Handling Missing Data

```
import pandas as pd
import numpy as np
df = pd.read_csv("weather_data (4).csv", parse_dates=["day"])
df.set_index('day',inplace=True)
df
```

day	temperature	windspeed	event
2017-01-01	32.0	6.0	Rain
2017-01-04	NaN	9.0	Sunny
2017-01-05	28.0	NaN	Snow
2017-01-06	NaN	7.0	NaN
2017-01-07	32.0	NaN	Rain
2017-01-08	NaN	NaN	Sunny
2017-01-09	NaN	NaN	NaN
2017-01-10	34.0	8.0	Cloudy
2017-01-11	40.0	12.0	Sunny

Replace method

```
new_df = df.replace(-99999, value=np.NaN)
new_df
```

day	temperature	windspeed	event
2017-01-01	32.0	6.0	Rain
2017-01-02	NaN	7.0	Sunny
2017-01-03	28.0	NaN	Snow
2017-01-04	NaN	7.0	0
2017-01-05	32.0	NaN	Rain
2017-01-06	31.0	2.0	Sunny
2017-01-06	34.0	5.0	0

```
new_df = df.replace({
        'temperature': -99999,
        'windspeed': -99999,
        'event': '0'
    }, np.nan)
new_df
```

day	temperature	windspeed	event
2017-01-01	32.0	6.0	Rain
2017-01-02	NaN	7.0	Sunny
2017-01-03	28.0	NaN	Snow
2017-01-04	NaN	7.0	NaN

```
2017-01-05         32.0        NaN   Rain
2017-01-06         31.0        2.0  Sunny
2017-01-06         34.0        5.0    NaN
```

Interpolate Method

```
new_df = df.interpolate()
new_df
```

```
            temperature  windspeed  event
day
2017-01-01           32          6   Rain
2017-01-02       -99999          7  Sunny
2017-01-03           28     -99999   Snow
2017-01-04       -99999          7      0
2017-01-05           32     -99999   Rain
2017-01-06           31          2  Sunny
2017-01-06           34          5      0
```

fillna Method

```
new_df = df.fillna(0)
new_df
```

```
            temperature  windspeed  event
day
2017-01-01           32          6   Rain
2017-01-02       -99999          7  Sunny
2017-01-03           28     -99999   Snow
2017-01-04       -99999          7      0
2017-01-05           32     -99999   Rain
2017-01-06           31          2  Sunny
2017-01-06           34          5      0
```

```
new_df = df.fillna(method="ffill",limit=1)
new_df
```

```
            temperature  windspeed  event
day
2017-01-01           32          6   Rain
2017-01-02       -99999          7  Sunny
2017-01-03           28     -99999   Snow
2017-01-04       -99999          7      0
2017-01-05           32     -99999   Rain
2017-01-06           31          2  Sunny
2017-01-06           34          5      0
```

```
new_df = df.fillna(method="bfill", axis="columns") # axis is either
"index" or "columns"
new_df
```

```
            temperature windspeed  event
day
2017-01-01           32         6   Rain
2017-01-02       -99999         7  Sunny
2017-01-03           28    -99999   Snow
2017-01-04       -99999         7      0
2017-01-05           32    -99999   Rain
2017-01-06           31         2  Sunny
2017-01-06           34         5      0
```

dropna method

```
new_df = df.dropna(thresh=2)
new_df
```

```
             temperature  windspeed  event
day
2017-01-01           32          6   Rain
2017-01-02       -99999          7  Sunny
2017-01-03           28     -99999   Snow
2017-01-04       -99999          7      0
2017-01-05           32     -99999   Rain
2017-01-06           31          2  Sunny
2017-01-06           34          5      0
```

```
new_df = df.dropna()
new_df
```

```
             temperature  windspeed  event
day
2017-01-01           32          6   Rain
2017-01-02       -99999          7  Sunny
2017-01-03           28     -99999   Snow
2017-01-04       -99999          7      0
2017-01-05           32     -99999   Rain
2017-01-06           31          2  Sunny
2017-01-06           34          5      0
```

Inserting Missing Dates

```
dt = pd.date_range("01-01-2017","01-11-2017")
idx = pd.DatetimeIndex(dt)
df.reindex(idx)
```

```
             temperature  windspeed   event
2017-01-01         32.0        6.0    Rain
2017-01-02          NaN        NaN     NaN
2017-01-03          NaN        NaN     NaN
2017-01-04          NaN        9.0   Sunny
2017-01-05         28.0        NaN    Snow
```

```
2017-01-06          NaN         7.0     NaN
2017-01-07         32.0         NaN    Rain
2017-01-08          NaN         NaN   Sunny
2017-01-09          NaN         NaN     NaN
2017-01-10         34.0         8.0  Cloudy
2017-01-11         40.0        12.0   Sunny
```

```
df.isnull()
```

```
            temperature  windspeed  event
day
2017-01-01        False      False  False
2017-01-04         True      False  False
2017-01-05        False       True  False
2017-01-06         True      False   True
2017-01-07        False       True  False
2017-01-08         True       True  False
2017-01-09         True       True   True
2017-01-10        False      False  False
2017-01-11        False      False  False
```

```
df.notnull()
```

```
            temperature  windspeed  event
day
2017-01-01         True       True   True
2017-01-04        False       True   True
2017-01-05         True      False   True
2017-01-06        False       True  False
2017-01-07         True      False   True
2017-01-08        False      False   True
2017-01-09        False      False  False
2017-01-10         True       True   True
2017-01-11         True       True   True
```