1) Filna to fill missing values using different ways

2) Interpolate to make a guess on missing values using interpolation

3) Dropna to drop rows with missing values

In [14]:
```
import pandas as pd
df = pd.read_csv("C:/Users/prasa/Desktop/ds projects/panda/weather_data5.csv",parse_dates=["day"])
df.set_index('day', inplace=True) #change date from string formate
df
```

Out[14]:

| 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 |

In [15]:
```
new_df=df.fillna(0)
new_df
```

Out[15]:

| | temperature | windspeed | event |
|---|---|---|---|

| day | temperature | windspeed | event |
|---|---|---|---|
| **day** | | | |
| **2017-01-01** | 32.0 | 6.0 | Rain |
| **2017-01-04** | 0.0 | 9.0 | Sunny |
| **2017-01-05** | 28.0 | 0.0 | Snow |
| **2017-01-06** | 0.0 | 7.0 | 0 |
| **2017-01-07** | 32.0 | 0.0 | Rain |
| **2017-01-08** | 0.0 | 0.0 | Sunny |
| **2017-01-09** | 0.0 | 0.0 | 0 |
| **2017-01-10** | 34.0 | 8.0 | Cloudy |
| **2017-01-11** | 40.0 | 12.0 | Sunny |

In [20]:
```
new_df = df.fillna({
    'temperature': 0,
    'windspeed': 0,
    'event' : 'no event'
})
new_df
```

Out[20]:

| | temperature | windspeed | event |
|---|---|---|---|
| **day** | | | |
| **2017-01-01** | 32.0 | 6.0 | Rain |
| **2017-01-04** | 0.0 | 9.0 | Sunny |
| **2017-01-05** | 28.0 | 0.0 | Snow |
| **2017-01-06** | 0.0 | 7.0 | no event |
| **2017-01-07** | 32.0 | 0.0 | Rain |
| **2017-01-08** | 0.0 | 0.0 | Sunny |
| **2017-01-09** | 0.0 | 0.0 | no event |

| day | temperature | windspeed | event |
|---|---|---|---|
| **2017-01-10** | 34.0 | 8.0 | Cloudy |
| **2017-01-11** | 40.0 | 12.0 | Sunny |

In [23]:
```
new_df = df.fillna(method='ffill' ) #'bfill' axis='columns' copy vertically
new_df
```

Out[23]:

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

Pandas fill na

In [27]:
```
new_df = df.fillna(method='ffill', limit=1 )
new_df
```

Out[27]:

| day | temperature | windspeed | event |
|---|---|---|---|

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

In [29]:
```
new_df = df.interpolate(method="time")
new_df
```

Out[29]:

| day | temperature | windspeed | event |
|---|---|---|---|
| **2017-01-01** | 32.000000 | 6.00 | Rain |
| **2017-01-04** | 29.000000 | 9.00 | Sunny |
| **2017-01-05** | 28.000000 | 8.00 | Snow |
| **2017-01-06** | 30.000000 | 7.00 | NaN |
| **2017-01-07** | 32.000000 | 7.25 | Rain |
| **2017-01-08** | 32.666667 | 7.50 | Sunny |
| **2017-01-09** | 33.333333 | 7.75 | NaN |
| **2017-01-10** | 34.000000 | 8.00 | Cloudy |
| **2017-01-11** | 40.000000 | 12.00 | Sunny |

In [31]:
```
new_df = df.dropna(how="all")
new_df
```

Out[31]:

| 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 |
| **2017-01-10** | 34.0 | 8.0 | Cloudy |
| **2017-01-11** | 40.0 | 12.0 | Sunny |

In [36]:
```
new_df = df.dropna(thresh=2)
new_df
```

Out[36]:

| 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 |
| **2017-01-07** | 32.0 | NaN | Rain |
| **2017-01-10** | 34.0 | 8.0 | Cloudy |
| **2017-01-11** | 40.0 | 12.0 | Sunny |

In [39]:
```
dt = pd.date_range("01-01-2017","01-11-2017")
idx = pd.DatetimeIndex(dt)
```

```
df=df.reindex(idx)
df
```

Out[39]:

| | 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 |