```
import pandas as pd
import numpy as np
```

```
# dictionary of lists
dict1 = {'Maths':[100, 85, np.nan, 90],'Science': [40, 55, 80, np.nan],'Social Science':[np.nan, 5(
# creating a dataframe from list
df = pd.DataFrame(dict1)
df
```

	Maths	Science	Social Science
0	100.0	40.0	NaN
1	85.0	55.0	50.0
2	NaN	80.0	70.0
3	90.0	NaN	98.0

```
# using isnull() function
df.isnull()
```

	Maths	Science	Social Science
0	False	False	True
1	False	False	False
2	True	False	False
3	False	True	False

```
# using notnull() function
df.notnull()
```

	Maths	Science	Social Science
0	True	True	False
1	True	True	True
2	False	True	True
3	True	False	True

```
df.isnull().sum()
```

```
Maths            1
Science          1
Social Science   1
dtype: int64
```

**Handling missing values

```
df.fillna(50)
```

	Maths	Science	Social Science
0	100.0	40.0	50.0
1	85.0	55.0	50.0
2	50.0	80.0	70.0
3	90.0	50.0	98.0

```
#Dropping missing values using dropna():
# dictionary of lists
dict1 = {'Maths':[100, 85, np.nan, 90],'Science': [40, 55, 80, np.nan],'Social Science':[np.nan, 5(
# creating a dataframe from list
df = pd.DataFrame(dict1)
print(df)
# dropping missing value using droppingna()
df.dropna()
```

```
   Maths  Science  Social Science
0  100.0     40.0             NaN
1   85.0     55.0            50.0
2    NaN     80.0            70.0
3   90.0      NaN            98.0
```

	Maths	Science	Social Science
1	85.0	55.0	50.0

```
df['Maths'].dropna()
```

```
0    100.0
1     85.0
3     90.0
Name: Maths, dtype: float64
```

```
df['Maths'].replace(np.nan,50,inplace=True)
df
```

	Maths	Science	Social Science
0	100.0	40.0	NaN
1	85.0	55.0	50.0
2	50.0	80.0	70.0
3	90.0	NaN	98.0

```
m=round(df['Science'].mean(),2)
df['Science'].fillna(m,inplace=True)
df
```

	Maths	Science	Social Science
0	100.0	40.00	NaN
1	85.0	55.00	50.0
2	NaN	80.00	70.0
3	90.0	58.33	98.0

```
df1=pd.read_csv('/content/weather_na.csv')
df2=df1.copy()
df2
```

```
df2.fillna(0, inplace=True)
df2.describe()
```

	avg_low	avg_high	record_high	record_low	avg_preci
count	12.000000	12.00000	12.000000	12.000000	12.000000
mean	62.166667	46.50000	76.750000	32.416667	1.265833
std	20.748859	15.82001	37.358521	8.240238	1.186396
min	0.000000	0.00000	0.000000	21.000000	0.000000
25%	60.250000	44.25000	77.000000	25.750000	0.155000
50%	66.000000	49.00000	88.000000	32.000000	0.915000
75%	73.500000	56.25000	99.250000	40.250000	2.395000
max	77.000000	59.00000	107.000000	44.000000	3.020000

```
df1.describe()
```

	avg_low	avg_high	record_high	record_low	avg_preci
count	11.000000	11.000000	10.000000	12.000000	12.000000
mean	67.818182	50.727273	92.100000	32.416667	1.265833
std	7.208581	6.278390	11.618472	8.240238	1.186396
min	58.000000	42.000000	74.000000	21.000000	0.000000
25%	62.500000	46.500000	84.000000	25.750000	0.155000
50%	67.000000	50.000000	94.000000	32.000000	0.915000
75%	74.000000	56.500000	101.750000	40.250000	2.395000
max	77.000000	59.000000	107.000000	44.000000	3.020000

```
df2
df2.describe()
```

	avg_low	avg_high	record_high	record_low	avg_preci
count	11.000000	11.000000	10.000000	12.000000	12.000000
std	7.208581	6.278590	11.618472	8.240256	1.186596

Write the python code to process all NA values in the given weather dataset and analyze its statistical measures.

Count the number of NA values in each attribute.

Drop all the NA values. How many records are retained?

Replace NA values in ‘record_high’ with its mean value

Replace NA values in avg_low with its minimum value.

Replace NA values in avg_high with its maximum value.

After replacing, compare its summary statistics with the original dataset.

Double-click (or enter) to edit

Double-click (or enter) to edit