

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
KHOA CÔNG NGHỆ THÔNG TÍN



TRỰC QUAN HÓA DỮ LIỆU

Báo cáo
Đồ án cuối kỳ

Nhóm 18:

21120579 - Phạm Thị Kiều Trinh
21120573 - Lê Phước Đoan Trang
21120568 - Lê Nguyễn Phương Thy
21120582 - Đinh Hoàng Trung

Giảng viên:
Thầy Bùi Tiến Lên

Mục lục

1 Phân công công việc	3
2 Tóm tắt đề bài	4
2.1 Nội dung	4
2.2 Yêu cầu về dữ liệu	4
2.3 Tổng quan về dữ liệu	4
2.3.1 Các tỉnh/thành phố được ghi nhận	4
2.3.2 Ý nghĩa các cột dữ liệu	4
2.3.3 Bảng hướng gió	5
2.3.4 Thông tin nổi bật	5
3 Phân tích dashboard	6
3.1 Tốc độ tăng trưởng nhiệt độ của các thành phố lớn	6
3.2 Biên độ nhiệt giữa các thành phố lớn	8
3.3 Áp suất của không khí theo khu vực	10
3.4 Tỉ lệ ngày có gió mạnh so với yếu	11
3.5 Sự biến động tốc độ gió giai đoạn từ 2009 - 2021	13
3.6 Mối tương quan giữa mây, mưa và độ ẩm	14
3.7 Đánh giá lượng mưa của từng tỉnh theo khu vực và khoảng thời gian trong năm	17
4 Tích hợp AI	23
4.1 Ý tưởng thực hiện	23
4.2 Thực hiện ý tưởng	23
4.2.1 Các yêu cầu cấu hình	23
4.2.2 Truy vấn dataset	23
4.2.3 Vẽ biểu đồ	26
4.2.4 Đánh giá phân tích dựa trên biểu đồ được vẽ	27

Danh sách hình vẽ

1	Biểu đồ thể hiện sự thay đổi nhiệt độ trung bình của các thành phố giai đoạn 2009 - 2021.	6
2	Biểu đồ thể hiện trung bình chênh lệch nhiệt độ của các thành phố	8
3	Lược đồ dải đồng bằng Duyên hải miền Trung	9
4	Biểu đồ thể hiện áp suất của không khí theo khu vực	10
5	Biểu đồ thể hiện tỉ lệ ngày có gió mạnh so với yếu	11
6	Biểu đồ thể hiện sự biến động tốc độ gió từ 2009-2021	13
7	Biểu đồ thể hiện mối tương quan giữa mây, mưa và độ ẩm	14
8	Trung bình lượng mưa của các tỉnh	17
9	Thống kê tổng lượng mưa của các tỉnh tại các thời điểm trong năm .	17
10	Địa hình Bắc bộ và Bắc Trung bộ	18
11	Địa hình khu vực Nam Trung Bộ, Tây Nguyên, Đông Nam Bộ và Đông Bằng Sông Cửu Long	20
12	Schema để lưu dataset	24
13	Câu lệnh import file csv vào table	24
14	Bảng dữ liệu trong database.	24
15	Câu prompt dành cho việc truy vấn	25
16	Giao diện chatbox	25
17	Phản hồi từ câu hỏi truy vấn.	26
18	Biểu đồ được vẽ bởi AI theo yêu cầu và dữ liệu vừa được truy vấn . .	27
19	Giao diện trang yêu cầu phân tích	27
20	Phản hồi từ yêu cầu phân tích	28

1 Phân công công việc

MSSV	Thành viên	Công việc	Đóng góp
21120658	Lê Nguyễn Phương Thy	Tìm kiếm dataset, vẽ và phân tích biểu đồ, vẽ dashboard	25%
21120573	Lê Phước Đoan Trang	Tìm kiếm dataset, vẽ và phân tích biểu đồ, vẽ dashboard, viết báo cáo	25%
21120579	Phạm Thị Kiều Trinh	Tìm kiếm dataset, tiền xử lý dữ liệu, vẽ dashboard, viết báo cáo và làm slide	25%
21120582	Đinh Hoàng Trung	Tìm kiếm dataset, vẽ và phân tích biểu đồ, tích hợp AI, viết báo cáo	25%

Bảng 1: Bảng phân công công việc cho các thành viên

Mốc thời gian thực hiện đồ án:

- 20/12/2024 - 27/12/2024:** Tìm kiếm và thống nhất chọn bộ dữ liệu.
- 27/12/2024 - 30/12/2024:** Quá trình tìm hiểu, tiền xử lí và khám phá dữ liệu.
- 01/01/2025 - 05/01/2025:** trực quan dữ liệu, dàn xen bổ sung phần tiền xử lí còn thiếu sót. Tích hợp AI.
- 05/01/2025 - 10/01/2025:** Viết báo cáo, làm slide. Chỉnh sửa bổ sung các nội dung đã làm.

2 Tóm tắt đề bài

2.1 Nội dung

- Trình bày dự án ứng dụng trực quan dùng dashboard để phân tích dữ liệu.
- Ngữ cảnh của dữ liệu là Việt Nam.

2.2 Yêu cầu về dữ liệu

Dữ liệu là dữ liệu thật có liên quan đất nước Việt Nam. Dữ liệu phải có tối thiểu 6 biến độc lập và tối thiểu 1000 dòng. Tổng số dòng dữ liệu liên quan đến Việt Nam phải trên 50%.

2.3 Tổng quan về dữ liệu

Bộ dữ liệu này ghi lại thông tin thời tiết hàng ngày tại **40 tỉnh/thành phố** ở Việt Nam trong khoảng thời gian từ **01/01/2009** đến **18/6/2021**. Dữ liệu bao gồm nhiều thông số quan trọng liên quan đến nhiệt độ, tốc độ gió, lượng mưa, độ ẩm và các yếu tố thời tiết khác, giúp cung cấp cái nhìn tổng quan về tình hình khí hậu tại các địa phương.

2.3.1 Các tỉnh/thành phố được ghi nhận

- **Miền Bắc:** Hà Nội, Hải Phòng, Thái Nguyên, Yên Bái, Hòa Bình, Nam Định, Việt Trì, Uông Bí, Cẩm Phả, Hòn Gai, Hải Dương.
- **Miền Trung:** Huế, Đà Lạt, Tam Kỳ, Tuy Hòa, Phan Rang, Phan Thiết, Nha Trang, Cam Ranh, Vinh, Thanh Hóa, Quy Nhơn, Pleiku, Buôn Ma Thuột.
- **Miền Nam:** TP. HCM, Biên Hòa, Cần Thơ, Bạc Liêu, Cà Mau, Long Xuyên, Bến Tre, Sóc Trăng, Mỹ Tho, Rạch Giá, Tân An, Vũng Tàu, Trà Vinh, Châu Đốc, Vĩnh Long.

2.3.2 Ý nghĩa các cột dữ liệu

Tên cột	Mô tả
province	Tên tỉnh hoặc thành phố nơi dữ liệu thời tiết được ghi nhận.
max	Nhiệt độ cao nhất ($^{\circ}\text{C}$) được ghi nhận trong ngày.
min	Nhiệt độ thấp nhất ($^{\circ}\text{C}$) được ghi nhận trong ngày.
wind	Tốc độ gió (km/h) trung bình trong ngày.
wind_d	Hướng gió theo ký hiệu la bàn. Chi tiết Bảng hướng gió dưới đây.

rain	Lượng mưa (mm) ghi nhận trong ngày.
humidi	Độ ẩm trung bình (%) trong ngày.
cloud	Mức độ mây che phủ trung bình (%) trong ngày.
pressure	Áp suất khí quyển trung bình trong ngày (hPa).
date	Ngày ghi nhận dữ liệu, định dạng YYYY-MM-DD.

2.3.3 Bảng hướng gió

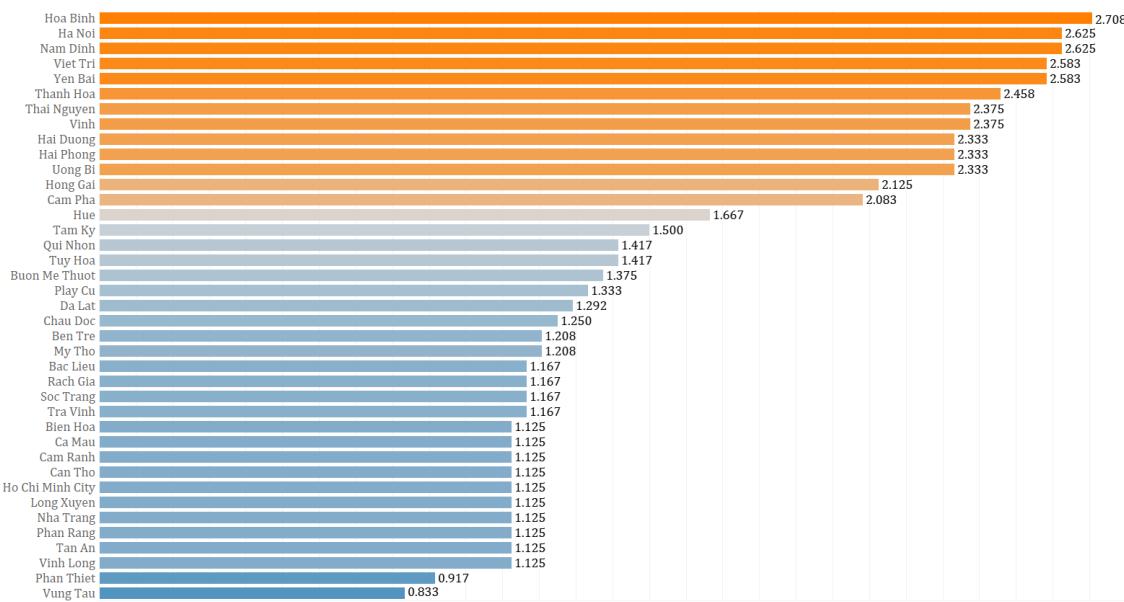
Hướng gió	Ký hiệu	Góc	Từ góc	Đến góc
Bắc	N	0°	348.75°	11.25°
Bắc Đông Bắc	NNE	22.5°	11.25°	33.75°
Đông Bắc	NE	45°	33.75°	56.25°
Đông Đông Bắc	ENE	67.5°	56.25°	78.75°
Đông	E	90°	78.75°	101.25°
Đông Đông Nam	ESE	112.5°	101.25°	123.75°
Đông Nam	SE	135°	123.75°	146.25°
Nam Đông Nam	SSE	157.5°	146.25°	168.75°
Nam	S	180°	168.75°	191.25°
Nam Tây Nam	SSW	202.5°	191.25°	213.75°
Tây Nam	SW	225°	213.75°	236.25°
Tây Tây Nam	WSW	247.5°	236.25°	258.75°
Tây	W	270°	258.75°	281.25°
Tây Tây Bắc	WNW	292.5°	281.25°	303.75°
Tây Bắc	NW	315°	303.75°	326.25°
Bắc Tây Bắc	NNW	337.5°	326.25°	348.75°

2.3.4 Thông tin nổi bật

- Thời gian:** 12 năm dữ liệu thời tiết hàng ngày (2009–2021).
- Địa điểm:** Bao gồm 40 tỉnh/thành phố đại diện cho các vùng khí hậu khác nhau tại Việt Nam.
- Ứng dụng:**
 - Phân tích các xu hướng thời tiết theo năm hoặc mùa.
 - Đánh giá tác động của biến đổi khí hậu.
 - Lập kế hoạch nông nghiệp và quản lý thiên tai.

3 Phân tích dashboard

3.1 Tốc độ tăng trưởng nhiệt độ của các thành phố lớn



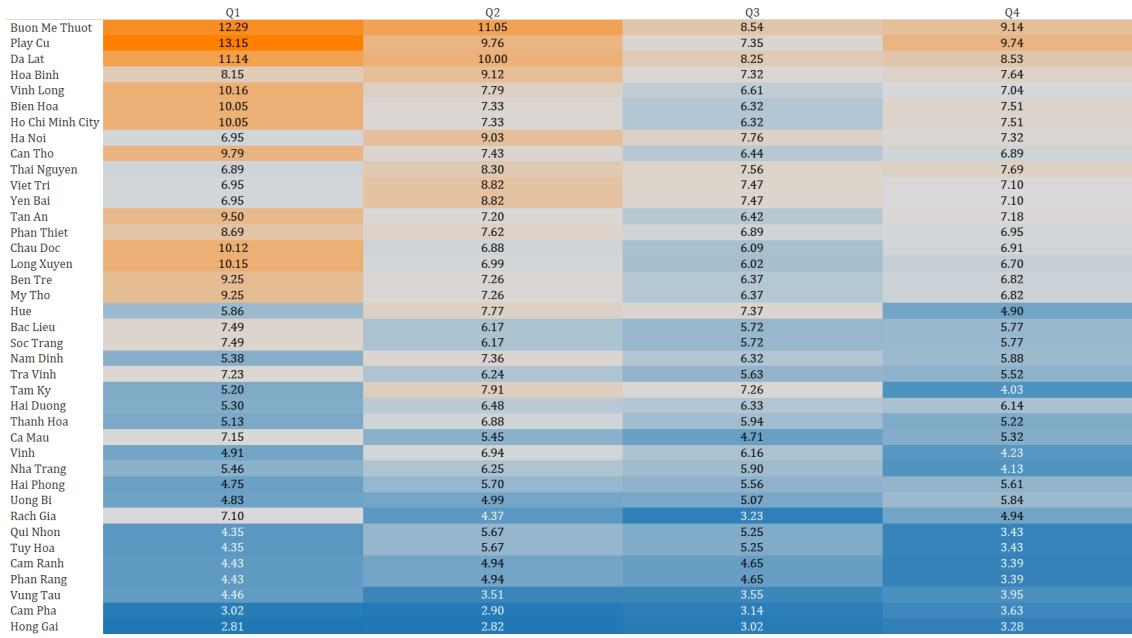
Hình 1: Biểu đồ thể hiện sự thay đổi nhiệt độ trung bình của các thành phố giai đoạn 2009 - 2021.

- Miền Bắc đã chứng kiến sự tăng trưởng nhiệt độ cao nhất cả nước trong vòng hơn 10 năm qua. Đặc biệt, các tỉnh thuộc khu vực đồng bằng sông Hồng và miền núi phía Bắc đứng đầu về tốc độ tăng nhiệt độ, với phần lớn nguyên nhân đến từ quá trình đô thị hóa và công nghiệp hóa mạnh mẽ tại Hà Nội cùng các địa bàn lân cận. Bên cạnh đó, sự suy giảm diện tích rừng tự nhiên ở các tỉnh miền núi cũng là yếu tố không thể không nhắc tới.
 - Trong đó, Hòa Bình nổi bật với mức tăng nhiệt độ lên tới 2.708%, là khu vực có tốc độ tăng nhanh nhất trên toàn quốc. Tình trạng này chủ yếu do việc khai thác tài nguyên khoáng sản và sự suy giảm diện tích rừng.
 - Hà Nội (2.625%) cũng không nằm ngoài xu hướng này, với vai trò là trung tâm kinh tế - chính trị lớn nhất cả nước, thành phố này đang phải đổi mới với hiệu ứng đảo nhiệt đô thị, khiến nhiệt độ tại đây tăng mạnh.
- Tại miền Trung, sự tăng nhiệt độ đạt mức trung bình, với sự khác biệt rõ rệt giữa khu vực ven biển và miền núi. Khu vực ven biển chứng kiến sự gia tăng nhiệt độ nhanh chóng, trong khi những vùng miền núi có sự thay đổi chậm hơn. Biến đổi khí hậu toàn cầu, nước biển dâng và tình trạng suy giảm diện tích rừng là những nguyên nhân chủ yếu dẫn đến sự thay đổi này.
 - Thanh Hóa (2.458%) là thành phố dẫn đầu về tốc độ tăng nhiệt độ, với sự gia tăng mạnh mẽ của các khu công nghiệp và quá trình đô thị hóa.

- Các thành phố Thái Nguyên và Vinh cũng đang phải đối mặt với áp lực nhiệt độ gia tăng, do sự phát triển nhanh chóng của nền kinh tế và các khu công nghiệp.
- Miền Nam có mức tăng trưởng thấp nhất cả nước, điều này chủ yếu nhờ vào hệ thống kênh rạch tự nhiên phong phú cùng các biện pháp bảo vệ môi trường hiệu quả, đặc biệt là ở khu vực đồng bằng sông Cửu Long. Việc duy trì diện tích rừng ngập mặn và các nỗ lực bảo vệ môi trường đã giúp khu vực này có khả năng điều hòa khí hậu tốt hơn.
 - Bà Rịa - Vũng Tàu, với tốc độ tăng nhiệt độ chỉ 0.833 - thấp nhất cả nước, là minh chứng điển hình cho sự hài hòa giữa vị trí địa lý ven biển và các nỗ lực bảo vệ môi trường. Nhờ được bao bọc bởi các luồng gió mát từ đại dương, khí hậu nơi đây duy trì sự ổn định, dễ chịu. Đồng thời, những chính sách phát triển bền vững, điển hình là Chiến lược tăng trưởng xanh đã giúp khu vực này trở thành hình mẫu trong việc giảm thiểu tác động của biến đổi khí hậu, giữ gìn vẻ đẹp thiên nhiên độc đáo.
 - Hồ Chí Minh (1.125), dù là trung tâm kinh tế lớn nhất miền Nam, nhưng đã có những nỗ lực đáng kể trong việc xanh hóa đô thị, qua đó giúp giảm tốc độ gia tăng nhiệt độ tại đây.

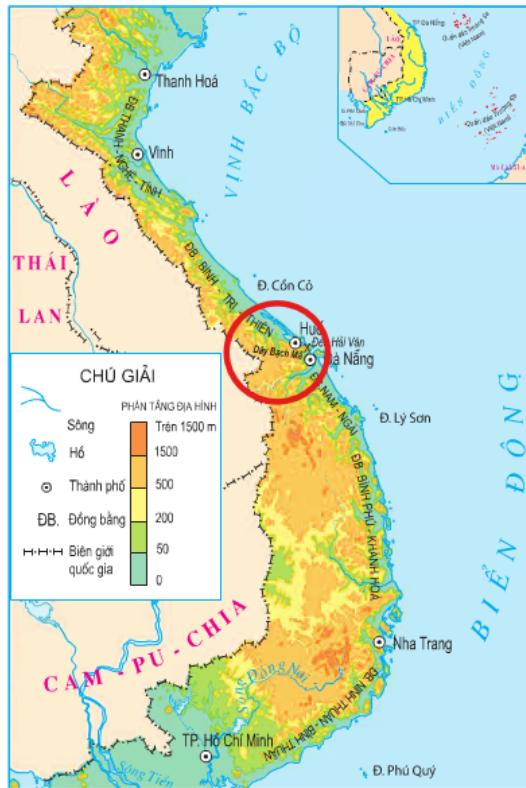
Tình trạng tăng trưởng nhiệt độ tại các khu vực của Việt Nam không chỉ phản ánh sự tác động của biến đổi khí hậu mà còn là hệ quả của quá trình đô thị hóa và phát triển kinh tế không bền vững. Các biện pháp bảo vệ môi trường, duy trì diện tích rừng và quy hoạch đô thị hợp lý sẽ là yếu tố quyết định để giảm thiểu tác động của biến đổi khí hậu trong tương lai.

3.2 Biên độ nhiệt giữa các thành phố lớn



Hình 2: Biểu đồ thể hiện trung bình chênh lệch nhiệt độ của các thành phố

- Tây Nguyên – Biên độ nhiệt lớn và sự phân hóa theo mùa
 - Tây Nguyên, đặc biệt là Pleiku và Buôn Ma Thuột, là khu vực có biên độ nhiệt trong ngày cao nhất quý 1, với Pleiku đạt 13,15°C và Buôn Ma Thuột là 12,29°C. Điều này phản ánh đặc điểm khí hậu cao nguyên với ngày khô nóng và đêm lạnh do bức xạ nhiệt mạnh vào ban đêm. Tuy nhiên, vào quý 3, nhiệt độ tại Tây Nguyên tương đối ổn định hơn, chỉ chênh khoảng 7°C đến 8°C, do bước vào mùa mưa với độ ẩm tăng cao làm giảm sự chênh lệch nhiệt độ giữa ngày và đêm. Sự thay đổi này tạo ra biên độ nhiệt lớn giữa các quý, ảnh hưởng rõ rệt đến nông nghiệp và đời sống, đặc biệt là các cây công nghiệp như cà phê và cao su, vốn phụ thuộc vào thời tiết theo mùa.
- Các khu vực miền Bắc, như Hà Nội, Thái Nguyên, và Việt Trì, thể hiện sự biến động nhiệt độ theo mùa rõ rệt hơn các khu vực khác.
 - Quý 1: Biên độ nhiệt trung bình từ 6,5°C đến 7°C, phản ánh mùa đông lạnh với sự chênh lệch nhỏ giữa ngày và đêm do nhiệt độ luôn thấp.
 - Quý 2: Điểm đỉnh của mùa hè, với nhiệt độ dao động khoảng 8°C đến 9°C, kết hợp với độ ẩm cao gây cảm giác oi bức.
 - Quý 3: Chênh lệch giảm dần, từ 7°C đến 8°C, phản ánh giai đoạn chuyển mùa sang mùa thu đông.
 - Quý 4: Sự giảm nhiệt khi mùa đông bắt đầu, từ 6°C đến 7°C. Biến động này tác động lớn đến nông nghiệp, đặc biệt là các loại cây vụ đông và lúa nước, cũng như nhu cầu sưởi ấm và tiêu thụ năng lượng của người dân.



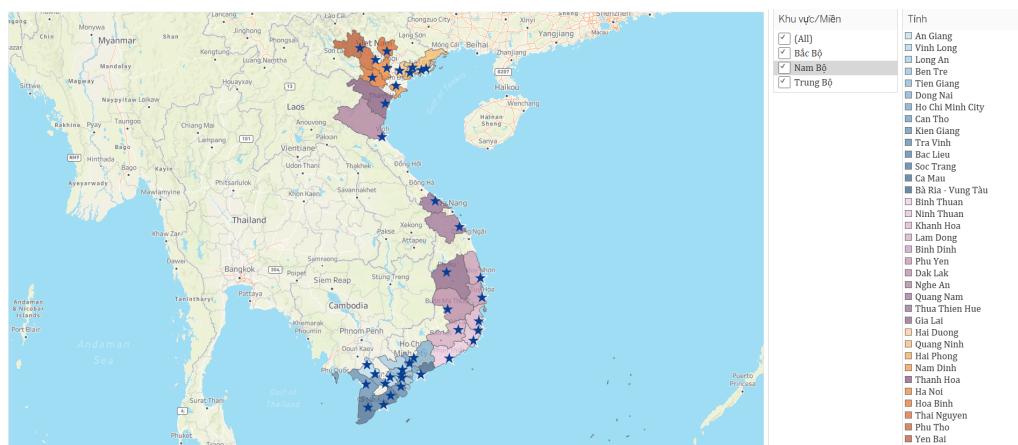
Hình 3: Lược đồ dải đồng bằng Duyên hải miền Trung

- Nhiệt độ ở miền Nam quanh năm luôn ở mức cao với biên độ nhiệt nhỏ hơn so với khu vực Bắc Bạch Mã. Các thành phố như Biên Hòa, TP. HCM, Cần Thơ, Long Xuyên, Bến Tre,... thể hiện một đặc điểm thú vị về biên độ nhiệt trong ngày:
 - Quý 1: Biên độ nhiệt cao nhất, dao động từ 7.1°C đến 10,16°C. Bởi lẽ thời điểm này đang rơi vào giữa mùa khô miền Nam (tháng 11 đến tháng 4 năm sau), ban ngày nắng nóng mạnh trong khi ban đêm giảm nhiệt đáng kể, tạo ra sự chênh lệch lớn giữa ngày và đêm.
 - Các quý còn lại: Biên độ nhiệt giảm xuống mức 6-7,7°C trong quý 2 và duy trì ổn định trong quý 3 và quý 4. Phần lớn nhờ vào lượng mưa trung bình hằng năm ở miền Nam thường rất lớn, khoảng từ 1500 - 2000mm. Độ ẩm tương đối cao và mức cân bằng ẩm luôn dương. Điều này thuận lợi cho nông nghiệp quanh năm, đặc biệt là trồng lúa và cây ăn quả. Tuy nhiên, các đợt mưa lớn trong mùa mưa (quý 2 và quý 3) có thể gây ngập lụt cục bộ, ảnh hưởng đến sinh hoạt và sản xuất.
- Các thành phố ven biển như Nha Trang, Phan Rang, Vũng Tàu,... nơi gió biển mang lại sự ổn định nhiệt độ quanh năm.
 - Khu vực tỉnh Quảng Ninh có chênh lệch nhiệt độ ngày và đêm nhỏ nhất và ổn định quanh năm. Hòn Gai và Cẩm Phả nằm trong khu vực ven Vịnh Hạ Long – một vùng biển kín, được bao quanh bởi các đảo đá vôi

và núi đồi. Điều này làm giảm tác động của nhiệt độ cực đoan từ gió mùa đông bắc (lạnh) và gió mùa hè (nóng).

- Ngược lại, các vùng ven biển miền Trung và miền Nam như Nha Trang, Phan Rang, và Vũng Tàu nằm trong khu vực có khí hậu nhiệt đới gió mùa, với nền nhiệt độ cao hơn và ít chịu ảnh hưởng từ biển hơn, vì đây là vùng biển mở, ít được che chắn bởi các yếu tố tự nhiên như núi và vịnh.
- Chênh lệch nhiệt độ thấp ở các vùng ven biển không chỉ giúp tạo môi trường sống dễ chịu mà còn mang lại nhiều lợi ích kinh tế và xã hội. Đây là lợi thế quan trọng để phát triển các ngành kinh tế chủ chốt như du lịch, nông nghiệp, thủy sản, và công nghiệp, đồng thời tạo nền tảng cho sự phát triển bền vững của khu vực ven biển.
- Tuy nhiên, biên độ nhiệt thấp có thể làm giảm hiệu suất một số ngành năng lượng, đặc biệt là nhiệt điện và năng lượng mặt trời. Nhiệt độ ổn định đi đôi với độ ẩm cao cũng tạo điều kiện cho sương mù dày đặc, đặc biệt vào sáng sớm và tối muộn, ảnh hưởng đến hoạt động giao thông vận tải. Nên việc hiểu rõ và khắc phục các hạn chế này là cần thiết để đảm bảo sự phát triển bền vững và toàn diện cho các khu vực ven biển.

3.3 Áp suất của không khí theo khu vực



Hình 4: Biểu đồ thể hiện áp suất của không khí theo khu vực

- **Xu hướng tổng quan về áp suất khí quyển:**

- Trong giai đoạn 2009–2021, áp suất khí quyển trung bình ở các khu vực miền Bắc thường cao hơn so với miền Trung và miền Nam.
- Đặc điểm địa hình ảnh hưởng đến áp suất: khu vực núi cao (Yên Bái, Thái Nguyên) có áp suất thấp hơn so với đồng bằng (Hà Nội, Hải Phòng).

- **Phân tích theo khu vực:**

- **Miền Bắc:** Áp suất khí quyển ổn định hơn, cao nhất tại khu vực đồng bằng sông Hồng (Hà Nội, Hải Dương). Các khu vực miền núi phía Bắc như Lào Cai, Yên Bái có áp suất thấp hơn vào cuối mùa đông.
- **Miền Trung:** Biến động áp suất lớn, do chịu ảnh hưởng của khí hậu gió mùa và bão. Các tỉnh ven biển (Đà Nẵng, Nha Trang) có áp suất thấp hơn trong mùa mưa.
- **Miền Nam:** Áp suất ổn định hơn, nhưng thấp hơn so với miền Bắc. Các tỉnh Cần Thơ, Bạc Liêu, và Cà Mau thường có áp suất thấp nhất.

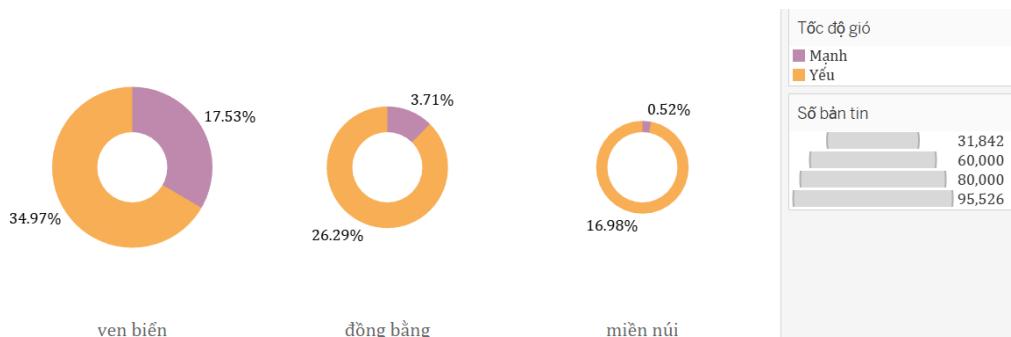
- **Biến động theo thời gian:**

- **Mùa đông (tháng 12 – tháng 2 năm sau):** Áp suất khí quyển tăng rõ rệt ở miền Bắc do ảnh hưởng của không khí lạnh.
- **Mùa hè (tháng 5–8):** Áp suất khí quyển giảm ở cả ba miền, đặc biệt ở miền Trung và miền Nam do ảnh hưởng của bão và gió mùa.

Nhận xét chung

- Áp suất khí quyển trung bình trong năm 2009–2021 thể hiện đặc điểm rõ rệt của khí hậu từng vùng miền tại Việt Nam.
- Sự phân bố áp suất theo không gian và thời gian phản ánh tác động của khí hậu gió mùa.

3.4 Tỉ lệ ngày có gió mạnh so với yếu



Hình 5: Biểu đồ thể hiện tỉ lệ ngày có gió mạnh so với yếu

- **Định nghĩa:**

- **Gió mạnh:** Thường được định nghĩa khi tốc độ gió lớn hơn hoặc bằng 10 m/s. Các trường hợp gió mạnh thường xảy ra khi có bão, áp thấp nhiệt đới hoặc gió mùa Đông Bắc.

– **Gió yếu:** Tốc độ gió thấp hơn 10 m/s, phô biến trong điều kiện thời tiết bình thường hoặc ít tác động của hệ thống khí hậu lớn.

- **Khu vực ven biển:**

- **Tỷ lệ gió mạnh:** 17.53% (cao nhất trong 3 khu vực).

- **Tỷ lệ gió yếu:** 34.97%.

- **Giải thích khí hậu:**

- * Khu vực ven biển chịu ảnh hưởng mạnh từ bão và áp thấp nhiệt đới, đặc biệt vào mùa mưa bão (tháng 6 đến tháng 11).

- * Ngoài ra, các đợt gió mùa Đông Bắc cũng ảnh hưởng rõ rệt, gây ra các cơn gió mạnh ven biển vào mùa đông.

- * Tỷ lệ gió yếu cao có thể phản ánh thời gian bình ổn khí hậu khi không có bão hoặc gió mùa.

- **Khu vực đồng bằng:**

- **Tỷ lệ gió mạnh:** 3.71%.

- **Tỷ lệ gió yếu:** 26.29%.

- **Giải thích khí hậu:**

- * Đồng bằng thường chịu ảnh hưởng ít hơn từ các hệ thống thời tiết mạnh như bão, áp thấp nhiệt đới so với khu vực ven biển. Do đó, tỷ lệ gió mạnh thấp hơn đáng kể.

- * Tỷ lệ gió yếu ở mức trung bình, phản ánh điều kiện khí hậu tương đối ổn định, đặc biệt vào mùa hè và mùa khô.

- **Khu vực miền núi:**

- **Tỷ lệ gió mạnh:** 0.52% (thấp nhất trong 3 khu vực).

- **Tỷ lệ gió yếu:** 16.98%.

- **Giải thích khí hậu:**

- * Do địa hình phức tạp, khu vực miền núi ít chịu ảnh hưởng trực tiếp từ bão và gió mùa hơn so với khu vực ven biển. Điều này lý giải tỷ lệ gió mạnh rất thấp.

- * Tỷ lệ gió yếu phản ánh đặc điểm khí hậu tương đối ổn định trong điều kiện không có các yếu tố thời tiết cực đoan.

- **So sánh và kết luận:**

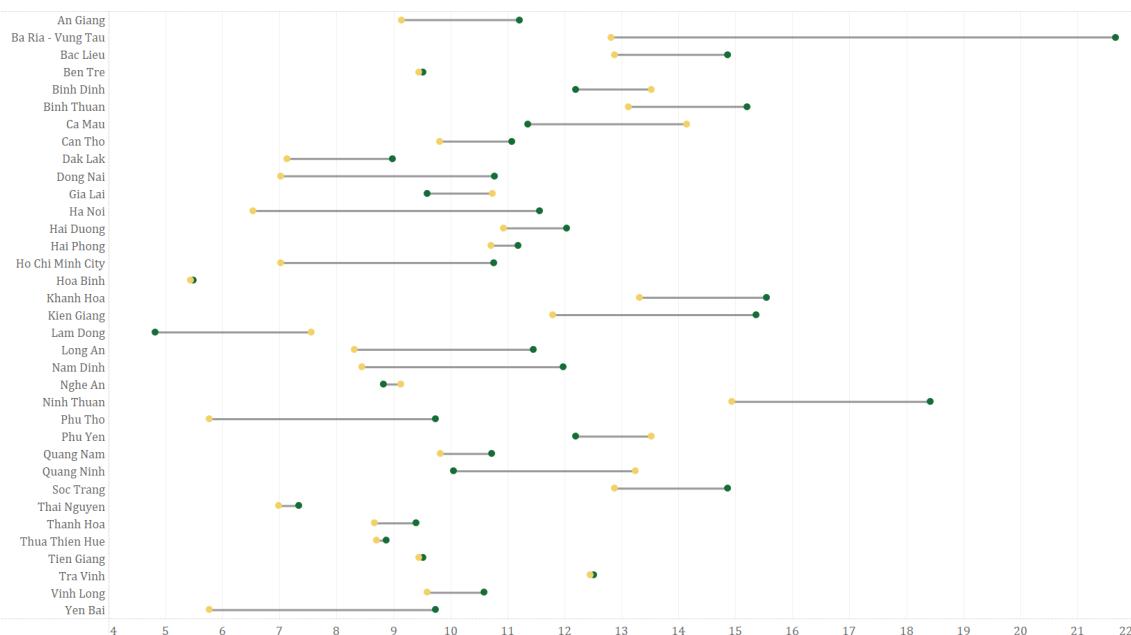
- **Tỷ lệ gió mạnh:** Giảm dần từ khu vực ven biển (17.53%) > đồng bằng (3.71%) > miền núi (0.52%), phản ánh sự giảm dần của mức độ chịu ảnh hưởng từ các hiện tượng thời tiết cực đoan khi đi sâu vào đất liền.

- **Tỷ lệ gió yếu:** Có xu hướng giảm dần từ ven biển > đồng bằng > miền núi, nhưng chênh lệch không đáng kể, cho thấy mức độ ổn định khí hậu tăng lên khi rời xa khu vực biển.

- **Liên hệ khí hậu Việt Nam:**

- **Liên hệ:** Việt Nam là quốc gia nhiệt đới gió mùa, chịu tác động mạnh từ gió mùa Đông Bắc và gió mùa Tây Nam. Khu vực ven biển có tỷ lệ gió mạnh cao nhất do tác động trực tiếp từ bão, áp thấp nhiệt đới, và gió mùa. Khu vực đồng bằng và miền núi ít chịu tác động từ gió mạnh hơn, đặc biệt miền núi do có địa hình chắn gió, giúp giảm thiểu sức gió mạnh từ các hệ thống khí hậu lớn.

3.5 Sự biến động tốc độ gió giai đoạn từ 2009 - 2021



Hình 6: Biểu đồ thê hiện sự biến động tốc độ gió từ 2009-2021

- **Chấm vàng ● :** Biểu thị tốc độ gió trung bình trong năm 2009. Một số khu vực có tốc độ gió cao, nhưng không đồng đều trên cả nước.
- **Chấm xanh ● :** Biểu thị tốc độ gió trung bình trong năm 2021. Có xu hướng biến động ở nhiều khu vực, với một số nơi tăng và một số nơi giảm so với năm 2009.

- **Sự khác biệt:**

- Một số vùng ven biển hoặc vùng núi (ví dụ: duyên hải miền Trung, Tây Nguyên) có tốc độ gió trung bình giảm.
- Ngược lại, một số khu vực đồng bằng hoặc nội địa có tốc độ gió tăng nhẹ.

Nguyên nhân dẫn đến sự tăng/giảm tốc độ gió:

1. **Biến đổi khí hậu toàn cầu:**

- Nhiệt độ toàn cầu tăng lên làm thay đổi chế độ gió và hoàn lưu khí quyển. Sự nóng lên ở một số khu vực gây thay đổi áp suất không khí, làm ảnh hưởng đến tốc độ gió trung bình.
- Việt Nam chịu ảnh hưởng rõ rệt từ biến đổi khí hậu, với các hiện tượng thời tiết cực đoan như bão mạnh hoặc các đợt không khí lạnh bất thường.

2. Thay đổi sử dụng đất và đô thị hóa:

- Ở các khu vực đồng bằng, tốc độ gió có thể tăng do hiệu ứng đô thị hóa, khi các thành phố lớn mở rộng khiến sự chênh lệch nhiệt độ giữa đất và không khí gia tăng. Điều này dẫn đến sự hình thành các cơn gió cục bộ mạnh hơn.
- Với các khu vực rừng núi, việc phá rừng và thay đổi lớp phủ bề mặt làm giảm sức cản của gió, khiến tốc độ gió giảm.

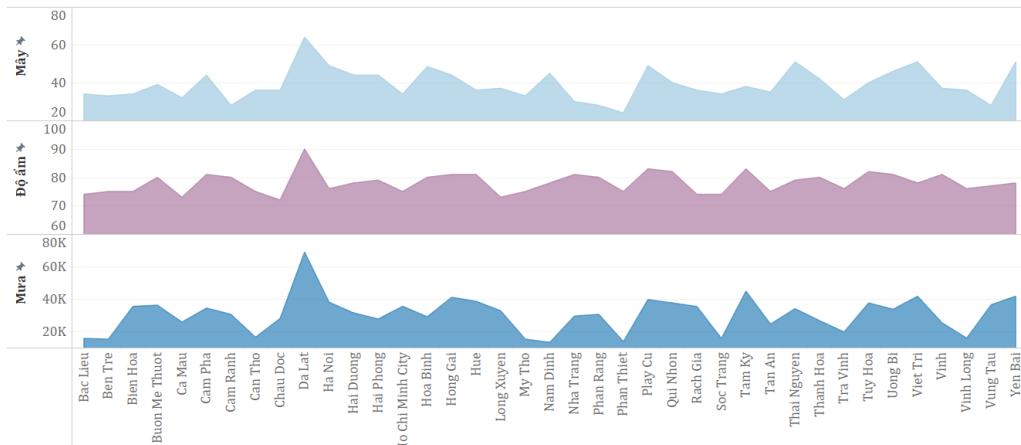
3. Ảnh hưởng của bão và áp thấp nhiệt đới:

- Trong những năm gần đây, tần suất bão và áp thấp nhiệt đới trên biển Đông đã tăng, đặc biệt vào mùa mưa bão (tháng 6–11). Tuy nhiên, do sự thay đổi quỹ đạo và cường độ, tác động của bão lên tốc độ gió trung bình tại các khu vực đất liền có sự biến động.
- Ví dụ: Bão mạnh hơn nhưng di chuyển nhanh hơn qua đất liền dẫn đến sự giảm tốc độ gió trung bình.

4. Thay đổi về hướng gió mùa:

- Hệ thống gió mùa Đông Bắc (mùa khô) và gió mùa Tây Nam (mùa mưa) có sự dịch chuyển trong thời gian và cường độ. Điều này ảnh hưởng lớn đến tốc độ gió ở các vùng duyên hải và đồng bằng ven biển.

3.6 Mối tương quan giữa mây, mưa và độ ẩm



Hình 7: Biểu đồ thể hiện mối tương quan giữa mây, mưa và độ ẩm

- Biểu đồ có 3 chart con đại diện cho các yếu tố:
 - **Mây:** Tần suất xuất hiện mây trung bình theo thời gian.
 - **Mưa:** Lượng mưa trung bình hằng năm.
 - **Độ ẩm:** Độ ẩm không khí trung bình hằng năm.
- Số liệu tính trung bình trong vòng 10 năm qua tổng quan cho thấy:
 - **Mây:** Tăng lên ở một số khu vực, đặc biệt là vùng núi và vùng ven biển.
 - **Mưa:** Có xu hướng giảm nhẹ ở một số khu vực đồng bằng và tăng ở khu vực miền núi.
 - **Độ ẩm:** Không thay đổi đáng kể, nhưng duy trì mức cao (trên 80%).

1. Tác động của mây:

- Mây ảnh hưởng trực tiếp đến nhiệt độ không khí thông qua việc cản bức xạ mặt trời. Ở những vùng có nhiều mây, nhiệt độ ban ngày có xu hướng giảm, trong khi nhiệt độ ban đêm giữ ở mức cao hơn do mây giữ lại bức xạ nhiệt.
- Sự gia tăng tần suất mây tại một số khu vực (như Tây Nguyên, duyên hải Nam Trung Bộ) có thể là do sự thay đổi dòng không khí ẩm từ biển và sự tăng cường của gió mùa Tây Nam.

2. Tác động của mưa:

- Mưa đóng vai trò quan trọng trong việc điều tiết nhiệt độ và độ ẩm không khí.
- Sự gia tăng lượng mưa ở vùng núi (Tây Bắc, Tây Nguyên) có thể do hiệu ứng địa hình khi gió ẩm bị chặn lại bởi các dãy núi, tạo mưa nhiều hơn.
- Đồng bằng sông Hồng và đồng bằng sông Cửu Long có xu hướng giảm lượng mưa do thay đổi về dòng chảy sông và xâm nhập mặn, ảnh hưởng đến chu trình nước.

3. Tác động của độ ẩm:

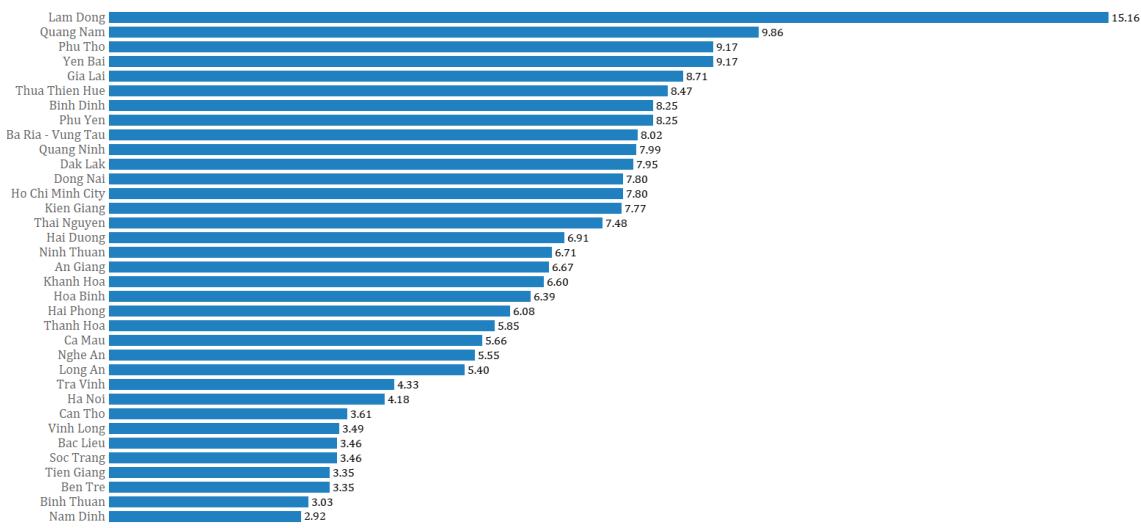
- Độ ẩm cao (80–90%) tại Việt Nam duy trì ổn định do vị trí địa lý nằm gần đường xích đạo và chịu ảnh hưởng của biển. Điều này dẫn đến thời tiết nóng ẩm đặc trưng, đặc biệt vào mùa hè.
- Tuy nhiên, độ ẩm cao kết hợp với nhiệt độ tăng có thể dẫn đến cảm giác oi bức hơn, tác động đến sức khỏe con người và các hoạt động kinh tế.

Phân tích theo khu vực

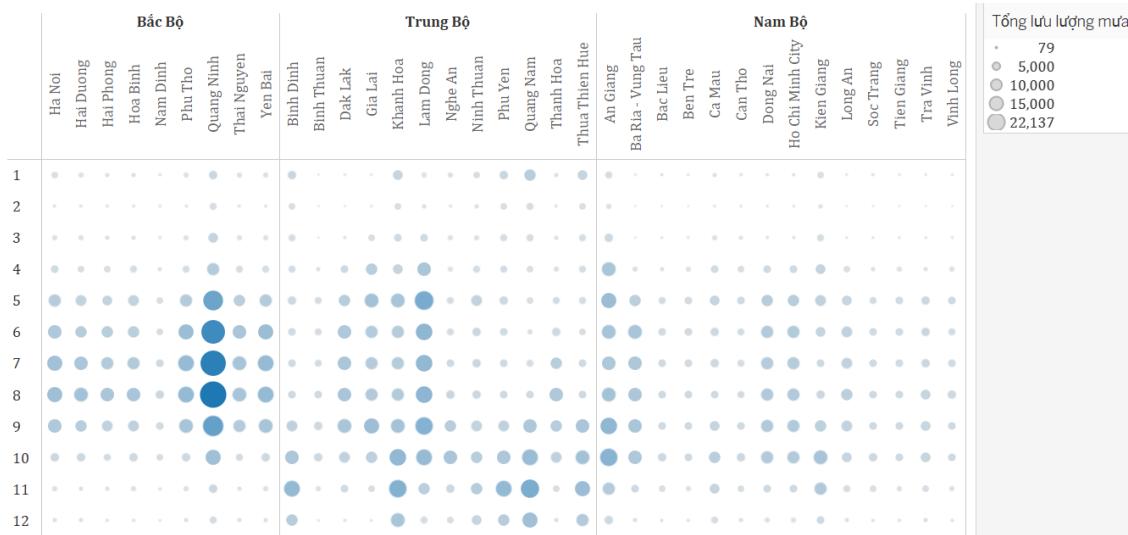
Khu vực	Mây	Mưa	Độ ẩm
Bắc	Hà Nội, Hải Phòng, Nam Định: Có tần suất mây trung bình, đặc biệt thấp hơn vào mùa hè do thời tiết khô và nóng. Yên Bái, Hòa Bình, Thái Nguyên: Tần suất mây cao hơn do ảnh hưởng của địa hình đồi núi và hơi ẩm từ biển mang đến.	Yên Bái, Hòa Bình: Lượng mưa lớn nhất, đặc biệt vào mùa mưa do hiệu ứng địa hình khi gió ẩm bị chặn bởi núi cao. Hà Nội, Nam Định, Hải Phòng: Mưa ít hơn vùng núi, nhưng vẫn ở mức trung bình cao do khí hậu nhiệt đới gió mùa.	Hải Phòng, Hòn Gai, Uông Bí: Độ ẩm rất cao (trên 85%) do gần biển, đặc biệt vào mùa hè khi gió biển Tây Nam hoạt động mạnh. Hà Nội, Việt Trì, Thái Nguyên: Độ ẩm tương đối cao quanh năm, nhưng mùa đông gió mùa Đông Bắc làm độ ẩm tăng đột biến.
	Đà Lạt: Tần suất mây cao nhất khu vực miền Trung, gần như quanh năm, do độ cao và khí hậu mát mẻ cận ôn đới. Phan Rang, Phan Thiết: Ít mây nhất khu vực do khí hậu khô hạn, đặc biệt vào mùa khô (tháng 3 - 8).	Huế, Tam Kỳ, Nha Trang: Lượng mưa lớn vào mùa mưa bão (tháng 9 - 12), đặc biệt Huế ghi nhận lượng mưa cao nhất cả nước do hiệu ứng địa hình. Phan Rang, Phan Thiết: Mưa ít nhất khu vực, đặc trưng bởi khí hậu khô hạn.	Huế, Nha Trang, Quy Nhơn: Độ ẩm cao quanh năm (trên 80%), đặc biệt vào mùa mưa. Phan Rang, Phan Thiết: Độ ẩm thấp hơn đáng kể (dưới 70%) do khí hậu khô cằn, nắng nhiều.
Nam	TP.HCM, Cần Thơ, Mỹ Tho: Tần suất mây cao vào mùa mưa (tháng 5 - 11), nhưng giảm rõ rệt vào mùa khô (tháng 12 - 4). Vũng Tàu, Bạc Liêu: Mây xuất hiện ổn định hơn do gần biển và gió mùa Tây Nam mang hơi ẩm.	Cần Thơ, Rạch Giá, Bạc Liêu: Lượng mưa cao nhất khu vực do ảnh hưởng từ gió mùa Tây Nam và mạng lưới sông ngòi dày đặc. TP.HCM, Vũng Tàu: Lượng mưa lớn vào mùa mưa nhưng phân hóa rõ rệt giữa mùa mưa và mùa khô.	Cần Thơ, Bạc Liêu, Sóc Trăng: Độ ẩm luôn duy trì trên 85% do mạng lưới sông ngòi, đặc biệt vào mùa mưa. TP.HCM, Tân An: Độ ẩm giảm nhẹ vào mùa khô, nhưng vẫn duy trì mức cao trong các tháng còn lại.

Bảng 4: Tần suất mây, mưa, và độ ẩm theo vùng miền

3.7 Đánh giá lượng mưa của từng tỉnh theo khu vực và khoảng thời gian trong năm



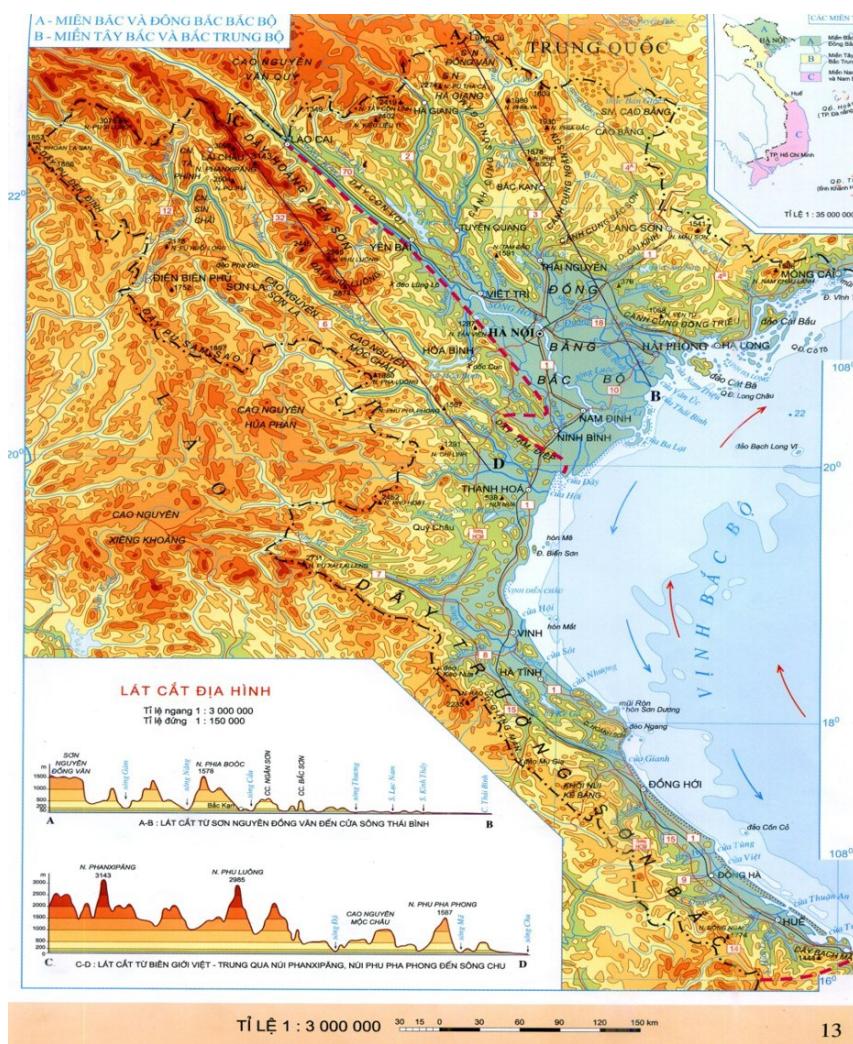
Hình 8: Trung bình lượng mưa của các tỉnh



Hình 9: Thông kê tổng lượng mưa của các tỉnh tại các thời điểm trong năm

- Nhìn chung ta thấy được xu hướng chung của cả nước:
 - Đặc điểm khí hậu ở Việt Nam là nhiệt đới ẩm gió mùa có 2 mùa đặc trưng là mùa khô và mùa mưa.
 - * Mùa mưa: bắt đầu vào tháng 5 và kéo dài đến tháng 10. Chiếm khoảng 80% tổng lượng mưa cả năm, đi kèm với thời tiết nóng ẩm. Do chịu ảnh hưởng của gió mùa Tây Nam (Tây ôn đới) và dải hội tụ nhiệt đới.

- * Mùa khô: Bắt đầu vào tháng 11 và kéo dài đến tháng 4 năm sau. Ít mưa, thời tiết hanh khô và lạnh vào mùa đông và se lạnh vào mùa xuân. Do chịu ảnh hưởng của gió mùa Đông Bắc (Tín phong).
 - Và xu hướng chung đó đã được thể hiện một cách khá rõ nét trong biểu đồ Hình 8.
- Trước tiên ta cần nhìn vào bản đồ địa hình chung và địa hình từng vùng, sau đó phân tích và so sánh lượng mưa của từng vùng theo các thời gian trong năm, bởi vì địa hình và gió theo từng thời điểm ảnh hưởng rất nhiều vào lượng mưa của một khu vực:
- Trước tiên ta nhìn vào các vùng ở Bắc bộ Hình 9 và bên phải Hình 8:



Hình 10: Địa hình Bắc bộ và Bắc Trung bộ

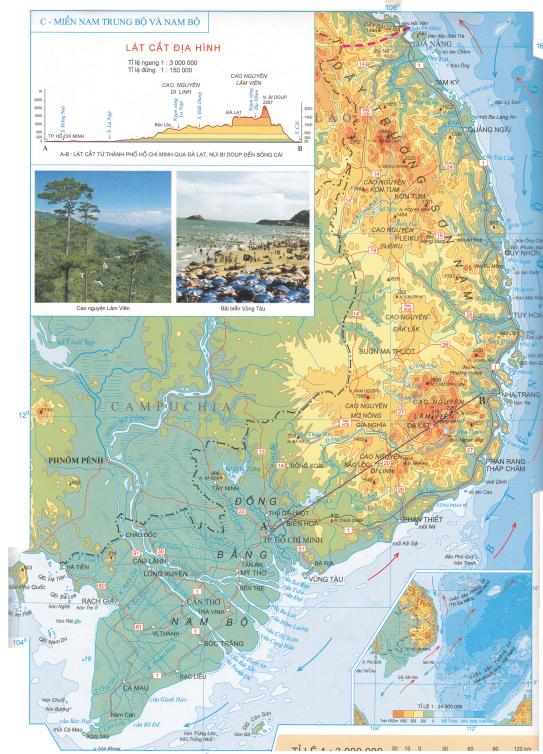
- Vào mùa khô, lượng mưa ở tất cả các tỉnh miền Bắc gần như bằng nhau:
 - * Từ tháng 11 đến tháng 4 năm sau là khoảng thời gian gió mùa Đông Bắc (Tín phong) hoạt động một cách mạnh mẽ mang đến thời tiết lạnh và khô.

- * Bên cạnh đó hãy nhìn vào địa hình "phẫu" tại các cánh cung như Cánh Cung Bắc Sơn, Cánh Cung Ngàn Sơn và Cánh Cung Sông Gam. Cả 3 cánh cung tạo nên địa hình đón gió từ phía Đông Nam, hội tụ sao đó đưa luồng gió lạnh và khô vào khu vực đồng bằng sông Hồng.
- * Vì vậy có thể nói vùng Bắc Bộ là nơi chịu ảnh hưởng nhiều nhất từ gió mùa Tây Nam. Đây là nguyên nhân gây ra tình trạng lượng mưa thấp trong khoảng thời gian này. Bên cạnh những nơi khác cũng chịu ảnh hưởng đến lượng mưa trong khoảng thời gian này.
- Ở mùa mưa, nơi đây có lượng mưa khá cao.
 - * Tuy nhiên, tỉnh Nam Định lại có lượng mưa tương đối ít so với các tỉnh còn lại. Vì vị trí địa lý khá đặc biệt - nằm sát phía nam và tiếp giáp với khu vực Bắc Trung Bộ, nên tính chất khí hậu nơi đây có đặc điểm của khu vực Bắc Trung Bộ nhiều hơn là Đồng Bằng sông Hồng.
 - * Bên cạnh đó ta thấy được sự phân bố không đều về lượng mưa ở các tỉnh: Các khu vực gần biển (Quảng Ninh, Hải Phòng) có nhiều mưa hơn so với khu vực sâu trong đất liền (Hà Nội, Hải Phòng).
- Các tỉnh miền Trung cho thấy một sự khác biệt khá lớn về tổng lượng mưa suốt các khoảng thời gian trong năm. Điều này có thể được lý giải do dựa vào địa hình và khí hậu thì khu vực miền trung được chia thành các vùng: Bắc Trung Bộ, Nam Trung Bộ và Tây Nguyên.
 - Khu vực Bắc Trung Bộ: Với đường ranh giới từ dãy Bạch Vân (tiếp giáp Nam Trung Bộ) đến phía Nam khu vực Đồng bằng sông Hồng. Gồm các tỉnh Thanh Hóa, Nghệ An, Hà Tĩnh, Quảng Bình, Quảng Trị, Thừa Thiên Huế (Thanh-Nghệ-Tĩnh, Bình-Trị-Thiên).
 - * Mùa mưa:
 - Lượng mưa rất lớn so với mùa khô, đặc biệt vào tháng 7, tháng 8 và tháng 9. Các cơn bão và áp thấp nhiệt đới có thể gây mưa lớn, tạo ra lượng mưa dồn dập trong thời gian ngắn.
 - Mưa trong mùa này chủ yếu là mưa rào, mưa giông, đôi khi xuất hiện mưa phùn, đặc biệt là vào tháng 8 và tháng 9. Lượng mưa có thể dao động mạnh trong khu vực và từng năm, tùy thuộc vào tác động của các hiện tượng thời tiết toàn cầu như El Niño hay La Niña.
 - Các tỉnh như Quảng Bình, Quảng Trị và Thừa Thiên Huế (trong biểu đồ chỉ có Thừa Thiên Huế) thường chịu ảnh hưởng của gió mùa Đông Bắc và các cơn bão, nên lượng mưa tại các khu vực này có thể cao hơn các khu vực nội địa trong mùa mưa.
 - Tuy nhiên lượng mưa ở khu vực Bắc Trung Bộ lại thấp hơn so với các khu vực khác rất nhiều kể cả vào mùa mưa là do sự hoạt động mạnh mẽ của gió mùa Tây Nam thổi theo hướng từ Lào, Thái Lan đi qua địa hình đồi núi là dãy Hoàng Liên Sơn (gió

Lào) mang đến khí hậu nóng oi bức và khô hạn. Từ đó dẫn đến nhiệt độ nơi đây trong khoảng thời gian này khá cao, đặc biệt là khô hạn kéo dài có thể gây ra cháy rừng. Ngoài ra, các khu vực miền Bắc cũng chịu ảnh hưởng nhỏ từ gió Lào.

* Mùa khô:

- Lượng mưa thấp, kéo dài từ tháng 11 đến tháng 4 năm sau, với tháng 12 và tháng 1 là những tháng ít mưa nhất. Tổng lượng mưa trong các tháng này qua nhiều năm chỉ khoảng 50-100 mm.
- Mưa thường xuất hiện dưới dạng mưa rào nhẹ hoặc mưa nhỏ vào cuối mùa đông, đặc biệt ở các khu vực ven biển như Thừa Thiên Huế. Tuy nhiên, lượng mưa trong mùa khô không đáng kể, và không ảnh hưởng nhiều đến sinh hoạt hàng ngày.



Hình 11: Địa hình khu vực Nam Trung Bộ, Tây Nguyên, Đông Nam Bộ và Đông Băng Sông Cửu Long

- Khu vực Nam Trung Bộ, bao gồm các tỉnh từ Đà Nẵng đến Bình Thuận, có sự phân hóa rõ rệt theo mùa, với đặc điểm khí hậu chịu ảnh hưởng của gió mùa và địa hình. Các tỉnh trong khu vực này bao gồm: Đà Nẵng, Quảng Nam, Quảng Ngãi, Bình Định, Phú Yên, Khánh Hòa, Ninh Thuận, và Bình Thuận.

* Mùa mưa: Mùa mưa ở Nam Trung Bộ thường bắt đầu muộn hơn so với Bắc Trung Bộ, kéo dài từ tháng 9 đến tháng 12, với đỉnh điểm vào tháng 10 và tháng 11. Có thể thấy lượng mưa của những nơi thuộc khu vực Nam Trung Bộ như Bình Định, Khánh Hòa, Ninh Thuận,

Phú Yên, Quảng Nam có sự phân bố của lượng mưa khác lêch so với những nơi khác.

- * Cho nên, thời gian mùa khô kéo dài từ tháng 1 đến tháng 8, với lượng mưa rất thấp ở hầu hết các tỉnh. Đây là giai đoạn chịu ảnh hưởng của gió mùa Đông Bắc khô và gió mùa Tây Nam từ tháng 5 trở đi.
- Khu vực Tây Nguyên: là khu vực nằm ở phía Tây của miền Trung Việt Nam, bao gồm 5 tỉnh: Kon Tum, Gia Lai, Đăk Lăk, Đăk Nông, và Lâm Đồng. Với địa hình cao nguyên và đặc điểm khí hậu nhiệt đới gió mùa (khác nhiệt đới ẩm gió mùa).
 - * Lượng mưa ở Tây Nguyên thường cao hơn so với các tỉnh duyên hải miền Trung (từ Thanh Hóa đến Bình Thuận) do đặc điểm địa hình và khí hậu khác biệt. Với tổng lượng mưa các tháng này rơi vào khoảng 10.00 mm - 20.000 mm, tập trung chủ yếu vào mùa mưa (tháng 5 đến tháng 10).
 - * Các khu vực như Đăk Lăk, Lâm Đồng nhận lượng mưa lớn hơn, nhờ vị trí gần cao nguyên và hiệu ứng gió mùa Tây Nam và địa hình phía Tây Nam dãy Hoàng Liên Sơn: địa hình đón gió mùa Tây Nam, mang hơi ẩm từ Án Độ Dương qua. Khi gió Tây Nam thổi đến Tây Nguyên, gặp địa hình cao nguyên và núi, hơi ẩm bị đẩy lên cao, gây mưa lớn theo hiệu ứng Foehn ngược (mưa nhiều ở sườn đón gió).
- Các khu vực miền Nam (Hình 10 và nhóm tỉnh bên phải Hình 8): Nhìn chung có thể thấy được lượng mưa ở miền Nam thấp hơn khá nhiều so với miền Trung và miền Bắc. Và nội bộ miền Nam thì ta thấy được tình trạng khô hạn xảy ra nhiều hơn ở các tỉnh Đồng Bằng Sông Cửu Long (DBSCL).
 - Khu vực Đông Nam Bộ bao gồm 6 tỉnh/thành phố: Thành phố Hồ Chí Minh, Bình Dương, Đồng Nai, Bà Rịa-Vũng Tàu, Tây Ninh, Bình Phước.
 - * Mùa mưa: Mùa mưa ở Đông Nam Bộ kéo dài từ tháng 5 đến tháng 10, và chiếm khoảng 85-90% tổng lượng mưa hàng năm. Mưa chủ yếu là mưa rào và có thể kèm theo dông. Mưa thường xảy ra vào buổi chiều và tối, đôi khi vào ban ngày nhưng thường không kéo dài lâu. Các cơn mưa có thể xuất hiện đột ngột và với lượng mưa lớn trong thời gian ngắn, gây ngập úng cục bộ. Những khu vực như TP.HCM và Bình Dương thường có lượng mưa cao hơn so với các khu vực xa biển. Do gió mùa Tây Nam từ Án Độ Dương thổi vào Việt Nam, mang theo hơi ẩm lớn. Khi gặp địa hình bằng phẳng của Đông Nam Bộ, hơi ẩm này dễ dàng ngưng tụ và gây mưa lớn. Các rãnh áp thấp và bão cũng có thể đi vào khu vực từ biển Đông, làm tăng lượng mưa trong mùa này.
 - * Mùa khô: Trong mùa khô, lượng mưa hầu như không có, nếu có mưa thì chỉ là những cơn mưa ngắn và không đáng kể. Gió mùa Đông Bắc thổi từ lục địa vào, mang theo không khí khô và ít ẩm. Dãy Trường Sơn chắn gió biển từ phía Đông, khiến khu vực này không nhận được nhiều hơi ẩm từ biển trong mùa khô.

– Khu vực Đồng Bằng Sông Cửu Long:

* Mùa mưa:

- Mưa ở DBSCL là mưa rào kéo dài, có thể kéo dài suốt ngày hoặc vài ngày liên tiếp, gây ngập úng tại nhiều khu vực, đặc biệt là vào các tháng 7-9.
- Các cơn mưa không chỉ gây ẩm ướt mà còn tạo ra hiện tượng mưa ngập vào mùa lũ, làm ảnh hưởng đến sinh hoạt và sản xuất. Từ đó tạo nên hiện tượng mưa nước nổi ở các tỉnh miền Tây Nam Bộ.
- Những khu vực có địa hình thấp như Cần Thơ, Sóc Trăng, Kiên Giang thường nhận lượng mưa lớn hơn, với thời gian mưa kéo dài hơn.
- Nguyên nhân có thể do gió mùa Tây Nam mang theo độ ẩm từ Ấn Độ Dương và gặp vùng đất trũng của DBSCL, gây mưa lớn trong mùa này. Hiện tượng hội tụ nhiệt đới và bão từ biển Đông cũng ảnh hưởng đến DBSCL, làm mưa lớn và kéo dài. DBSCL có hệ thống sông ngòi dày đặc, nên độ ẩm trong không khí tăng cao và tạo điều kiện cho mưa denses hình thành.

- * Mùa khô: mùa khô ở DBSCL thường có hạn hán kéo dài dẫn đến hiện tượng xâm nhập mặn, đặc biệt ở những vùng ven biển và khu vực thấp trũng gần các cửa sông lớn. Lượng mưa trong mùa khô rất ít, gần như không có mưa, khiến tình trạng hạn hán càng trở nên nghiêm trọng, đặc biệt là các tỉnh ven biển như Bến Tre, Trà Vinh. Đặc biệt là khi thủy triều lên xuồng và sự tác động của hệ thống sông ngòi khiến mùa khô tại DBSCL càng trở nên khắc nghiệt, gây hiện tượng xâm nhập mặn sâu vào đất liền.

4 Tích hợp AI

4.1 Ý tưởng thực hiện

- Model AI được sử dụng: API được cung cấp bởi Gemini (gemini-1.5-flash).
- Trước tiên dùng API để truy vấn vào database đang lưu trữ dataset theo yêu cầu của câu hỏi truyền vào.
- Tiếp theo có thể dùng data vừa được truy vấn và vẽ biểu đồ theo yêu cầu được nhập vào.
- Biểu đồ sẽ được ghi nhận lại và dùng vào một chatbox khác để gửi một yêu cầu phân tích dựa trên biểu đồ.

4.2 Thực hiện ý tưởng

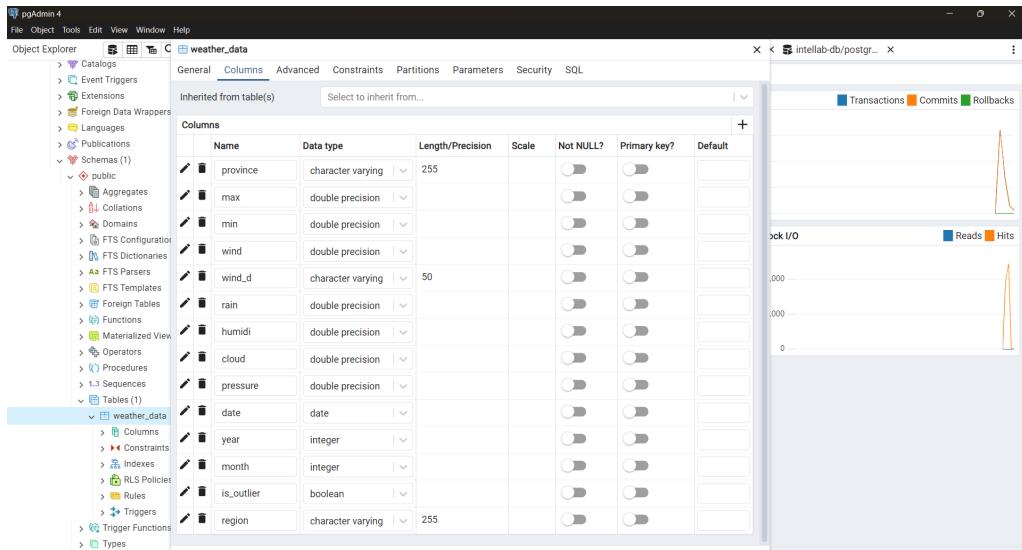
4.2.1 Các yêu cầu cấu hình

Cần có các thư viện của python như sau:

- **streamlit**: Dùng để host web cho dashboard và chatbox.
- **pandas**: Dùng để xử lý dữ liệu lấy từ dataset.
- **base64**: Dùng để mã hóa và xử lý hình ảnh.
- **pg8000**: Dùng để kết nối với database postgresql.
- **dotenv**: Dùng để load các biến môi trường lưu trong file .env.
- **os**: Dùng để hỗ trợ cho việc đọc file .env.
- **google.generativelai**: Dùng để gửi request, nhận response từ API của gemini.

4.2.2 Truy vấn dataset

- Trước tiên cần import dataset vào database:
 - Tạo một database và schema để lưu dataset như sau:



Hình 12: Schema để lưu dataset

- Dataset được lưu vào một file .csv. Vậy nên để import vào thì ta có thể dùng nhiều cách nhưng cách đơn giản nhất là dùng lệnh.

```
COPY weather_data(province, max, min, wind, wind_d, rain, humidi, cloud,
                  pressure, date, year, month, is_outlier, region)
FROM '/path/to/weather_data.csv'
DELIMITER ','
CSV HEADER;
```

Hình 13: Câu lệnh import file csv vào table

- Sau đó ta được bảng dữ liệu sau:

	province	max	min	wind	wind_d	rain	humidi	cloud	pressure	date	year	month	is_outlier	region
1	Bac Lieu	27	22	17	NNE	0	6.9	90						
2	Bac Lieu	31	25	20	ENE	0	64							
3	Bac Lieu	29	24	14	E	0	75							
4	Bac Lieu	30	24	30	E	0	79							
5	Bac Lieu	31	25	20	ENE	0	70							
6	Bac Lieu	28	23	14	ENE	0	75							
7	Bac Lieu	29	23	10	ENE	0.4	75							
8	Bac Lieu	32	24	22	ENE	0	63							
9	Bac Lieu	30	24	20	ENE	0.5	76							
10	Bac Lieu	29	23	16	E	0	70							
11	Bac Lieu	27	22	19	WSW	4.5	73							
12	Bac Lieu	32	23	22	ENE	0	61							
13	Bac Lieu	28	21	9	WSW	0	74							
14	Bac Lieu	28	21	12	ENE	0	74							
15	Bac Lieu	31	25	20	E	0.6	68							
16	Bac Lieu	31	23	16	E	0	69							
17	Bac Lieu	30	24	17	E	0.5	73							

Hình 14: Bảng dữ liệu trong database.

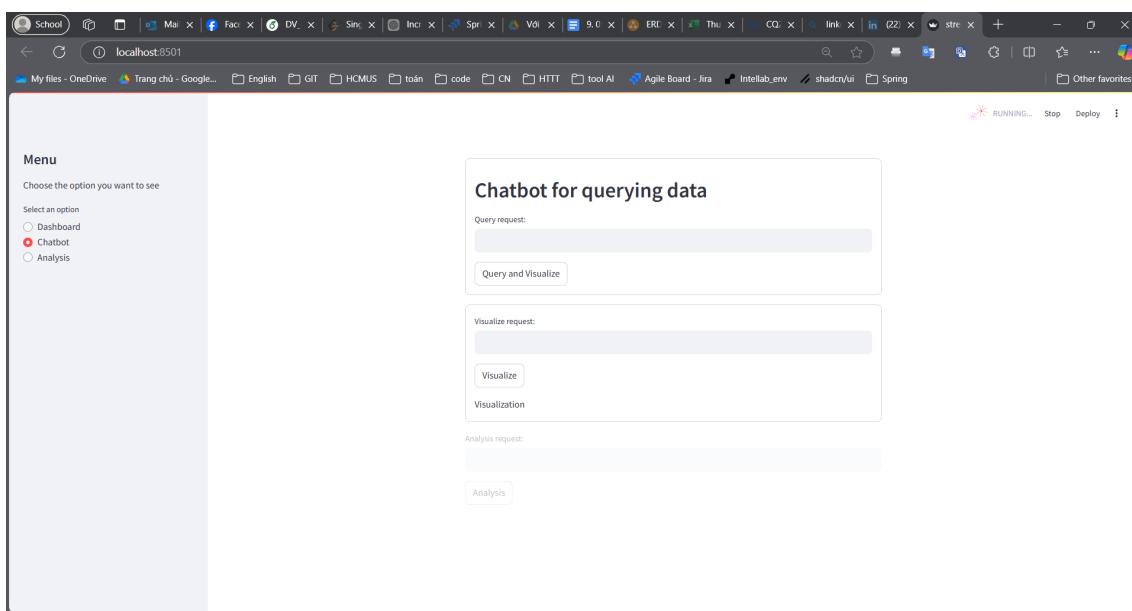
- Thiết kế câu prompt để nhận được response theo ý muốn.

- Khi gửi request thì sẽ gửi câu prompt + câu hỏi để mô tả trỏ nên cụ thể hóa mà không cần phải nhập liệu quá nhiều.
- Câu prompt sẽ được thiết kế sẵn như sau:

```
base_prompt = """
You are an expert in converting English questions to postgres SQL query!\n
The SQL database has the name "weather" and the table name is "weather_data".\n
CREATE TABLE weather_data (\n
    province VARCHAR(255),\n
    max FLOAT,\n
    min FLOAT,\n
    wind FLOAT,\n
    wind_d VARCHAR(50),\n
    rain FLOAT,\n
    humidi FLOAT,\n
    cloud FLOAT,\n
    pressure FLOAT,\n
    date DATE,\n
    year INT,\n
    month INT,\n
    is_outlier BOOLEAN,\n
    region VARCHAR(255)\n);
\n
max - maximum temperature in Celsius, in a day\n
min - minimum temperature in Celsius, in a day\n
When I ask a question for the max in Hanoi in 2021, the SQL query should be "SELECT year, max(max) as "max_temperature_of_year" FROM weather_data WHERE province = 'Hanoi' AND year = 2021".\n
When I ask a question for the min during the year 2021, the SQL query should be "SELECT date, province, min FROM weather_data WHERE year = 2021"\n
For example:\n
EX1: if the question is "What is the maximum temperature in Hanoi? in 2021",\n
the SQL query should be "SELECT year, max(max) FROM weather_data WHERE province = 'Hanoi' AND year = 2021".\n
EX2: if the question is "What is the minimum temperature in Hanoi in 2021?",\n
the SQL query should be "SELECT year, min(min) FROM weather data WHERE province = 'Hanoi' AND year = 2021".\n
EX3: if the question is "What is the humidi in Hanoi in 2021?",\n
the SQL query should be "SELECT date, humidi FROM weather_data WHERE province = 'Hanoi' AND year = 2021".\n
If using any function in the SQL query, please define the name of the column name in the output.\n
Also, the Postgres code should not have ``` in the beginning or end and the Postgres word in the output.\n
"""

```

Hình 15: Câu prompt dành cho việc truy vấn



Hình 16: Giao diện chatbox

- Ví dụ: Câu hỏi được đặt là What is the humidity of Bac Lieu in 2015?

Chatbot for querying data

Query request:

What is the humidity of Bac Lieu for each month in 2015 ?

Query

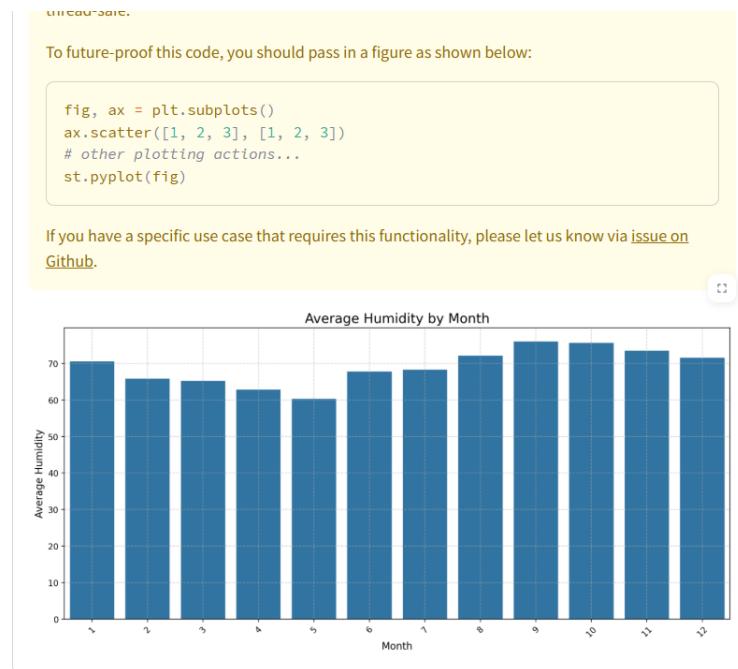
Response: SELECT month, AVG(humidity) AS average_humidity FROM weather_data WHERE province = 'Bac Lieu' AND year = 2015 GROUP BY month;

	month	average_humidity
0	1	70.6129
1	2	65.8571
2	3	65.2258
3	4	62.8667
4	5	60.3226
5	6	67.8
6	7	68.2903
7	8	72.129
8	9	76
9	10	75.6452

Hình 17: Phản hồi từ câu hỏi truy vấn.

4.2.3 Vẽ biểu đồ

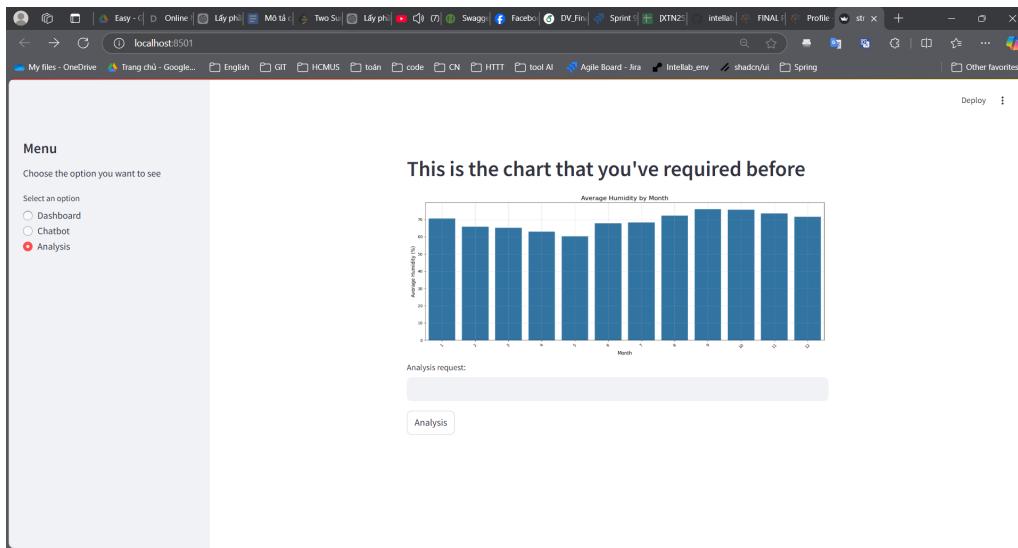
- Tương tự như truy vấn thì ở đây ta thiết kế câu prompt để dùng API tạo script python cho việc vẽ biểu đồ sau đó bằng cách run script python đó.
- Bên cạnh đó thì ta đọc dữ liệu đã được truy vấn trước đó để vẽ sau đó in ra màn hình và lưu lại.



Hình 18: Biểu đồ được vẽ bởi AI theo yêu cầu và dữ liệu vừa được truy vấn

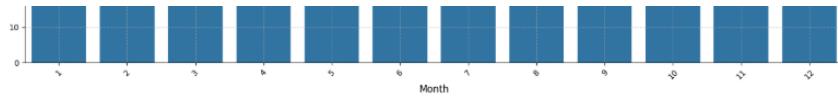
4.2.4 Đánh giá phân tích dựa trên biểu đồ được vẽ

- Từ biểu đồ vừa được vẽ ta có thể chuyển qua tab Analysis để gửi yêu cầu phân tích dựa trên biểu đồ đó.



Hình 19: Giao diện trang yêu cầu phân tích

- Ví dụ: Ta gửi yêu cầu phân tích xu hướng về độ ẩm của năm đó và một số yêu tố có thể là nguyên nhân tạo ra xu hướng đó.



Analysis request:

Give me some analysis about the trend of the humidity of that year and some reason lead that trend

Analysis

Response: Here's an analysis of the average humidity trend throughout the year, along with possible contributing factors:

Trend:

The data shows a general U-shaped pattern in average humidity. Humidity is highest in the late summer and early autumn (August-October), with a peak in September. It gradually decreases towards winter (November-December) before bottoming out in spring (April-May) and then gradually increasing again towards the summer months.

Possible Reasons for this Trend:

Several factors can contribute to this humidity pattern:

- **Seasonal Temperature Changes:** Temperature is the most dominant influence on humidity. Warmer months (June-August) generally have higher humidity due to increased evaporation from bodies of water, soil, and vegetation. The decrease in temperature during autumn and winter leads to less evaporation, hence lower humidity.
- **Precipitation:** The amount and timing of rainfall significantly impacts humidity. Periods with higher rainfall often lead to higher humidity, assuming sufficient evaporation also occurs. The specific

Hình 20: Phản hồi từ yêu cầu phân tích