

TRƯỜNG ĐẠI HỌC THĂNG LONG
KHOA CÔNG NGHỆ THÔNG TIN



BÀI TẬP LỚN
NHẬP MÔN KHOA HỌC DỮ LIỆU

ĐỀ TÀI: THU THẬP VÀ PHÂN TÍCH DỮ LIỆU VỀ THỜI TIẾT MIỀN BẮC

Sinh viên thực hiện: **Phạm Hải Đăng – A41308**
 Tạ Quang Dũng – A42865
Nhóm: **7**
Chuyên ngành: **Trí tuệ nhân tạo**
Giáo viên hướng dẫn: **TS. Nguyễn Thị Huyền Châu – PAI001**

HÀ NỘI – 2023

MỤC LỤC

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT	2
1.1. Quy trình làm việc trong Khoa học dữ liệu	2
1.2. Tổng quan về đề tài thực hiện.....	3
<i>1.2.1. Đối tượng phân tích.....</i>	<i>3</i>
<i>1.2.2. Hướng tiếp cận và thực hiện đề tài.....</i>	<i>3</i>
CHƯƠNG 2. THU THẬP DỮ LIỆU VÀ LƯU TRỮ DỮ LIỆU	4
2.1. Thông tin về nguồn dữ liệu.....	4
2.2. Thu thập dữ liệu	4
2.3. Lưu trữ dữ liệu	7
CHƯƠNG 3. TIỀN XỬ LÝ DỮ LIỆU.....	8
3.1. Kiểm tra các vấn đề của dữ liệu.....	8
<i>3.1.1. Đọc dữ liệu.....</i>	<i>8</i>
CHƯƠNG 4. PHÂN TÍCH DỮ LIỆU – EXPLORATORY DATA ANALYSIS... 11	
4.1. Thống kê mô tả	11
<i>4.1.1. Đồng bằng sông Hồng</i>	<i>11</i>
<i>4.1.2. Tây Bắc Bộ.....</i>	<i>14</i>
<i>4.1.3. Đông Bắc Bộ.....</i>	<i>15</i>
4.2. Trực quan hóa dữ liệu	17
<i>4.2.1. Đồng bằng sông Hồng</i>	<i>17</i>
<i>4.2.2. Tây Bắc Bộ.....</i>	<i>24</i>
<i>4.2.3. Đông Bắc Bộ.....</i>	<i>28</i>
CHƯƠNG 5. TỔNG KẾT, NHẬN XÉT VÀ ĐÁNH GIÁ.....	33
5.1. Tổng kết các kết quả đạt được và những hạn chế còn sót lại sau khi thực hiện xong đề tài	33
5.2. Bảng phân công công việc và mức độ hoàn thành của từng cá nhân trong nhóm	33
5.3. Nhận xét, đánh giá và xếp loại của giáo viên	33

DANH MỤC MINH HỌA

Hình 2.1. Giao diện lúc tạo tài khoản thành công và lấy được API Key	4
Hình 2.2. Giao diện phân nhập thông tin cần lấy dữ liệu	5
Hình 4.1. Bảng thống kê mô tả của tập dữ liệu về thời tiết của Hà Nội.....	11
Hình 4.2. Biểu đồ thể hiện sự tương quan giữa tốc độ gió tối đa, lượng mưa và độ ẩm trung bình của Đồng bằng sông Hồng	13
Hình 4.3. Bảng thống kê mô tả của tập dữ liệu về thời tiết của tỉnh Sơn La	14
Hình 4.4. Bảng thống kê mô tả của tập dữ liệu về thời tiết của tỉnh Yên Bái.....	15
Hình 4.5. Biểu đồ nhiệt độ lớn nhất trong tháng 5/2023 của Hà Nội.....	17
Hình 4.6. Biểu đồ nhiệt độ lớn nhất trong tháng 6/2023 của Hà Nội.....	17
Hình 4.7. Biểu đồ tổng lượng mưa trung bình trong tháng 5/2023 của Hà Nội.....	19
Hình 4.8. Biểu đồ tổng lượng mưa trung bình trong tháng 6/2023 của Hà Nội.....	19
Hình 4.9. Biểu đồ nhiệt độ cao nhất và thấp nhất trung bình trong mỗi tháng của các tỉnh Đồng bằng sông Hồng	20
Hình 4.10. Biểu đồ tổng lượng mưa trung bình mỗi tháng của các tỉnh Đồng bằng sông Hồng	21
Hình 4.11. Biểu đồ độ ẩm trung bình theo tháng của các tỉnh Đồng bằng sông Hồng	22
Hình 4.12. Biểu đồ mức độ ảnh hưởng trung bình của tia UV theo tháng của Đồng bằng sông Hồng.....	23
Hình 4.13. Biểu đồ thể hiện sự tương quan giữa tốc độ gió tối đa, lượng mưa và độ ẩm trung bình của vùng Tây Bắc Bộ	24
Hình 4.14. Biểu đồ mức nhiệt cao nhất và thấp nhất trung bình theo tháng của các tỉnh vùng Tây Bắc Bộ	25
Hình 4.15. Biểu đồ thể hiện lượng mưa trung bình theo tháng của vùng Tây Bắc Bộ	26
Hình 4.16. Biểu đồ thể hiện độ ẩm trung bình theo tháng của vùng Tây Bắc Bộ.....	27
Hình 4.17. Biểu đồ thể hiện mối tương quan giữa tốc độ gió lớn nhất, lượng mưa và độ ẩm trung bình của vùng Đông Bắc Bộ	28
Hình 4.18. Biểu đồ mức nhiệt cao nhất và thấp nhất trung bình theo tháng của các tỉnh vùng Đông Bắc Bộ	29
Hình 4.19. Biểu đồ thể hiện lượng mưa trung bình theo tháng của vùng Đông Bắc Bộ ..	30
Hình 4.20. Biểu đồ thể hiện độ ẩm trung bình theo tháng của vùng Đông Bắc Bộ	31

LỜI NÓI ĐẦU

Thời tiết là một yếu tố quan trọng và không thể thiếu trong cuộc sống hàng ngày của mỗi chúng ta. Nó ảnh hưởng trực tiếp đến mọi khía cạnh của đời sống xã hội từ việc lên kế hoạch du lịch, sản xuất phát triển nông nghiệp, đến việc quyết định trang phục và hoạt động ngoài trời... Đặc biệt, thời tiết miền Bắc còn được mẹ thiên nhiên ưu ái ban tặng cho những đặc điểm rất khác biệt, đặc trưng chỉ ở miền Bắc mới có và quan trọng hơn hết là do ảnh hưởng của kiểu khí hậu nhiệt đới gió mùa.

Để hiểu sâu hơn, cũng như nghiên cứu và đưa ra những dự đoán về tình hình thời tiết của miền Bắc hiện nay, việc thu thập và phân tích dữ liệu về thời tiết là một công việc cần thiết. Nằm trong khuôn khổ của đề tài này, nhóm chúng em xin được tập trung vào việc thu thập và phân tích dữ liệu về nhiệt độ, độ ẩm, lượng mưa và một vài chỉ số liên quan khác trong thời tiết của miền Bắc Việt Nam. Chúng em tin rằng việc nắm bắt thông tin về thời tiết trong khu vực này không chỉ giúp chúng ta hiểu rõ hơn về môi trường sống mà qua đó còn hỗ trợ ta đưa ra quyết định và lập kế hoạch một cách hiệu quả.

Với mục tiêu là thu thập và phân tích dữ liệu về thời tiết miền Bắc, nhóm em hy vọng rằng đề tài này sẽ cung cấp cho bạn đọc một cái nhìn toàn diện và sâu sắc hơn về đặc điểm khí hậu và tình hình thời tiết hiện nay trong khu vực này.

Nhóm em xin chân thành cảm ơn!

CHƯƠNG 1. CƠ SỞ LÝ THUYẾT

1.1. Quy trình làm việc trong Khoa học dữ liệu

- Bước 1: Xác định vấn đề: Khi kết thúc bước đầu tiên này của quy trình khoa học dữ liệu, bạn sẽ có những ý tưởng và xác định rõ ràng về cách tiến hành. Phác thảo này sẽ giúp bạn điều hướng sự phức tạp của dữ liệu và đạt được mục tiêu của mình. Việc đưa ra các câu hỏi đặt vấn đề phù hợp là rất quan trọng để nâng cao hiệu quả công việc và cuối cùng sẽ tiết kiệm thời gian cho bạn trong những bước sau.
- Bước 2: Thu thập và lưu trữ dữ liệu: Sau bước xác định vấn đề thì bây giờ bạn đã có một bộ câu hỏi rõ ràng, và đã đến lúc bạn bắt tay vào làm nó. Trước tiên, bạn cần thu thập và lưu trữ dữ liệu của mình ở một nơi an toàn để phân tích dữ liệu đó. Dữ liệu có thể được tồn tại từ trước đó hoặc là tải trực tiếp từ trên mạng. Còn không, ta có thể cào dữ liệu thông qua HTML hoặc API (Giao diện lập trình ứng dụng), hoặc mua của một công ty hay xin của tổ chức nào đó nhưng phải đảm bảo những yêu cầu nhất định.
- Bước 3: Tiền xử lý dữ liệu (Làm sạch và chuẩn bị dữ liệu): Như cái tên của nó, sau khi bạn đã thu thập và lưu trữ dữ liệu của mình xong, bước tiếp theo chính là đánh giá chất lượng của nó. Điều quan trọng cần nhớ là sự thành công của phân tích dữ liệu phụ thuộc rất nhiều vào chất lượng dữ liệu của bạn. Những hiểu biết của bạn sẽ sai hoặc gây hiểu lầm nếu thông tin của bạn không chính xác, không đầy đủ hoặc không nhất quán. Đó là lý do tại sao việc dành thời gian cho việc tiền xử lý dữ liệu là bắt buộc. Trong bước này ta sẽ phải xử lý lọc, làm sạch và chuẩn hóa dữ liệu để loại bỏ các sai sót, giảm nhiễu và đạt tiêu chuẩn để đưa vào phân tích và xây dựng mô hình phía sau.
- Bước 4: Phân tích dữ liệu: Bây giờ dữ liệu của bạn đã được làm sạch và sẵn sàng để phân tích. Tùy thuộc vào mục tiêu của phân tích, loại dữ liệu và các kỹ thuật khác nhau có sẵn. Thông qua việc sử dụng các phương pháp thống kê mô tả và công cụ trực quan hóa dữ liệu, ta có thể dễ dàng tiếp cận và hiểu rõ hơn về dữ liệu, truyền đạt thông tin chi tiết của bạn một cách hiệu quả.
- Bước 5: Xây dựng mô hình: Sau khi đã có những đánh giá, phân tích sơ bộ về dữ liệu. Tiếp theo ta sẽ sử dụng một vài thuật toán hoặc các mô hình có sẵn để nhét dữ liệu vào cho máy học. Mô hình tạo ra có thể được thử nghiệm với tập dữ liệu có trước để đánh giá độ hiệu quả của mô hình, rồi sau đó tinh chỉnh, cải thiện làm sao để có thể thu được mô hình tối ưu và hiệu quả nhất.

- Bước 6: Giao tiếp với dữ liệu: Sau khi hoàn thành tất cả các bước trên, ta gần như đã có một mô hình khá hoàn chỉnh. Tuy nhiên, ta cần phải giao tiếp và tương tác với chúng để có thể đưa ra các quyết định, dự đoán dựa trên các số liệu phân tích, đồ thị và biểu đồ trực quan.

1.2. Tổng quan về đề tài thực hiện

1.2.1. Đối tượng phân tích

Đối tượng phân tích của đề tài này là dữ liệu liên quan về thời tiết của miền Bắc Việt Nam. Dữ liệu này bao gồm các thông tin về nhiệt độ (cao nhất, thấp nhất, trung bình), độ ẩm, tổng lượng mưa, lực gió, chỉ số UV và một vài dự báo được ghi nhận tại các tỉnh trong khu vực miền Bắc trong một khoảng thời gian nhất định. Cụ thể ở đây là từ 29/06/2022 đến 30/06/2023.

1.2.2. Hướng tiếp cận và thực hiện đề tài

Nằm trong khuôn khổ của đề tài và nội dung trong môn học này, nhóm em sẽ chỉ thực hiện 4/6 bước của quy trình làm việc trong khoa học dữ liệu:

- Bước 1: Xác định vấn đề: Vấn đề đặt ra ở đây là thu thập dữ liệu về thời tiết của miền Bắc qua các trang trực tuyến và phân tích về dữ liệu đó.
- Bước 2: Thu thập và lưu trữ dữ liệu:
 - + Ở bước này, sau khi nhóm em có thử tìm và cào dữ liệu ở một số trang web có dữ liệu về thời tiết, tuy nhiên dữ liệu thu được ở những trang này khá là ít, và nếu muốn nhiều hơn thì mình phải mất phí thì nhóm em đã quyết định cào dữ liệu thông qua một trang web có tạo ra API cho mình kết hợp với việc code để lấy ra nguồn dữ liệu mình mong muốn.
 - + Sau đó, nhóm em sẽ gom dữ liệu của các tỉnh thu được theo 3 vùng như trên bản đồ, rồi tổng hợp lại và lưu trữ dưới dạng file .xlsx gồm nhiều sheet và mỗi sheet chứa data của 1 tỉnh.
- Bước 3: Tiền xử lý dữ liệu: Trong bước này nhóm em sẽ chia ra làm 3 file theo 3 vùng rồi áp dụng các kỹ thuật trong tiền xử lý dữ liệu để xử lý theo từng file dữ liệu vùng.
- Bước 4: Phân tích dữ liệu: Cũng theo 3 file tiền xử lý ở trên, tiếp tục ở bước này, nhóm em sẽ phân tích dữ liệu của một vài tỉnh đại diện trong vùng đó rồi lấy trung bình tất cả dữ liệu của các tỉnh rồi đem vào sử dụng các phương pháp thống kê mô tả, công cụ trực quan hóa dữ liệu để phân tích, vẽ biểu đồ và nhận xét, đưa ra các đánh giá.

CHƯƠNG 2. THU THẬP DỮ LIỆU VÀ LƯU TRỮ DỮ LIỆU

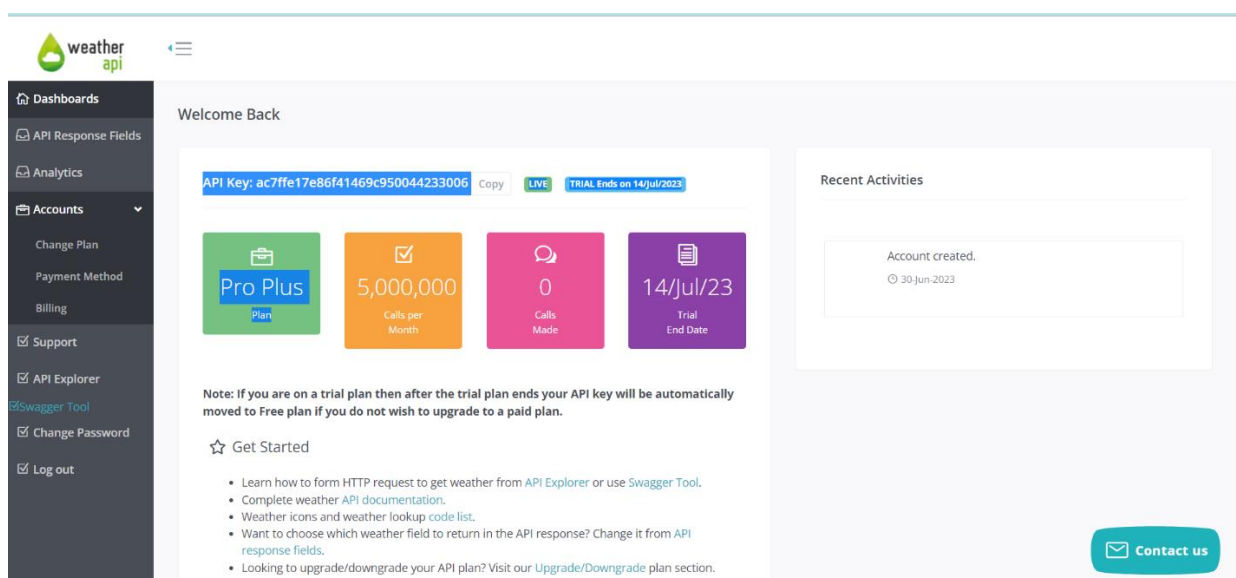
2.1. Thông tin về nguồn dữ liệu

Để thu thập dữ liệu, nhóm em có thực hiện cào ở một số trang web có thông tin dữ liệu về thời tiết như: OpenWeather, AccuWeather, The Weather Channel... Tuy nhiên, dữ liệu thu được rất ít và nó yêu cầu phải trả thêm phí nếu muốn lấy nhiều thêm. Nhưng thật may mắn khi nhóm em tìm được một trang web có tên là Weather API, nó cung cấp cho mình toàn bộ lịch sử dữ liệu thời tiết theo từng giờ trong một ngày nào đó tại địa điểm mà mình yêu cầu.

Địa chỉ của trang web: <https://www.weatherapi.com/>

2.2. Thu thập dữ liệu

- Đầu tiên ta sẽ đăng nhập vào địa chỉ trang web trên và sẽ tạo một tài khoản. Rồi sau đó, trang web sẽ cung cấp cho mình một API Key để mình có thể tìm kiếm dữ liệu.



Hình 2.1. Giao diện lúc tạo tài khoản thành công và lấy được API Key

- Tiếp theo mình sẽ chọn mục API Explorer để tìm kiếm lịch sử dữ liệu trong quá khứ.
- Và cuối cùng chọn mục History và nhập API Key trên, tên địa điểm và thời gian mình muốn lấy dữ liệu.

The screenshot shows the weather API web interface. At the top, there's a navigation bar with links: Features, Pricing, API Explorer, Docs, Weather, Contact, and My Account. Below this, there's a section for 'Your API Key' with a text input field containing 'c7ffe17e86f41469c950044233006'. Below that, there's a 'Protocol' dropdown menu set to 'HTTP' and a 'Format' dropdown menu set to 'JSON'. A table with columns 'Parameter', 'Value', 'Type', 'Location', and 'Description' is shown. The first row has 'q' as the parameter, 'Ha Noi' as the value, 'string' as the type, 'query' as the location, and a description about passing US Zipcode, UK Postcode, etc. Below this table, there's another section with tabs: Current, Forecast, Search/Autocomplete, History, Future, Astronomy, Time Zone, and Sports. The 'Search/Autocomplete' tab is active. Below the tabs, there's another table with columns 'Parameter', 'Value', 'Type', 'Location', and 'Description'. The first row has 'dt' as the parameter, '2022-06-29' as the value, 'string' as the type, 'query' as the location, and a description about the date format. A 'Contact us' button is visible on the right side.

Hình 2.2. Giao diện phần nhập thông tin cần lấy dữ liệu

- Và trang web sẽ cung cấp cho mình một địa chỉ API chứa thông tin về lịch sử thời tiết trong ngày đó
- Tiếp theo ta sẽ sử dụng code để lấy ra dữ liệu nhiều hơn:
 - + Import các thư viện cần thiết
 - + Tạo ra 2 biến kiểu string ‘base_url’ có chứa link URL của trang web trên, ‘api_key’ là API Key của mình và 1 mảng có tên là ‘city_list’ chứa 25 tỉnh của miền Bắc Việt Nam.
 - + Viết một hàm có tên là ‘find_position’ để tìm vị trí của một thành phố trong danh sách ‘city_list’. Nếu thành phố được tìm thấy, vị trí của nó sẽ được gán cho biến ‘position’. Nếu thành phố không được tìm thấy trong danh sách, một thông báo lỗi sẽ được in ra màn hình (`print(f'City '{city}' not found in the list')`) và hàm trả về giá trị -1 để chỉ ra rằng thành phố không tồn tại trong danh sách.
 - + Tạo ra 1 vòng lặp để duyệt qua từng thành phố trong ‘city_list’.
 - + Bên trong vòng lặp, tạo ra các mảng để lưu trữ thông tin về dữ liệu mình muốn lấy.
 - + Biến ‘time’ được đặt là 29-06-2022, và sau đó được chuyển đổi thành đối tượng ‘datetime’ bằng `datetime.datetime.strptime(time, "%d-%m-%Y").date()`, để xác định thời gian bắt đầu lấy thông tin thời tiết.
 - + Chạy một vòng lặp for nữa sẽ lặp qua 366 ngày tiếp theo để lấy ra dữ liệu trong 1 năm tính từ ngày 29/06/2022 và 1 ngày sau đó nữa là 30/06/2023.

- + Bên trong vòng lặp, biến 'time_str' được đặt để chuyển biến 'time_dt' thành chuỗi thời gian tiêu chuẩn để gán vào URL `http://api.weatherapi.com/v1/history.json?key={api_key}&q={city}&dt={time_str}`
- + Sau khi gửi yêu cầu HTTP và nhận được phản hồi, kết quả được chuyển thành dữ liệu dạng JSON bằng 'response.json()'.
- + Sau khi lặp qua tất cả các ngày, một mảng 'summary' được tạo ra để tập hợp lại tất cả các dữ liệu thu thập được theo từng mảng riêng ở trên và chuyển thành dạng DataFrames.
- + Sau đó, vị trí của thành phố hiện tại trong 'city_list' được tìm bằng cách gọi hàm 'find_position(city_list, item)'
- + Nếu thành phố được tìm thấy 'position != -1' tên file CSV được tạo bằng cách ghép tên thành phố và đuôi '.csv', 'filename = f"{city_list[position]}.csv"'
- + Cuối cùng, DataFrame 'df' được chuyển thành kiểu dữ liệu .csv và lưu vào file với tên tương ứng.

```

1 for item in city_list:
2     city = item
3     maxtemp = []
4     mintemp = []
5     avgtemp = []
6     maxwind_kph = []
7     totalprecip_mm = []
8     avgvis_km = []
9     avghumidity = []
10    comment = []
11    UV_index = []
12    date = []
13    time = "29-06-2022"
14    time_dt = datetime.datetime.strptime(time, "%d-%m-%Y").date()
15    for i in range(367):
16        time_str = time_dt.isoformat()
17        api_url = f"http://api.weatherapi.com/v1/history.json?key={api_key}&q={city_list}&dt={time_str}"
18        response = requests.get(api_url)
19        data = response.json()
20        dataday = data['forecast']['forecastday'][0]
21        day = dataday['day']
22        date.append(time_dt)
23        maxtemp.append(day['maxtemp_c'])
24        mintemp.append(day['mintemp_c'])
25        avgtemp.append(day['avgtemp_c'])
26        maxwind_kph.append(day['maxwind_kph'])
27        totalprecip_mm.append(day['totalprecip_mm'])
28        avghumidity.append(day['avghumidity'])
29        comment.append(day['condition']['text'])
30        UV_index.append(day['uv'])
31        time_dt += datetime.timedelta(days=1)
32    summary = np.array([date, maxtemp, mintemp, avgtemp, maxwind_kph, totalprecip_mm, avghumidity, comment, UV_index])
33    column_values = list(range(1, 368))
34    index_values = ['Day', 'Temperature Max', 'Temperature Min', 'Temperature Avg', 'MaxWind_kmh', 'TotalPrecip_mm', 'Humidity Avg', 'Comment', 'UV Index']

```

```

35    df = pd.DataFrame(data = summary,
36                      index = index_values,
37                      columns = column_values)
38    position = find_position(city_list, item)
39    if position != -1:
40        filename = f"{city_list[position]}.csv"
41        df.T.to_csv(filename, index=False)

```

2.3. Lưu trữ dữ liệu

Đầu tiên ta sẽ import thêm thư viện cần sử dụng và kết nối với Google Drive để ta không phải đọc dữ liệu lên nhiều lần, làm tốn thời gian.

```
[ ] 1 import pandas as pd
    2 import glob
    3 import os
```

```
[ ] 1 from google.colab import drive
    2 drive.mount('/content/drive')
    3 os.chdir('/content/drive/MyDrive/DS_PROJECT_FINAL TEST')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Tiếp theo, thực hiện tổng hợp lại dữ liệu của các tỉnh theo vùng để giúp cho việc quản lý dữ liệu được dễ dàng hơn cũng như là phục vụ cho việc sử dụng sau này nếu cần thiết, để không bị thất lạc dữ liệu. Đồng thời, cũng giúp ta có thể đọc dữ liệu của các tỉnh nhanh và tiết kiệm thời gian hơn.

```
1 from openpyxl import Workbook
2 from openpyxl.utils.dataframe import dataframe_to_rows
3
4 tay_bac_bo = ['Lao Cai', 'Yen Bai', 'Dien Bien', 'Lai Chau', 'Son La', 'Hoa Binh']
5 dong_bac_bo = ['Ha Giang', 'Cao Bang', 'Lang Son', 'Tuyen Quang', 'Bac Kan', 'Thai Nguyen', 'Yen Bai', 'Phu Tho', 'Bac Giang', 'Quang Ninh']
6 dong_bang_song_hong = ['Ha Noi', 'Hai Phong', 'Hung Yen', 'Nam Dinh', 'Thai Binh', 'Hai Duong', 'Hoa Binh', 'Ninh Binh', 'Vinh Phuc', 'Bac Ninh', 'Ha Nam']
7 northernVN = ['Tay Bac Bo', 'Dong Bac Bo', 'Dong bang song Hong']
8
9 for vùng, area in zip([tay_bac_bo, dong_bac_bo, dong_bang_song_hong], northernVN):
10     workbook = Workbook()
11     for tỉnh in vùng:
12         filename = f"{tỉnh}.csv"
13         data = pd.read_csv(filename)
14         sheet_name = tỉnh.replace(" ", "_")
15         sheet = workbook.create_sheet(sheet_name)
16         for row in dataframe_to_rows(data, index=False, header=True):
17             sheet.append(row)
18
19     workbook.save(f"{area}.xlsx")
20 data_list = []
21 for filename in glob.glob("*.xlsx"):
22     data = pd.read_excel(filename, sheet_name=None)
23     for sheet_name, sheet_data in data.items():
24         sheet_data['Province'] = sheet_name
25         data_list.append(sheet_data)
26
27 all_data = pd.concat(data_list)
28 with pd.ExcelWriter("Data Weather in Northern VietNam.xlsx") as writer:
29     for province, province_data in all_data.groupby('Province'):
30         province_data.to_excel(writer, sheet_name=province, index=False)
```

CHƯƠNG 3. TIỀN XỬ LÝ DỮ LIỆU

3.1. Kiểm tra các vấn đề của dữ liệu

3.1.1. Đọc dữ liệu

Đọc các sheet trong file dữ liệu theo vùng .xlsx thành bảng DataFrames thông qua hàm 'load_dataframes_from_excel' với tham số truyền vào là 1 file có đuôi là .xlsx.

Lấy ra một bảng dữ liệu đại diện để xem các thuộc tính của chúng và kiểm tra các vấn đề của dữ liệu.

	Day	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	Comment	UV Index
0	2022-06-29	34.5	27.2	30.7	10.8	4.00	70	Heavy rain at times	7
1	2022-06-30	30.4	25.6	27.5	11.5	17.00	82	Moderate or heavy rain shower	7
2	2022-07-01	33.5	25.3	28.9	11.2	8.50	72	Moderate or heavy rain shower	7
3	2022-07-02	36.4	27.0	31.1	18.4	0.90	67	Patchy rain possible	8
4	2022-07-03	34.3	26.9	29.3	12.6	10.00	85	Heavy rain at times	7
5	2022-07-04	34.5	28.5	31.4	11.9	0.00	71	Partly cloudy	8
6	2022-07-05	34.3	27.3	30.7	11.9	1.17	75	Cloudy	7
7	2022-07-06	29.8	26.2	27.9	10.4	6.16	87	Moderate rain at times	6
8	2022-07-07	28.8	26.1	27.2	9.7	19.30	86	Moderate or heavy rain shower	6
9	2022-07-08	31.0	25.2	27.7	10.4	5.30	80	Patchy rain possible	7

In ra một vài thông tin cơ bản về dữ liệu: bảng gồm bao nhiêu hàng, bao nhiêu cột; số lượng giá trị định tính và định lượng, các thuộc tính của dữ liệu là gì, kiểu dữ liệu của các thuộc tính trong bảng đó là gì.

```
[10] 1 dataFrames['Ha_Noi'].shape
(367, 9)

[11] 1 dataFrames['Ha_Noi'].info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Day                   367 non-null   object
1   Temperature Max       367 non-null   float64
2   Temperature Min       367 non-null   float64
3   Temperature Avg       367 non-null   float64
4   MaxWind_kmh/h         367 non-null   float64
5   TotalPrecip_mm        367 non-null   float64
6   Humidity Avg          367 non-null   int64
7   Comment               367 non-null   object
8   UV Index              367 non-null   int64
dtypes: float64(5), int64(2), object(2)
memory usage: 25.9+ KB

[12] 1 dataFrames['Ha_Noi'].columns
Index(['Day', 'Temperature Max', 'Temperature Min', 'Temperature Avg',
       'MaxWind_kmh/h', 'TotalPrecip_mm', 'Humidity Avg', 'Comment',
       'UV Index'],
      dtype='object')
```

Ta rút ra được nhận xét ban đầu cho dữ liệu như sau:

- Thông qua bảng data đại diện của Hà Nội, thì ta có thể rút ra nhận xét, mỗi bảng dữ liệu đều gồm có 367 dòng (thời gian từ 29/06/2022 - 30/06/2023) và 9 cột thuộc tính gồm (2 biến định tính và 7 biến định lượng)
- Tuy nhiên kiểu dữ liệu cột 'Date' ở đây đang lưu chưa đúng kiểu dữ liệu

=> **Chỉnh sửa về đúng kiểu dữ liệu của thời gian là 'datetime'**

- Ý nghĩa của các thuộc tính:
 - + `Day`: Ngày quan sát lấy dữ liệu
 - + `Temperature Max`: Nhiệt độ cao nhất(độ C) trong ngày đó.
 - + `Temperature Min`: Nhiệt độ thấp nhất(độ C) trong ngày đó.
 - + `Temperature Avg`: Nhiệt độ trung bình(độ C) trong ngày đó.
 - + `MaxWind_kmh`: Sức(Tốc độ) gió lớn nhất(km/h) trong ngày đó.
 - + `TotalPrecip_mm`: Tổng lượng mưa(mm) trong ngày đó.
 - + `Humidity Avg`: Độ ẩm trung bình(%) trong ngày đó.
 - + `Comment`: Các ghi chú hoặc thông tin dự báo khác về ngày đó.
 - + `UV Index`: Chỉ số tia tử ngoại trong ngày đó.

3.1.2. Kiểm tra và phát hiện các vấn đề của dữ liệu:

Trong bước này, ta sẽ đi thực hiện kiểm tra một vài lỗi điển hình trong dữ liệu như là: Kiểm tra xem dữ liệu có chứa NULL không, kiểm tra dữ liệu có bị lỗi logic không, có bị trùng lặp dữ liệu không...

▼ Kiểm tra giá trị NULL

```
[15] 1 dataFrames['Ha_Noi'].isnull().sum()

Day                0
Temperature Max    0
Temperature Min    0
Temperature Avg    0
MaxWind_kmh        0
TotalPrecip_mm     0
Humidity Avg       0
Comment            0
UV Index           0
dtype: int64
```

Dữ liệu không có giá trị Null

▼ Kiểm tra lỗi trùng lặp dữ liệu

```
[16] 1 dataFrames['Ha_Noi']['Day'].duplicated().sum()

0
```

Dữ liệu về ngày không có ngày nào bị lặp lại => Không có ngày nào được quan sát nhiều hơn 1 lần

▼ Kiểm tra lỗi Logic

```
[17] 1 (dataFrames['Ha_Noi']['Temperature Max'] < dataFrames['Ha_Noi']['Temperature Min']).sum()

0
```

Dữ liệu về nhiệt độ không có lỗi về logic, nhiệt độ lớn nhất nhỏ hơn nhiệt độ thấp nhất

✓

0

giây

[18]

1 dataFrames['Ha_Noi'][(dataFrames['Ha_Noi']['Temperature Max'] + dataFrames['Ha_Noi']['Temperature Min'])/2 != dataFrames['Ha_Noi']['Temperature Avg']]

	Day	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	Comment	UV Index
0	2022-06-29	34.5	27.2	30.7	10.8	4.0	70	Heavy rain at times	7
1	2022-06-30	30.4	25.6	27.5	11.5	17.0	82	Moderate or heavy rain shower	7
2	2022-07-01	33.5	25.3	28.9	11.2	8.5	72	Moderate or heavy rain shower	7
3	2022-07-02	36.4	27.0	31.1	18.4	0.9	67	Patchy rain possible	8
4	2022-07-03	34.3	26.9	29.3	12.6	10.0	85	Heavy rain at times	7
...
362	2023-06-26	30.5	24.1	26.5	12.6	32.6	88	Moderate or heavy rain shower	7
363	2023-06-27	32.4	24.7	28.2	17.6	9.3	81	Light rain shower	7
364	2023-06-28	33.6	26.0	29.5	19.4	2.6	78	Patchy rain possible	7
365	2023-06-29	35.9	26.5	30.6	20.9	1.7	75	Light rain shower	8
366	2023-06-30	38.4	27.0	31.8	18.7	5.5	73	Light rain shower	8

358 rows x 9 columns

Nguyên nhân xảy ra vấn đề này có thể là do ở đây mình chỉ lấy dữ liệu nhiệt độ lớn nhất và nhỏ nhất trong ngày, nhưng dữ liệu nhiệt độ trung bình lại có thể là trung bình cộng của nhiệt độ lớn nhất và nhỏ nhất hàng giờ của ngày đó. Do vậy nó sẽ có một vài sai số.

✓

0

giây

[19]

1 # Đếm số lượng giá trị trong cột 'MaxWind_kmh' nhỏ hơn 0
2 maxwind_below_zero = (dataFrames['Ha_Noi']['MaxWind_kmh'] < 0).sum()
3
4 # Đếm số lượng giá trị trong cột 'TotalPrecip_mm' nhỏ hơn 0
5 precip_below_zero = (dataFrames['Ha_Noi']['TotalPrecip_mm'] < 0).sum()
6
7 # Đếm số lượng giá trị trong cột 'Humidity Avg' nhỏ hơn 0
8 humidity_below_zero = (dataFrames['Ha_Noi']['Humidity Avg'] < 0).sum()
9
10 # Đếm số lượng giá trị trong cột 'UV Index' nhỏ hơn 0
11 uv_below_zero = (dataFrames['Ha_Noi']['UV Index'] < 0).sum()

Dữ liệu tại các cột `MaxWind_kmh`, `TotalPrecip_mm`, `Humidity Avg` và `UV Index` không có giá trị nào bất thường và lỗi logic

Nhận xét:

- Dữ liệu thu thập được ở đây gần như đã được xử lý và làm sạch từ trước cho nên nó tương đối là sạch sẽ và không có lỗi, giá trị NULL hay trùng lặp trong các thuộc tính dữ liệu.
- Số lượng quan sát đủ lớn và đủ khách quan, chân thực để đưa vào phân tích và đánh giá kết quả.
- Tên của các thuộc tính được đặt khá dễ hiểu và rõ ràng, giúp người phân tích có thể dễ dàng định hình vấn đề

CHƯƠNG 4. PHÂN TÍCH DỮ LIỆU – EXPLORATORY DATA ANALYSIS

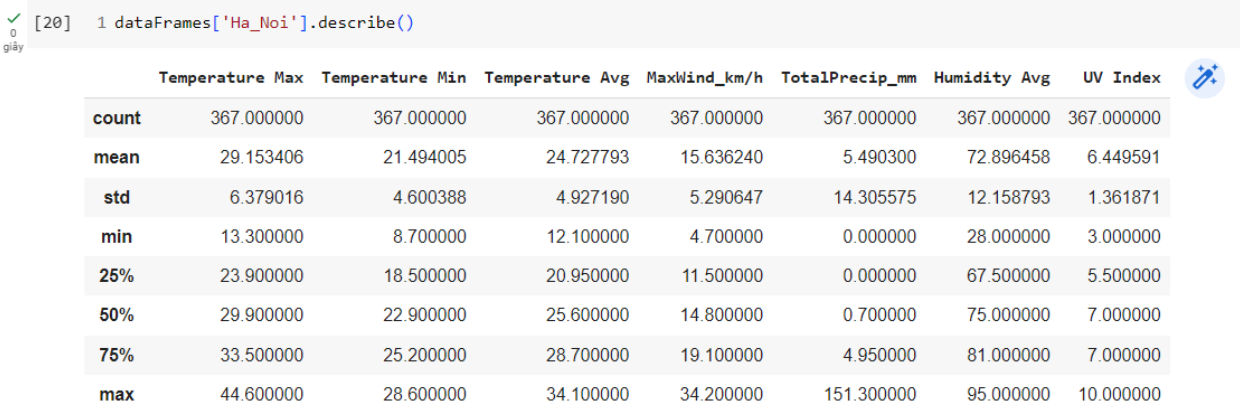
4.1. Thống kê mô tả

Trong phần phân tích dữ liệu này, ta sẽ tiếp cận vấn đề theo cách sẽ đi phân tích một vài tỉnh đại diện trong vùng đó và phân tích dữ liệu của cả vùng đó bằng bảng dữ liệu trung bình của các tỉnh.

4.1.1. Đồng bằng sông Hồng

❖ Hà Nội:

Đầu tiên, ta sẽ sử dụng hàm `describe()` để có được bảng thống kê mô tả cơ bản về dữ liệu thời tiết của thành phố Hà Nội như sau:



```
[20] 1 dataFrames['Ha_Noi'].describe()
```

	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	UV Index
count	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000
mean	29.153406	21.494005	24.727793	15.636240	5.490300	72.896458	6.449591
std	6.379016	4.600388	4.927190	5.290647	14.305575	12.158793	1.361871
min	13.300000	8.700000	12.100000	4.700000	0.000000	28.000000	3.000000
25%	23.900000	18.500000	20.950000	11.500000	0.000000	67.500000	5.500000
50%	29.900000	22.900000	25.600000	14.800000	0.700000	75.000000	7.000000
75%	33.500000	25.200000	28.700000	19.100000	4.950000	81.000000	7.000000
max	44.600000	28.600000	34.100000	34.200000	151.300000	95.000000	10.000000

Hình 4.1. Bảng thống kê mô tả của tập dữ liệu về thời tiết của Hà Nội

Sau đó, ta rút ra được một vài nhận xét:

Nhìn sơ qua bảng thống kê mô tả trên về tập dữ liệu của Hà Nội thì ta có thể rút ra một vài đánh giá ban đầu như sau:

- Nhiệt độ cao nhất hàng ngày trung bình khoảng 29.2 độ C với khoảng biến động là 6.38 độ C, ngày có nhiệt độ cao nhất thu được là 44.6 độ C.
- Nhiệt độ thấp nhất hàng ngày trung bình khoảng 21.5 độ C với khoảng biến động là 4.6 độ C, ngày có nhiệt độ thấp nhất thu được là 8.7 độ C.
- Nhiệt độ trung bình hàng ngày là khoảng 24.73 độ C, với khoảng biến động là 4.93 độ C.
- Tốc độ gió trung bình hàng ngày là khoảng 15.64 km/h, với độ lệch chuẩn là khoảng 5.29 km/h. Giá trị nhỏ nhất là 4.7 km/h và lớn nhất là 34.2 km/h.
- Tổng lượng mưa trung bình trong thời gian trên là khoảng 5.49 mm, và độ lệch chuẩn là khoảng 14.31 mm. Giá trị nhỏ nhất là 0.0 mm là những ngày không có mưa và giá trị lớn nhất là 151.3 mm.

- Độ ẩm trung bình hàng ngày là khoảng 72.9%, với độ lệch chuẩn là khoảng 12.16%. Giá trị nhỏ nhất là 28% và giá trị lớn nhất là 95%.
- Chỉ số tia UV hàng ngày có giá trị trung bình là khoảng 6.45, với độ lệch chuẩn là khoảng 1.36. Giá trị nhỏ nhất là 3 và giá trị lớn nhất là 10. Điều này cho thấy rằng là nhiệt độ của Trái Đất đang ngày một nóng lên rất nhiều, khiến cho chỉ số tia UV đang ngày càng tăng cao, thấp nhất đã là mức nguy hại trung bình(3-5) và cao nhất đã đạt tới gần đỉnh điểm là nguy hại rất cao là 10.

Ngoài ra, nhóm em còn có lấy ra một vài dữ liệu có thông tin đặc biệt ở trong file code, hơi dài nên nhóm em sẽ chỉ nhặt lại những cái gì quan trọng để viết báo cáo.

❖ Các tỉnh đồng bằng sông Hồng:

Để phân tích dữ liệu cho tất cả các tỉnh trong vùng đồng bằng sông Hồng được khách quan và chính xác nhất thì nhóm em sẽ lấy trung bình tất cả các giá trị trong từng dataFrames của mỗi tỉnh theo ngày và thu được bảng như sau:

	Day	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	UV Index
0	2022-06-29	38.34	30.48	33.80	13.89	5.500	79.6	8.3
1	2022-06-30	35.18	28.32	31.03	12.23	25.120	88.0	7.6
2	2022-07-01	35.87	28.05	31.43	12.21	15.800	84.1	7.6
3	2022-07-02	39.13	29.26	33.44	18.09	9.970	77.1	8.7
4	2022-07-03	35.66	28.53	31.14	16.34	13.331	97.5	7.8
...
362	2023-06-26	33.68	27.30	29.91	14.68	28.270	93.5	7.6
363	2023-06-27	34.77	27.30	30.55	18.97	23.570	92.3	7.7
364	2023-06-28	36.58	28.88	32.20	21.93	7.540	88.4	7.7
365	2023-06-29	38.36	29.38	33.14	21.38	3.880	86.6	8.6
366	2023-06-30	41.53	30.08	34.39	19.44	3.490	83.9	8.8

367 rows x 8 columns

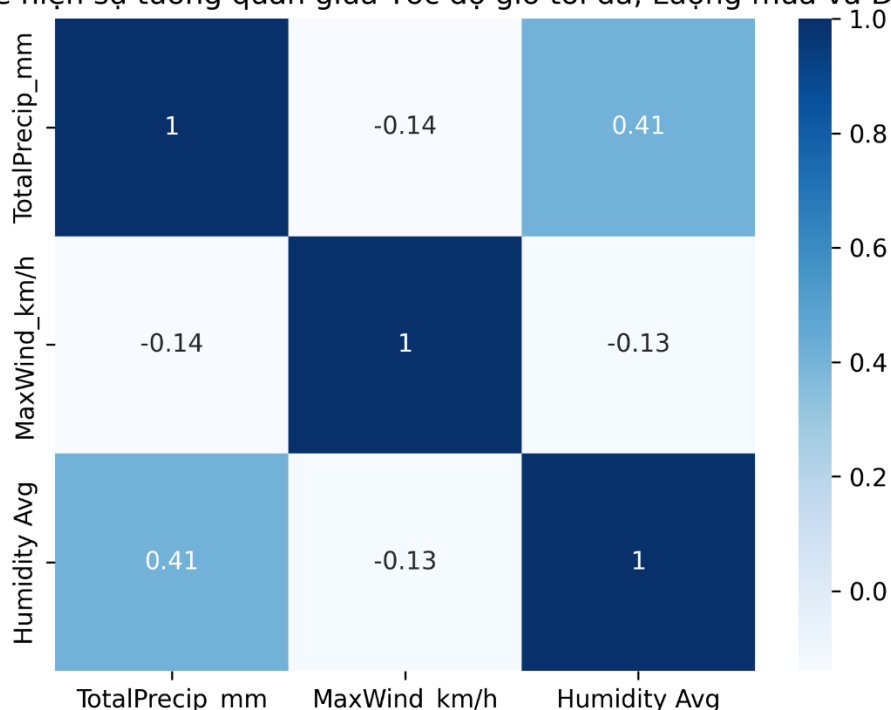
Kiểm tra kiểu dữ liệu của các thuộc tính xem có chính xác hay chưa:

```
1 average_df.dtypes
```

```
Day                datetime64[ns]
Temperature Max    float64
Temperature Min    float64
Temperature Avg    float64
MaxWind_kmh       float64
TotalPrecip_mm    float64
Humidity Avg      float64
UV Index          float64
dtype: object
```

Tính một vài giá trị thống kê mô tả, từ đó vẽ biểu đồ hệ số tương quan để đánh giá xem các biến có mối tương quan với nhau như thế nào.

Biểu đồ thể hiện sự tương quan giữa Tốc độ gió tối đa, Lượng mưa và Độ ẩm



Hình 4.2. Biểu đồ thể hiện sự tương quan giữa tốc độ gió tối đa, lượng mưa và độ ẩm trung bình của Đồng bằng sông Hồng

Nhận xét:

Các giá trị trong biểu đồ thể hiện mức độ tương quan tuyến tính giữa các biến.

- Nếu giá trị càng gần 1 (màu xanh lam thẫm) thì cho thấy quan hệ tuyến tính dương giữa hai biến càng mạnh, hai biến đồng biến với nhau.
- Ngược lại, nếu giá trị càng gần -1 (màu trắng) cho thấy quan hệ tuyến tính âm càng mạnh, hai biến nghịch biến với nhau
- Giá trị gần bằng 0 cho thấy mối tương quan yếu hoặc không có quan hệ tuyến tính giữa 2 biến.

Từ đó ta có thể kết luận về mối tương qua giữa 3 biến này như sau:

- TotalPrecip_mm' (Tổng lượng mưa) và 'HumidityAvg' (Độ ẩm trung bình) có mối tương quan dương mạnh với nhau (0.41). Tức là là khi lượng mưa càng tăng, thì độ ẩm cũng tăng.
- 'MaxWind_km/h' (Tốc độ gió tối đa) gần như không có mối tương quan với 2 biến còn lại 'TotalPrecip_mm' (Tổng lượng mưa) (-0.14) và 'HumidityAvg' (Độ ẩm trung bình) (-0.13). Điều này cho thấy rằng không có mối liên hệ rõ ràng nào giữa tốc độ gió tối đa với lượng mưa và độ ẩm.

4.1.2. Tây Bắc Bộ

❖ Sơn La:

Trong số các tỉnh này, có thể coi Sơn La là một trong những tỉnh quan trọng nhất. Sơn La không chỉ có vị trí chiến lược trên tuyến giao thông quan trọng từ phía Tây sang Hà Nội và các khu vực khác, mà còn là một tỉnh có diện tích lớn và đa dạng về địa hình, tài nguyên tự nhiên và đa dạng các dân tộc anh em đang sinh sống ở đây.

```
[151] 1 f5.describe()
```

	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	UV Index
count	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000
mean	28.875204	17.868392	22.444414	6.342779	6.160490	73.419619	6.517711
std	5.658183	3.857706	4.072273	2.702083	12.739744	17.541627	1.374668
min	13.000000	6.500000	10.900000	2.200000	0.000000	23.000000	3.000000
25%	25.350000	14.900000	20.100000	4.600000	0.000000	64.000000	6.000000
50%	28.600000	19.100000	22.700000	5.400000	0.400000	76.000000	6.000000
75%	31.750000	20.900000	25.200000	7.600000	5.450000	87.000000	7.000000
max	44.700000	24.400000	32.900000	19.100000	80.300000	100.000000	10.000000

Hình 4.3. Bảng thống kê mô tả của tập dữ liệu về thời tiết của tỉnh Sơn La

Nhận xét:

Nhìn sơ qua bảng thống kê mô tả trên về tập dữ liệu của Sơn La thì ta có thể rút ra một vài đánh giá ban đầu như sau:

- Nhiệt độ:
 - + Nhiệt độ cao nhất hàng ngày trung bình khoảng 28.9 độ C, có thể đạt tới 44.7 độ C. Điều này cho thấy Sơn La Có thể trải qua những ngày thời tiết nóng gay gắt nhiều.
 - + Nhiệt độ thấp nhất hàng ngày trung bình khoảng 17.9 độ C, có ngày có nhiệt giảm xuống thấp nhất là 6.5 độ C.
 - + Nhiệt độ trung bình hàng ngày là khoảng 22.4 độ C, với khoảng biến động là 4.07 độ C. Các con số này cho thấy khá ổn định và đồng đều về mức độ nhiệt độ trung bình.
- Tốc độ gió: Tốc độ gió trung bình hàng ngày là khoảng 6.34 km/h, với độ lệch chuẩn là khoảng 2.70 km/h. Giá trị nhỏ nhất là 2.2 km/h và lớn nhất là 19.1 km/h. Điều này cho ta thấy tốc độ gió không quá mạnh, ổn định trong khoảng giá trị này.
- Lượng mưa: Tổng lượng mưa trung bình trong thời gian trên là khoảng 6.16 mm, và độ lệch chuẩn là khoảng 12.74 mm. Giá trị nhỏ nhất là 0.0 mm là những

ngày không có mưa và giá trị lớn nhất là 80.3 mm, điều này cho thấy Lạng Sơn đã có thể trải qua những trận mưa lớn.

- Độ ẩm: Độ ẩm trung bình hàng ngày là khoảng 73.42%, với độ lệch chuẩn là khoảng 17.54%. Giá trị nhỏ nhất là 23% và giá trị lớn nhất là 100%, qua đó ta thấy rằng độ ẩm của Lạng Sơn có sự biến đổi lớn theo ngày.
- Chỉ số tia UV: Chỉ số tia cực tím hàng ngày có giá trị trung bình là khoảng 6.52, với độ lệch chuẩn là khoảng 1.37. Giá trị nhỏ nhất là 3 và giá trị lớn nhất là 10.

Tổng quan, dữ liệu này cho thấy Lạng Sơn có khí hậu nóng ẩm và có thể trải qua những ngày nóng nhiều, những ngày có mưa và sự biến đổi về nhiệt độ và độ ẩm. Chỉ số tia cực tím (UV Index) cũng cho thấy tình trạng nguy hại với mức độ tăng dần.

❖ Các tỉnh vùng Tây Bắc Bộ:

Để phân tích dữ liệu cho tất cả các tỉnh được khách quan và chính xác nhất thì nhóm em sẽ lấy trung bình tất cả các giá trị trong từng dataFrames của mỗi tỉnh theo ngày và tính một vài giá trị thống kê mô tả, từ đó vẽ biểu đồ hệ số tương quan để đánh giá xem các biến có mối tương quan với nhau như thế nào.

4.1.3. Đông Bắc Bộ

❖ Yên Bái:

Trong số các tỉnh này, Yên Bái là một tỉnh có nền kinh tế chủ yếu dựa vào ngành nông nghiệp. Nổi tiếng với đất canh tác rộng lớn vì vậy việc phân tích khí hậu ở tỉnh này giúp ta nắm được khí hậu thời tiết thích hợp để nuôi trồng

```
1 f7.describe()
```

	Temperature Max	Temperature Min	Temperature Avg	MaxWind_kmh	TotalPrecip_mm	Humidity Avg	UV Index
count	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000	367.000000
mean	28.670572	21.099728	24.343052	12.630518	6.017411	75.223433	6.365123
std	6.219725	4.187690	4.659716	4.775724	13.288482	12.107181	1.333750
min	12.900000	9.700000	12.500000	3.800000	0.000000	27.000000	3.000000
25%	23.450000	18.250000	20.900000	9.000000	0.000000	68.500000	5.000000
50%	29.500000	22.500000	25.200000	11.900000	1.000000	78.000000	6.000000
75%	33.000000	24.500000	28.000000	15.500000	5.150000	83.000000	7.000000
max	43.700000	27.300000	33.400000	36.400000	107.600000	96.000000	10.000000

Hình 4.4. Bảng thống kê mô tả của tập dữ liệu về thời tiết của tỉnh Yên Bái

Nhận xét:

Nhìn sơ qua bảng thống kê mô tả trên về tập dữ liệu của Yên Bái thì ta có thể rút ra một vài đánh giá ban đầu như sau:

- Nhiệt độ:
 - + Nhiệt độ cao nhất hàng ngày trung bình khoảng 28.6 độ C, có thể đạt tới 43.7 độ C. Điều này cho thấy Yên Bái có thể trải qua những ngày nắng nóng nhiều.
 - + Nhiệt độ thấp nhất hàng ngày trung bình khoảng 21.09 độ C trong ngày có nhiệt giảm xuống thấp nhất là 4.1 độ C.
 - + Nhiệt độ trung bình hàng ngày là khoảng 24,3 độ C, với khoảng biến động là 4.6 độ C. Các con số này cho thấy khá ổn định và thoải mái về mức độ nhiệt độ trung bình.
- Tốc độ gió: Tốc độ gió trung bình hàng ngày là khoảng 12.63 km/h, với độ lệch chuẩn là khoảng 4.7 km/h. Giá trị nhỏ nhất là 3.8 km/h và lớn nhất là 36.4 km/h. Điều này cho ta thấy tốc độ gió không quá mạnh, ổn định trong khoảng giá trị này
- Lượng mưa: Tổng lượng mưa trung bình trong thời gian trên là khoảng 6.01mm, và độ lệch chuẩn là khoảng 13.28 mm. Giá trị nhỏ nhất là 0.0 mm là những ngày không có mưa và giá trị lớn nhất là 107 mm. Điều này cho thấy Lạng Sơn đã có thể trải qua những trận mưa lớn trong năm.
- Độ ẩm: Độ ẩm trung bình hàng ngày là khoảng 75.2%, với độ lệch chuẩn là khoảng 12.1%. Giá trị nhỏ nhất là 27% và giá trị lớn nhất là 96%. Điều này cho thấy rằng độ ẩm Lạng Sơn có sự biến đổi lớn.
- Chỉ số tia UV: Chỉ số tia cực tím hàng ngày có giá trị trung bình là khoảng 6.3, với độ lệch chuẩn là khoảng 1.3. Giá trị nhỏ nhất là 3 và giá trị lớn nhất là 10.

Tổng quan, dữ liệu này cho thấy Yên Bái có khí hậu nóng ẩm và có thể trải qua những ngày nóng nhiều, những ngày có mưa và sự biến đổi về nhiệt độ và độ ẩm. Chỉ số tia cực tím (UV Index) cũng cho thấy tình trạng nguy hại với mức độ tăng dần.

❖ Các tỉnh vùng Đông Bắc Bộ:

Ta cũng xử lý vấn đề như 2 vùng trên, cụ thể trong file code.

4.2. Trực quan hóa dữ liệu

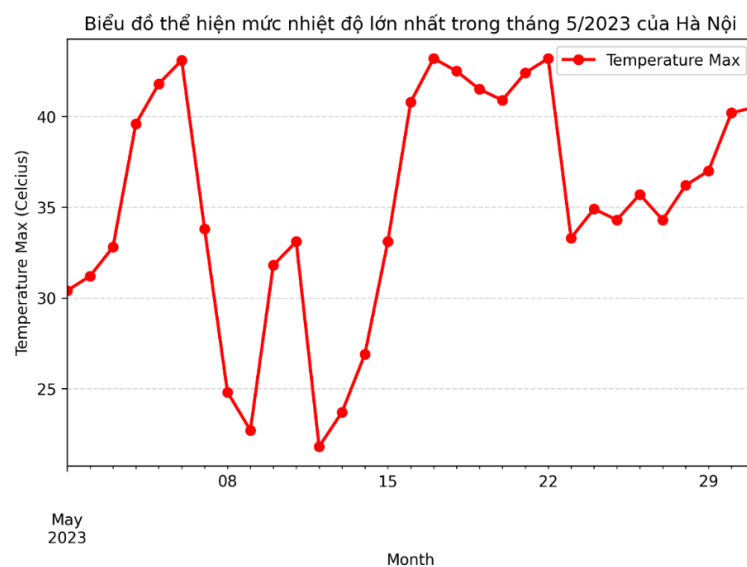
Trong phần này, ta sẽ sử dụng các thư viện vẽ biểu đồ để quan sát, trực quan và từ đó rút ra các nhận xét, đánh giá về dữ liệu.

4.2.1. Đồng bằng sông Hồng

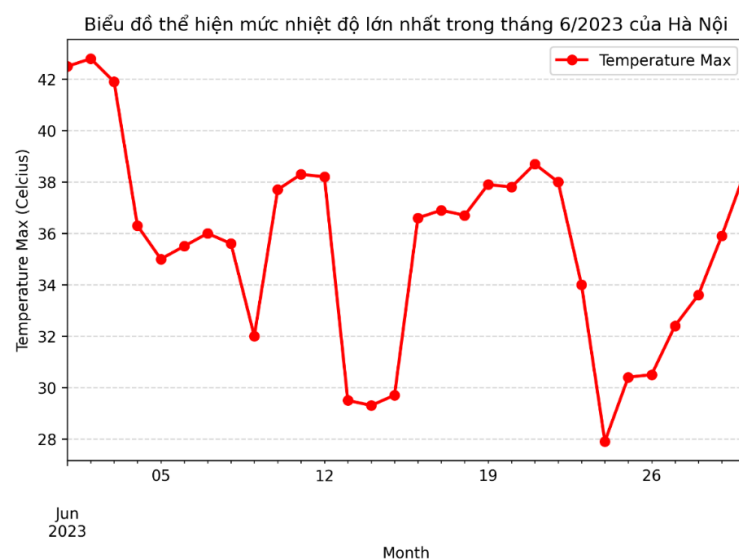
- ❖ **Phân tích và biểu diễn dữ liệu về nhiệt độ cao nhất gần đây trong 2 tháng Tháng 5 và 6 của Hà Nội để có những cái nhìn trực quan hơn ban đầu về thời tiết của miền Bắc hiện tại:**

Trước tiên ta sẽ lọc ra dữ liệu của 2 tháng này trước, rồi sau đó chuyển thành bảng DataFrames gồm thông tin về thời tiết của tháng 5 và 6 của Hà Nội.

Sau đó ta sẽ đi vẽ biểu đồ nhiệt độ của Tháng 5 và 6:



Hình 4.5. Biểu đồ nhiệt độ lớn nhất trong tháng 5/2023 của Hà Nội



Hình 4.6. Biểu đồ nhiệt độ lớn nhất trong tháng 6/2023 của Hà Nội

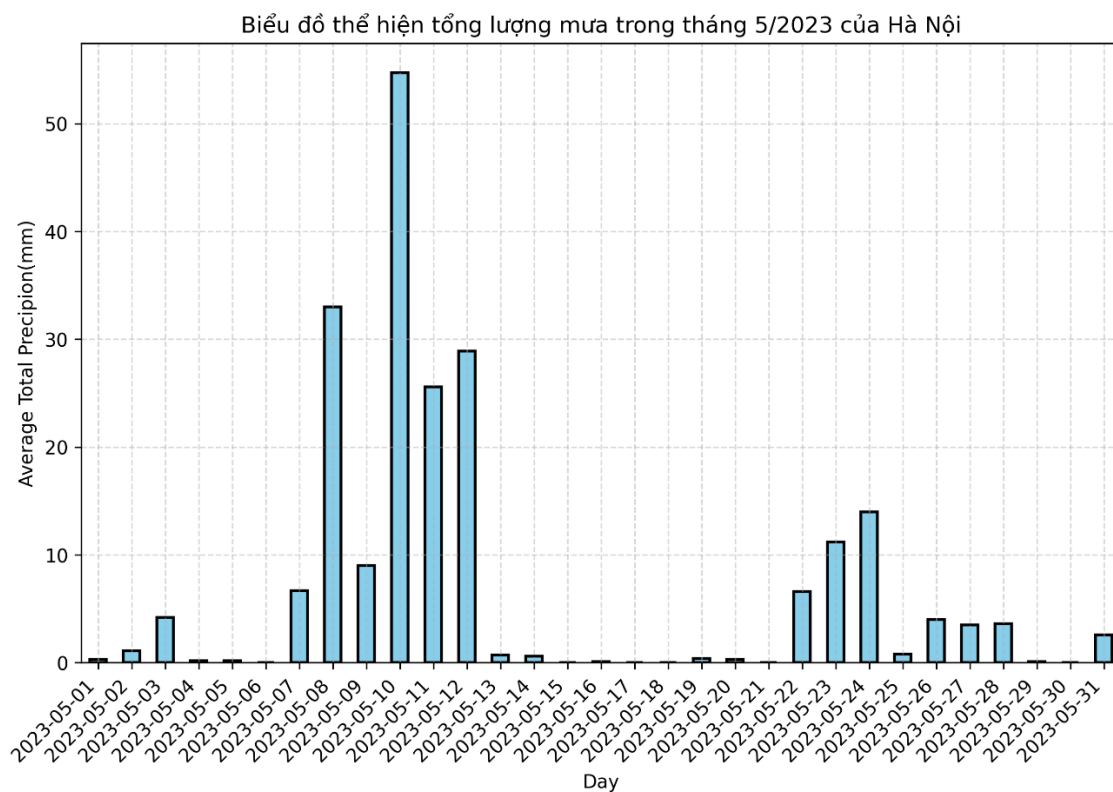
Nhận xét: Nhìn vào biểu đồ nhiệt độ lớn nhất ở 2 tháng nóng nhất của mùa hè tháng 5 và 6 của Hà Nội ta có thể rút ra một vài nhận xét sau:

- Tháng 6 là tháng có mức nhiệt độ cao nhất, hơn 60% nhiệt độ trong tháng đều ≥ 35 độ C, nhưng nhiệt độ cao nhất trong tháng chỉ là 42.8 độ C.
- Tháng 5 cũng là tháng có mức nhiệt độ cao, khoảng 50% nhiệt độ trong tháng đều ≥ 35 độ C, nhưng có ngày lại có nhiệt độ cao hơn cả lên đến 43.2 độ C (22/5/2023).

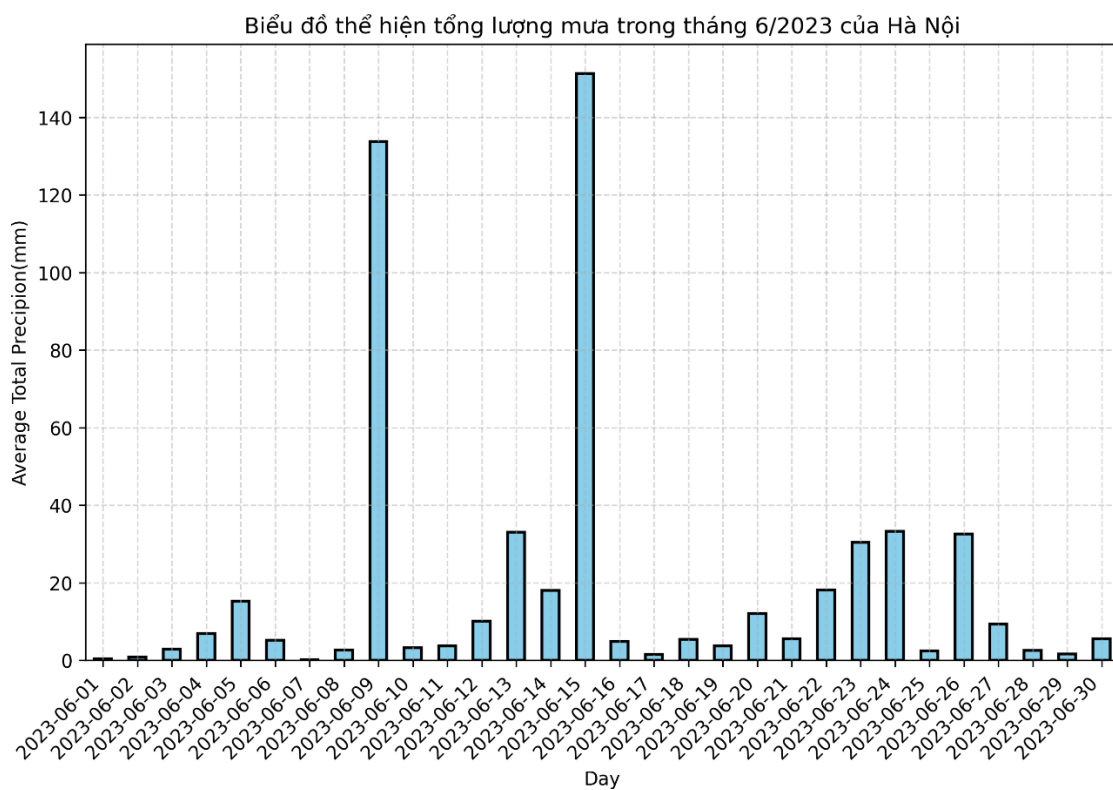
=> Điều này cho thấy rằng nhiệt độ của chúng ta đang ngày càng tăng lên do tác động của hiện tượng nóng lên toàn cầu, nhiệt độ cao cũng khiến cho chỉ số UV ngày càng tăng hơn và nguy hiểm hơn. Do vậy, chúng ta cần phải hành động gì đó để làm sao hạn chế được tác động của hiện tượng trên:

- Giảm lượng khí thải: Tăng cường sử dụng và phát triển các nguồn năng lượng tái tạo như năng lượng mặt trời, gió, thủy điện, hạt nhân thay vì năng lượng từ các nguồn hóa thạch.
- Bảo vệ rừng và điều chỉnh việc sử dụng đất: Rừng, đặc biệt là rừng nhiệt đới, có khả năng hấp thụ lượng lớn khí CO₂ và giảm hiệu ứng nhà kính. Việc bảo vệ và phục hồi rừng giúp giảm lượng CO₂ trong khí quyển. Đồng thời, ứng dụng các phương pháp bền vững trong việc sử dụng đất, như trồng cây bảo vệ, quản lý đất và hồ, có thể giảm lượng khí thải và bảo vệ hệ sinh thái.
- Sử dụng công nghệ năng lượng hiệu quả: Tăng cường công nghệ sử dụng năng lượng hiệu quả để giảm lượng năng lượng tiêu thụ và khí thải CO₂. Điển hình có thể kể đến là việc sử dụng đèn LED thay thế đèn huỳnh quang, sử dụng thiết bị điện tiết kiệm năng lượng, cải tiến năng suất nhiệt và điện trong các ngành công nghiệp...
- Khuyến khích sử dụng vận tải công cộng và giao thông bền vững: Giảm sự phụ thuộc vào xe hơi cá nhân bằng cách khuyến khích sử dụng phương tiện công cộng, đạp xe, đi bộ và điều chỉnh chính sách về giao thông công cộng. Đồng thời, xây dựng và phát triển giao thông công cộng và hạ tầng giao thông bền vững.
- Khuyến khích sử dụng năng lượng xanh: Khuyến khích sử dụng năng lượng tái tạo như năng lượng mặt trời, gió và thủy điện. Đồng thời, hỗ trợ và đầu tư vào nghiên cứu và phát triển công nghệ năng lượng xanh để giảm giá thành và tăng hiệu suất của các nguồn năng lượng tái tạo.

Tiếp nữa là biểu đồ tổng lượng mưa của tháng này:



Hình 4.7. Biểu đồ tổng lượng mưa trung bình trong tháng 5/2023 của Hà Nội



Hình 4.8. Biểu đồ tổng lượng mưa trung bình trong tháng 6/2023 của Hà Nội

Nhận xét: Như nhận xét phần nhiệt độ ở trên kết hợp với quan sát biểu đồ lượng mưa, ta có thể thấy:

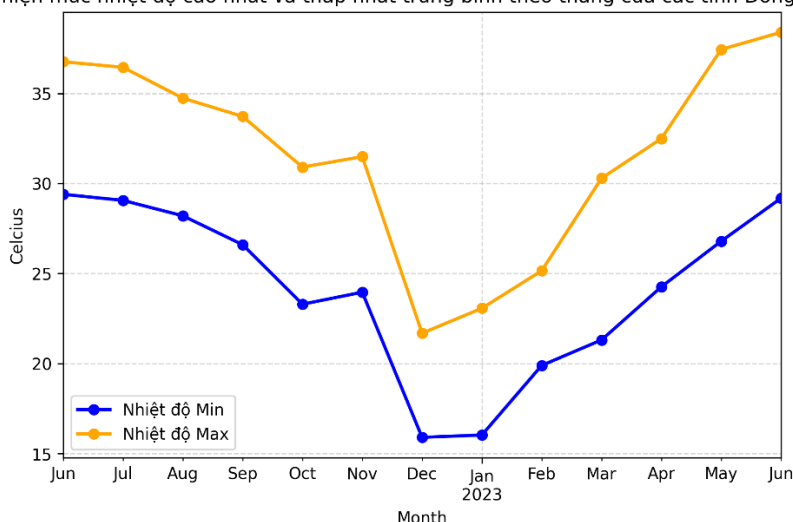
Tháng 5 là tháng có lượng mưa ít nhất trong các tháng, nhìn vào biểu đồ ta có thể thấy rất khô hạn, lượng mưa lớn nhất trong tháng cũng chỉ khoảng 60mm, lượng nước không đáng kể.

=> Chính vì thế mà ta có thể thấy rằng trong tháng 5 vừa qua nắng nóng, khô hạn kéo dài + mưa ít khiến cho nguồn điện từ các nhà máy thủy điện tạo ra khá là ít, khiến cho Hà Nội trong khoảng thời gian này bị cắt điện gián đoạn.

Tháng 6 là tháng có nền nhiệt độ cả tháng cao nhất trong các tháng và cũng ít mưa, chỉ có duy nhất 2 ngày có lượng mưa cao nhất là trên 130mm, còn lại thì vẫn có mưa nhưng không đáng kể dưới 40mm.

❖ **Biểu đồ thể hiện mức nhiệt độ cao nhất và thấp nhất trung bình theo tháng của các tỉnh đồng bằng sông Hồng:**

Biểu đồ thể hiện mức nhiệt độ cao nhất và thấp nhất trung bình theo tháng của các tỉnh Đồng bằng sông Hồng



Hình 4.9. Biểu đồ nhiệt độ cao nhất và thấp nhất trung bình trong mỗi tháng của các tỉnh Đồng bằng sông Hồng

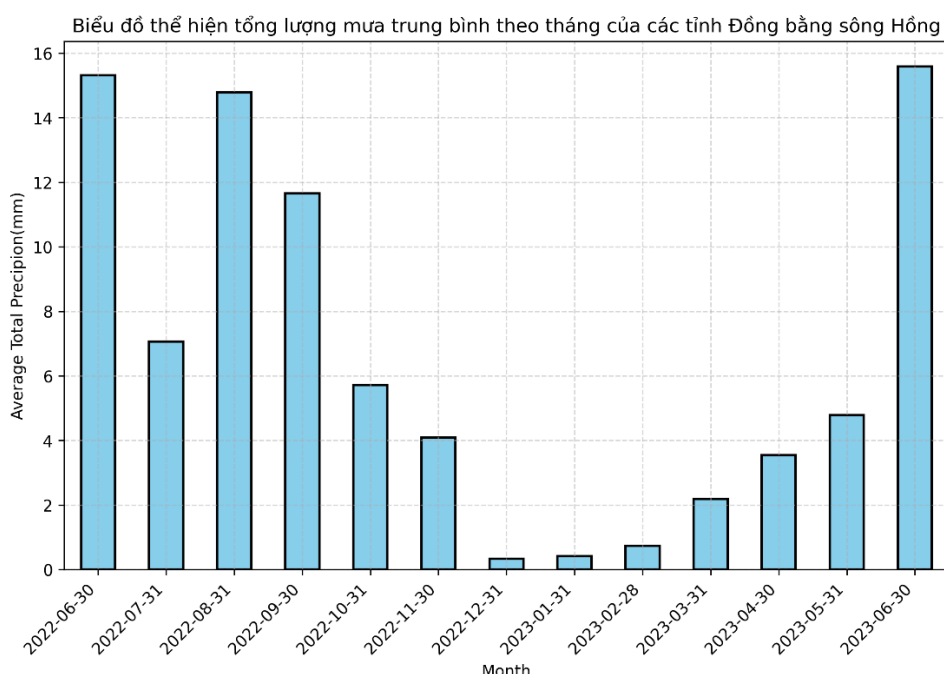
Nhận xét:

- Nhiệt độ thấp nhất rơi vào khoảng tháng 12 và tháng 1. Đây chính là khoảng thời gian mùa Đông đang diễn ra ở miền Bắc. Tuy nhiên nhiệt độ thấp nhất ở đây chưa dừng lại ở đó, nếu ta quan sát kỹ hơn về nhiệt độ của những tỉnh ở vùng núi Tây Bắc Bộ như Sơn La, Lai Châu thì có những thời điểm nếu năm đó có rét đậm rét hại thì nhiệt độ thấp nhất còn có thể dưới âm độ C, nhưng ở đây do ta đã lấy trung bình nên nhiệt độ thấp nhất chung cho toàn miền Bắc sẽ là như khoảng trên. Nói sâu hơn về mặt địa lý, đặc điểm địa hình của các tỉnh vùng Tây Bắc Bộ này ta thấy đây là nơi có địa hình cao nhất ở nước ta với những dãy

núi cao đồ sộ thuộc dãy Hoàng Liên Sơn ở phía Đông, có đỉnh Phan-xi-păng cao 3143m được coi là nóc nhà của Đông Dương, phía Tây là địa hình đồi núi trung bình với các sơn nguyên và cao nguyên đá vôi hùng vĩ. Do tác động ảnh hưởng của địa hình như vậy nên các tỉnh phía Tây Bắc sẽ có sự phân hóa khí hậu theo độ cao. Địa hình như thời tiết của Sa Pa, ta có thể được trải nghiệm du lịch có 4 mùa trong 1 ngày chính là vì điều đó.

- Biểu đồ cho thấy sự biến đổi nhiệt độ theo các tháng trong năm. Có sự chênh lệch rõ rệt giữa nhiệt độ cao nhất và thấp nhất trong mùa Đông (tháng 12 - tháng 1) vào mùa Hè (tháng 5 – tháng 6 – tháng 7). Điều này cho thấy miền Bắc có mùa Đông lạnh và mùa hè nóng, với sự chênh lệch nhiệt độ lớn giữa các mùa.

❖ **Biểu đồ thể hiện tổng lượng mưa trung bình theo tháng của các tỉnh đồng bằng sông Hồng:**

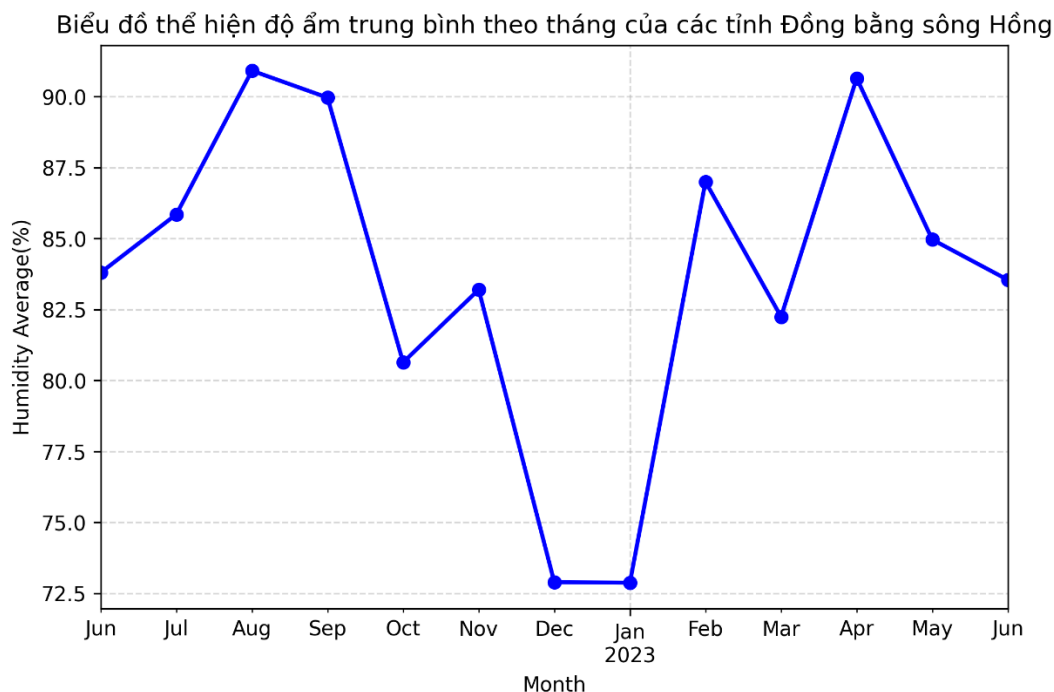


Hình 4.10. Biểu đồ tổng lượng mưa trung bình mỗi tháng của các tỉnh Đồng bằng sông Hồng

Nhận xét:

- Tháng 6/2022, 8/2022 và 6/2023 là 3 tháng có lượng mưa lớn nhất trong tập quan sát thu thập giai đoạn từ 29/02/2022 - 30/06/2023.
- Mùa mưa ở miền Bắc thường diễn ra vào khoảng từ Tháng 6 - Tháng 9. Tháng 10, 11 tuy vẫn còn nhưng số lượng sẽ ít hơn, vì lúc này thời tiết đang là mùa Thu và sắp chuyển giao sang mùa Đông.
- Từ các Tháng 12 - Tháng 2 do là thời tiết lạnh, hanh khô và đã là mùa đông lên lượng mưa giảm.

❖ **Biểu đồ thể hiện độ ẩm trung bình theo tháng của các tỉnh Đồng bằng sông Hồng:**



Hình 4.11. Biểu đồ độ ẩm trung bình theo tháng của các tỉnh Đồng bằng sông Hồng

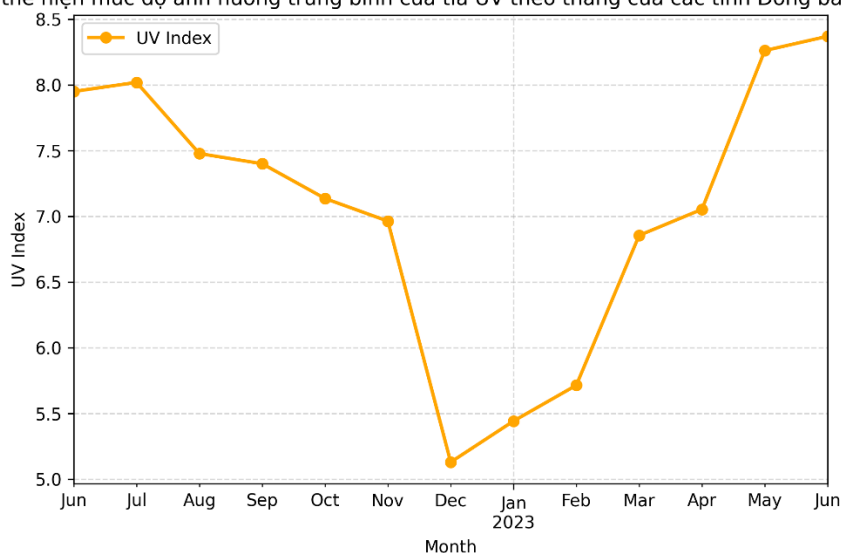
Nhận xét:

- Tháng có độ ẩm lớn nhất thường rơi vào những tháng 8 và 9. Tuy nhiên trong những khoảng thời gian này, thời tiết ở các tỉnh Đồng bằng sông Hồng lại không xảy ra những hiện tượng nồm ẩm như khoảng tháng Giêng và ngoài Giêng tầm Tháng 2 - 4.
- Lý giải hiện tượng này ta có thể giải thích như sau bởi vì khoảng này là thời tiết vừa chuyển từ Hè - Thu nên nền nhiệt độ vẫn còn cao mặc dù đã không còn gay gắt như khoảng tháng 5, 6, mát mẻ, không quá lạnh, hơi nước không có đủ khả năng tồn tại trong không khí dưới dạng nước, và do đó, không gây ra hiện tượng nồm ẩm.
- Tháng 10 - 1 năm sau, đây là khoảng thời tiết mùa đông, nhiệt độ thấp, lạnh và hanh khô. Khi nhiệt độ giảm, khả năng của không khí để chứa hơi nước cũng giảm đi. Điều này có nghĩa là không khí trong mùa đông có khả năng chứa ít hơi nước hơn, dẫn đến độ ẩm thấp hơn. Thêm nữa, mùa đông ít mưa hoặc không có mưa. Thiếu mưa làm giảm nguồn cung cấp hơi nước từ nguồn nước như sông, hồ và đất, dẫn đến độ ẩm cũng thấp hơn trong không khí.
- Mùa đông thường có sự hiện diện của gió lạnh và khô. Gió lạnh từ các vùng cận nhiệt đới thổi vào khu vực, mang theo không khí khô và không chứa nhiều hơi nước. Điều này làm tăng sự bay hơi và làm giảm độ ẩm trong không khí.

- Tháng Xuân 2 - 4: Bởi vì đây là thời điểm chuyển giao giữa mùa đông và mùa xuân ở miền Bắc. Trong giai đoạn này, sự không ổn định của khí hậu và các luồng gió từ phía Bắc có thể mang theo hơi nước từ các vùng biển và sông lớn, tạo ra độ ẩm cao trong không khí. Sự giao thoa giữa không khí lạnh từ phía Bắc và ẩm từ phía Nam gặp nhau cũng có thể gia tăng thêm độ ẩm không khí và gây ra hiện tượng nồm ẩm. Tác động của hướng gió: tùy thuộc vào hướng thổi gió, miền Bắc có thể nhận được luồng không khí ẩm từ các vùng biển nhiệt đới hoặc châu Á. Điều này khiến cho độ ẩm tăng lên và góp phần vào sự xuất hiện nồm ẩm.

❖ **Biểu đồ thể hiện xu hướng mức độ ảnh hưởng của tia UV theo tháng của các tỉnh Đồng bằng sông Hồng:**

Biểu đồ thể hiện mức độ ảnh hưởng trung bình của tia UV theo tháng của các tỉnh Đồng bằng sông Hồng



Hình 4.12. Biểu đồ mức độ ảnh hưởng trung bình của tia UV theo tháng của Đồng bằng sông Hồng

Nhận xét:

Nhìn ở biểu đồ này, ta cũng cần đáng cảnh báo về hiện tượng nóng lên toàn cầu hiện nay đc gây nên bởi hiệu ứng El Nino mà Trái Đất đang phải đối mặt, nó khiến cho cường độ ánh sáng trở nên mạnh và nguy hiểm hơn. Nhìn trên biểu đồ kia ta cũng dễ dàng thấy là mức độ nguy hiểm trung bình của tia UV thấp nhất ở đây cũng là 5 - mức trung bình, mặc dù là vẫn có những ngày ở mức thấp hơn thế nhưng tính theo trung bình tháng thì cũng đang là ở mức trung bình.

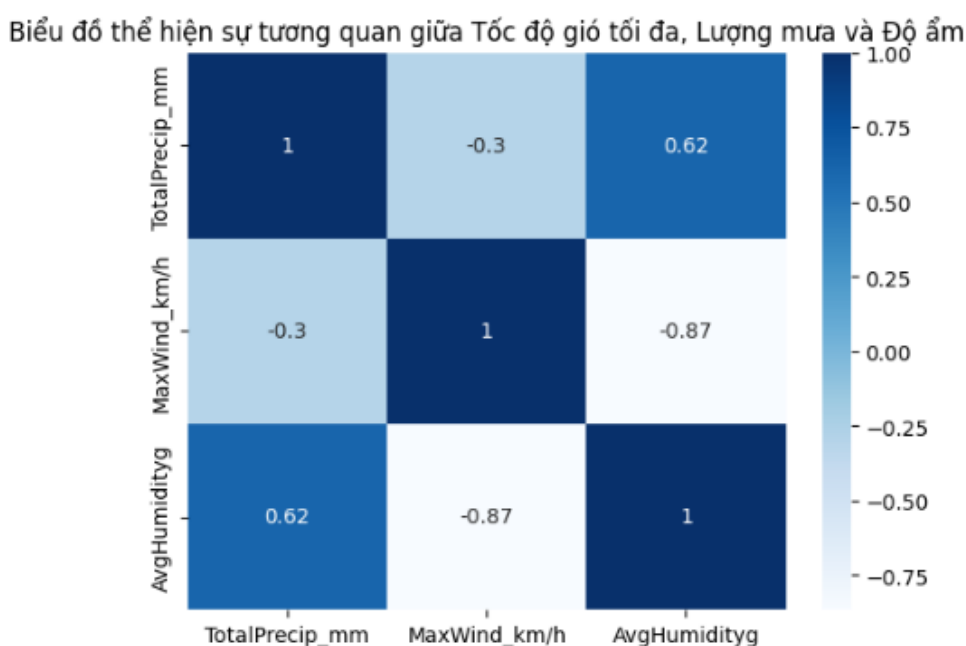
- Mùa đông (Tháng 12 - Tháng 2): Mức độ tia UV thường thấp hơn vào mùa đông (biểu đồ có xu hướng đi xuống nhanh). Nhiệt độ thấp và có nắng nhưng mức độ ánh sáng không mạnh, gắt và thời tiết có thể khá mây mù, dẫn đến mức độ nguy hiểm của tia UV cũng chỉ ở mức trung bình.

- Mùa xuân (Tháng 2 - Tháng 4): Khi thời tiết bắt đầu chuyển sang mùa xuân thời tiết trở nên ấm áp hơn, vẫn có rét tuy nhiên có sự đan xen của những ngày mưa xuân nên mức độ tia UV cũng tăng lên và đang ở mức cao nên khi ra ngoài, tiếp xúc trực tiếp với ánh nắng mặt trời ta cũng cần phải che chắn kín đáo và sử dụng kem chống nắng.
- Mùa hè (Tháng 5 - Tháng 8): Trong mùa hè, mức độ tia UV thường cao nhất trong năm. Ánh sáng mặt trời mạnh mẽ, gay gắt và dữ dội, trời ít mây. Nhìn vào biểu đồ ta cũng có thể thấy tăng lên rõ ràng từ mức cao -> rất cao. Đặc biệt nếu quan sát kỹ những ngày có nền nhiệt độ cao nhất trong tháng nóng ở trên bảng số liệu thì còn có những ngày chỉ số UV lên tới 10.
- Mùa thu (Tháng 9 - Tháng 11): Mức độ tia UV có xu hướng giảm đi khi mùa thu đến. Ánh sáng mặt trời không còn quá mạnh và thời tiết có thể mát mẻ hơn. Tuy nhiên, việc bảo vệ da vẫn cần thiết để ngăn chặn các vấn đề như lão hóa da và ung thư da.

=> Mức độ tia UV trong các tháng trong năm có thể ảnh hưởng đáng kể đến sức khỏe của con người. Vì vậy, việc bảo vệ da khỏi tác động của tia UV là rất quan trọng, bất kể thời gian nào trong năm. Sử dụng kem chống nắng, đeo mũ, áo chống nắng và kính râm, cũng như tránh tiếp xúc trực tiếp với ánh sáng mặt trời vào giữa ngày có thể giúp giảm nguy cơ bị tổn thương do tia UV.

4.2.2. Tây Bắc Bộ

1. Nhận xét về mối tương quan giữa tốc độ gió lớn nhất, lượng mưa và độ ẩm:



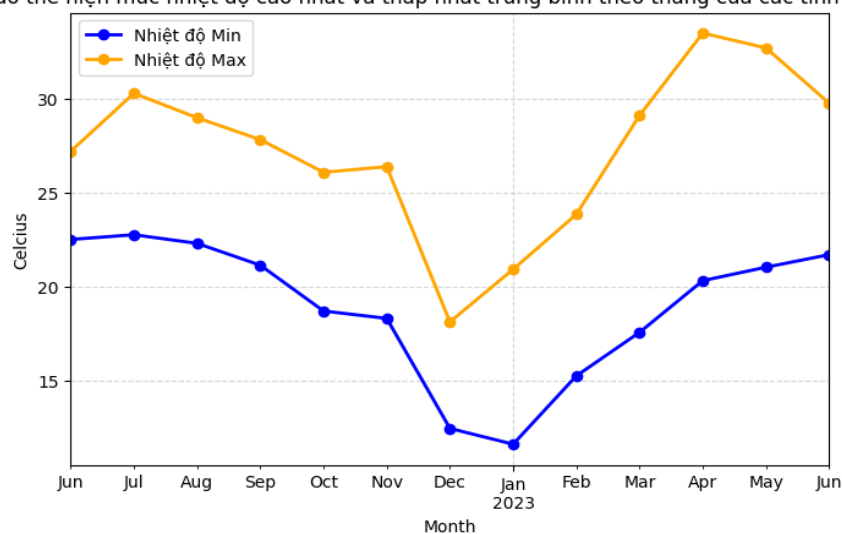
Hình 4.13. Biểu đồ thể hiện sự tương quan giữa tốc độ gió tối đa, lượng mưa và độ ẩm trung bình của vùng Tây Bắc Bộ

Nhận xét:

- Các giá trị trong biểu đồ thể hiện mức độ tương quan giữa các cặp biến. Một giá trị càng gần 1 (gần màu xanh dương) cho thấy mối tương quan dương mạnh, tức là khi một biến tăng, biến kia cũng tăng. Ngược lại, một giá trị càng gần -1 (gần màu đỏ) cho thấy mối tương quan âm mạnh, tức là khi một biến tăng, biến kia giảm. Giá trị gần 0 (gần màu trắng) cho thấy mối tương quan yếu hoặc không có tương quan.
- Biểu đồ giúp xác định mối quan hệ tuyến tính giữa các biến. Điều này giúp ta hiểu hơn về mối quan hệ giữa các biến và có thể sử dụng dự đoán hoặc phân tích tương quan giữa chúng
- Biểu đồ hệ số tương quan có thể hỗ trợ quá trình lựa chọn biến đầu vào trong mô hình hóa dữ liệu hoặc phân tích dữ liệu, bằng cách xem xét mức độ tương quan giúp ta chọn ra biến quan trọng nhất để sử dụng trong mô hình hoặc phân tích giúp cải thiện hiệu suất và hiểu rõ tác động của các biến lên biến mục tiêu
- 'TotalPrecip_mm' (lượng mưa tổng cộng) và 'Humidity Avg' (độ ẩm trung bình) có mối tương quan dương mạnh. 'MaxWind_kmh' (tốc độ gió tối đa) không có tương quan mạnh với 'Total Precip mm' (lượng mưa tổng cộng) hoặc 'Humidity Avg' (độ ẩm trung bình). Điều này cho thấy không có mối liên hệ rõ ràng giữa tốc độ gió tối đa và lượng mưa hoặc độ ẩm.
- Có thể nói tốc độ gió sẽ không ảnh hưởng gì nhiều đến các biến còn lại, còn về lượng mưa thì có một sự mật thiết đối với độ ẩm hay(nhiệt độ) .

2. Nhận xét về mức nhiệt độ trung bình theo tháng của các tỉnh vùng Tây Bắc Bộ:

Biểu đồ thể hiện mức nhiệt độ cao nhất và thấp nhất trung bình theo tháng của các tỉnh Tây Bắc Bộ



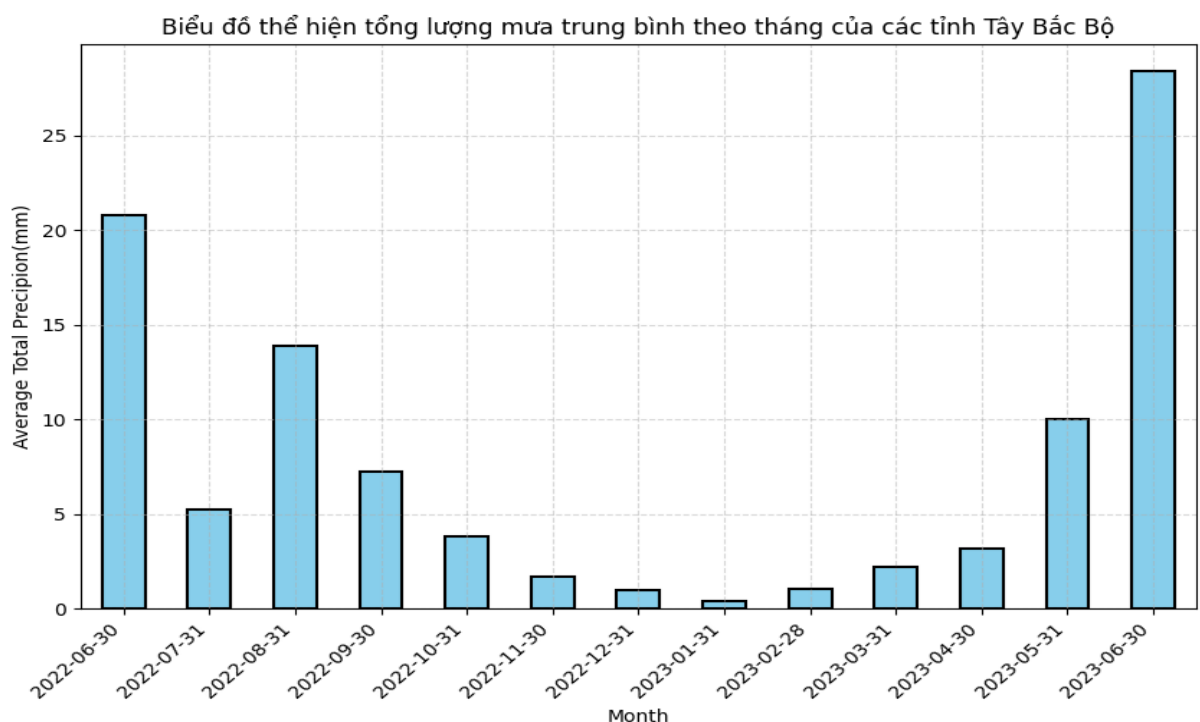
Hình 4.14. Biểu đồ mức nhiệt cao nhất và thấp nhất trung bình theo tháng của các tỉnh vùng Tây Bắc Bộ

Nhận xét:

- Ta thấy nhiệt độ cao nhất trong năm thuộc về tháng 4/2023 và tiếp đến là tháng 7/2022 có thể thấy nhiệt độ cao thường tập trung vào những tháng mùa hè.
- Nhiệt độ thấp là từ tháng 12/2022 đến tháng 1/2023 đây là những tháng đông và cuối đông nên việc nhiệt độ hạ thấp đến dưới 10 độ C, như vậy không có gì bất thường.
- Các tháng từ 10/2022 đến 02/2023 có một sự biến động khá rõ rệt bản đồ như đang bị trũng xuống.
- Có thể thấy tháng 12 là tháng có nhiệt độ duy trì khá thấp trong khoảng từ 10 đến 20 độ nhiệt độ này có thể khiến cây trồng vật nuôi chịu một số ảnh hưởng nhất định.

Nhìn chung nhiệt độ của các tỉnh vùng Tây Bắc Bộ có luôn có một sự phân chia nhiệt độ ổn định mùa đông lạnh và mùa hè nóng có một số tháng với sự chênh lệch nhiệt độ tương đối giữa các mùa.

3. Nhận xét về lượng mưa trung bình theo tháng của các tỉnh vùng Tây Bắc Bộ:



Hình 4.15. Biểu đồ thể hiện lượng mưa trung bình theo tháng của vùng Tây Bắc Bộ

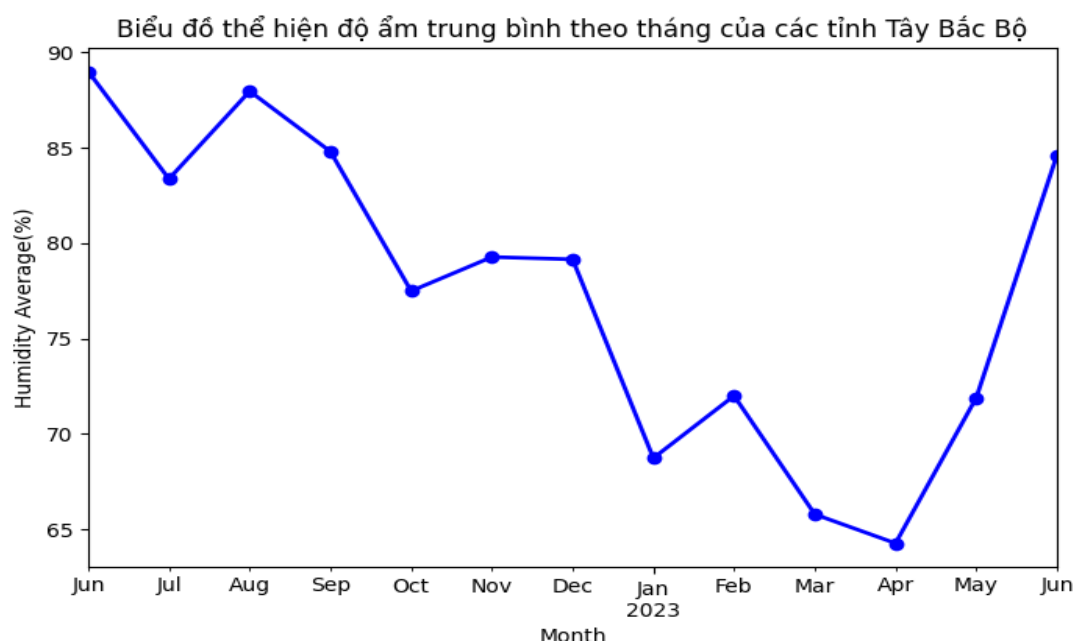
Nhận xét:

- Từ biểu đồ ta có thể thấy lượng mưa cao nhất nằm ở tháng 6 của năm 2022 và tháng 6 của năm 2023 lượng mưa đạt được là trên 20mm đỉnh điểm lên đến gần 30mm tuy chỉ là biểu đồ trung bình nhưng ta thấy lượng mưa của những tháng trên là tương đối cao.

- Lượng mưa thấp nhất ghi nhận được là vào các tháng 11,12 của năm 2022 và tháng 1,2,3 của năm 2023 chưa đạt tới 5mm.
- + Vì sao lại có sự chênh lệch lớn đến như vậy trong một năm?
- + Có thể đưa ra nhận định vì các tháng có lượng mưa thấp đa phần là các tháng mùa Đông và đầu Xuân vì nhiệt độ các tháng này tương đối thấp nên lượng hơi nước có được là tương đối thấp.
- + Những tháng có lượng mưa đều rơi vào các tháng mùa hè. Như từ trước đến nay ta luôn biết mùa hè là mùa mà thường xuyên có nhiều bão đổ bộ nhất nên việc có một lượng mưa lớn hơn rõ rệt so với các tháng cuối đông đầu xuân là điều có căn cứ.

Kết luận có thể thấy các tháng cuối đông đầu xuân là những tháng dễ bị khô hạn do lượng mưa phân bố trong năm không đồng đều, còn các tháng đầu hè cuối thu sẽ năm giữ lượng mưa khá lớn. Nhìn chung với lượng mưa lớn hay bé nhất thì nó cũng sẽ ảnh hưởng ít nhiều đến đời sống người dân về nuôi trồng và sản xuất.

4. Nhận xét về độ ẩm trung bình theo tháng của các tỉnh vùng Tây Bắc Bộ:



Hình 4.16. Biểu đồ thể hiện độ ẩm trung bình theo tháng của vùng Tây Bắc Bộ

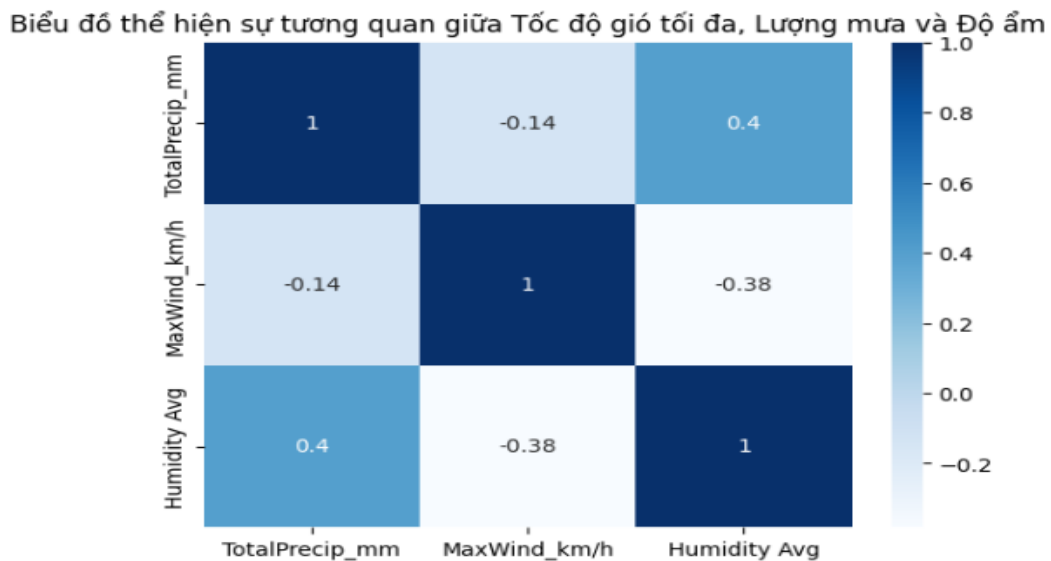
Nhận xét:

- Về độ ẩm phân bố không đồng đều ở các tháng, tháng 6 là tháng có độ ẩm cao nhất gần 90% và tháng 4 là tháng có độ ẩm thấp nhất chỉ dưới 65%.

- Độ ẩm cũng có một mối liên hệ đến thời tiết với các tháng có mức nhiệt cao thì đa phần độ ẩm sẽ cao vì sự bốc hơi từ sông suối cũng như mặt đất. Các yếu tố này đóng vai trò trong việc tạo ra môi trường ẩm ướt.
- Với các tháng có nhiệt độ thấp vào mùa đông kéo theo gió lạnh sẽ làm cho độ ẩm trong không khí giảm đi. Điều này có thể do sự khô hanh của không khí trong điều kiện lạnh, cũng như mức độ ẩm thấp hơn trong đất và các nguồn nước.
- Độ ẩm không khí cũng ảnh hưởng đến sự thoải mái và sức khỏe con người. Độ ẩm cao có thể tạo cảm giác ướt và nặng nề, trong khi ngược lại có thể gây ra khô hạn da.
- Độ ẩm cũng có các tác động đến các hoạt động, đời sống hàng ngày và sinh thái tự nhiên. Với độ ẩm phù hợp, cây trồng có thể phát triển tốt hơn, trong khi độ ẩm quá cao hoặc thấp gây khó khăn và ảnh hưởng đến sản xuất nông nghiệp và sinh thái.

4.2.3. Đông Bắc Bộ

1. Nhận xét về mối tương quan giữa tốc độ gió lớn nhất, lượng mưa và độ ẩm:



Hình 4.17. Biểu đồ thể hiện mối tương quan giữa tốc độ gió lớn nhất, lượng mưa và độ ẩm trung bình của vùng Đông Bắc Bộ

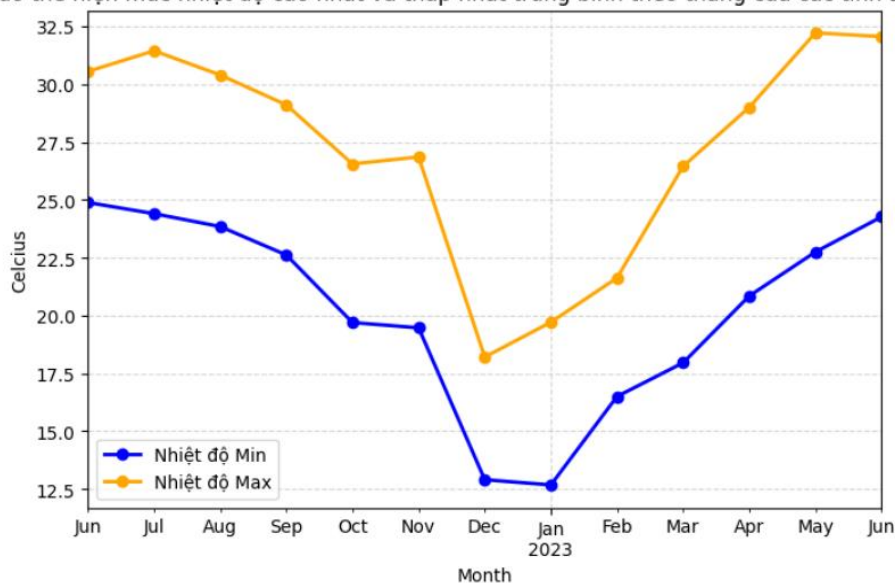
Nhận xét:

- Các giá trị trong biểu đồ thể hiện mức độ tương quan giữa các cặp biến. Một giá trị càng gần 1 (gần màu xanh dương) cho thấy mối tương quan dương mạnh, tức là khi một biến tăng, biến kia cũng tăng. Ngược lại, một giá trị càng gần -1 (gần màu đỏ) cho thấy mối tương quan âm mạnh, tức là khi một biến tăng, biến kia giảm. Giá trị gần 0 (gần màu trắng) cho thấy mối tương quan yếu hoặc không có tương quan.

- Biểu đồ giúp xác định mối quan hệ tuyến tính giữa các biến. Điều này giúp ta hiểu hơn về mối quan hệ giữa các biến và có thể sử dụng dự đoán hoặc phân tích tương quan giữa chúng
- Biểu đồ hệ số tương quan có thể hỗ trợ quá trình lựa chọn biến đầu vào trong mô hình hóa dữ liệu hoặc phân tích dữ liệu, bằng cách xem xét mức độ tương quan giúp ta chọn ra biến quan trọng nhất để sử dụng trong mô hình hoặc phân tích giúp cải thiện hiệu suất và hiểu rõ tác động của các biến lên biến mục tiêu
- TotalPrecip_mm' (lượng mưa tổng cộng) và 'Humidity Avg' (độ ẩm trung bình) có mối tương quan dương khá mạnh. 'MaxWind_kmh' (tốc độ gió tối đa) không có tương quan mạnh với 'Total Precip mm' (lượng mưa tổng cộng) hoặc 'Humidity Avg' (độ ẩm trung bình). Điều này cho thấy không có mối liên hệ rõ ràng giữa tốc độ gió tối đa và lượng mưa hoặc độ ẩm.
- Có thể nói tốc độ gió sẽ không ảnh hưởng gì nhiều đến các biến còn lại, còn về lượng mưa thì có một sự mật thiết đối với độ ẩm hay(nhiệt độ)

2. Nhận xét về mức nhiệt độ trung bình theo tháng của các tỉnh vùng Đông Bắc Bộ:

Biểu đồ thể hiện mức nhiệt độ cao nhất và thấp nhất trung bình theo tháng của các tỉnh Đông Bắc Bộ



Hình 4.18. Biểu đồ mức nhiệt cao nhất và thấp nhất trung bình theo tháng của các tỉnh vùng Đông Bắc Bộ

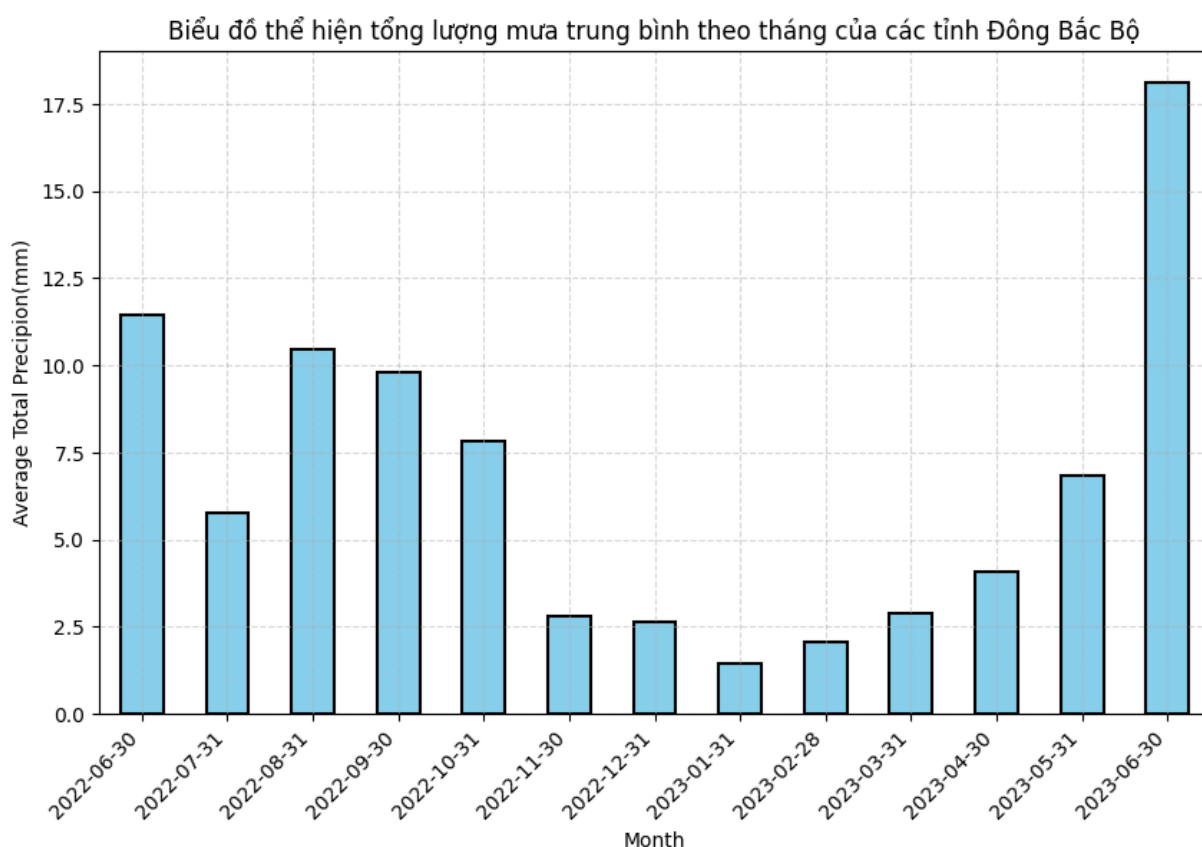
Nhận xét:

- Nhiệt độ thấp nhất rơi vào khoảng tháng 12 và tháng 1. Đây chính là khoảng thời gian mùa Đông ở miền Bắc. Tuy nhiên nhiệt độ thấp nhất ở đây không chỉ dừng ở đó, mà những tỉnh ở vùng núi Tây Bắc Bộ như là Sơn La, Lai Châu có những thời điểm năm đó rét đậm rét hại nhiệt độ thấp nhất còn có thể dưới âm

độ C, nhưng ở đây do ta đã lấy trung bình nên là nhiệt độ thấp nhất chung cho toàn miền Bắc sẽ là như khoảng trên.

- Biểu đồ cho thấy sự biến đổi nhiệt độ theo các tháng trong năm. Có sự chênh lệch rõ rệt giữa nhiệt độ cao nhất và thấp nhất trong mùa Đông (tháng 12 - tháng 1) và mùa Hè (tháng 5-tháng 6 - tháng 7). Điều này cho thấy miền Bắc có mùa đông lạnh và mùa hè nóng, với sự chênh lệch nhiệt độ lớn giữa các mùa.
- Có sự khác biệt về nhiệt độ giữa các tỉnh. Các tỉnh ở vùng núi Tây Bắc Bộ như Sơn La, Lai Châu có nhiệt độ thấp hơn so với các tỉnh Đồng bằng sông Hồng. Điều này có lẽ liên quan đến độ cao của khu vực, vì các tỉnh núi cao sẽ có nhiệt độ thấp hơn do tác động của địa hình núi cao.
- Một điều khác là sự biến đổi nhiệt độ trong một năm là khá rõ rệt nhất là từ tháng 10/2022 đến tháng 3/2023 biểu đồ lên xuống liên tục cho ta rút ra nhận xét là cuối những tháng của mùa thu cũng như xuân nhiệt độ sẽ biến đổi rất nhiều để bắt đầu chuyển giao sang một trường nhiệt mới

3. Nhận xét về lượng mưa trung bình theo tháng của các tỉnh vùng Đông Bắc Bộ:

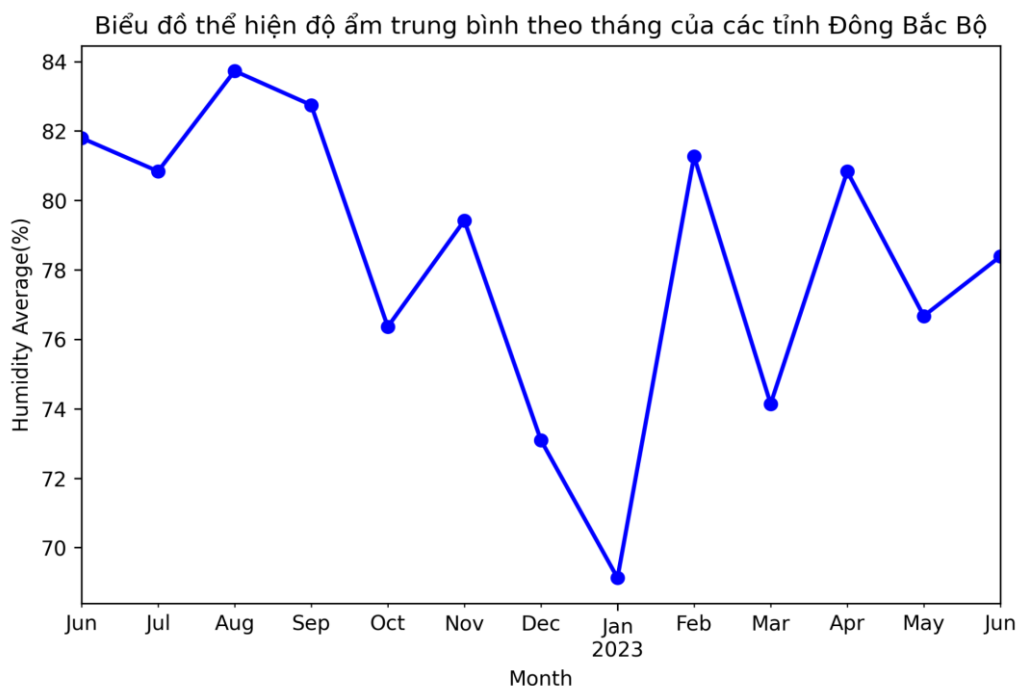


Hình 4.19. Biểu đồ thể hiện lượng mưa trung bình theo tháng của vùng Đông Bắc Bộ

Nhận xét:

- Tháng 6/2022, 8/2022, 9/2022 và 6/2023 là 3 tháng có lượng mưa lớn nhất trong tập quan sát thu thập giai đoạn từ 29/02/2022 - 30/06/2023.
- Các tháng 11/2022-03/2023 lượng mưa tương đối ít chỉ khoảng 2.5 mm.
- Mùa mưa ở miền Bắc thường diễn ra vào khoảng từ Tháng 6 - Tháng 9 tuy vẫn có nhưng số lượng sẽ ít hơn, vì lúc này thời tiết đang là mùa Thu và sắp chuyển giao sang mùa Đông. Tháng 5 vẫn có một lượng mưa nhất định vì bắt đầu vào mùa mưa nhưng không qua nhiều.
- Từ các Tháng 11 - Tháng 4 do là thời tiết lạnh, hanh khô và đã là mùa đông lên lượng mưa giảm. Điều này cho thấy sự tương quan giữa mùa mưa và mùa khô ở miền bắc.
- Lượng mưa trong miền Bắc có thể có sự biến đổi và dao động từ năm này sang năm khác, tùy thuộc vào yếu tố khí hậu và các yếu tố thời tiết khác việc theo dõi và phân tích lượng mưa là rất quan trọng để đánh giá tình hình nguồn nước và quản lý tài nguyên tự nhiên hiệu quả.
- Kết luận có thể thấy các tháng cuối đông đầu xuân là những tháng dễ bị khô hạn do lượng mưa phân bố trong năm không đồng đều, còn các tháng đầu hè cuối thu sẽ nắm giữ lượng mưa khá lớn. Nhìn chung với lượng mưa lớn hay bé nhất thì nó cũng sẽ ảnh hưởng ít nhiều đến đời sống người dân về nuôi trồng và sản xuất.

4. Nhận xét về độ ẩm trung bình theo tháng của các tỉnh vùng Đông Bắc Bộ:



Hình 4.20. Biểu đồ thể hiện độ ẩm trung bình theo tháng của vùng Đông Bắc Bộ

Nhận xét:

- Tháng 8 và 9 thường là thời điểm có độ ẩm lớn nhất trong năm, tuy nhiên, trong khu vực Đông Bắc Bộ, không xảy ra hiện tượng nồm ẩm như trong khoảng thời gian từ tháng Giêng đến tháng 4. Lý giải cho hiện tượng này có thể được giải thích như sau:
- Chuyển từ Hè sang Thu: Trong khoảng thời gian này, nhiệt độ vẫn cao, nhưng không quá gay gắt như trong tháng 5 và 6. Mặt khác, không khí cũng không quá lạnh. Do đó, hơi nước không có đủ khả năng tồn tại trong không khí dưới dạng nước và không gây ra hiện tượng nồm ẩm.
- Mùa Đông: Trái ngược với tháng 8 và 9, mùa đông có nhiệt độ thấp, lạnh và hanh khô. Khi nhiệt độ giảm, khả năng của không khí để chứa hơi nước cũng giảm đi. Điều này dẫn đến độ ẩm thấp hơn trong không khí. Ngoài ra, mùa đông ít mưa hoặc không có mưa, làm giảm nguồn cung cấp hơi nước từ nguồn nước như sông, hồ và đất, cũng đóng góp vào độ ẩm thấp trong không khí.
- Gió lạnh và khô: Mùa đông thường có sự hiện diện của gió lạnh và khô. Gió lạnh từ các vùng cận nhiệt đới thổi vào khu vực, mang theo không khí khô và không chứa nhiều hơi nước. Điều này làm tăng sự bay hơi và làm giảm độ ẩm trong không khí.
- Chuyển giao từ mùa Đông sang mùa Xuân (tháng 2 - 4): Trong giai đoạn này, sự không ổn định của khí hậu và sự giao thoa giữa không khí lạnh từ phía Bắc và không khí ấm từ phía Nam có thể tạo ra độ ẩm cao trong không khí và góp phần vào hiện tượng nồm ẩm. Các luồng gió từ các vùng biển và sông lớn cũng có thể mang theo hơi nước, tăng thêm độ ẩm không khí.

CHƯƠNG 5. TỔNG KẾT, NHẬN XÉT VÀ ĐÁNH GIÁ

5.1. Tổng kết các kết quả đạt được và những hạn chế còn sót lại sau khi thực hiện xong đề tài

❖ Các kết quả đạt được:

- Thông qua sự nghiên cứu và thực hiện đề tài này, nhóm chúng em đã được mở rộng hơn vốn hiểu biết về các đặc điểm khí hậu, thời tiết và các vấn đề hiện nay mà miền Bắc chúng ta đang phải đối mặt nói riêng và trên toàn thế giới nói chung.
- Biết cách quan sát và phân tích, đánh giá biểu đồ thông qua những thay đổi theo từng thời kỳ khác nhau, từ đó có thể xác định cũng như dự đoán trước được thời tiết của bất kỳ ngày nào đó nếu ta có được những thông tin về dữ liệu liên quan.
- Từ đó, góp phần làm nền tảng, cung cấp những cái nhìn, hướng tiếp cận ban đầu cho việc nghiên cứu và làm khóa luận sau này.

❖ Một số mặt còn hạn chế của đề tài:

- Dữ liệu thu thập được khá là sạch sẽ, gần như không có vấn đề gì. Điều này sẽ khiến cho mình chưa có nhiều những trải nghiệm thực tế hơn trong công việc.

5.2. Bảng phân công công việc và mức độ hoàn thành của từng cá nhân trong nhóm

Công việc	Người phụ trách
- Thu thập dữ liệu + Chuẩn bị dữ liệu + Tiền xử lý và phân tích dữ liệu của vùng Đồng bằng sông Hồng - Viết báo cáo	Tạ Quang Dũng
- Thu thập dữ liệu + Tiền xử lý và phân tích dữ liệu của vùng Tây Bắc Bộ và Đông Bắc Bộ -Viết báo cáo	Phạm Hải Đăng

5.3. Nhận xét, đánh giá và xếp loại của giáo viên

