

[Thực hành] Phân tích sự biến đổi nhiệt độ theo thời gian - daily-min-temperatures.csv

Mục tiêu: Vẽ biểu đồ đường, biểu đồ Histogram để tìm hiểu các giá trị thống kê như max, min, mode, tần suất và xu hướng thay đổi nhiệt độ theo thời gian

Bộ dữ liệu daily-min-temperatures.csv mô tả nhiệt độ hàng ngày từ 1/1/1981 đến 31/12/1990. Trong buổi thực hành này, chúng ta sẽ sử dụng thư viện Matplotlib vừa học được để vẽ một số biểu đồ nhằm phân tích sự biến đổi nhiệt độ trên bộ dữ liệu này. Điều kiện:

- Dữ liệu đã được tiền xử lý.
- Thành thạo các câu lệnh đọc dữ liệu từ file, trích xuất dữ liệu mong muốn.
- Nắm được các khái niệm cơ bản về Data Visualization. Các bước thực hiện:

1. Chọn mục tiêu Vẽ biểu đồ Histogram phân tích các giá trị thống kê như max, min, mode, tần suất. Vẽ biểu đồ đường phân tích xu hướng thay đổi nhiệt độ theo thời gian.
2. Khai báo thư viện cần dùng

```
import pandas as pd
import matplotlib.pyplot as plt
```

3. Lấy dữ liệu

```
df = pd.read_csv("daily-min-temperatures.csv")
```

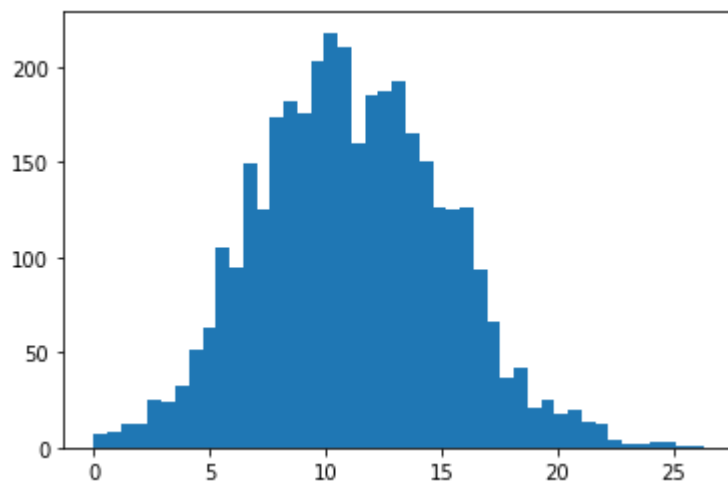
4. Hiển thị các giá trị thống kê của dữ liệu nhiệt độ

```
df['Temp'].describe()
```

```
count    3650.000000
mean      11.177753
std        4.071837
min        0.000000
25%        8.300000
50%       11.000000
75%       14.000000
max       26.300000
Name: Temp, dtype: float64
```

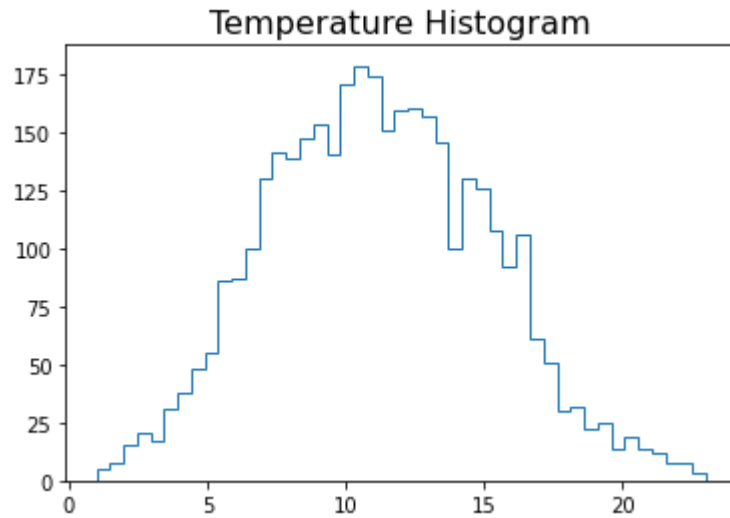
5. Vẽ biểu đồ Histogram

```
plt.hist(df['Temp'], bins = 45)
plt.show()
```



6. Tinh chỉnh biểu đồ

```
plt.hist(df['Temp'], bins = 45, range = (1, 23), histtype = 'step')
plt.title('Temperature Histogram', fontsize = 16)
plt.show()
```

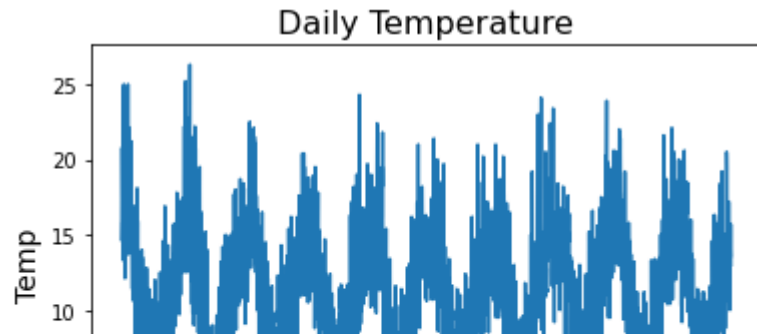


Nhận xét:

- Giá trị xuất hiện nhiều nhất là trong khoảng 10-11 độ.
- Có một số giá trị bất thường ở cả 2 phía. Một số rất ít ngày có nhiệt độ cao bất thường trên 23 độ và thấp bất thường dưới 1 độ.
- Nhìn chung, nếu loại bỏ các ngày có giá trị bất thường, có thể nói nhiệt độ tuân theo phân phối chuẩn.

7. Vẽ biểu đồ đường thể hiện xu hướng nhiệt độ

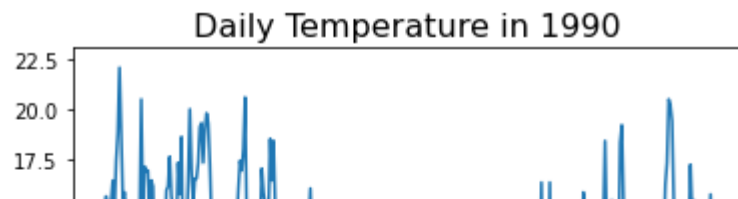
```
plt.plot(df['Date'], df['Temp'])  
plt.title('Daily Temperature', fontsize = 16)  
plt.xlabel('Date', fontsize = 14)  
plt.ylabel('Temp', fontsize = 14)  
plt.show()
```



Nhận xét:

- Có thể thấy xu hướng thay đổi nhiệt độ có tính chu kỳ.
- Dựa vào đồ thị, ta phán đoán chu kỳ thay đổi của nhiệt độ là theo năm.
- Tuy nhiên, do khoảng thời gian biểu diễn dài nên khó có thể nhìn chi tiết xu hướng biến đổi nhiệt độ trong năm. Do đó ta vẽ bổ sung biểu đồ đường thể hiện xu hướng nhiệt độ theo từng năm quan tâm.

```
df['Date'] = pd.to_datetime(df['Date'])
bounds = ['1/1/1990', '12/31/1990']
bounds = pd.to_datetime(bounds)
d1 = df[(df['Date'] >= bounds[0]) & (df['Date'] <= bounds[1])]
plt.plot(d1['Date'], d1['Temp'])
plt.title('Daily Temperature in 1990', fontsize = 16)
plt.xlabel('Date', fontsize = 14)
plt.ylabel('Temp', fontsize = 14)
plt.show()
```



Nhận xét:

- Nhiệt độ có xu hướng giảm từ tháng 5 và bắt đầu tăng trở lại vào tháng 9.
- 3 tháng đầu năm có nhiệt độ trung bình cao nhất.

5.0 10.0 15.0 20.0 25.0 30.0 35.0 40.0 45.0 50.0 55.0 60.0 65.0 70.0 75.0 80.0 85.0 90.0 95.0 100.0

▼ Tổng kết

Qua bài thực hành này, chúng ta đã biết cách trực quan hóa dữ liệu nhiệt độ bằng một số biểu đồ. Chúng ta cũng biết cách sử dụng thư viện Matplotlib để vẽ biểu đồ dựa trên mục tiêu, dữ liệu và đưa ra một số thông tin dựa trên trực quan hóa dữ liệu.

[Các sản phẩm có tính phí của Colab](#) - [Huỷ hợp đồng tại đây](#)

