In [25]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [34]:

```
weather = pd.read_csv("weatherAUS.csv")
# netflix dataset from www.kaggle.com
weather.head(5)
```

Out[34]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGu
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	

5 rows × 23 columns

In [27]:

```
weather.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145460 entries, 0 to 145459
Data columns (total 23 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   Date           145460 non-null  object 
 1   Location       145460 non-null  object 
 2   MinTemp        143975 non-null  float64
 3   MaxTemp        144199 non-null  float64
 4   Rainfall       142199 non-null  float64
 5   Evaporation    82670 non-null   float64
 6   Sunshine       75625 non-null   float64
 7   WindGustDir    135134 non-null  object 
 8   WindGustSpeed  135197 non-null  float64
 9   WindDir9am     134894 non-null  object 
 10  WindDir3pm     141232 non-null  object 
 11  WindSpeed9am   143693 non-null  float64
 12  WindSpeed3pm   142398 non-null  float64
 13  Humidity9am    142806 non-null  float64
 14  Humidity3pm    140953 non-null  float64
 15  Pressure9am    130395 non-null  float64
 16  Pressure3pm    130432 non-null  float64
 17  Cloud9am       89572 non-null   float64
 18  Cloud3pm       86102 non-null   float64
 19  Temp9am        143693 non-null  float64
 20  Temp3pm        141851 non-null  float64
 21  RainToday      142199 non-null  object 
 22  RainTomorrow   142193 non-null  object 
dtypes: float64(16), object(7)
memory usage: 25.5+ MB
```

In [28]:

```
weather.describe(include='all')
```

Out[28]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGu
count	145460	145460	143975.000000	144199.000000	142199.000000	82670.000000	75625.000000	1
unique	3436	49	NaN	NaN	NaN	NaN	NaN	
top	2013-11-12	Canberra	NaN	NaN	NaN	NaN	NaN	
freq	49	3436	NaN	NaN	NaN	NaN	NaN	
mean	NaN	NaN	12.194034	23.221348	2.360918	5.468232	7.611178	
std	NaN	NaN	6.398495	7.119049	8.478060	4.193704	3.785483	
min	NaN	NaN	-8.500000	-4.800000	0.000000	0.000000	0.000000	
25%	NaN	NaN	7.600000	17.900000	0.000000	2.600000	4.800000	
50%	NaN	NaN	12.000000	22.600000	0.000000	4.800000	8.400000	

In [33]:

```
print(weather.std())
print('-------------------------------------')
print(weather.median())
```

```
MinTemp           6.398495
MaxTemp           7.119049
Rainfall          8.478060
Evaporation       4.193704
Sunshine          3.785483
WindGustSpeed    13.607062
WindSpeed9am      8.915375
WindSpeed3pm      8.809800
Humidity9am      19.029164
Humidity3pm      20.795902
Pressure9am       7.106530
Pressure3pm       7.037414
Cloud9am          2.887159
Cloud3pm          2.720357
Temp9am           6.488753
Temp3pm           6.936650
dtype: float64
-------------------------------------
MinTemp           12.0
MaxTemp           22.6
Rainfall           0.0
Evaporation        4.8
Sunshine           8.4
WindGustSpeed     39.0
WindSpeed9am      13.0
WindSpeed3pm      19.0
Humidity9am       70.0
Humidity3pm       52.0
Pressure9am     1017.6
Pressure3pm     1015.2
Cloud9am           5.0
Cloud3pm           5.0
Temp9am           16.7
Temp3pm           21.1
dtype: float64

C:\Users\gv7lo\AppData\Local\Temp/ipykernel_1520/3980635267.py:1: FutureWarn
ing: Dropping of nuisance columns in DataFrame reductions (with 'numeric_onl
y=None') is deprecated; in a future version this will raise TypeError.  Sele
ct only valid columns before calling the reduction.
  print(weather.std())
C:\Users\gv7lo\AppData\Local\Temp/ipykernel_1520/3980635267.py:3: FutureWarn
ing: Dropping of nuisance columns in DataFrame reductions (with 'numeric_onl
y=None') is deprecated; in a future version this will raise TypeError.  Sele
ct only valid columns before calling the reduction.
  print(weather.median())
```

In [30]:

```
weather.shape
```

Out[30]:

```
(145460, 23)
```

In [31]:

```
weather.isnull().sum()
```

Out[31]:

```
Date                0
Location            0
MinTemp          1485
MaxTemp          1261
Rainfall         3261
Evaporation     62790
Sunshine        69835
WindGustDir     10326
WindGustSpeed   10263
WindDir9am      10566
WindDir3pm       4228
WindSpeed9am     1767
WindSpeed3pm     3062
Humidity9am      2654
Humidity3pm      4507
Pressure9am     15065
Pressure3pm     15028
Cloud9am        55888
Cloud3pm        59358
Temp9am          1767
Temp3pm          3609
RainToday        3261
RainTomorrow     3267
dtype: int64
```