

```
import pandas as pd

weather = pd.read_csv("3701635.csv", index_col="DATE")

weather
```


	STATION	NAME	ACMH	ACSH	AWND	FMTM	FRGT	PGTM	PRCP	PS
DATE										
1960-01-01	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	712.0	0.00	N
1960-01-02	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1618.0	0.00	N
1960-01-03	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1000.0	0.92	N
1960-01-04	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	936.0	0.00	N
1960-01-05	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1412.0	0.00	N
...	...	...	...	...	...	...	...	...	...	...
2023-12-27	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	5.14	NaN	NaN	NaN	0.02	N
2023-12-28	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	14.76	NaN	NaN	NaN	0.56	N

```
weather.loc["1960-01-01"]
```


STATION	USW00014739
NAME	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US
ACMH	NaN
ACSH	NaN
AWND	NaN
FMTM	NaN
FRGT	NaN
PGTM	712.0
PRCP	0.0
PSUN	NaN
SNOW	0.0
SNWD	0.0
TAVG	NaN
TMAX	37
TMIN	24
TSUN	NaN
WDF1	NaN
WDF2	NaN
WDF5	NaN
WDFG	338.0
WDFM	NaN
WESD	NaN
WSF1	NaN
WSF2	NaN
WSF5	NaN
WSFG	27.5
WSFM	NaN
WT01	NaN
WT02	NaN
WT03	NaN
WT04	NaN

```
WT05
WT06
WT07
WT08
WT09
WT10
WT11
WT13
WT14
WT15
WT16
WT17
WT18
WT19
WT21
WT22
Name: 1960-01-01, dtype: object


weather.loc["1960-01-01":"1960-01-31"]
```

	01-17	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1430.0	0.00	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-18	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	236.0	0.18	NaN	...	NaN	NaN	NaN	NaN	1.0	NaN
	1960-01-19	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	2318.0	0.38	NaN	...	NaN	NaN	NaN	NaN	1.0	NaN
	1960-01-20	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1054.0	0.00	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-21	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	18.0	0.07	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-22	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1024.0	0.01	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-23	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	454.0	0.00	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-24	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	30.0	0.00	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-25	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	2318.0	0.07	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-26	USW00014739	BOSTON LOGAN INTERNATIONAL AIRPORT, MA US	NaN	NaN	NaN	NaN	NaN	1324.0	0.00	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
	1960-01-27		BOSTON LOGAN INTERNATIONAL AIRPORT, MA US															

```
weather.apply(pd.isnull).sum()
```

STATION0  
NAME0  
ACMH11966  
ACSH11963  
AWND8766  
FMTM13232  
FRGT23375  
PGTM5559  
PRCP2  
PSUN16439  
SNOW0  
SNWD8016  
TAVG16780  
TMAX0  
TMIN0  
TSUN11630  
WDF119788  
WDF213243  
WDF513264  
WDFG10490  
WDFM15578  
WESD16535  
WSF119788  
WSF213243  
WSF513264  
WSFG10479  
WSFM15578  
WT0114484  
WT0222146  
WT0322287  
WT0422864  
WT0522967  
WT0623077  
WT0723314  
WT0819133  
WT0923151  
WT1023375  
WT1123357  
WT1321047  
WT1422138  
WT1523307  
WT1615080  
WT1723304  
WT1820949  
WT1923338  
WT2123374  
WT2223284  
dtype: int64


```
weather.apply(pd.isnull).sum()/weather.shape[0]
```



```
STATION    0.000000
NAME       0.000000
ACMH       0.511893
ACSH       0.511764
AWNND      0.375000
FMTM       0.566051
FRGT       0.999957
PGTM       0.237808
PRCP       0.000086
PSUN       0.703243
SNOW       0.000000
SNWD       0.342916
TAVG       0.717830
TMAX       0.000000
TMIN       0.000000
TSUN       0.497519
WDF1       0.846509
WDF2       0.566521
WDF5       0.567420
WDFG       0.448751
WDFM       0.666410
WESD       0.707349
WSF1       0.846509
WSF2       0.566521
WSF5       0.567420
WSFG       0.448280
WSFM       0.666410
WT01       0.619610
WT02       0.947382
WT03       0.953414
WT04       0.978097
WT05       0.982503
WT06       0.987209
WT07       0.997348
WT08       0.818489
WT09       0.990375
WT10       0.999957
WT11       0.999187
WT13       0.900368
WT14       0.947040
WT15       0.997048
WT16       0.645106
WT17       0.996920
WT18       0.896176
WT19       0.998374
WT21       0.999914
WT22       0.996064
dtype: float64
```

```
core_weather = weather[["PRCP", "SNOW", "SNWD", "TMAX", "TMIN"]].copy()
```

core\_weather



	PRCP	SNOW	SNWD	TMAX	TMIN
DATE					
1960-01-01	0.00	0.0	0.0	37	24
1960-01-02	0.00	0.0	0.0	47	23
1960-01-03	0.92	0.0	0.0	57	37
1960-01-04	0.00	0.0	0.0	41	33
1960-01-05	0.00	0.0	0.0	37	27
...	...	...	...	...	...
2023-12-27	0.02	0.0	NaN	53	41
2023-12-28	0.56	0.0	NaN	45	42
2023-12-29	0.02	0.0	NaN	46	40
2023-12-30	0.09	0.0	NaN	46	35
2023-12-31	0.00	0.0	NaN	41	33

23376 rows x 5 columns

Next steps:

[Generate code with core\\_weather](#)

 [View recommended plots](#)

```
core_weather.columns = ["prcp", "snow", "snwd", "tmax", "tmin"]
```

core\_weather

	prcp	snow	snwd	tmax	tmin
DATE					
1960-01-01	0.00	0.0	0.0	37	24
1960-01-02	0.00	0.0	0.0	47	23
1960-01-03	0.92	0.0	0.0	57	37
1960-01-04	0.00	0.0	0.0	41	33
1960-01-05	0.00	0.0	0.0	37	27
...	...	...	...	...	...
2023-12-27	0.02	0.0	NaN	53	41
2023-12-28	0.56	0.0	NaN	45	42
2023-12-29	0.02	0.0	NaN	46	40
2023-12-30	0.09	0.0	NaN	46	35
2023-12-31	0.00	0.0	NaN	41	33

23376 rows × 5 columns

Next steps: [Generate code with core\\_weather](#) [View recommended plots](#)

```
core_weather.apply(pd.isnull).sum()/weather.shape[0]
```

prcp	0.000086
snow	0.000000
snwd	0.342916
tmax	0.000000
tmin	0.000000
dtype:	float64

```
core_weather["snow"].value_counts()
```

snow	
0.0	21916
0.1	176
0.2	134
0.3	123
0.4	79
...	
7.7	1
8.1	1
22.4	1
7.5	1
11.7	1

Name: count, Length: 128, dtype: int64

```
core_weather[pd.isnull(core_weather["prcp"])]
# here we index the core weather by isnull function, this will only select the rows where prcp is null
```

	prcp	snow	snwd	tmax	tmin
DATE					
2003-08-23	NaN	0.0	0.0	84	65
2003-08-24	NaN	0.0	0.0	76	60



```
core_weather["prcp"] = core_weather['prcp'].fillna(0)
```

```
core_weather["snwd"].value_counts()
```


snwd	
0.0	13755
1.0	385
2.0	268
3.0	228
4.0	131
5.0	111
6.0	95
7.0	83
8.0	62
9.0	42
12.0	42
10.0	33
11.0	21
13.0	16
14.0	14

```
15.0      12
18.0      12
16.0      11
17.0      11
19.0       9
20.0       4
22.0       3
29.0       3
26.0       2
25.0       2
21.0       2
24.0       1
30.0       1
31.0       1
Name: count, dtype: int64
```

```
core_weather[pd.isnull(core_weather["snwd"])]
```




DATE	prcp	snow	snwd	tmax	tmin
1996-04-01	0.00	0.0	NaN	55	34
1996-04-02	0.54	0.0	NaN	48	37
1996-04-03	0.00	0.0	NaN	50	35
1996-04-04	0.00	0.0	NaN	54	32
1996-04-05	0.00	0.0	NaN	44	39
...	...	...	...	...	...
2023-12-27	0.02	0.0	NaN	53	41
2023-12-28	0.56	0.0	NaN	45	42
2023-12-29	0.02	0.0	NaN	46	40
2023-12-30	0.09	0.0	NaN	46	35
2023-12-31	0.00	0.0	NaN	41	33



8016 rows x 5 columns


```
core_weather["snwd"] = core_weather['snwd'].fillna(0)
```

```
core_weather.apply(pd.isnull).sum()/weather.shape[0]
```




```
prcp      0.0
snow      0.0
snwd      0.0
tmax      0.0
tmin      0.0
dtype: float64
```

```
core_weather.dtypes
```



```
prcp      float64
snow      float64
snwd      float64
tmax      int64
tmin      int64
dtype: object
```


```
core_weather.index
```



```
Index(['1960-01-01', '1960-01-02', '1960-01-03', '1960-01-04', '1960-01-05',
      '1960-01-06', '1960-01-07', '1960-01-08', '1960-01-09', '1960-01-10',
      ...,
      '2023-12-22', '2023-12-23', '2023-12-24', '2023-12-25', '2023-12-26',
      '2023-12-27', '2023-12-28', '2023-12-29', '2023-12-30', '2023-12-31'],
      dtype='object', name='DATE', length=23376)
```

```
core_weather.index = pd.to_datetime(core_weather.index)
```

```
core_weather.index
```



```
DatetimeIndex(['1960-01-01', '1960-01-02', '1960-01-03', '1960-01-04',
              '1960-01-05', '1960-01-06', '1960-01-07', '1960-01-08',
              '1960-01-09', '1960-01-10',
              ...,
              '2023-12-22', '2023-12-23', '2023-12-24', '2023-12-25',
              '2023-12-26', '2023-12-27', '2023-12-28', '2023-12-29',
              '2023-12-30', '2023-12-31'],
              dtype='datetime64[ns]', freq='D')
```

```
'2023-12-30', '2023-12-31'],
dtype='datetime64[ns]', name='DATE', length=23376, freq=None)
```

```
core_weather.apply(lambda x: (x==9999).sum())
```

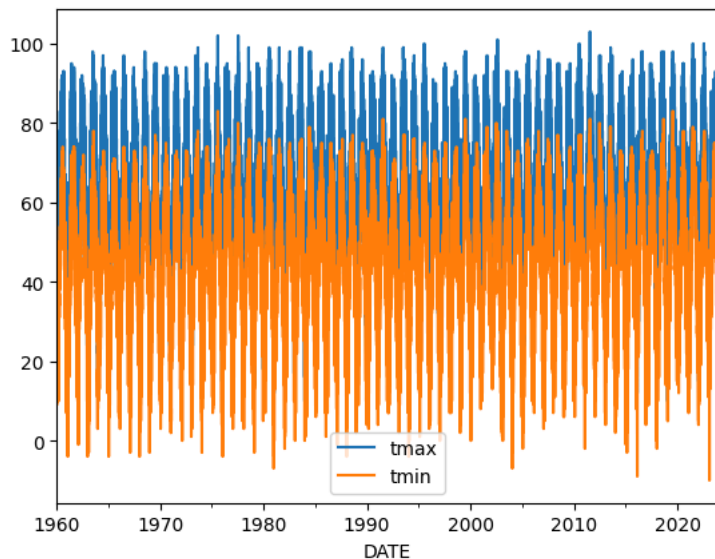
```
prcp    0
snow    0
snowd   0
tmax    0
tmin    0
dtype: int64
```

```
core_weather.index.year.value_counts().sort_index()
```

```
DATE
1960    366
1961    365
1962    365
1963    365
1964    366
...
2019    365
2020    366
2021    365
2022    365
2023    365
Name: count, Length: 64, dtype: int64
```

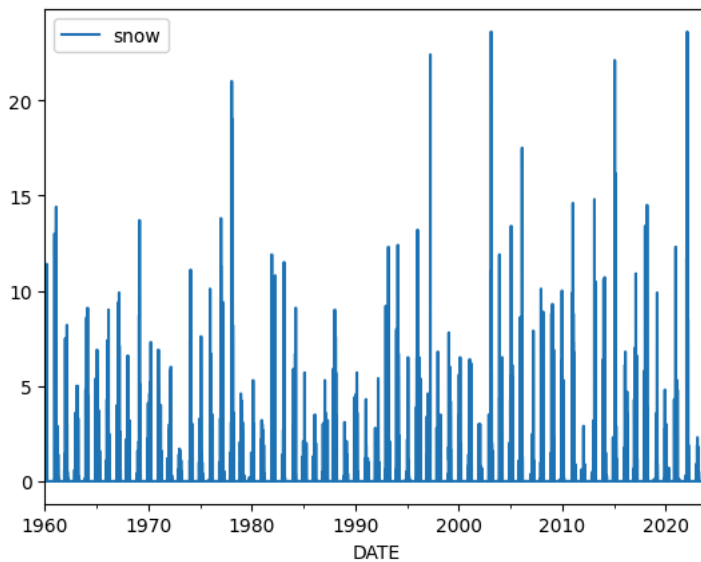
```
core_weather[["tmax", "tmin"]].plot()
```

```
<Axes: xlabel='DATE'>
```



```
core_weather[["snow"]].plot()
```

```
<Axes: xlabel='DATE'>
```



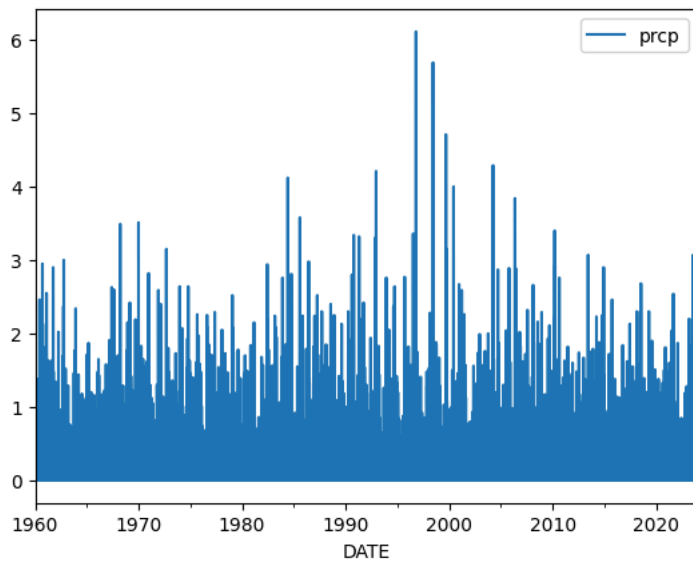


```
core_weather.groupby(core_weather.index.year).sum()['snow']
```

```
↵ DATE
1960    51.7
1961    56.9
1962    38.6
1963    42.4
1964    57.5
1965    40.5
1966    51.7
1967    67.1
1968    33.4
1969    60.9
1970    64.1
1971    40.1
1972    40.7
1973     6.4
1974    42.5
1975    41.4
1976    45.4
1977    46.2
1978    89.2
1979    19.7
1980    18.5
1981    31.9
1982    49.7
1983    29.8
1984    44.1
1985    27.2
1986    20.7
1987    52.1
1988    39.8
1989    22.5
1990    29.7
1991    23.7
1992    26.5
1993    85.2
1994    86.3
1995    41.5
1996    86.2
1997    56.8
1998    14.7
1999    35.6
2000    29.4
2001    46.4
2002    24.8
2003    77.7
2004    29.0
2005    87.3
2006    28.9
2007    43.2
2008    49.6
2009    55.9
2010    42.4
2011    60.0
2012    12.1
2013    71.3
2014    50.1
2015   108.6
2016    41.1
2017     5.0
```

```
core_weather[["prcp"]].plot()
```

```
<Axes: xlabel='DATE'>
```



```
core_weather.groupby(core_weather.index.year).sum()['prcp']
```

```

1967    47.60
1968    42.28
1969    47.78
1970    41.91
1971    35.67
1972    53.11
1973    42.75
1974    40.24
1975    45.79
1976    36.72
1977    44.17
1978    37.64
1979    44.17
1980    29.39
1981    35.71
1982    44.61
1983    53.60
1984    50.24
1985    36.59
1986    44.33
1987    45.48
1988    34.78
1989    42.42
1990    46.50
1991    42.25
1992    43.72
1993    43.21
1994    47.62
1995    35.10
1996    52.52
1997    30.39
1998    53.67
1999    37.91
2000    45.60
2001    30.72
2002    41.07
2003    44.37
2004    44.57
2005    43.67
2006    52.89
2007    39.47
2008    54.46
2009    43.49
2010    49.66
2011    52.39
2012    36.73
2013    40.36
2014    45.25
2015    34.80
2016    33.05
2017    43.45
2018    53.32
2019    50.38
2020    37.54
2021    52.33
2022    31.25
2023    49.64
Name: prcp, dtype: float64

```

```
core_weather["target"] = core_weather.shift(-1) ["tmax"]
```

```
core_weather
```



2023-11-30	0.00	0.0	0.0	47	30	53.0
2023-12-01	0.06	0.0	0.0	53	41	50.0
2023-12-02	0.00	0.0	0.0	50	43	46.0
2023-12-03	0.79	0.0	0.0	46	42	48.0
2023-12-04	0.11	0.0	0.0	48	39	43.0
2023-12-05	0.00	0.0	0.0	43	31	35.0
2023-12-06	0.00	0.2	0.0	35	25	34.0
2023-12-07	0.00	0.0	0.0	34	21	42.0
2023-12-08	0.00	0.0	0.0	42	27	52.0
2023-12-09	0.00	0.0	0.0	52	33	62.0
2023-12-10	0.52	0.0	0.0	62	44	64.0
2023-12-11	1.75	0.0	0.0	64	37	44.0
2023-12-12	0.00	0.0	0.0	44	34	45.0
2023-12-13	0.00	0.0	0.0	45	32	36.0
2023-12-14	0.00	0.0	0.0	36	28	56.0
2023-12-15	0.00	0.0	0.0	56	32	51.0
2023-12-16	0.00	0.0	0.0	51	40	56.0
2023-12-17	0.37	0.0	0.0	56	40	63.0
2023-12-18	1.68	0.0	0.0	63	45	48.0
2023-12-19	0.00	0.0	0.0	48	35	44.0
2023-12-20	0.00	0.0	0.0	44	30	36.0
2023-12-21	0.00	0.0	0.0	36	24	37.0
2023-12-22	0.00	0.0	0.0	37	21	42.0
2023-12-23	0.00	0.0	0.0	42	24	45.0
2023-12-24	0.01	0.0	0.0	45	37	49.0
2023-12-25	0.00	0.0	0.0	49	38	49.0
2023-12-26	0.00	0.0	0.0	49	37	53.0
2023-12-27	0.02	0.0	0.0	53	41	45.0
2023-12-28	0.56	0.0	0.0	45	42	46.0
2023-12-29	0.02	0.0	0.0	46	40	46.0
2023-12-30	0.09	0.0	0.0	46	35	41.0
2023-12-31	0.00	0.0	0.0	41	33	NaN

