

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('rainfall.csv')

In [2]: df.head()
Out[2]:
   day  pressure  maxtemp  mintemp  dewpoint  humidity  cloud  rainfall  sunshine  winddirection  windspeed
0    1    1025.9    19.9    18.3    16.8    13.1    72    49    9.3    80.0    26.3
1    2    1022.0    21.7    18.9    17.2    15.6    81    83    yes    0.6    50.0    15.3
2    3    1019.7    20.3    19.3    18.0    18.4    95    91    yes    1.0    40.0    14.2
3    4    1018.9    22.3    20.6    19.1    18.8    90    88    yes    1.0    50.0    16.9
4    5    1015.9    21.3    20.7    20.2    19.9    95    81    yes    1.0    40.0    13.7

In [3]: df['rainfall'] = df['rainfall'].replace({'yes':1,'no':0})
Out[3]:
   day  pressure  maxtemp  mintemp  dewpoint  humidity  cloud  rainfall  sunshine  winddirection  windspeed
0    1    1025.9    19.9    18.3    16.8    13.1    72    49    1    9.3    80.0    26.3
1    2    1022.0    21.7    18.9    17.2    15.6    81    83    1    0.6    50.0    15.3
2    3    1019.7    20.3    19.3    18.0    18.4    95    91    1    1.0    40.0    14.2
3    4    1018.9    22.3    20.6    19.1    18.8    90    88    1    1.0    50.0    16.9
4    5    1015.9    21.3    20.7    20.2    19.9    95    81    1    1.0    40.0    13.7

In [5]: df['rainfall'].value_counts()
Out[5]:
1    249
0    117
Name: rainfall, dtype: int64

In [6]: df.info()
Out[6]:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ---
 0   day                 366 non-null    int64
 1   pressure            366 non-null    float64
 2   maxtemp             366 non-null    float64
 3   temperature         366 non-null    float64
 4   mintemp             366 non-null    float64
 5   dewpoint            366 non-null    float64
 6   humidity            366 non-null    float64
 7   cloud              366 non-null    int64
 8   rainfall            366 non-null    float64
 9   sunshine            366 non-null    float64
10  winddirection       366 non-null    int64(4)
11  windspeed           366 non-null    float64(48), int64(4)
memory usage: 34.4 KB

In [7]: df.drops()
Out[7]:
   day  pressure  maxtemp  temperature  mintemp  dewpoint  humidity  cloud  rainfall  sunshine  winddirection  windspeed
0    1    1025.9    19.9    18.3    16.8    13.1    72    49    1    9.3    80.0    26.3
1    2    1022.0    21.7    18.9    17.2    15.6    81    83    1    0.6    50.0    15.3
2    3    1019.7    20.3    19.3    18.0    18.4    95    91    1    1.0    40.0    14.2
3    4    1018.9    22.3    20.6    19.1    18.8    90    88    1    1.0    50.0    16.9
4    5    1015.9    21.3    20.7    20.2    19.9    95    81    1    1.0    40.0    13.7

...

361  27    1022.7    18.8    17.7    16.9    15.0    84    90    1    0.0    30.0    18.4
362  28    1026.6    18.6    17.3    16.3    12.8    75    85    1    1.0    20.0    25.9
363  29    1025.9    18.9    17.7    16.4    13.3    75    78    1    4.6    70.0    33.4
364  30    1025.3    19.2    17.3    15.2    13.3    78    86    1    1.2    20.0    20.9
365  31    1026.4    20.5    17.8    15.5    13.0    74    66    0    5.7    20.0    23.3

365 rows x 12 columns

In [8]: df.info()
Out[8]:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ---
 0   day                 366 non-null    int64
 1   pressure            366 non-null    float64
 2   maxtemp             366 non-null    float64
 3   temperature         366 non-null    float64
 4   mintemp             366 non-null    float64
 5   dewpoint            366 non-null    float64
 6   humidity            366 non-null    float64
 7   cloud              366 non-null    int64
 8   rainfall            366 non-null    float64(48), int64(4)
 9   sunshine            366 non-null    float64
10  winddirection       366 non-null    int64(4)
11  windspeed           366 non-null    float64(48), int64(4)
memory usage: 34.4 KB

In [9]: df.isnull().sum()
Out[9]:
day                0
pressure           0
maxtemp            0
temperature         0
mintemp            0
dewpoint           0
humidity           0
cloud              0
rainfall           0
sunshine           0
winddirection      1
windspeed          1
dtype: int64

In [10]: df.drops(inplace = True)
Out[10]:
day                0
pressure           0
maxtemp            0
temperature         0
mintemp            0
dewpoint           0
humidity           0
cloud              0
rainfall           0
sunshine           0
winddirection      0
windspeed          0
dtype: int64

In [12]: df.shape
Out[12]:
(365, 12)

In [13]: df.describe()
Out[13]:
<bound method NDFrame.describe of
   day  pressure  maxtemp  temperature  mintemp  dewpoint  humidity  \
0    1    1025.9    19.9    18.3    16.8    13.1    72
1    2    1022.0    21.7    18.9    17.2    15.6    81
2    3    1019.7    20.3    19.3    18.0    18.4    95
3    4    1018.9    22.3    20.6    19.1    18.8    90
4    5    1015.9    21.3    20.7    20.2    19.9    95

...

361  27    1022.7    18.8    17.7    16.9    15.0    84
362  28    1026.6    18.6    17.3    16.3    12.8    75
363  29    1025.9    18.9    17.7    16.4    13.3    75
364  30    1025.3    19.2    17.3    15.2    13.3    78
365  31    1026.4    20.5    17.8    15.5    13.0    74

...

cloud  rainfall  sunshine  winddirection  windspeed
0    49         1         9.3          80.0    26.3
1    83         1         0.6          50.0    15.3
2    91         1         1.0          40.0    14.2
3    88         1         1.0          50.0    16.9
4    81         1         1.0          40.0    13.7

...

361    90         1         0.0          30.0    18.4
362    85         1         1.0          20.0    25.9
363    78         1         4.6          70.0    33.4
364    86         1         1.2          20.0    20.9
365    66         0         5.7          20.0    23.3

[365 rows x 12 columns]>

In [52]: %pip install xgboost
Out[52]:
Defaulting to user installation because normal site-packages is not writeable
Collecting xgboost
  Downloading xgboost-2.1.2-py3-none-win_amd64.whl (124.9 MB)
    0.0/124.9 MB 9.8 MB/s eta 0:00:13
    1.3/124.9 MB 13.5 MB/s eta 0:00:09
    1.9/124.9 MB 13.6 MB/s eta 0:00:10
    2.6/124.9 MB 13.9 MB/s eta 0:00:09
    3.2/124.9 MB 9.1 MB/s eta 0:00:18
    3.9/124.9 MB 8.5 MB/s eta 0:00:15
    4.5/124.9 MB 8.2 MB/s eta 0:00:15
    5.2/124.9 MB 6.7 MB/s eta 0:00:18
    5.8/124.9 MB 6.9 MB/s eta 0:00:18
    6.5/124.9 MB 6.5 MB/s eta 0:00:18
    7.1/124.9 MB 7.5 MB/s eta 0:00:17
    7.8/124.9 MB 7.7 MB/s eta 0:00:16
    8.4/124.9 MB 7.4 MB/s eta 0:00:17
    9.1/124.9 MB 7.2 MB/s eta 0:00:18
    9.7/124.9 MB 6.9 MB/s eta 0:00:18
    10.4/124.9 MB 6.7 MB/s eta 0:00:18
    11.1/124.9 MB 6.5 MB/s eta 0:00:19
    11.7/124.9 MB 6.5 MB/s eta 0:00:19
    12.4/124.9 MB 6.7 MB/s eta 0:00:19
    13.1/124.9 MB 6.8 MB/s eta 0:00:18
    13.7/124.9 MB 6.8 MB/s eta 0:00:18
    14.4/124.9 MB 6.6 MB/s eta 0:00:18
    15.1/124.9 MB 6.6 MB/s eta 0:00:18
    15.7/124.9 MB 6.6 MB/s eta 0:00:18
    16.4/124.9 MB 6.6 MB/s eta 0:00:18
    17.1/124.9 MB 6.6 MB/s eta 0:00:18
    17.7/124.9 MB 6.7 MB/s eta 0:00:18
    18.4/124.9 MB 6.7 MB/s eta 0:00:18
    19.1/124.9 MB 6.7 MB/s eta 0:00:18
    19.7/124.9 MB 6.7 MB/s eta 0:00:18
    20.4/124.9 MB 6.7 MB/s eta 0:00:18
    21.1/124
```