```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix,accuracy_score
```

```
df=pd.read_csv(r"/content/Weather-related disease prediction.csv")
df.head(5)
```

	Age	Gender	Temperature (C)	Humidity	Wind Speed (km/h)	nausea	joint_pain	abdominal_pain	high_fever	chills	...	facial_pain	s
0	4	1	25.826	0.740000	8.289000	1	0	0	0	0	...	0	
1	55	0	21.628	0.600000	15.236000	0	0	0	0	1	...	0	
2	45	0	13.800	0.817083	4.291992	0	0	0	0	0	...	0	
3	6	0	37.254	0.610000	18.009000	1	0	0	1	0	...	0	
4	70	0	18.162	0.870000	17.916000	0	0	0	0	0	...	1	

5 rows × 51 columns

```
df.tail(5)
```

	Age	Gender	Temperature (C)	Humidity	Wind Speed (km/h)	nausea	joint_pain	abdominal_pain	high_fever	chills	...	facial_pain
5195	91	1	21.714000	0.870000	5.450000	0	1	0	0	0	...	0
5196	83	0	10.491000	0.920000	9.165000	0	0	0	0	0	...	0
5197	77	0	19.942824	0.744583	9.305129	0	0	0	0	0	...	0
5198	84	0	20.069000	0.850000	17.971000	0	0	1	0	0	...	0
5199	100	1	38.021000	0.750000	18.300000	0	0	0	0	0	...	0

5 rows × 51 columns

```
#Data Info
df.describe()
df.info()
df.shape
df.columns
```

```
8   high_fever              5200 non-null   int64
9   chills                  5200 non-null   int64
10  fatigue                 5200 non-null   int64
11  runny_nose              5200 non-null   int64
12  pain_behind_the_eyes    5200 non-null   int64
13  dizziness               5200 non-null   int64
14  headache                5200 non-null   int64
15  chest_pain              5200 non-null   int64
16  vomiting                5200 non-null   int64
17  cough                   5200 non-null   int64
18  shivering               5200 non-null   int64
19  asthma_history          5200 non-null   int64
20  high_cholesterol        5200 non-null   int64
21  diabetes                5200 non-null   int64
22  obesity                 5200 non-null   int64
23  hiv_aids                5200 non-null   int64
```

```
 38  swollen_glands           5200 non-null   int64
 39  rashes                   5200 non-null   int64
 40  sinus_headache           5200 non-null   int64
 41  facial_pain              5200 non-null   int64
 42  shortness_of_breath      5200 non-null   int64
 43  reduced_smell_and_taste  5200 non-null   int64
 44  skin_irritation          5200 non-null   int64
 45  itchiness                5200 non-null   int64
 46  throbbing_headache       5200 non-null   int64
 47  confusion                5200 non-null   int64
 48  back_pain                5200 non-null   int64
 49  knee_ache                5200 non-null   int64
 50  prognosis                5200 non-null   object
dtypes: float64(3), int64(47), object(1)
memory usage: 2.0+ MB
Index(['Age', 'Gender', 'Temperature (C)', 'Humidity', 'Wind Speed (km/h)',
       'nausea', 'joint_pain', 'abdominal_pain', 'high_fever', 'chills',
       'fatigue', 'runny_nose', 'pain_behind_the_eyes', 'dizziness',
       'headache', 'chest_pain', 'vomiting', 'cough', 'shivering',
       'asthma_history', 'high_cholesterol', 'diabetes', 'obesity', 'hiv_aids',
       'nasal_polyps', 'asthma', 'high_blood_pressure', 'severe_headache',
       'weakness', 'trouble_seeing', 'fever', 'body_aches', 'sore_throat',
       'sneezing', 'diarrhea', 'rapid_breathing', 'rapid_heart_rate',
       'pain_behind_eyes', 'swollen_glands', 'rashes', 'sinus_headache',
       'facial_pain', 'shortness_of_breath', 'reduced_smell_and_taste',
       'skin_irritation', 'itchiness', 'throbbing_headache', 'confusion',
       'back_pain', 'knee_ache', 'prognosis'],
      dtype='object')
```

```
#Removed Null Values
df.isna().sum()
df.dropna(inplace=True)
```

```
#Removed Duplicates
df.duplicated().sum()
df.drop_duplicates(inplace=True)
```

```
#EDA
print("\nShape (rows, columns):", df.shape)
```

```
Shape (rows, columns): (4981, 51)
```

```
print("\nColumns:\n", df.columns.tolist())
```

```
Columns:
 ['Age', 'Gender', 'Temperature (C)', 'Humidity', 'Wind Speed (km/h)', 'nausea', 'joint_pain', 'abdominal_pain', 'high_f
```

```
print("\nData Types:\n", df.dtypes)
```

```
Data Types:
 Age                         int64
Gender                       int64
Temperature (C)            float64
Humidity                   float64
Wind Speed (km/h)          float64
nausea                       int64
joint_pain                   int64
abdominal_pain               int64
high_fever                   int64
chills                       int64
fatigue                      int64
runny_nose                   int64
pain_behind_the_eyes         int64
dizziness                    int64
headache                     int64
chest_pain                   int64
vomiting                     int64
cough                        int64
shivering                    int64
asthma_history               int64
high_cholesterol             int64
diabetes                     int64
obesity                      int64
hiv_aids                     int64
nasal_polyps                 int64
asthma                       int64
high_blood_pressure          int64
severe_headache              int64
weakness                     int64
trouble_seeing               int64
fever                        int64
```

```
body_aches                  int64
sore_throat                 int64
sneezing                    int64
diarrhea                    int64
rapid_breathing             int64
rapid_heart_rate            int64
pain_behind_eyes            int64
swollen_glands              int64
rashes                      int64
sinus_headache              int64
facial_pain                 int64
shortness_of_breath         int64
reduced_smell_and_taste     int64
skin_irritation             int64
itchiness                   int64
throbbing_headache          int64
confusion                   int64
back_pain                   int64
knee_ache                   int64
prognosis                  object
dtype: object
```

```
print("\nMissing Values:\n", df.isnull().sum())
```

```
Missing Values:
 Age                       0
Gender                     0
Temperature (C)            0
Humidity                   0
Wind Speed (km/h)          0
nausea                     0
joint_pain                 0
abdominal_pain             0
high_fever                 0
chills                     0
fatigue                    0
runny_nose                 0
pain_behind_the_eyes       0
dizziness                  0
headache                   0
chest_pain                 0
vomiting                   0
cough                      0
shivering                  0
asthma_history             0
high_cholesterol           0
diabetes                   0
obesity                    0
hiv_aids                   0
nasal_polyps               0
asthma                     0
high_blood_pressure        0
severe_headache            0
weakness                   0
trouble_seeing             0
fever                      0
body_aches                 0
sore_throat                0
sneezing                   0
diarrhea                   0
rapid_breathing            0
rapid_heart_rate           0
pain_behind_eyes           0
swollen_glands             0
rashes                     0
sinus_headache             0
facial_pain                0
shortness_of_breath        0
reduced_smell_and_taste    0
skin_irritation            0
itchiness                  0
throbbing_headache         0
confusion                  0
back_pain                  0
knee_ache                  0
prognosis                  0
dtype: int64
```

```
print("\nDuplicate Rows:", df.duplicated().sum())
```

```
Duplicate Rows: 0
```

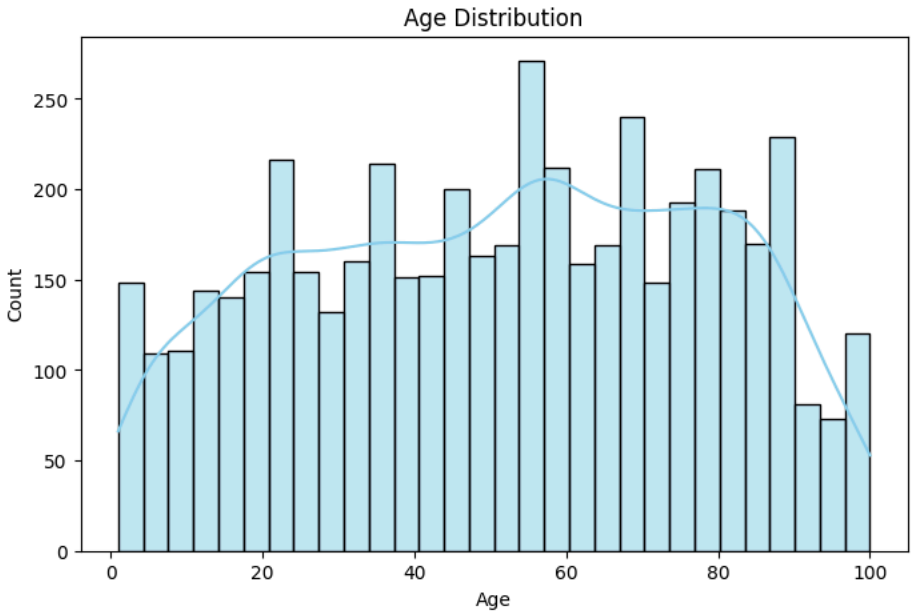
```
print("\nSummary Statistics:\n", df.describe())
```

```
Summary Statistics:
                Age       Gender  Temperature (C)     Humidity  \
count  4981.000000  4981.000000      4981.000000  4981.000000
mean     51.290905     0.513150        19.962910     0.745075
std      26.624880     0.499877        11.439338     0.143060
min       1.000000     0.000000       -15.125000     0.370833
25%      29.000000     0.000000        12.212269     0.624167
50%      53.000000     1.000000        20.081713     0.750000
75%      74.000000     1.000000        28.540000     0.860000
max     100.000000     1.000000        40.996000     1.000000

       Wind Speed (km/h)       nausea   joint_pain  abdominal_pain  \
count        4981.000000  4981.000000  4981.000000     4981.000000
mean           10.706016     0.152981     0.066051        0.040153
std             5.672566     0.360006     0.248396        0.196337
min             0.008000     0.000000     0.000000        0.000000
25%             6.415179     0.000000     0.000000        0.000000
50%            10.185933     0.000000     0.000000        0.000000
75%            15.062000     0.000000     0.000000        0.000000
max            31.303096     1.000000     1.000000        1.000000

        high_fever       chills  ...  sinus_headache  facial_pain  \
count  4981.000000  4981.000000  ...     4981.000000  4981.000000
mean      0.102590     0.063441  ...        0.031118     0.034933
std       0.303453     0.243779  ...        0.173655     0.183628
min       0.000000     0.000000  ...        0.000000     0.000000
25%       0.000000     0.000000  ...        0.000000     0.000000
50%       0.000000     0.000000  ...        0.000000     0.000000
75%       0.000000     0.000000  ...        0.000000     0.000000
max       1.000000     1.000000  ...        1.000000     1.000000

       shortness_of_breath  reduced_smell_and_taste  skin_irritation  \
count          4981.000000              4981.000000      4981.000000
mean              0.032323                 0.033327         0.032122
std               0.176874                 0.179506         0.176342
min               0.000000                 0.000000         0.000000
25%               0.000000                 0.000000         0.000000
50%               0.000000                 0.000000         0.000000
75%               0.000000                 0.000000         0.000000
max               1.000000                 1.000000         1.000000

         itchiness  throbbing_headache    confusion    back_pain    knee_ache
count  4981.000000         4981.000000  4981.000000  4981.000000  4981.000000
mean      0.029713            0.037141     0.067055     0.035535     0.038346
std       0.169811            0.189127     0.250142     0.185146     0.192049
min       0.000000            0.000000     0.000000     0.000000     0.000000
25%       0.000000            0.000000     0.000000     0.000000     0.000000
50%       0.000000            0.000000     0.000000     0.000000     0.000000
75%       0.000000            0.000000     0.000000     0.000000     0.000000
max       1.000000            1.000000     1.000000     1.000000     1.000000

[8 rows x 50 columns]
```

```
print("\nUnique values in 'prognosis':", df['prognosis'].nunique())
```

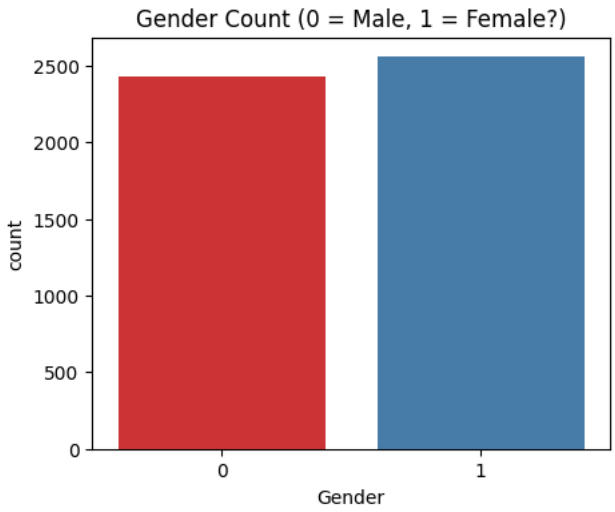
```
Unique values in 'prognosis': 11
```

```
# Age Distribution
plt.figure(figsize=(8,5))
sns.histplot(df['Age'], bins=30, kde=True, color='skyblue')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

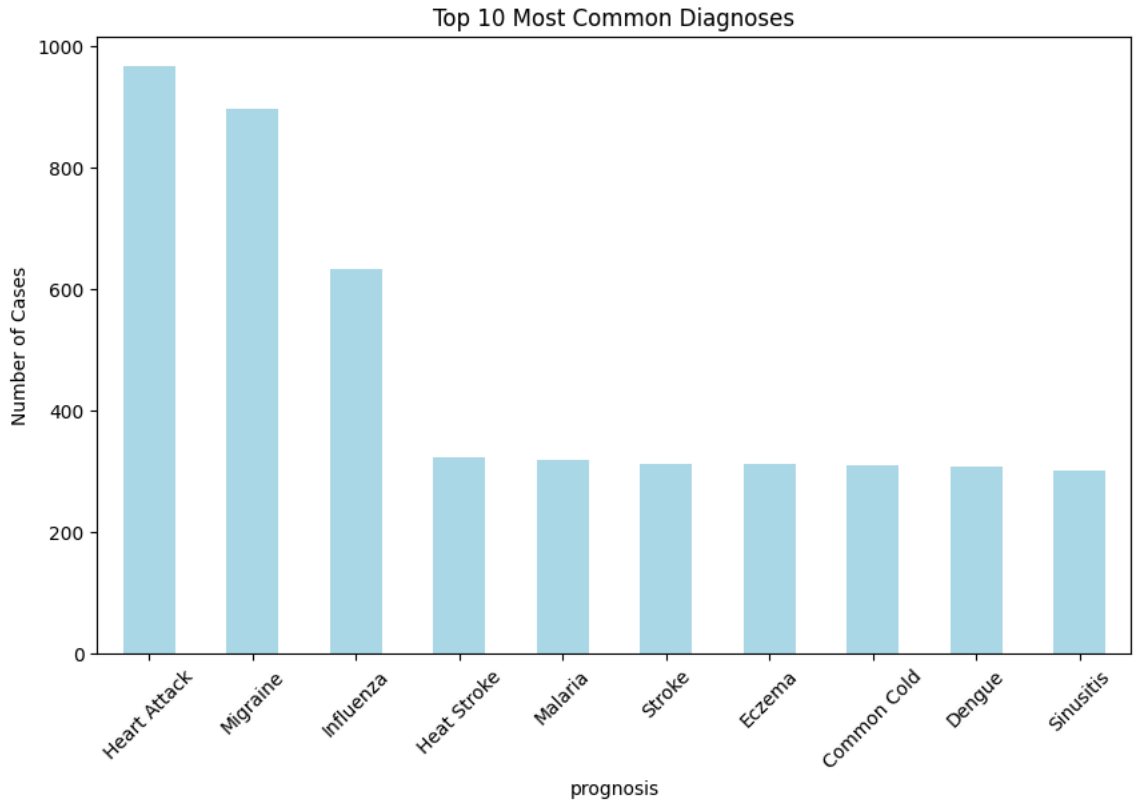



```
# Gender Count
plt.figure(figsize=(5,4))
sns.countplot(x='Gender', data=df, palette='Set1')
plt.title('Gender Count (0 = Male, 1 = Female?)')
plt.show()
```

```
/tmp/ipython-input-19-204277392.py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue`

  sns.countplot(x='Gender', data=df, palette='Set1')
```




```
# Top 10 Most Common Diseases
plt.figure(figsize=(10,6))
df['prognosis'].value_counts().head(10).plot(kind='bar', color='lightblue')
plt.title('Top 10 Most Common Diagnoses')
plt.ylabel('Number of Cases')
plt.xticks(rotation=45)
plt.show()
```




```
# 5. Data Preprocessing
# Features and Target
X = df.drop('prognosis', axis=1)
y = df['prognosis']


# Encode Target
le = LabelEncoder()
y = le.fit_transform(y)


# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)



# 6. Build the Model
model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
# 7. Predict and Evaluate
y_pred = model.predict(X_test)


print("\nAccuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

```
Accuracy: 0.9899699097291875

Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        71
           1       0.93      0.96      0.95        56
           2       1.00      1.00      1.00        54
           3       1.00      1.00      1.00        63
           4       0.99      1.00      0.99       188
           5       0.97      0.97      0.97        61
           6       0.98      0.97      0.98       128
           7       1.00      1.00      1.00        57
           8       1.00      1.00      1.00       191
           9       1.00      1.00      1.00        62
          10       1.00      0.97      0.98        66
```

```
              precision    recall  f1-score   support

    accuracy                         0.99       997
   macro avg       0.99      0.99      0.99       997
weighted avg       0.99      0.99      0.99       997
```