Weather Forecasting

Weather forecasting is the task of forecasting climate conditions for a given location and time. With the use of weather data and algorithms, it is possible to predict weather conditions for the next n number of days.

For forecasting weather using Python, we need a dataset containing historical weather data based on a particular location. Using the following data from kaggle that demonstrate the weather on Delhi from 2013 until 2017

"https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data"

The data contains the following table

- Date: of format YYYY-MM-DD
- Mean temperature averaged out from multiple 3 hour intervals in a day.
- Humidity value for the day (units are grams of water vapor per cubic meter volume of air).
- Wind speed measured in kmph.
- Pressure reading of weather (measure in atm)

Questions:

1- Write a python code that reads the data from CSV file
2- Print the first 5 records
3- Print the count, mean, std, min and max values for every attribute
4- Print the column name, Non Null count and Dtypes
5- Draw the histogram and scatter of the wind speed of Delhi
6- The date column is not having a datetime data type. Change it
7- Plot the Mean temperatures in Delhi over the years
8- Now Plot the Humidity over the years
9- Also Plot the wind speed over the years
10- Is there any good correlations between attributes?
11- Have a look at the relationship between temperature and humidity (using Scatter)
12- Draw a linear regression line in red between temperature and humidity
13- Predict the Temp when the humidity is 50, Accuracy is good?
14- Have a look now at the relationship between temperature and wind speed (using Scatter), describe the result
15- Now try to calculate R-square of Temp based on humidity and wind speed using Multiple regression

```python
import pandas as pd
import numpy as np
from scipy import stats
from sklearn import linear_model
import matplotlib.pyplot as plt

df = pd.read_csv("weather/DailyDelhiClimateTrain.csv")
print(df.head())
print(df.describe())
print(df.info())

ws = df['wind_speed'] # or ws= df.loc[:,'wind_speed']
plt.hist(ws, 100)
plt.show()

df['date'] = pd.to_datetime(df['date'])

plt.plot(df['date'], df['meantemp'])
plt.xlabel('Date')
plt.ylabel('Mean Temperature')
plt.title('Mean Temperature in Delhi Over the Years')
plt.show()

plt.plot(df['date'], df['humidity'])
plt.xlabel('Date')
plt.ylabel('humidity')
plt.title('Humidity in Delhi Over the Years')
plt.show()

plt.plot(df['date'], df['wind_speed'])
plt.xlabel('Date')
plt.ylabel('wind speed')
plt.title('Wind Speed in Delhi Over the Years')
plt.show()


print(df.corr())


df.plot(kind = 'scatter', x = 'humidity', y = 'meantemp')
plt.show()
```

```python
x=df['humidity']
y=df['meantemp']

slope, intercept, r, p, std_err = stats.linregress(x, y)

def getTemp(x):
    return slope * x + intercept

line = list(map(getTemp, x))
print(r)
print('The temp of 50 humidity is', getTemp(50))
plt.scatter(x, y)
plt.plot(x, line, color='red')
plt.show()


df.plot(kind = 'scatter', x = 'wind_speed', y = 'meantemp')
plt.show()

X = df[['humidity', 'wind_speed']]
y = df['meantemp']
regr = linear_model.LinearRegression()
regr.fit(X, y)
new_data = pd.DataFrame([[50, 20]], columns=['humidity','wind_speed'])
predictedtemp = regr.predict(new_data)
print('Temp =', predictedtemp)
print("R-squared= ", regr.score(X, y))

newline = regr.predict(X)
plt.scatter(df['humidity'], df['wind_speed'], c=df['meantemp'],
cmap='coolwarm')
plt.plot(np.array(X), newline)
plt.xlabel('Humidity')
plt.ylabel('Wind Speed')
plt.title('Mean Temperature in Delhi')
plt.colorbar(label='Mean Temperature')
plt.show()
```