Installing required Libraries and dependencies **bold text**

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.utils import resample
from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
import pickle
```

```
data = pd.read_csv(r"/content/Rainfall.csv")
```

```
data.head()
```

	day	pressure	maxtemp	temparature	mintemp	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspeed
0	1	1025.9	19.9	18.3	16.8	13.1	72	49	yes	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	0.0	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	0.0	40.0	13.7

Next steps: Generate code with `data` New interactive sheet

```
data.tail()
```

	day	pressure	maxtemp	temparature	mintemp	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspee
361	27	1022.7	18.8	17.7	16.9	15.0	84	90	yes	0.0	30.0	18.
362	28	1026.6	18.6	17.3	16.3	12.8	75	85	yes	1.0	20.0	25.
363	29	1025.9	18.9	17.7	16.4	13.3	75	78	yes	4.6	70.0	33.
364	30	1025.3	19.2	17.3	15.2	13.3	78	86	yes	1.2	20.0	20.
365	31	1026.4	20.5	17.8	15.5	13.0	74	66	no	5.7	20.0	23.

```
data.shape
```

```
(366, 12)
```

```
data["day"].unique()
```

```
array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31])
```

```
data.info()
```

```
<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   temparature            366 non-null    float64
 4   mintemp                366 non-null    float64
 5   dewpoint               366 non-null    float64
 6   humidity               366 non-null    int64
 7   cloud                  366 non-null    int64
 8   rainfall               366 non-null    object
 9   sunshine               366 non-null    float64
 10          winddirection  365 non-null    float64
 11  windspeed              365 non-null    float64
dtypes: float64(8), int64(3), object(1)
memory usage: 34.4+ KB
```

```
data.columns = data.columns.str.strip()
```

```
data = data.drop(columns = ["day"])
```

```
data.isnull().sum()
```

	0
pressure	0
maxtemp	0
temparature	0
mintemp	0
dewpoint	0
humidity	0
cloud	0
rainfall	0
sunshine	0
winddirection	1
windspeed	1

dtype: int64

```
data["winddirection"].unique()
```

```
array([ 80.,  50.,  40.,  20.,  30.,  60.,  70.,  10., 200., 220., 120.,
       190., 210., 300., 240., 180., 230.,  90., 170., 150., 100., 130.,
        nan, 160., 270., 280., 250., 260., 290., 350., 110., 140.])
```

```
data["winddirection"] = data["winddirection"].fillna(data["winddirection"].mode()[0])
```

```
data["windspeed"] = data["windspeed"].fillna(data["windspeed"].median())
```

```
data.isnull().sum().sum()
```

```
np.int64(0)
```

```
data["rainfall"].unique()
```

```
array(['yes', 'no'], dtype=object)
```

```
data["rainfall"] = data["rainfall"].map({"yes": 1, "no": 0})
```

```
data.head()
```

	pressure	maxtemp	temparature	mintemp	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspeed
0	1025.9	19.9	18.3	16.8	13.1	72	49	1	9.3	80.0	26.3
1	1022.0	21.7	18.9	17.2	15.6	81	83	1	0.6	50.0	15.3
2	1019.7	20.3	19.3	18.0	18.4	95	91	1	0.0	40.0	14.2
3	1018.9	22.3	20.6	19.1	18.8	90	88	1	1.0	50.0	16.9
4	1015.9	21.3	20.7	20.2	19.9	95	81	1	0.0	40.0	13.7

Next steps: Generate code with `data` New interactive sheet

```
sns.set(style = "whitegrid")
```

```
data.describe()
```

	pressure	maxtemp	temparature	mintemp	dewpoint	humidity	cloud	rainfall	sunshine	winddire
count	366.000000	366.000000	366.000000	366.000000	366.000000	366.000000	366.000000	366.000000	366.000000	366.0
mean	1013.742623	26.191257	23.747268	21.894536	19.989071	80.177596	71.128415	0.680328	4.419399	101.2
std	6.414776	5.978343	5.632813	5.594153	5.997021	10.062470	21.798012	0.466988	3.934398	81.7
min	998.500000	7.100000	4.900000	3.100000	-0.400000	36.000000	0.000000	0.000000	0.000000	10.0
25%	1008.500000	21.200000	18.825000	17.125000	16.125000	75.000000	58.000000	0.000000	0.500000	40.0
50%	1013.000000	27.750000	25.450000	23.700000	21.950000	80.500000	80.000000	1.000000	3.500000	70.0
75%	1018.100000	31.200000	28.600000	26.575000	25.000000	87.000000	88.000000	1.000000	8.200000	190.0
max	1034.600000	36.300000	32.400000	30.000000	26.700000	98.000000	100.000000	1.000000	12.100000	350.0

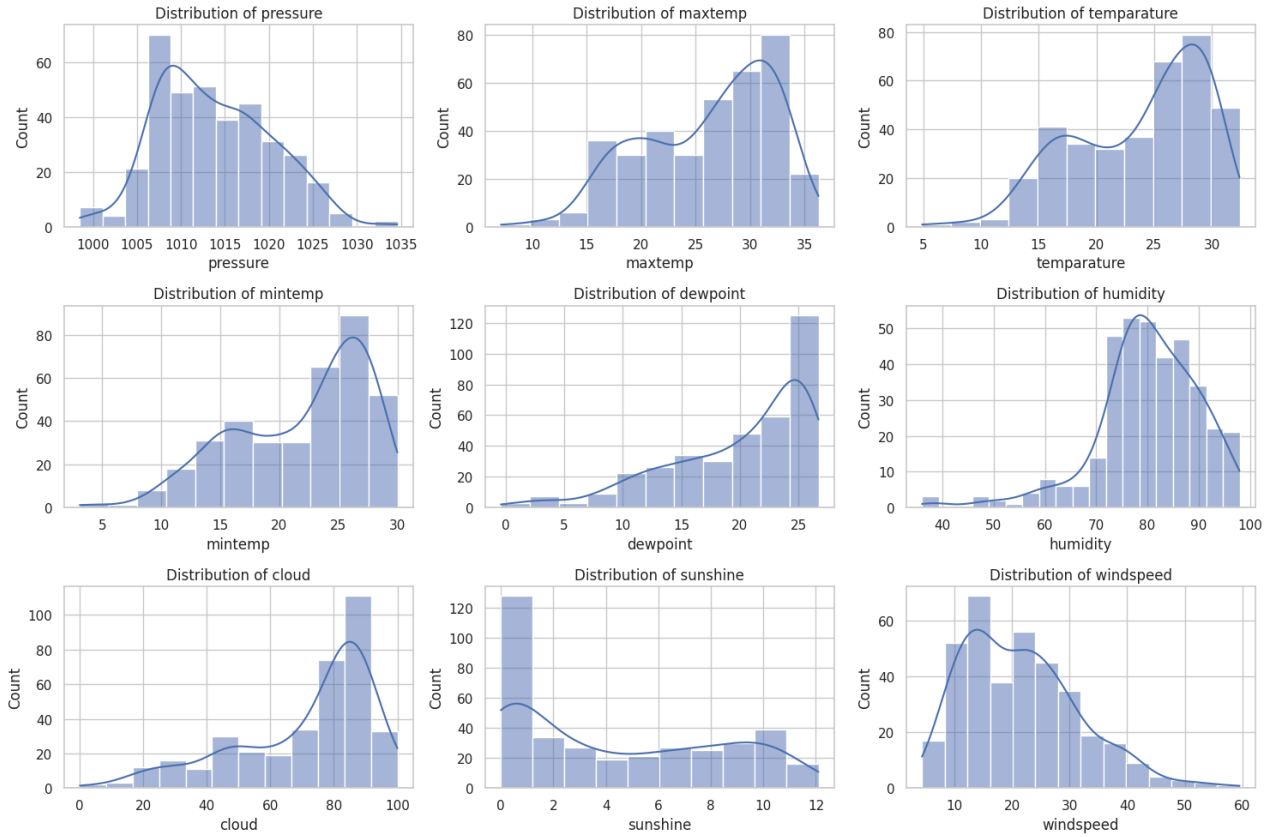
```
data.columns
```

```
Index(['pressure', 'maxtemp', 'temparature', 'mintemp', 'dewpoint', 'humidity',
       'cloud', 'rainfall', 'sunshine', 'winddirection', 'windspeed'],
      dtype='object')
```

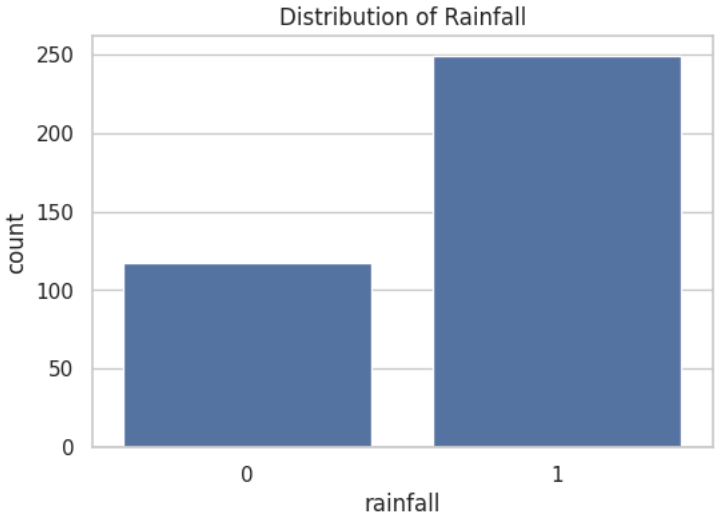
```
plt.figure(figsize=(15, 10))

for i, column in enumerate(['pressure', 'maxtemp', 'temparature', 'mintemp', 'dewpoint', 'humidity','cloud', 'sunshine
  plt.subplot(3, 3, i)
  sns.histplot(data[column], kde=True)
  plt.title(f"Distribution of {column}")

plt.tight_layout()
plt.show()
```

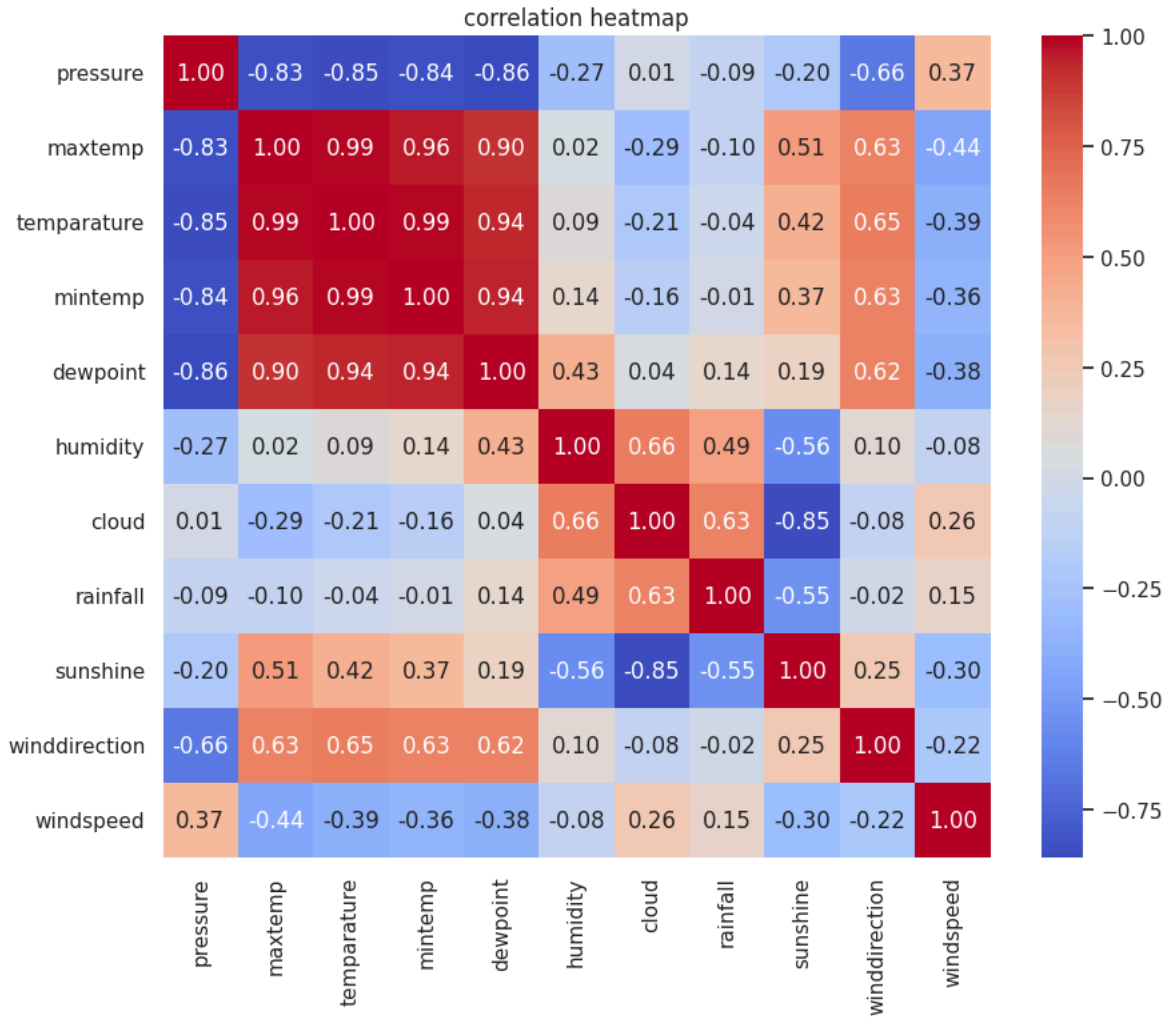



```
plt.figure(figsize=(6, 4))
sns.countplot(x="rainfall", data=data)
plt.title("Distribution of Rainfall")
plt.show()
```

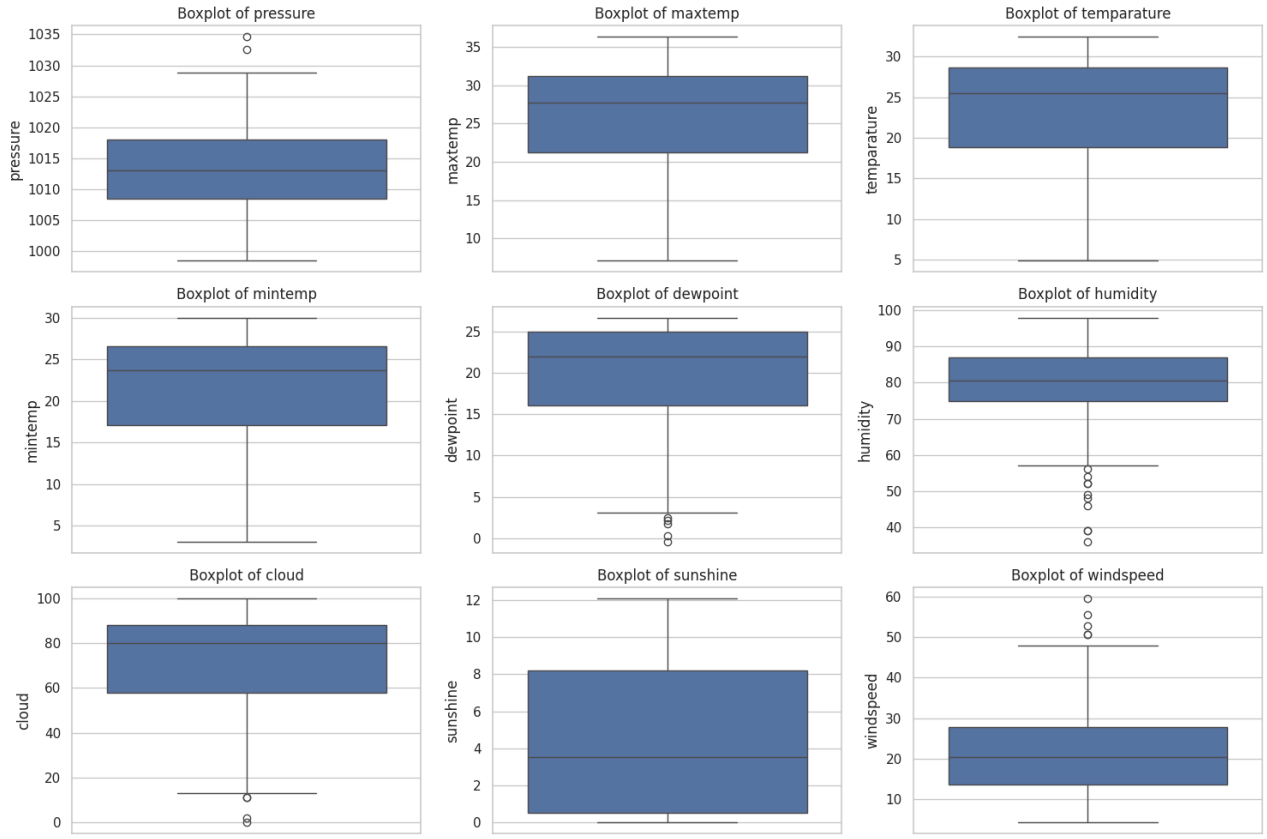



```
plt.figure(figsize=(10, 8))
sns.heatmap(data.corr(), annot = True, cmap = "coolwarm", fmt = ".2f")
```

```
plt.title("correlation heatmap")
plt.show()
```




```
plt.figure(figsize = (15, 10))
for i, column in enumerate(['pressure', 'maxtemp', 'temparature', 'mintemp', 'dewpoint', 'humidity','cloud', 'sunshine
  plt.subplot(3,3,i)
  sns.boxplot(data[column])
  plt.title(f"Boxplot of {column}")
plt.tight_layout()
plt.show()
```




```
data = data.drop(columns = ['pressure', 'maxtemp',  'mintemp'])
data.head()
```

	temparature	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspeed
0	18.3	13.1	72	49	1	9.3	80.0	26.3
1	18.9	15.6	81	83	1	0.6	50.0	15.3
2	19.3	18.4	95	91	1	0.0	40.0	14.2
3	20.6	18.8	90	88	1	1.0	50.0	16.9
4	20.7	19.9	95	81	1	0.0	40.0	13.7

Next steps: Generate code with `data` New interactive sheet

```
data["rainfall"].value_counts()
```

rainfall	count
1	249
0	117

dtype: int64

```
df_majority = data[data["rainfall"] == 1]
df_minority = data[data["rainfall"] == 0]
```

```
print(df_majority.shape, df_minority.shape)
```

```
(249, 8) (117, 8)
```

```
df_majority_downsampled = resample(df_majority, replace = False, n_samples = len(df_minority),random_state = 42 )
```

```
df_majority_downsampled.shape
```

```
(117, 8)
```

```
df_downsampled = pd.concat([df_majority_downsampled, df_minority])
```

```
df_downsampled.shape
```

```
(234, 8)
```

```
df_downsampled.head()
```

	temparature	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspeed
188	30.2	25.6	77	53	1	10.5	270.0	11.3
9	18.0	15.5	85	91	1	0.0	70.0	37.7
137	23.7	20.1	80	86	1	0.3	80.0	39.5
89	20.0	16.3	79	89	1	2.4	40.0	14.8
157	26.2	24.7	91	80	1	2.2	20.0	11.2

Next steps: Generate code with `df_downsampled` New interactive sheet

```
df_downsampled = df_downsampled.sample(frac = 1, random_state = 42).reset_index(drop = True)
```

```
df_downsampled.head()
```

	temparature	dewpoint	humidity	cloud	rainfall	sunshine	winddirection	windspeed
0	18.0	14.1	78	90	1	0.0	30.0	28.5
1	26.2	19.5	69	17	0	10.5	70.0	12.4
2	29.6	24.4	74	27	0	10.8	220.0	8.7
3	27.6	24.8	85	84	1	1.8	70.0	34.8
4	14.8	8.4	66	18	0	10.1	20.0	24.4

Next steps: Generate code with `df_downsampled` New interactive sheet

```
df_downsampled["rainfall"].value_counts()
```

rainfall	count
1	117
0	117

dtype: int64

```
x = df_downsampled.drop(columns = ["rainfall"])
y = df_downsampled["rainfall"]
```

```
print(x, y)
```

```
     temparature  dewpoint  humidity  cloud  sunshine  winddirection  \
0           18.0      14.1        78     90       0.0           30.0
1           26.2      19.5        69     17      10.5           70.0
2           29.6      24.4        74     27      10.8          220.0
3           27.6      24.8        85     84       1.8           70.0
4           14.8       8.4        66     18      10.1           20.0
```

```
..           ...       ...       ...    ...       ...           ...
229         28.1      25.4        86     75       5.7          20.0
230         21.5      19.9        91     89       0.0          70.0
231         16.1      14.7        91     88       0.3          50.0
232         29.4      24.1        74     29       5.7          10.0
233         28.6      26.1        87     82       2.2         160.0

     windspeed
0         28.5
1         12.4
2          8.7
3         34.8
4         24.4
..         ...
229        9.5
230       31.8
231       24.4
232        4.4
233       12.6

[234 rows x 7 columns] 0      1
1      0
2      0
3      1
4      0
      ..
229    1
230    1
231    1
232    0
```