32      No      Travel_Rarely      334      Research & Development
17     22      No      Non-Travel      1123      Research & Development
18     53      No      Travel_Rarely      1219      Sales
19     38      No      Travel_Rarely      371      Research & Development
20     24      No      Non-Travel      673      Research & Development
21     36      Yes      Travel_Rarely      1218      Sales
22     34      No      Travel_Rarely      419      Research & Development
23     21      No      Travel_Rarely      391      Research & Development
24     34      Yes      Travel_Rarely      699      Research & Development
25     53      No      Travel_Rarely      1282      Research & Development
26     32      Yes      Travel_Frequently      1125      Research & Development
27     42      No      Travel_Rarely      691      Sales
28     44      No      Travel_Rarely      477      Research & Development
29     46      No      Travel_Rarely      705      Sales
...     ...     ...     ...     ...     ...
1440   36      No      Travel_Frequently      688      Research & Development
1441   56      No      Non-Travel      667      Research & Development
1442   29      Yes      Travel_Rarely      1092      Research & Development
1443   42      No      Travel_Rarely      300      Research & Development
1444   56      Yes      Travel_Rarely      310      Research & Development
1445   41      No      Travel_Rarely      582      Research & Development
1446   34      No      Travel_Rarely      704      Sales
1447   36      No      Non-Travel      301      Sales
1448   41      No      Travel_Rarely      930      Sales
1449   32      No      Travel_Rarely      529      Research & Development
1450   35      No      Travel_Rarely      1146      Human Resources
1451   38      No      Travel_Rarely      345      Sales
1452   50      Yes      Travel_Frequently      878      Sales
1453   36      No      Travel_Rarely      1120      Sales
1454   45      No      Travel_Rarely      374      Sales
1455   40      No      Travel_Rarely      1322      Research & Development
1456   35      No      Travel_Frequently      1199      Research & Development
1457   40      No      Travel_Rarely      1194      Research & Development
1458   35      No      Travel_Rarely      287      Research & Development
1459   29      No      Travel_Rarely      1378      Research & Development
1460   29      No      Travel_Rarely      468      Research & Development
1461   50      Yes      Travel_Rarely      410      Sales
1462   39      No      Travel_Rarely      722      Sales
1463   31      No      Non-Travel      325      Research & Development

```

1464	26	No	Travel_Rarely	1167	Sales
1465	36	No	Travel_Frequently	884	Research & Development
1466	39	No	Travel_Rarely	613	Research & Development
1467	27	No	Travel_Rarely	155	Research & Development
1468	49	No	Travel_Frequently	1023	Sales
1469	34	No	Travel_Rarely	628	Research & Development

	DistanceFromHome	Education	EducationField	EmployeeCount	\
0	1	2	Life Sciences	1	
1	8	1	Life Sciences	1	
2	2	2	Other	1	
3	3	4	Life Sciences	1	
4	2	1	Medical	1	
5	2	2	Life Sciences	1	
6	3	3	Medical	1	
7	24	1	Life Sciences	1	
8	23	3	Life Sciences	1	
9	27	3	Medical	1	
10	16	3	Medical	1	
11	15	2	Life Sciences	1	
12	26	1	Life Sciences	1	
13	19	2	Medical	1	
14	24	3	Life Sciences	1	
15	21	4	Life Sciences	1	
16	5	2	Life Sciences	1	
17	16	2	Medical	1	
18	2	4	Life Sciences	1	
19	2	3	Life Sciences	1	
20	11	2	Other	1	
21	9	4	Life Sciences	1	
22	7	4	Life Sciences	1	
23	15	2	Life Sciences	1	
24	6	1	Medical	1	
25	5	3	Other	1	
26	16	1	Life Sciences	1	
27	8	4	Marketing	1	
28	7	4	Medical	1	
29	2	4	Marketing	1	
...	
1440	4	2	Life Sciences	1	
1441	1	4	Life Sciences	1	
1442	1	4	Medical	1	
1443	2	3	Life Sciences	1	
1444	7	2	Technical Degree	1	
1445	28	4	Life Sciences	1	
1446	28	3	Marketing	1	
1447	15	4	Marketing	1	
1448	3	3	Life Sciences	1	
1449	2	3	Technical Degree	1	
1450	26	4	Life Sciences	1	
1451	10	2	Life Sciences	1	
1452	1	4	Life Sciences	1	
1453	11	4	Marketing	1	
1454	20	3	Life Sciences	1	
1455	2	4	Life Sciences	1	
1456	18	4	Life Sciences	1	
1457	2	4	Medical	1	

1458	1	4	Life Sciences	1
1459	13	2	Other	1
1460	28	4	Medical	1
1461	28	3	Marketing	1
1462	24	1	Marketing	1
1463	5	3	Medical	1
1464	5	3	Other	1
1465	23	2	Medical	1
1466	6	1	Medical	1
1467	4	3	Life Sciences	1
1468	2	3	Medical	1
1469	8	3	Medical	1

	EmployeeNumber	...	RelationshipSatisfaction	\
0	1	...	1	
1	2	...	4	
2	4	...	2	
3	5	...	3	
4	7	...	4	
5	8	...	3	
6	10	...	1	
7	11	...	2	
8	12	...	2	
9	13	...	2	
10	14	...	3	
11	15	...	4	
12	16	...	4	
13	18	...	3	
14	19	...	2	
15	20	...	3	
16	21	...	4	
17	22	...	2	
18	23	...	3	
19	24	...	3	
20	26	...	4	
21	27	...	2	
22	28	...	3	
23	30	...	4	
24	31	...	3	
25	32	...	4	
26	33	...	2	
27	35	...	4	
28	36	...	4	
29	38	...	4	
...	
1440	2025	...	2	
1441	2026	...	1	
1442	2027	...	2	
1443	2031	...	1	
1444	2032	...	4	
1445	2034	...	3	
1446	2035	...	4	
1447	2036	...	1	
1448	2037	...	3	
1449	2038	...	4	
1450	2040	...	3	
1451	2041	...	3	

1452	2044	...	4
1453	2045	...	1
1454	2046	...	3
1455	2048	...	4
1456	2049	...	4
1457	2051	...	2
1458	2052	...	4
1459	2053	...	1
1460	2054	...	2
1461	2055	...	2
1462	2056	...	1
1463	2057	...	2
1464	2060	...	4
1465	2061	...	3
1466	2062	...	1
1467	2064	...	2
1468	2065	...	4
1469	2068	...	1

	StandardHours	StockOptionLevel	TotalWorkingYears \
0	80	0	8
1	80	1	10
2	80	0	7
3	80	0	8
4	80	1	6
5	80	0	8
6	80	3	12
7	80	1	1
8	80	0	10
9	80	2	17
10	80	1	6
11	80	0	10
12	80	1	5
13	80	1	3
14	80	0	6
15	80	1	10
16	80	2	7
17	80	2	1
18	80	0	31
19	80	0	6
20	80	1	5
21	80	0	10
22	80	0	13
23	80	0	0
24	80	0	8
25	80	1	26
26	80	0	10
27	80	1	10
28	80	1	24
29	80	0	22
...
1440	80	3	18
1441	80	1	13
1442	80	3	4
1443	80	0	24
1444	80	1	14
1445	80	1	21

1446	80	2	8
1447	80	1	15
1448	80	1	14
1449	80	0	4
1450	80	0	9
1451	80	1	10
1452	80	2	12
1453	80	1	8
1454	80	0	8
1455	80	0	8
1456	80	2	10
1457	80	3	20
1458	80	1	4
1459	80	1	10
1460	80	0	5
1461	80	1	20
1462	80	1	21
1463	80	0	10
1464	80	0	5
1465	80	1	17
1466	80	1	9
1467	80	1	6
1468	80	0	17
1469	80	0	6

	TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany \
0	0	1	6
1	3	3	10
2	3	3	0
3	3	3	8
4	3	3	2
5	2	2	7
6	3	2	1
7	2	3	1
8	2	3	9
9	3	2	7
10	5	3	5
11	3	3	9
12	1	2	5
13	2	3	2
14	4	3	4
15	1	3	10
16	5	2	6
17	2	2	1
18	3	3	25
19	3	3	3
20	5	2	4
21	4	3	5
22	4	3	12
23	6	3	0
24	2	3	4
25	3	2	14
26	5	3	10
27	2	3	9
28	4	3	22
29	2	2	2
...

1440	3	3	4
1441	2	2	13
1442	3	4	2
1443	2	2	22
1444	4	1	10
1445	3	3	20
1446	2	3	8
1447	4	2	15
1448	5	3	5
1449	4	3	4
1450	2	3	9
1451	1	3	10
1452	3	3	6
1453	2	2	6
1454	3	3	5
1455	2	3	2
1456	2	4	10
1457	2	3	5
1458	5	3	4
1459	2	3	4
1460	3	1	5
1461	3	3	3
1462	2	2	20
1463	2	3	9
1464	2	3	4
1465	3	3	5
1466	5	3	7
1467	0	3	6
1468	3	2	9
1469	3	4	4

	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	4	0	5
1	7	1	7
2	0	0	0
3	7	3	0
4	2	2	2
5	7	3	6
6	0	0	0
7	0	0	0
8	7	1	8
9	7	7	7
10	4	0	3
11	5	0	8
12	2	4	3
13	2	1	2
14	2	0	3
15	9	8	8
16	2	0	5
17	0	0	0
18	8	3	7
19	2	1	2
20	2	1	3
21	3	0	3
22	6	2	11
23	0	0	0
24	2	1	3

25	13	4	8
26	2	6	7
27	7	4	2
28	6	5	17
29	2	2	1
...
1440	2	0	2
1441	12	1	9
1442	2	2	2
1443	6	4	14
1444	9	9	8
1445	7	0	10
1446	7	1	7
1447	12	11	11
1448	4	0	4
1449	2	1	2
1450	0	1	7
1451	7	1	9
1452	3	0	1
1453	3	0	0
1454	3	0	1
1455	2	2	2
1456	2	0	2
1457	3	0	2
1458	3	1	1
1459	3	0	3
1460	4	0	4
1461	2	2	0
1462	9	9	6
1463	4	1	7
1464	2	0	0
1465	2	0	3
1466	7	1	7
1467	2	0	3
1468	6	0	8
1469	3	1	2

[1470 rows x 35 columns]>

In [3]: df.sum()

```
Out[3]: Age                                     54278
Attrition                                     YesNoYesNoNoNoNoNoNoNoNoNoNoYesNoNoNoNoNoNoY...
BusinessTravel                             Travel_RarelyTravel_FrequentlyTravel_RarelyTra...
DailyRate                                  1179654
Department                               SalesResearch & DevelopmentResearch & Developm...
DistanceFromHome                           13513
Education                                  4282
EducationField                             Life SciencesLife SciencesOtherLife SciencesMe...
EmployeeCount                              1470
EmployeeNumber                             1506552
EnvironmentSatisfaction                     4001
Gender                                     FemaleMaleMaleFemaleMaleMaleFemaleMaleMaleMale...
HourlyRate                                 96860
JobInvolvement                             4013
JobLevel                                   3034
JobRole                                   Sales ExecutiveResearch ScientistLaboratory Te...
JobSatisfaction                             4011
MaritalStatus                             SingleMarriedSingleMarriedMarriedSingleMarried...
MonthlyIncome                             9559309
MonthlyRate                               21040262
NumCompaniesWorked                         3959
Over18                                    YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY...
OverTime                                   YesNoYesYesNoNoYesNoNoNoNoYesNoNoYesNoYesYesNo...
PercentSalaryHike                          22358
PerformanceRating                          4636
RelationshipSatisfaction                     3987
StandardHours                              117600
StockOptionLevel                           1167
TotalWorkingYears                          16581
TrainingTimesLastYear                      4115
WorkLifeBalance                            4059
YearsAtCompany                             10302
YearsInCurrentRole                          6217
YearsSinceLastPromotion                     3216
YearsWithCurrManager                       6061
dtype: object
```

```
In [4]: df.mean()
```

```
Out[4]: Age                36.923810
        DailyRate          802.485714
        DistanceFromHome    9.192517
        Education           2.912925
        EmployeeCount        1.000000
        EmployeeNumber      1024.865306
        EnvironmentSatisfaction 2.721769
        HourlyRate          65.891156
        JobInvolvement       2.729932
        JobLevel            2.063946
        JobSatisfaction      2.728571
        MonthlyIncome        6502.931293
        MonthlyRate         14313.103401
        NumCompaniesWorked   2.693197
        PercentSalaryHike    15.209524
        PerformanceRating    3.153741
        RelationshipSatisfaction 2.712245
        StandardHours        80.000000
        StockOptionLevel     0.793878
        TotalWorkingYears    11.279592
        TrainingTimesLastYear 2.799320
        WorkLifeBalance      2.761224
        YearsAtCompany       7.008163
        YearsInCurrentRole   4.229252
        YearsSinceLastPromotion 2.187755
        YearsWithCurrManager 4.123129
        dtype: float64
```

```
In [5]: df.median()
```

```
Out[5]: Age                36.0
        DailyRate          802.0
        DistanceFromHome    7.0
        Education           3.0
        EmployeeCount        1.0
        EmployeeNumber      1020.5
        EnvironmentSatisfaction  3.0
        HourlyRate           66.0
        JobInvolvement        3.0
        JobLevel              2.0
        JobSatisfaction        3.0
        MonthlyIncome        4919.0
        MonthlyRate          14235.5
        NumCompaniesWorked    2.0
        PercentSalaryHike     14.0
        PerformanceRating     3.0
        RelationshipSatisfaction  3.0
        StandardHours         80.0
        StockOptionLevel      1.0
        TotalWorkingYears     10.0
        TrainingTimesLastYear  3.0
        WorkLifeBalance        3.0
        YearsAtCompany         5.0
        YearsInCurrentRole     3.0
        YearsSinceLastPromotion  1.0
        YearsWithCurrManager   3.0
        dtype: float64
```



```
In [6]: df.count()
```

```
Out[6]: Age                1470
Attrition                1470
BusinessTravel          1470
DailyRate              1470
Department             1470
DistanceFromHome       1470
Education              1470
EducationField         1470
EmployeeCount          1470
EmployeeNumber         1470
EnvironmentSatisfaction 1470
Gender                 1470
HourlyRate             1470
JobInvolvement         1470
JobLevel               1470
JobRole                1470
JobSatisfaction        1470
MaritalStatus          1470
MonthlyIncome          1470
MonthlyRate            1470
NumCompaniesWorked     1470
Over18                 1470
OverTime               1470
PercentSalaryHike      1470
PerformanceRating      1470
RelationshipSatisfaction 1470
StandardHours          1470
StockOptionLevel       1470
TotalWorkingYears      1470
TrainingTimesLastYear  1470
WorkLifeBalance        1470
YearsAtCompany         1470
YearsInCurrentRole     1470
YearsSinceLastPromotion 1470
YearsWithCurrManager   1470
dtype: int64
```

In [7]: `df.min()`

```
Out[7]: Age                                18
Attrition                                No
BusinessTravel                          Non-Travel
DailyRate                              102
Department                            Human Resources
DistanceFromHome                        1
Education                              1
EducationField                          Human Resources
EmployeeCount                           1
EmployeeNumber                           1
EnvironmentSatisfaction                  1
Gender                                  Female
HourlyRate                              30
JobInvolvement                           1
JobLevel                                1
JobRole                                Healthcare Representative
JobSatisfaction                           1
MaritalStatus                            Divorced
MonthlyIncome                           1009
MonthlyRate                             2094
NumCompaniesWorked                       0
Over18                                  Y
OverTime                                 No
PercentSalaryHike                         11
PerformanceRating                         3
RelationshipSatisfaction                   1
StandardHours                             80
StockOptionLevel                         0
TotalWorkingYears                         0
TrainingTimesLastYear                     0
WorkLifeBalance                           1
YearsAtCompany                           0
YearsInCurrentRole                        0
YearsSinceLastPromotion                   0
YearsWithCurrManager                      0
dtype: object
```

```
In [8]: df.std()
```

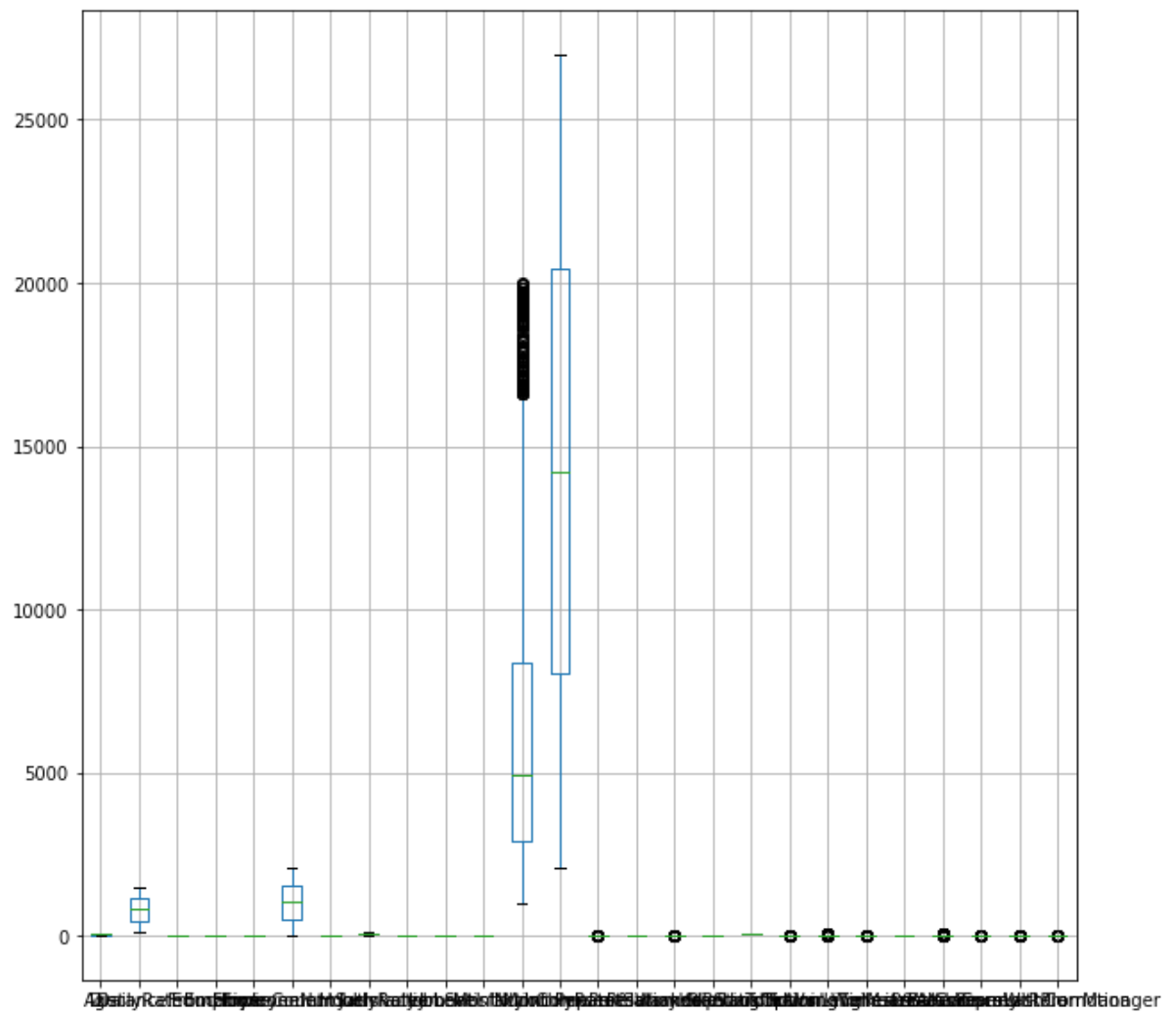
```
Out[8]: Age                9.135373
        DailyRate          403.509100
        DistanceFromHome    8.106864
        Education           1.024165
        EmployeeCount        0.000000
        EmployeeNumber       602.024335
        EnvironmentSatisfaction 1.093082
        HourlyRate           20.329428
        JobInvolvement        0.711561
        JobLevel             1.106940
        JobSatisfaction       1.102846
        MonthlyIncome         4707.956783
        MonthlyRate          7117.786044
        NumCompaniesWorked    2.498009
        PercentSalaryHike     3.659938
        PerformanceRating     0.360824
        RelationshipSatisfaction 1.081209
        StandardHours         0.000000
        StockOptionLevel      0.852077
        TotalWorkingYears     7.780782
        TrainingTimesLastYear 1.289271
        WorkLifeBalance       0.706476
        YearsAtCompany        6.126525
        YearsInCurrentRole    3.623137
        YearsSinceLastPromotion 3.222430
        YearsWithCurrManager  3.568136
        dtype: float64
```

```
In [9]: df.isnull().any()
```

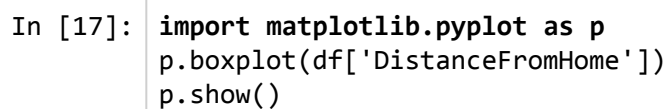
```
Out[9]: Age                False
Attrition                 False
BusinessTravel            False
DailyRate                 False
Department                False
DistanceFromHome          False
Education                 False
EducationField             False
EmployeeCount              False
EmployeeNumber             False
EnvironmentSatisfaction   False
Gender                    False
HourlyRate                 False
JobInvolvement             False
JobLevel                  False
JobRole                   False
JobSatisfaction            False
MaritalStatus              False
MonthlyIncome              False
MonthlyRate                False
NumCompaniesWorked         False
Over18                     False
OverTime                   False
PercentSalaryHike          False
PerformanceRating          False
RelationshipSatisfaction   False
StandardHours              False
StockOptionLevel           False
TotalWorkingYears          False
TrainingTimesLastYear      False
WorkLifeBalance            False
YearsAtCompany             False
YearsInCurrentRole         False
YearsSinceLastPromotion    False
YearsWithCurrManager       False
dtype: bool
```

```
In [19]: df.boxplot(figsize='10,10')
```

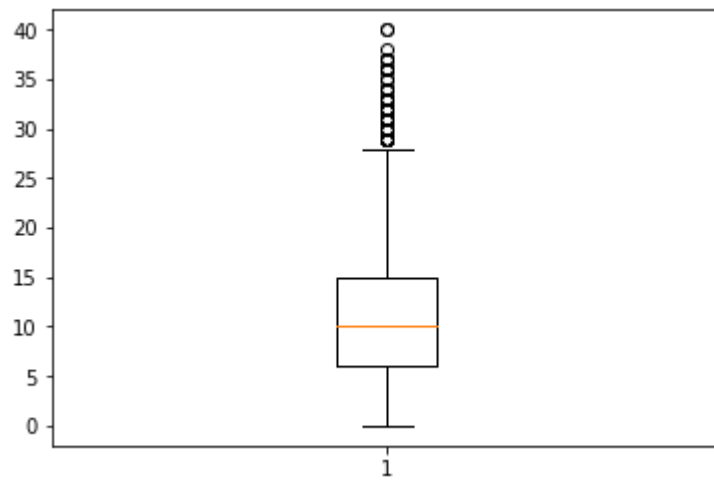
```
Out[19]: <matplotlib.axes._subplots.AxesSubplot at 0xd825630>
```



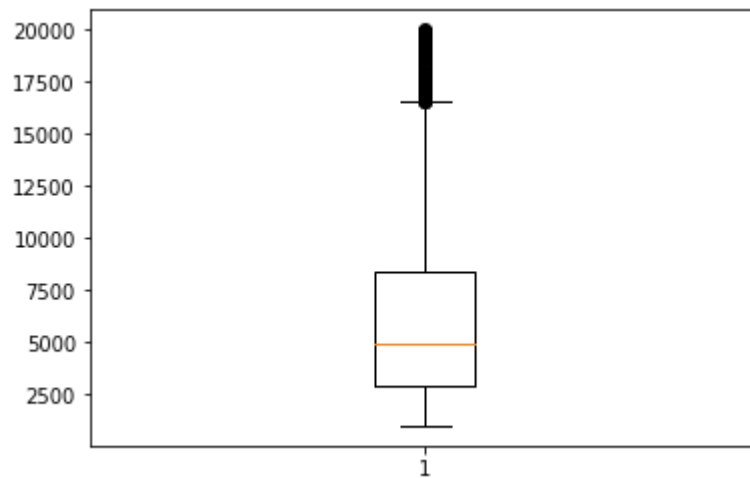
```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7373930>
```



```
In [18]: import matplotlib.pyplot as p
p.boxplot(df['TotalWorkingYears'])
p.show()
```



```
In [15]: import matplotlib.pyplot as p
p.boxplot(df['MonthlyIncome'])
p.show()
```



```
In [20]: df.boxplot(figsize='6,5')
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0xdb162f0>
```

