p12

April 17, 2024

[3]:
```
import pandas as pd
```

[10]:
```
df = pd.read_csv('sample_weather.txt', sep='\s+', header=None)
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96 entries, 0 to 95
Data columns (total 19 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   0       96 non-null     int64  
 1   1       96 non-null     int64  
 2   2       96 non-null     object 
 3   3       96 non-null     float64
 4   4       96 non-null     float64
 5   5       96 non-null     int64  
 6   6       96 non-null     float64
 7   7       96 non-null     int64  
 8   8       96 non-null     float64
 9   9       96 non-null     int64  
 10  10      96 non-null     float64
 11  11      96 non-null     int64  
 12  12      96 non-null     float64
 13  13      96 non-null     int64  
 14  14      96 non-null     float64
 15  15      96 non-null     float64
 16  16      96 non-null     object 
 17  17      96 non-null     float64
 18  18      96 non-null     int64  
dtypes: float64(9), int64(8), object(2)
memory usage: 14.4+ KB
None
```

[11]:
```
df.columns = ['Station_ID', 'WMO_Number', 'Timestamp', 'Temperature',␣
 ↪'Dew_Point', 'Wind_Speed', 'Pressure', 'Humidity', 'Visibility',␣
 ↪'Cloud_Cover', 'Wind_Direction', 'Precipitation', 'Snow_Depth', 'Radiation',␣
 ↪'UV_Index', 'Weather_Code', 'Sunrise_Sunset', 'Sea_Level_Pressure',␣
 ↪'Ground_Level_Pressure']
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 96 entries, 0 to 95
Data columns (total 19 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Station_ID            96 non-null     int64
 1   WMO_Number            96 non-null     int64
 2   Timestamp             96 non-null     object
 3   Temperature           96 non-null     float64
 4   Dew_Point             96 non-null     float64
 5   Wind_Speed            96 non-null     int64
 6   Pressure              96 non-null     float64
 7   Humidity              96 non-null     int64
 8   Visibility            96 non-null     float64
 9   Cloud_Cover           96 non-null     int64
 10  Wind_Direction        96 non-null     float64
 11  Precipitation         96 non-null     int64
 12  Snow_Depth            96 non-null     float64
 13  Radiation             96 non-null     int64
 14  UV_Index              96 non-null     float64
 15  Weather_Code          96 non-null     float64
 16  Sunrise_Sunset        96 non-null     object
 17  Sea_Level_Pressure    96 non-null     float64
 18  Ground_Level_Pressure 96 non-null     int64
dtypes: float64(9), int64(8), object(2)
memory usage: 14.4+ KB
```

[13]:
```
df.head()
```

[13]:

	Station_ID	WMO_Number	Timestamp	Temperature	Dew_Point	Wind_Speed	\
0	690190	13910	20060201_0	51.75	33.0	24	
1	690190	13910	20060201_1	54.74	33.0	24	
2	690190	13910	20060201_2	50.59	33.0	24	
3	690190	13910	20060201_3	51.67	33.0	24	
4	690190	13910	20060201_4	65.67	33.0	24	

	Pressure	Humidity	Visibility	Cloud_Cover	Wind_Direction	Precipitation	\
0	1006.3	24	943.9	24	15.0	24	
1	1006.3	24	943.9	24	15.0	24	
2	1006.3	24	943.9	24	15.0	24	
3	1006.3	24	943.9	24	15.0	24	
4	1006.3	24	943.9	24	15.0	24	

	Snow_Depth	Radiation	UV_Index	Weather_Code	Sunrise_Sunset	\
0	10.7	24	22.0	28.9	0.00I	

```
1        10.7         24      22.0          28.9          0.00I
2        10.7         24      22.0          28.9          0.00I
3        10.7         24      22.0          28.9          0.00I
4        10.7         24      22.0          28.9          0.00I

   Sea_Level_Pressure  Ground_Level_Pressure
0               999.9                      0
1               999.9                      0
2               999.9                      0
3               999.9                      0
4               999.9                      0
```

[14]:
```
df.isnull().sum()
```

[14]:
```
Station_ID              0
WMO_Number              0
Timestamp               0
Temperature             0
Dew_Point               0
Wind_Speed              0
Pressure                0
Humidity                0
Visibility              0
Cloud_Cover             0
Wind_Direction          0
Precipitation           0
Snow_Depth              0
Radiation               0
UV_Index                0
Weather_Code            0
Sunrise_Sunset          0
Sea_Level_Pressure      0
Ground_Level_Pressure   0
dtype: int64
```

[15]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Assuming 'Temperature', 'Dew_Point', and 'Wind_Speed' are the correct column␣
 ↪names
avg_temperature = df['Temperature'].mean()
avg_dew_point = df['Dew_Point'].mean()
avg_wind_speed = df['Wind_Speed'].mean()

print(f"Average Temperature: {avg_temperature}")
print(f"Average Dew Point: {avg_dew_point}")
print(f"Average Wind Speed: {avg_wind_speed}")
```

```
# Now let's create some plots
plt.figure(figsize=(15, 5))

# Temperature plot
plt.subplot(1, 3, 1)
plt.hist(df['Temperature'], bins=20, color='skyblue', edgecolor='black')
plt.title('Temperature Distribution')
plt.xlabel('Temperature')
plt.ylabel('Frequency')

# Dew point plot
plt.subplot(1, 3, 2)
plt.hist(df['Dew_Point'], bins=20, color='skyblue', edgecolor='black')
plt.title('Dew Point Distribution')
plt.xlabel('Dew Point')
plt.ylabel('Frequency')

# Wind speed plot
plt.subplot(1, 3, 3)
plt.hist(df['Wind_Speed'], bins=20, color='skyblue', edgecolor='black')
plt.title('Wind Speed Distribution')
plt.xlabel('Wind Speed')
plt.ylabel('Frequency')

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

```
Average Temperature: 53.58260416666667
Average Dew Point: 25.900000000000006
Average Wind Speed: 24.0
```