

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM
KHOA CÔNG NGHỆ THÔNG TIN



HCMUTE

BÁO CÁO CUỐI KÌ

MÔN: LẬP TRÌNH PYTHON

Đề tài:

**PHÂN TÍCH VÀ TRỰC QUAN HÓA DỮ LIỆU
THỜI TIẾT Ở NƯỚC ANH**

GV: ThS. Trần Quang Khải

SVTH: Bùi Thanh Tùng 23110357

Phạm Phúc Tiến 23110339

Nguyễn Tấn Tài 23110307

Thành phố Hồ Chí Minh, tháng 11 năm 2024

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM
KHOA CÔNG NGHỆ THÔNG TIN



HCMUTE

BÁO CÁO CUỐI KÌ

MÔN: LẬP TRÌNH PYTHON

Đề tài:

THỜI TIẾT Ở ANH

GV: *ThS. Trần Quang Khải*

SVTH: Bùi Thanh Tùng 23110357

 Phạm Phúc Tiến 23110339

 Nguyễn Tấn Tài 23110307

Thành phố Hồ Chí Minh, tháng 11 năm 2024

DANH SÁCH THÀNH VIÊN THAM GIA ĐỒ ÁN

Đề tài: Thời tiết ở Anh

STT	HỌ VÀ TÊN THÀNH VIÊN	MÃ SỐ SINH VIÊN	TỶ LỆ THAM GIA
1	Bùi Thanh Tùng	23110357	100%
2	Phạm Phúc Tiến	23110339	100%
3	Nguyễn Tấn Tài	23110307	100%

Ghi chú:

Tỷ lệ %: Mức độ phân trăm hoàn thành của từng sinh viên tham gia.

Trưởng nhóm: Bùi Thanh Tùng

Nhận xét của giảng viên

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Tp. Hồ Chí Minh - Tháng 11 năm 2024

LỜI CẢM ƠN

Kính gửi thầy Trần Quang Khải,

Em xin gửi lời cảm ơn chân thành đến thầy vì những kiến thức quý báu và sự tận tâm trong giảng dạy môn Lập trình Python. Nhờ thầy mà em đã học được rất nhiều kỹ năng và hiểu sâu hơn về lập trình, giúp em có thể ứng dụng trong học tập và công việc sau này.

Thầy sinh ngày 08/11/1991 và là cựu học sinh của trường Bưu chính Viễn thông Miền Nam. Những kinh nghiệm và kiến thức mà thầy chia sẻ đã giúp chúng em hiểu rõ hơn về môn học và tạo động lực để chúng em không ngừng cố gắng và phát triển. Thầy luôn luôn dành thời gian giải đáp mọi thắc mắc của chúng em, dù là những vấn đề nhỏ nhất. Sự tận tình và nhiệt huyết của thầy đã truyền cảm hứng rất nhiều cho chúng em.

Em cũng rất cảm ơn thầy vì những bài giảng phong phú và sáng tạo, giúp chúng em không chỉ hiểu lý thuyết mà còn có thể áp dụng thực tế. Thầy luôn khuyến khích chúng em tự khám phá và thực hành, điều này đã giúp em tự tin hơn trong việc xử lý các vấn đề phức tạp và phát triển kỹ năng lập trình của mình. Thầy không chỉ là một người thầy, mà còn là một người hướng dẫn, một người truyền cảm hứng thực sự.

Em xin kính chúc thầy luôn mạnh khỏe, hạnh phúc và thành công trong sự nghiệp giảng dạy. Mong thầy sẽ tiếp tục dẫn dắt và truyền đạt kiến thức cho các thế hệ học sinh sau này. Sự cống hiến và tâm huyết của thầy là nguồn động lực lớn lao cho chúng em, và em tin chắc rằng những bài học từ thầy sẽ luôn là hành trang quý báu trên con đường sự nghiệp của mỗi chúng em. Trân trọng.

Đại diện nhóm Bùi Thanh Tùng

MỤC LỤC

LỜI CẢM ƠN	iii
MỤC LỤC.....	iv
MỤC LỤC HÌNH ẢNH.....	vi
CHƯƠNG 1. PHẦN MỞ ĐẦU	1
1.1. Giới thiệu chung	1
1.2. Mục tiêu nghiên cứu	1
CHƯƠNG 2. TỔNG QUAN LÝ THUYẾT	2
2.1. Thời tiết.....	2
2.1.1. Định nghĩa	2
2.1.2. Các yếu tố thời tiết chính.....	2
2.1.3 Các phương pháp dự báo thời tiết hiện tại	2
2.2. Khoa học dữ liệu.....	3
2.2.1. Định nghĩa	3
2.2.2. Làm sạch dữ liệu	3
2.2.3. Trực quan hóa dữ liệu.....	3
2.2.4. Cách thức áp dụng trong việc phân tích và dự báo thời tiết.....	3
CHƯƠNG 3. PHƯƠNG PHÁP NGHIÊN CỨU	4
3.1. Thu thập dữ liệu	4
3.1.1. Mô tả nguồn dữ liệu	4
3.1.2. Tiêu chí chọn mẫu nghiên cứu	4
3.2. Xử lý dữ liệu	5
3.3. Phát triển giao diện quá lý và trực quan hóa dữ liệu	6
3.3.1 Lựa chọn công cụ và ngôn ngữ lập trình.....	6
3.3.2 Chức năng của giao diện	6
3.3.3 Hiện thị và trực quan hóa dữ liệu	7
CHƯƠNG 4. KẾT QUẢ VÀ THẢO LUẬN	8
4.1. Trình bày kết quả.....	8
4.2. Phân tích kết quả.....	13
4.2.1. Đánh giá mức độ ứng dụng của giao diện.....	13
4.2.2. Thảo luận về các yếu tố ảnh hưởng đến kết quả	13
4.2.3. Đề xuất các cải tiến cho phần mềm.....	14
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TƯƠNG LAI.....	16
5.1. Tổng kết	16
5.1.1. Tóm tắt các phát hiện chính	16

5.1.2. Đánh giá về mức độ thành công của bài báo cáo	16
5.2. Hướng phát triển tương lai.....	16
5.2.1. Các khía cạnh còn hạn chế và cách khắc phục.....	16
5.2.2. Đề xuất các nghiên cứu tiếp theo	16
5.3. Tác động của nghiên cứu	17
5.3.1. Tác động đến lĩnh vực phân tích dữ liệu thời tiết.....	17
5.3.2. Khả năng ứng dụng công nghệ trong phân tích và trực quan hóa dữ liệu thời tiết.....	17
Tài liệu tham khảo.....	18

MỤC LỤC HÌNH ẢNH

Hình 4.1. Giao diện màn hình chờ	8
Hình 4.2. Giao diện ban đầu của các dạng biểu đồ	8
Hình 4.3. Biểu đồ Summary Count.....	9
Hình 4.4. Biểu đồ Month Count.....	9
Hình 4.5. Biểu đồ Precip Type Count.....	10
Hình 4.6. Biểu đồ Average Humidity by Summary.....	10
Hình 4.7. Biểu đồ Correlation Heatmap	11
Hình 4.8. Giao diện tương tác các số liệu thống kê thời tiết.....	11

CHƯƠNG 1. PHẦN MỞ ĐẦU

1.1. Giới thiệu chung

Trong thời đại 4.0, Công nghệ thông tin đã trở thành một trong những lĩnh vực mũi nhọn, với sự phát triển mạnh mẽ và nhanh chóng trong việc ứng dụng vào mọi mặt của đời sống. Công nghệ không chỉ là nền tảng cho sự phát triển xã hội văn minh mà còn góp phần quan trọng trong công cuộc công nghiệp hóa và hiện đại hóa đất nước. Máy tính và các phần mềm hỗ trợ đã giúp tối ưu hóa quy trình quản lý, tổ chức, sắp xếp và xử lý công việc một cách hiệu quả và nhanh chóng.

Theo thống kê hiện nay, hàng tỷ người trên toàn cầu sử dụng Internet, với Việt Nam đứng thứ 20 trên thế giới về tỷ lệ người truy cập mạng. Một trong những ứng dụng thiết yếu của Internet chính là theo dõi và dự báo thời tiết. Với hơn 70% dân số Việt Nam phụ thuộc vào các hoạt động ngoài trời như nông nghiệp, ngư nghiệp và du lịch, thông tin dự báo thời tiết chính xác và kịp thời trở nên vô cùng quan trọng. Nhận thấy tiềm năng lớn từ xu hướng này, nhóm chúng em đã cùng nhau phát triển một phần mềm phân tích và cung cấp thông tin để có thể đưa ra các dự báo về thời tiết nhằm hỗ trợ người dùng tiếp cận nhanh chóng với các dữ liệu khí tượng cập nhật theo thời gian thực.

Với sự thuận lợi từ môi trường công nghệ phát triển và nhu cầu lớn từ người dân trong việc tiếp cận thông tin thời tiết, chúng em tin tưởng rằng đây là một đề tài giàu triển vọng. Đồng thời, dự án này không chỉ giúp chúng em nâng cao kiến thức về Python, trí tuệ nhân tạo và các ngôn ngữ lập trình khác mà còn mang lại một sản phẩm hữu ích, giúp người dùng đưa ra các quyết định quan trọng trong công việc và đời sống hàng ngày.

1.2. Mục tiêu nghiên cứu

Xây dựng Phần Mềm Phân Tích Trực Quan Hóa Dữ Liệu Về Thời Tiết

Nghiên cứu, học các Module: Numpy, Pandas, Matplotlib, Seaborn

Vận dụng kiến thức ngôn ngữ Python

CHƯƠNG 2. TỔNG QUAN LÝ THUYẾT

2.1. Thời tiết

2.1.1. Định nghĩa

Thời tiết là trạng thái tức thời của khí quyển ở một địa điểm cụ thể, được đặc trưng bởi các đại lượng đo được, như nhiệt độ, độ ẩm, gió, lượng mưa,... hoặc các hiện tượng quan trắc được, như sương mù, đông, mưa, nắng,...

2.1.2. Các yếu tố thời tiết chính

Những yếu tố được sử dụng để biểu hiện thời tiết: nhiệt độ, độ ẩm, mưa và gió.

2.1.3 Các phương pháp dự báo thời tiết hiện tại

Phương pháp phân tích Synop dựa trên việc xác định các hình thế thời tiết và hoàn lưu khí quyển gây mưa và mưa lớn như bão, áp thấp nhiệt đới, rãnh áp thấp, dải hội tụ nhiệt đới, gió mùa đông bắc, gió mùa tây nam để dự báo định tính sự xuất hiện của mưa cũng như mưa lớn. Việc dự báo bằng phương pháp phân tích synop dựa hoàn toàn vào thống kê kinh nghiệm các hình thế đã gây mưa và mưa lớn trong quá khứ cho từng khu vực. Ngoài ra, hạn chế chính của phương pháp phân tích synop là dựa vào tính quán tính của khí quyển, dẫn tới khả năng mở rộng hạn dự báo là rất hạn chế, chỉ có độ chính xác cao ở thời đoạn từ 1-2 ngày.

Dự báo thời tiết bằng phương pháp thống kê là phương pháp áp dụng toán học thống kê trên chuỗi số liệu thời tiết trong nhiều năm, từ đó tìm ra các mối tương quan thống kê giữa các yếu tố khí tượng và hiện tượng thời tiết đã xuất hiện trong quá khứ để dự báo sự xuất hiện trong tương lai. Tuy nhiên, để đảm bảo tính chính xác cao thì phương pháp dự báo này đòi hỏi chuỗi số liệu quan trắc phải đủ dài và thông thường là 30 năm, tối thiểu phải từ 10 - 15 năm. Với hạn chế không mô tả được bản chất vật lý của các hiện tượng thời tiết và do chịu ảnh hưởng của biến đổi khí hậu toàn cầu nên các hiện tượng thời tiết có diễn biến khác với quy luật nhiều năm, do đó dự báo thời tiết bằng phương pháp thống kê cho kết quả dự báo không cao.

Radar thời tiết là một công cụ vô cùng quan trọng đối với dự báo tức thời -

Nowcasting (00h - 12h) các hiện tượng thời tiết nguy hiểm như: Đông, lốc, mưa đá, mưa lớn... Dựa vào vị trí, khoảng cách, hình dạng, cường độ phản hồi vô tuyến của sóng điện từ mà Radar thời tiết thu được khi gặp các giọt nước, giọt nước siêu lạnh, tinh thể đá trong mây, các Dự báo viên sẽ sớm đưa ra các cảnh báo, dự báo nhanh chóng, chính xác và kịp thời các hiện tượng thời tiết nguy hiểm sắp xảy ra khá chính xác.

2.2. Khoa học dữ liệu

2.2.1. Định nghĩa

Khoa học dữ liệu là khoa học về việc quản trị và phân tích dữ liệu để tìm ra các hiểu biết, các tri thức hành động, các quyết định dẫn dắt hành động.

2.2.2. Làm sạch dữ liệu

Làm sạch dữ liệu là quá trình xác định và giải quyết dữ liệu bị hỏng, không chính xác hoặc không liên quan. Giai đoạn xử lý dữ liệu quan trọng này — còn được gọi là dọn dẹp dữ liệu hoặc làm sạch dữ liệu — giúp tăng cường tính nhất quán, độ tin cậy và giá trị của dữ liệu.

2.2.3. Trực quan hóa dữ liệu

Trực quan hóa dữ liệu là hoạt động biểu diễn dữ liệu theo cách trực quan thông qua đồ họa, chẳng hạn như biểu đồ hoặc hoạt hình. Những hình ảnh trực quan này giúp dữ liệu phức tạp trở nên dễ hiểu hơn.

2.2.4. Cách thức áp dụng trong việc phân tích và dự báo thời tiết

Trong dự báo thời tiết, dữ liệu thường có khối lượng lớn và độ phức tạp cao do thu thập từ nhiều nguồn như cảm biến, vệ tinh, trạm khí tượng. Do đó, làm sạch dữ liệu là bước thiết yếu để đảm bảo tính chính xác và đáng tin cậy của các dự báo.

Trực quan hóa dữ liệu đóng vai trò quan trọng trong việc phân tích và trình bày kết quả dự báo thời tiết. Nó giúp nhà nghiên cứu và người dùng cuối dễ dàng hiểu được xu hướng và mô hình thời tiết từ dữ liệu thô.

CHƯƠNG 3. PHƯƠNG PHÁP NGHIÊN CỨU

3.1. Thu thập dữ liệu

3.1.1. Mô tả nguồn dữ liệu

Tập dữ liệu: lấy từ Kaggle

(<https://www.kaggle.com/code/melissamonfared/england-weather-analysis-eda/input>)

Tổng quan về tập dữ liệu: bộ dữ liệu về các thông tin thời tiết theo ngày của nước Anh bao gồm các chỉ số về độ ẩm, lượng mưa, ...

Mô tả dữ liệu:

Dữ liệu được lưu trong file EnglandWeather.csv

Dữ liệu được thu thập từ ngày 01/04/2006 đến 09/09/2016

Dữ liệu gồm 8 cột và 96453 dòng

Ý nghĩa của các cột:

Formatted Date: Hiển thị dấu thời gian (timestamp) cho mỗi mẫu trong tập dữ liệu.

Summary: Cung cấp mô tả ngắn gọn về tình trạng thời tiết.

Precip Type: Chỉ định xem có mưa hay tuyết vào thời điểm đó.

Temperature (C): Chỉ ra nhiệt độ tính bằng độ C.

Wind Speed (km/h): Đại diện cho tốc độ gió tính bằng km/h.

Pressure (millibars): Hiển thị áp suất không khí tính bằng millibars.

Humidity: Minh họa độ ẩm tương đối.

Date: Ngày với thông tin thời tiết tương ứng

3.1.2. Tiêu chí chọn mẫu nghiên cứu

Phạm vi không gian: mẫu dữ liệu cần bao quát nhiều khu vực địa lý khác nhau để đảm bảo tính đại diện và khả năng phân tích toàn diện. Ví dụ, dữ liệu thu thập không chỉ từ các thành phố lớn mà còn từ khu vực nông thôn, vùng ven biển, và các khu vực có địa hình đặc thù như vùng núi hoặc cao nguyên. Điều này giúp đánh giá được sự khác biệt về thời tiết giữa các vùng miền và cung cấp thông tin phù hợp cho các nhu cầu cụ thể như dự báo nông nghiệp, du lịch, hoặc quản lý thiên tai.

Phạm vi thời gian: tập dữ liệu nên bao gồm các bản ghi thời gian liên tục, tối

thiếu theo ngày, và kéo dài trong một khoảng thời gian đủ dài (vài năm hoặc vài thập kỷ) để phân tích xu hướng thời tiết dài hạn. Ngoài ra, dữ liệu thời gian thực hoặc gần thời gian thực cũng rất cần thiết để đáp ứng các nhu cầu dự báo và ứng phó nhanh với các hiện tượng thời tiết bất thường như bão, lũ, hay hạn hán.

Độ tin cậy của nguồn dữ liệu: nguồn dữ liệu phải được lấy từ các tổ chức uy tín và đã qua kiểm chứng, như các cơ quan khí tượng quốc gia, các tổ chức nghiên cứu khoa học quốc tế, hoặc các nền tảng nổi tiếng như Kaggle. Độ chính xác và chất lượng dữ liệu được ưu tiên để đảm bảo các phân tích và dự báo dựa trên dữ liệu đó đáng tin cậy. Đồng thời, việc sử dụng dữ liệu từ nhiều nguồn khác nhau cũng giúp so sánh và kiểm tra tính nhất quán của thông tin.

Tính ứng dụng thực tiễn: mẫu dữ liệu cần chứa đựng các yếu tố thời tiết thường được người dân quan tâm, chẳng hạn như nhiệt độ, lượng mưa, độ ẩm, gió, và áp suất khí quyển. Thông tin về thời tiết cực đoan (như bão, sương giá, hạn hán) cũng rất quan trọng vì ảnh hưởng trực tiếp đến các quyết định hàng ngày và các lĩnh vực kinh tế như nông nghiệp, vận tải, và xây dựng. Tính ứng dụng của dữ liệu không chỉ nằm ở việc dự báo mà còn hỗ trợ lập kế hoạch dài hạn và quản lý rủi ro.

Mức độ liên quan đến hoạt động con người: thời tiết ảnh hưởng mạnh mẽ đến nhiều khía cạnh của đời sống và kinh tế. Do đó, dữ liệu được chọn nên hỗ trợ cho các lĩnh vực cụ thể, chẳng hạn như dự báo mùa vụ trong nông nghiệp, kế hoạch du lịch theo mùa, quản lý giao thông trong điều kiện thời tiết xấu, hoặc dự báo năng lượng dựa trên khả năng phát điện từ năng lượng gió và mặt trời. Việc đảm bảo dữ liệu phản ánh mối quan hệ giữa thời tiết và hoạt động con người giúp các nghiên cứu và ứng dụng có giá trị thực tiễn cao hơn.

3.2. Xử lý dữ liệu

Chúng tôi đã tách riêng thông tin ngày từ cột Formatted Date sang một cột mới với tên Date, nhằm tối ưu hóa việc quản lý dữ liệu. Việc này không chỉ giúp dữ liệu trở nên dễ dàng theo dõi mà còn hỗ trợ tìm kiếm thông tin theo ngày một cách nhanh chóng và hiệu quả hơn đồng thời tạo điều kiện cho việc thực hiện các thao tác phân tích chuyên sâu như so sánh xu hướng theo ngày, nhóm dữ liệu theo khoảng thời gian, lập báo cáo chi tiết. Đây là một bước quan trọng trong việc chuẩn hóa và nâng cao chất lượng dữ liệu phục vụ mục đích nghiên cứu dữ liệu.

3.3. Phát triển giao diện quá lý và trực quan hóa dữ liệu

3.3.1 Lựa chọn công cụ và ngôn ngữ lập trình

Để xây dựng giao diện quản lý và trực quan hóa dữ liệu thời tiết, chúng tôi sử dụng Python làm ngôn ngữ lập trình chính. Python không chỉ phổ biến trong cộng đồng khoa học dữ liệu mà còn cung cấp một hệ sinh thái phong phú các thư viện hỗ trợ việc phát triển giao diện và trực quan hóa dữ liệu.

Thư viện hỗ trợ giao diện

Tkinter: Tạo giao diện thời tiết đơn giản và trực quan.

Thư viện xử lý dữ liệu và đồ thị

Pandas: Đọc, chỉnh sửa và quản lý file CSV thời tiết.

Matplotlib và Seaborn: Trực quan hóa dữ liệu dưới dạng biểu đồ tĩnh, khả năng tùy chỉnh linh hoạt, phù hợp với yêu cầu hiển thị dữ liệu thời tiết.

3.3.2 Chức năng của giao diện

Đọc dữ liệu

Tải file CSV thời tiết (ví dụ: nhiệt độ, lượng mưa, độ ẩm, tốc độ gió).

Hiển thị dữ liệu trong bảng để người dùng dễ dàng theo dõi.

Thao tác với dữ liệu

Thêm: Người dùng có thể thêm các bản ghi mới vào file CSV.

Xóa: Cung cấp tính năng xóa các bản ghi không cần thiết.

Sửa: Cho phép chỉnh sửa thông tin sai lệch hoặc cần cập nhật.

Lưu dữ liệu:

Lưu các thay đổi vào file CSV gốc hoặc xuất ra file mới.

Tìm kiếm dữ liệu và in thông tin của dòng dữ liệu:

Tạo ô tìm kiếm giúp truy cập nhanh đến dữ liệu cần tìm bằng cột dữ liệu

Date tương ứng, nhấp đúp chuột vào dòng bất kỳ để xem mọi thông tin của dòng đó giúp dễ theo dõi số liệu.

Sắp xếp dữ liệu:

Nhấp vào cột tương ứng của dữ liệu để tự động sắp xếp tăng dần hoặc giảm dần giá trị của cột dữ liệu đã chọn.

3.3.3 Hiển thị và trực quan hóa dữ liệu

Loại biểu đồ giao diện hỗ trợ:

Biểu đồ cột (Bar chart): Cho phép so sánh các yếu tố thời tiết giữa các khoảng thời gian như tháng hoặc năm.

Biểu đồ đếm (Countplot): Thống kê số lượng bản ghi theo từng giá trị cụ thể, cung cấp góc nhìn tổng quan về tần suất của các điều kiện thời tiết trong tập dữ liệu.

Biểu đồ nhiệt (Heatmap): Minh họa mối tương quan giữa các yếu tố thời tiết, giúp người dùng dễ dàng nhận ra các mối liên hệ hoặc mẫu hình nổi bật trong dữ liệu.

Tương tác:

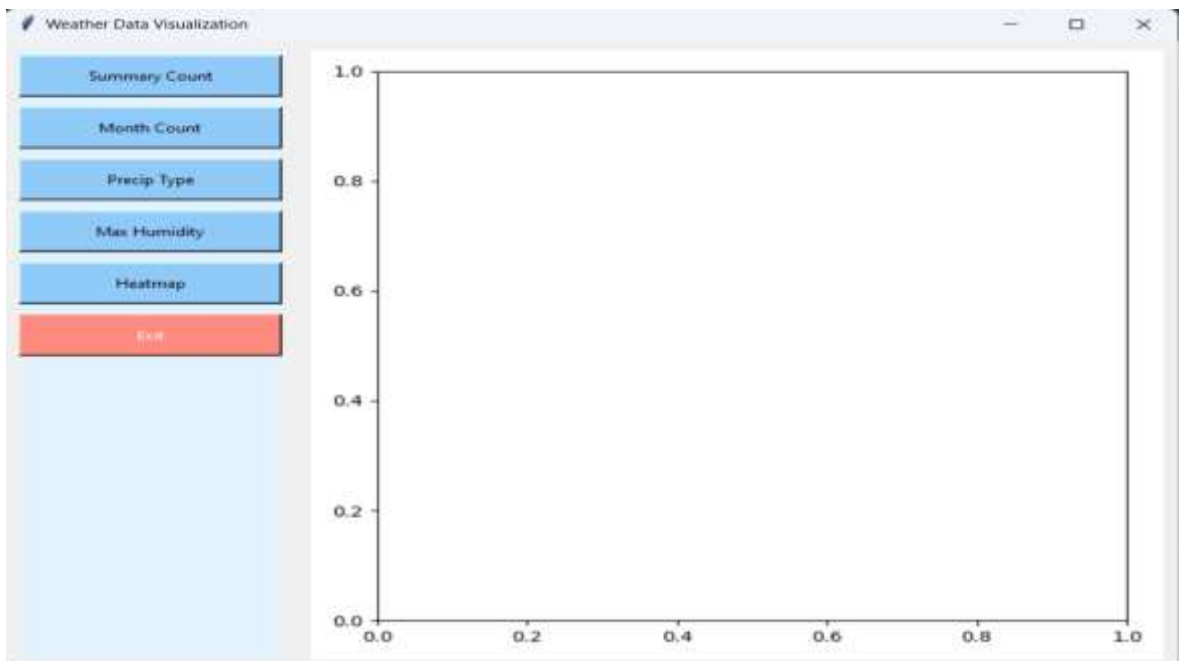
Lọc và nhóm dữ liệu trực tiếp trên giao diện trước khi vẽ đồ thị.

CHƯƠNG 4. KẾT QUẢ VÀ THẢO LUẬN

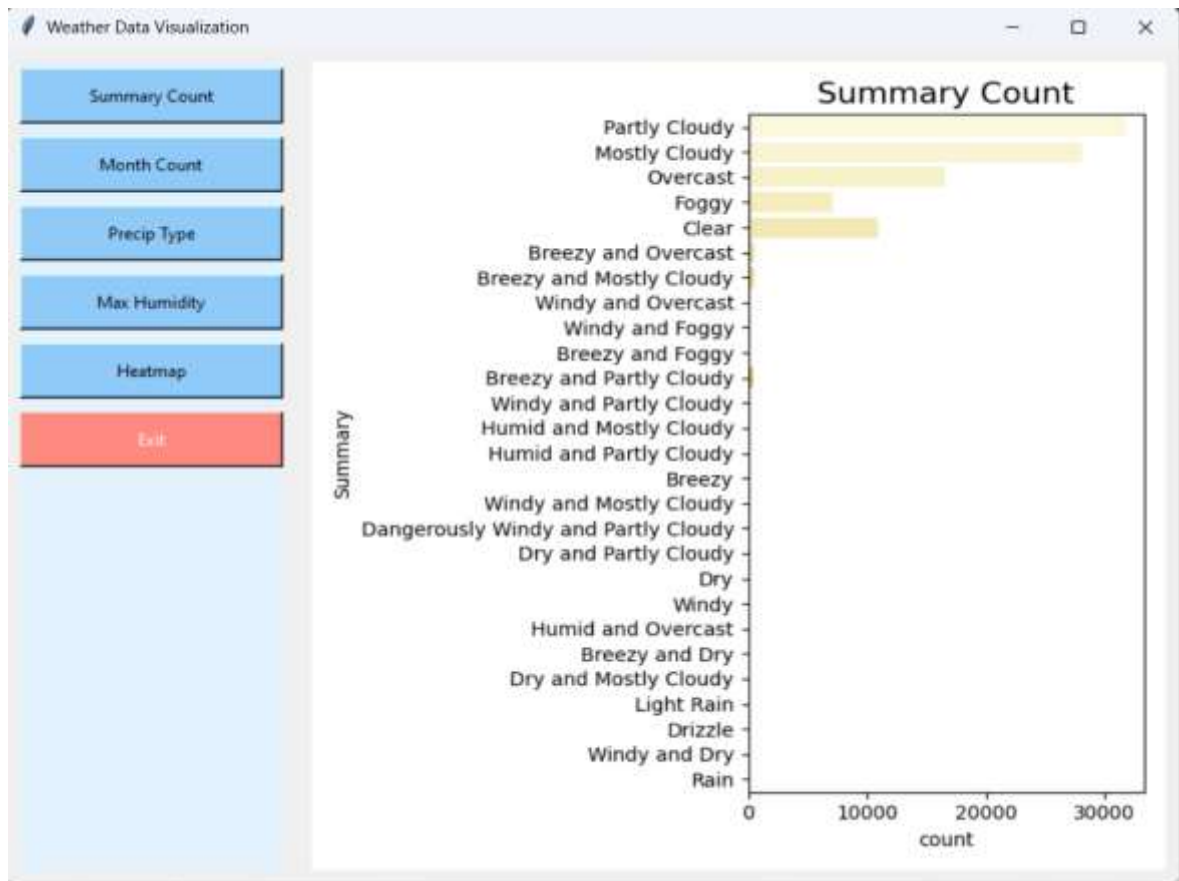
4.1. Trình bày kết quả



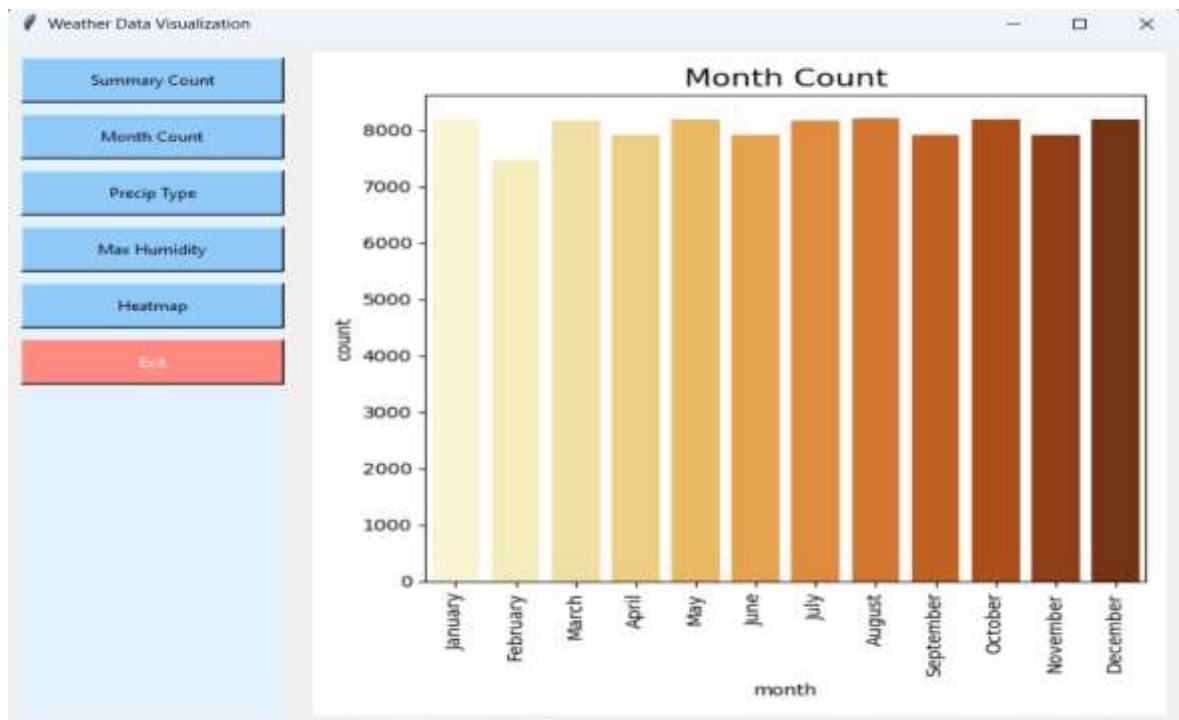
Hình 4.1. Giao diện màn hình chờ



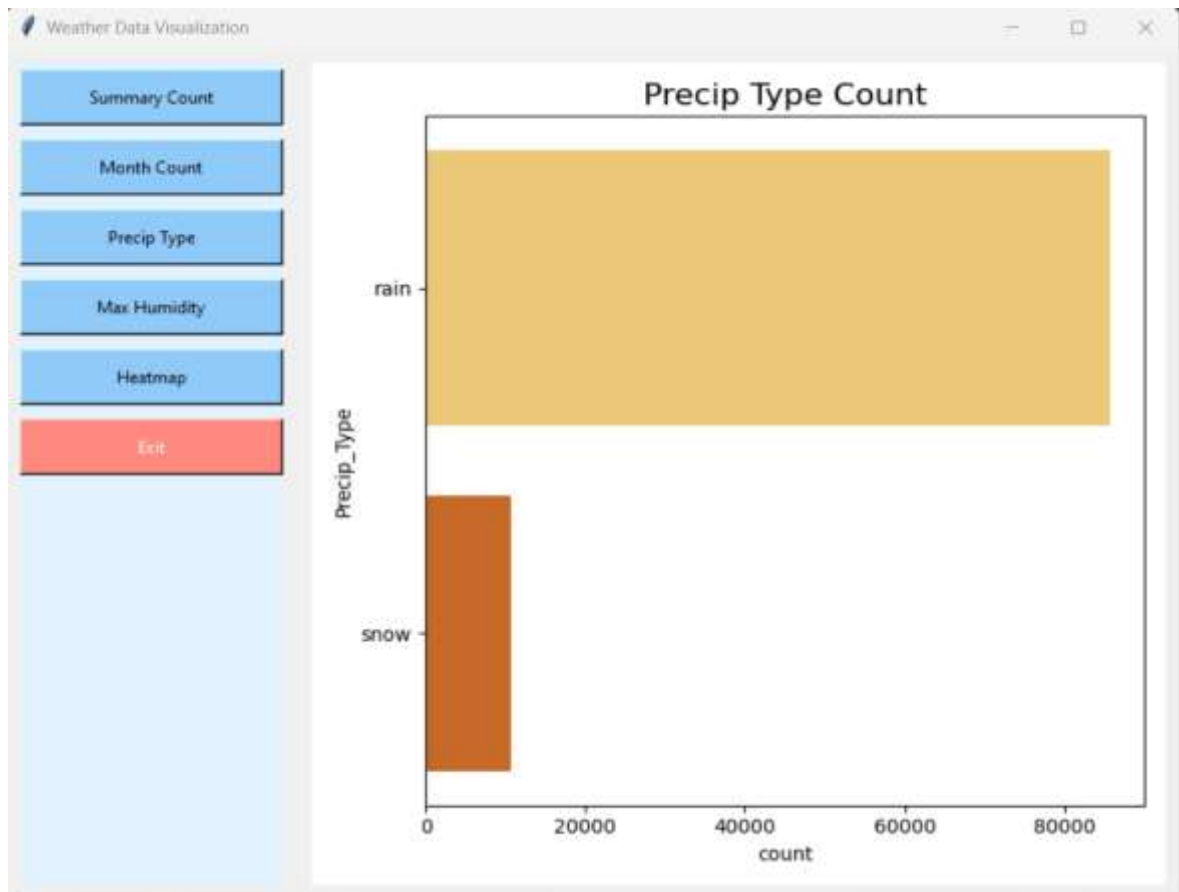
Hình 4.2. Giao diện ban đầu của các dạng biểu đồ



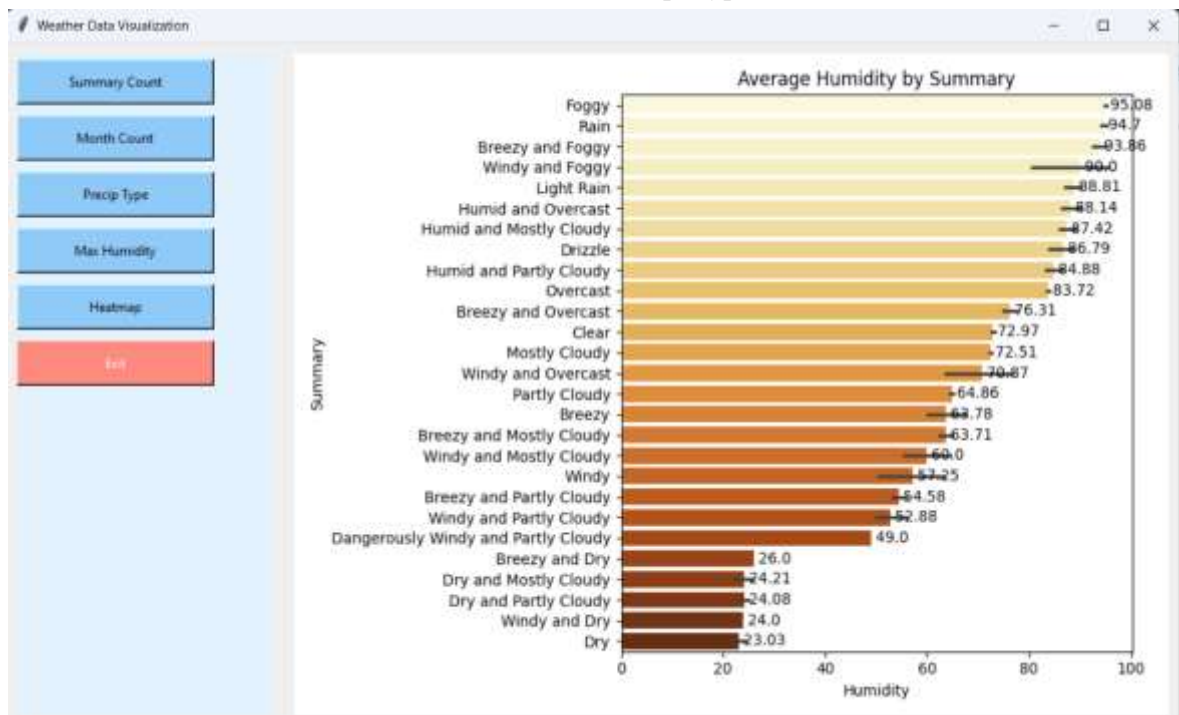
Hình 4.3. Biểu đồ Summary Count



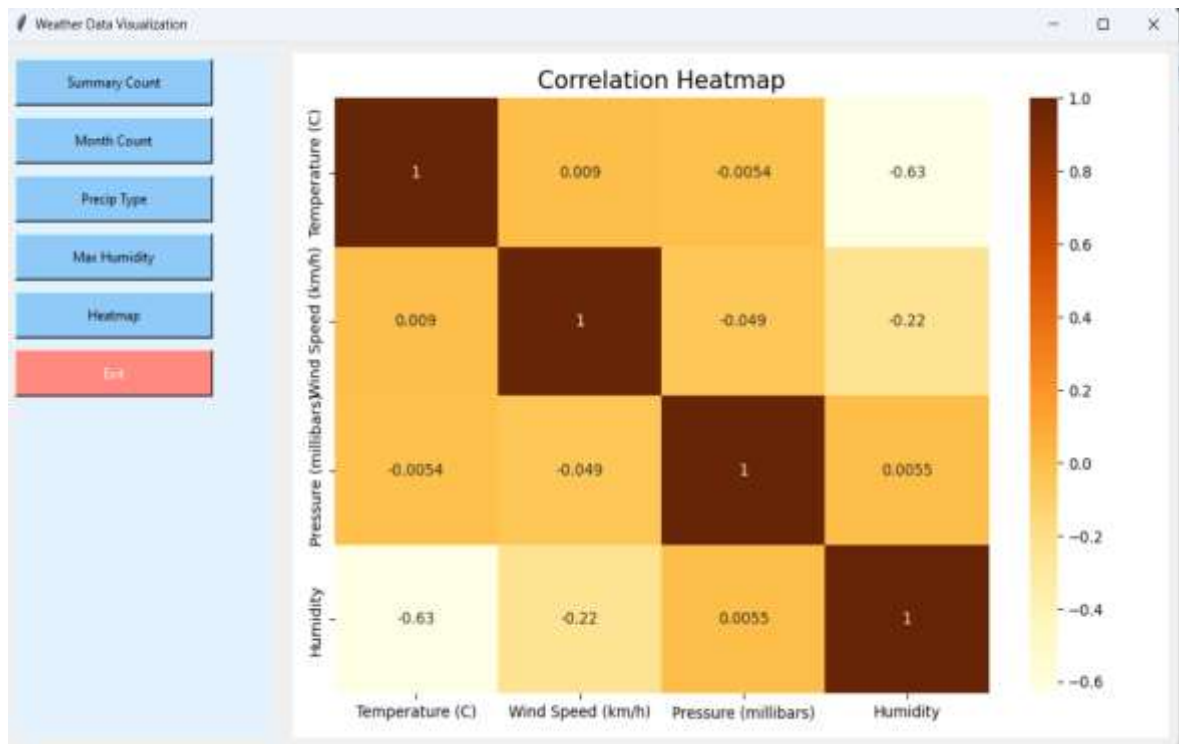
Hình 4.4. Biểu đồ Month Count



Hình 4.5. Biểu đồ Precip Type Count



Hình 4.6. Biểu đồ Average Humidity by Summary



Hình 4.7. Biểu đồ Correlation Heatmap

Weather Data Management

Index	Formatted Dat	Summary	Precip_Type	Temperature	Wind Speed (t)	Pressure (milli)	Humidity	Date
0	2006-03-31 22:00	Partly Cloudy	rain	9.47222222	14.1197	1015.13	0.89	2006-03-31
1	2006-03-31 23:00	Partly Cloudy	rain	9.35555556	14.2646	1015.63	0.86	2006-03-31
2	2006-04-01 00:00	Mostly Cloudy	rain	9.37777778	3.9284	1015.94	0.89	2006-04-01
3	2006-04-01 01:00	Partly Cloudy	rain	8.28888889	14.1036	1016.41	0.83	2006-04-01
4	2006-04-01 02:00	Mostly Cloudy	rain	8.75555556	11.0446	1016.51	0.83	2006-04-01
5	2006-04-01 03:00	Partly Cloudy	rain	9.22222222	13.9587	1016.66	0.85	2006-04-01
6	2006-04-01 04:00	Partly Cloudy	rain	7.73333333	12.3648	1016.72	0.95	2006-04-01
7	2006-04-01 05:00	Partly Cloudy	rain	8.77222222	14.1519	1016.84	0.89	2006-04-01
8	2006-04-01 06:00	Partly Cloudy	rain	10.82222222	11.3183	1017.37	0.82	2006-04-01
9	2006-04-01 07:00	Partly Cloudy	rain	13.77222222	12.5258	1017.22	0.72	2006-04-01
10	2006-04-01 08:00	Partly Cloudy	rain	16.01666667	17.5651	1017.42	0.67	2006-04-01
11	2006-04-01 09:00	Partly Cloudy	rain	17.14444444	19.7869	1017.74	0.54	2006-04-01
12	2006-04-01 10:00	Partly Cloudy	rain	17.8	21.9443	1017.59	0.55	2006-04-01
13	2006-04-01 11:00	Partly Cloudy	rain	17.33333333	20.6885	1017.48	0.51	2006-04-01
14	2006-04-01 12:00	Partly Cloudy	rain	18.87777778	15.3755	1017.17	0.47	2006-04-01

Find date: Find Show all

Create Update Delete Info Exit

Tiếng mưa rơi ở nước Anh cực chill Cảnh vật nước Anh dưới tuyết tuyệt đẹp

Hình 4.8. Giao diện tương tác các số liệu thống kê thời tiết

Biểu đồ Summary Count hiển thị tần suất xuất hiện của các tình trạng thời tiết khác nhau trong dữ liệu, cho phép chúng ta nhận ra các mẫu và xu hướng liên quan đến các trạng thái thời tiết như overcast, mostly cloudy, partly cloudy, v.v. Thông qua

đó xác định được các trạng thái thời tiết phổ biến có thể cho thấy sự thay đổi khí hậu hay biến động về môi trường.

Biểu đồ Month Count hiển thị tần suất xuất hiện của các tháng trong dữ liệu, cho phép chúng ta nhận ra một số xu hướng và mẫu quan trọng. Số lượng dữ liệu được ghi nhận trong từng tháng, ta có thể thấy số liệu khá đồng đều dù có một số tháng có nhiều bản ghi hơn các tháng khác. Biểu đồ này còn cho thấy sự tin cậy của dữ liệu và phản ánh tình trạng thời tiết của nước Anh theo tháng.

Biểu đồ Precip Type Count cho thấy tần suất xuất hiện của các loại giáng thủy khác nhau trong dữ liệu, cho phép chúng ta nhận ra các mẫu và xu hướng liên quan đến các loại giáng thủy như mưa, tuyết, sương mù, v.v. Như ta thấy biểu chỉ thu thập dữ mưa và tuyết rơi từ đó rút ra mưa và tuyết là chủ yếu, trong đó mưa xuất hiện nhiều nhất. Dựa trên những phát hiện từ biểu đồ Precip Type Count, chúng ta có thể đưa ra các dự đoán và kế hoạch ứng phó với các hiện tượng thời tiết. Và mưa là hiện tượng phổ biến nhất, các biện pháp phòng chống ngập lụt và quản lý nước mưa sẽ cần được ưu tiên.

Biểu đồ Average Humidity by Summary hiển thị độ ẩm trung bình cao nhất theo các trạng thái thời tiết khác nhau. Hiểu rõ mối liên hệ giữa tình trạng thời tiết và độ ẩm có thể giúp cải thiện độ chính xác của dự báo thời tiết. Như khi độ ẩm tăng cao, khả năng cao sẽ xảy ra mưa và có nhiều sương mù v.v.

Biểu đồ Correlation Heatmap là biểu đồ heatmap tạo ma trận tương quan để so sánh tương quan giữa các cột trong tập dữ liệu. Nhiệt độ và độ ẩm có một mối quan hệ tương quan tiêu cực nghĩa là khi nhiệt độ tăng, độ ẩm có xu hướng giảm, và ngược lại, phản ánh rằng không khí ẩm có khả năng giữ nhiều hơi nước hơn, do đó khi nhiệt độ tăng, độ ẩm tương đối có thể giảm. Tốc độ gió và áp suất có một mối quan hệ tương quan tiêu cực, phản ánh hiện tượng thời tiết bão, nơi áp suất thấp thường đi kèm với gió mạnh. Nhiệt độ và áp suất có một mối quan hệ tương quan yếu, thể hiện rằng chúng không ảnh hưởng nhiều đến nhau trong dữ liệu này. Tuy nhiên, sự thay đổi của nhiệt độ vẫn có thể ảnh hưởng đến áp suất không khí ở một mức độ nào đó.

Các biểu đồ và phân tích dữ liệu thời tiết đã cung cấp một bức tranh tổng thể về các yếu tố khí hậu quan trọng và mối quan hệ giữa chúng. Các phát hiện từ các

biểu đồ này rất hữu ích trong việc dự báo thời tiết, quản lý môi trường, và nghiên cứu khoa học.

4.2. Phân tích kết quả

4.2.1. Đánh giá mức độ ứng dụng của giao diện

Giao diện được xây dựng nhằm cung cấp các chức năng chính như xem biểu đồ thời tiết, thao tác với bảng dữ liệu, và hiển thị bảng số liệu từ file CSV về thời tiết. Quá trình kiểm tra và đánh giá giao diện đã tập trung vào các yếu tố sau:

Về tính năng, các thao tác như thêm, sửa, xóa và lưu dữ liệu trong bảng được kiểm tra trên nhiều tập dữ liệu khác nhau để đảm bảo tính chính xác và ổn định. Biểu đồ thời tiết hiển thị các yếu tố như nhiệt độ, độ ẩm, và lượng mưa lấy dữ liệu thực từ file data để luôn cập nhật chính xác theo các thay đổi trong dữ liệu đầu vào.

Về giao diện và trải nghiệm người dùng (UI/UX), nhóm tạo giao diện khá thân thiện với người dùng, hỗ trợ tốt việc thao tác trực tiếp trên bảng dữ liệu và dễ dàng quan sát biểu đồ. Trong quá trình thực hiện nhóm đã phát hiện một số điểm cần cải thiện chẳng hạn như sắp xếp các nút chức năng hợp lý hoặc bổ sung thông tin của các dòng dữ liệu để người dùng dễ nắm thông tin hơn.

Về hiệu suất, khi làm việc với các file CSV có kích thước và số lượng dữ liệu lớn giao diện đã được tối ưu hóa để đảm bảo quá trình tải và xử lý dữ liệu diễn ra mượt mà và nhanh chóng. Bên cạnh đó, các thuật toán vẽ biểu đồ và xử lý dữ liệu cũng được nhóm điều chỉnh để giảm thiểu độ trễ và nâng cao hiệu quả hoạt động. Với những cải tiến trên đã giúp giao diện người dùng dễ dàng thao tác và quan sát dữ liệu thời tiết, cung cấp công cụ trực quan hóa mạnh mẽ cho việc phân tích xu hướng thời tiết một cách nhanh chóng và chính xác dựa trên dữ liệu thực.

4.2.2. Thảo luận về các yếu tố ảnh hưởng đến kết quả

Có nhiều yếu tố ảnh hưởng đến kết quả của giao diện quản lý và phân tích dữ liệu thời tiết được nhóm rút ra trong quá trình thực hiện:

Thứ nhất, chất lượng dữ liệu đầu vào đóng vai trò quan trọng vì các file CSV không đầy đủ hoặc chứa dữ liệu lỗi có thể gây sai lệch biểu đồ và số liệu. Công đoạn xử lý và làm sạch dữ liệu như loại bỏ giá trị ngoại lai và định dạng lại các trường dữ liệu không hợp lệ là bước cần thiết để đảm bảo kết quả chính xác.

Thứ hai, thuật toán xử lý và vẽ biểu đồ cũng tác động đáng kể đến hiệu suất của chương trình. Các thuật toán hiện tại mặc dù đã được tối ưu hóa nhưng vẫn có thể gặp khó khăn khi xử lý các file dữ liệu lớn khiến thời gian tải chậm.

Cuối cùng, các yếu tố môi trường như cấu hình thiết bị và trình duyệt cũng ảnh hưởng lớn đến trải nghiệm người dùng. Trên các thiết bị có cấu hình thấp, tốc độ xử lý và hiển thị biểu đồ có thể bị giảm. Điều này đòi hỏi giao diện cần được thiết kế linh hoạt và tối ưu hóa để hoạt động tốt trên nhiều nền tảng khác nhau.

4.2.3. Đề xuất các cải tiến cho phần mềm

Đầu tiên, việc nâng cấp giao diện người dùng (UI/UX) là yếu tố quan trọng để cải thiện trải nghiệm. Giao diện hiện tại chủ yếu dựa trên các hộp thoại nhập liệu đơn giản nên được thay thế bằng một thiết kế hiện đại và trực quan hơn. Sử dụng các công cụ như Tkinter hoặc PyQt để xây dựng các giao diện chuyên nghiệp với menu điều hướng rõ ràng, các ô nhập liệu được tổ chức logic và các tùy chọn lựa chọn nhanh như danh sách thả xuống (dropdown). Một bảng hiển thị dữ liệu tương tác, cho phép chỉnh sửa trực tiếp và lọc dữ liệu theo ý muốn.

Thứ hai, phần mềm cần hỗ trợ nhập nhiều tệp CSV cùng lúc thay vì chỉ làm việc với một tệp duy nhất. Từ đó có thể tích hợp nhiều nguồn dữ liệu khác nhau vào cùng một hệ thống mà không cần ghép thủ công. Chức năng này có thể được triển khai bằng cách cho phép người dùng chọn nhiều tệp từ hệ thống hoặc thậm chí kéo-thả (drag-and-drop) tệp vào giao diện. Dữ liệu từ các tệp này sẽ được hợp nhất và xử lý một cách tự động, đảm bảo không bị trùng lặp hoặc thiếu thông tin.

Thứ ba, phần mềm cần có khả năng vẽ nhiều biểu đồ chi tiết hơn, giúp người dùng hiểu sâu hơn về dữ liệu. Thay vì chỉ hiển thị các biểu đồ cơ bản, hệ thống có thể cung cấp các loại biểu đồ nâng cao như biểu đồ cột nhóm, biểu đồ hình tròn, hoặc biểu đồ phân tán để phân tích mối quan hệ giữa các yếu tố như nhiệt độ và tốc độ gió. Sử dụng matplotlib, seaborn, hoặc plotly tạo ra các biểu đồ tương tác, giúp người dùng dễ dàng thay đổi thông số và theo dõi trực quan các xu hướng.

Thứ tư, một tính năng quan trọng khác cần được bổ sung là mô phỏng kiểu thời tiết dựa trên dữ liệu hiện có. Chức năng sẽ dự đoán các giá trị như "Summary" hoặc "Precip_Type" dựa vào các thông số như nhiệt độ, độ ẩm, áp suất, và tốc độ gió. Để thực hiện điều này, phần mềm có thể áp dụng các thuật toán Machine Learning

như hồi quy logistic (Logistic Regression) hoặc cây quyết định (Decision Tree). Sử dụng thư viện scikit-learn hoặc TensorFlow huấn luyện mô hình dự đoán dựa trên tập dữ liệu hiện tại. Kết quả mô phỏng có thể được thêm trực tiếp vào bảng dữ liệu, hỗ trợ phân tích xu hướng thời tiết mà không cần dữ liệu đầu vào bổ sung.

Cuối cùng, dữ liệu hiện tại chỉ được lưu dưới dạng tệp CSV, điều này gây hạn chế khi người dùng muốn sử dụng dữ liệu cho mục đích báo cáo hoặc trình bày chuyên nghiệp. Do đó, phần mềm nên bổ sung khả năng xuất dữ liệu sang các định dạng khác như Excel hoặc PDF. Thư viện pandas hỗ trợ xuất dữ liệu sang Excel thông qua hàm `to_excel`, trong khi các thư viện như `fpdf` hoặc `ReportLab` sẽ được sử dụng để tạo báo cáo PDF với nội dung và biểu đồ trực quan từ dữ liệu từ đó giúp người dùng dễ dàng tạo các báo cáo chất lượng cao phục vụ công việc.

CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TƯƠNG LAI

5.1. Tổng kết

5.1.1. Tóm tắt các phát hiện chính

Giao diện quản lý và phân tích dữ liệu thời tiết đã đạt được những kết quả tích cực trong việc xử lý và hiển thị thông tin thời tiết từ file CSV. Các chức năng chính như thêm, sửa, xóa và lưu dữ liệu hoạt động chính xác, đồng thời biểu đồ thời tiết được cập nhật đúng với dữ liệu đầu vào. Mặc dù có một số hạn chế về hiệu suất khi xử lý các file dữ liệu lớn, giao diện vẫn hoạt động mượt mà và cung cấp trải nghiệm người dùng tốt, đặc biệt trong việc trực quan hóa các xu hướng thời tiết. Các công cụ trực quan hóa và thao tác với bảng số liệu giúp người dùng dễ dàng phân tích và hiểu rõ hơn về dữ liệu.

5.1.2. Đánh giá về mức độ thành công của bài báo cáo

Bài báo cáo đã thành công trong việc phát triển một giao diện hiệu quả để người dùng có thể thao tác với dữ liệu thời tiết và quan sát các biểu đồ một cách dễ dàng. Tuy nhiên, mức độ thành công này vẫn cần được đánh giá thêm qua các thử nghiệm thực tế với dữ liệu lớn và đa dạng hơn, đồng thời cần cải thiện một số yếu tố liên quan đến tối ưu hóa hiệu suất và khả năng tương thích trên nhiều nền tảng.

5.2. Hướng phát triển tương lai

5.2.1. Các khía cạnh còn hạn chế và cách khắc phục

Mặc dù giao diện đã hoàn thành tốt các chức năng cơ bản, nhưng vẫn còn một số hạn chế về khả năng xử lý các file CSV lớn và khả năng tương thích với các thiết bị cấu hình thấp. Để khắc phục, cần tối ưu hóa thuật toán xử lý dữ liệu, nâng cao hiệu suất và giảm thiểu thời gian tải. Ngoài ra, cần tiếp tục cải thiện khả năng tương thích của giao diện trên các trình duyệt và thiết bị khác nhau.

5.2.2. Đề xuất các nghiên cứu tiếp theo

Nghiên cứu tiếp theo nên tập trung vào việc nâng cao khả năng xử lý và hiển thị các file dữ liệu lớn hơn, tối ưu hóa giao diện người dùng để hỗ trợ tốt hơn cho các

phân tích thời gian thực. Cũng có thể nghiên cứu việc tích hợp thêm các tính năng như dự báo thời tiết tự động hoặc phân tích xu hướng dữ liệu sâu hơn thông qua các công cụ học máy, giúp người dùng có thêm thông tin hữu ích trong việc phân tích thời tiết.

5.3. Tác động của nghiên cứu

5.3.1. Tác động đến lĩnh vực phân tích dữ liệu thời tiết

Giao diện này mang lại một công cụ hữu ích trong việc phân tích và trực quan hóa dữ liệu thời tiết, giúp người dùng dễ dàng theo dõi và nhận biết các xu hướng thời tiết. Việc tích hợp các tính năng phân tích và dự báo cũng giúp người dùng, đặc biệt là các chuyên gia và nhà nghiên cứu, có được cái nhìn sâu sắc hơn về các biến động thời tiết.

5.3.2. Khả năng ứng dụng công nghệ trong phân tích và trực quan hóa dữ liệu thời tiết

Công nghệ này có tiềm năng ứng dụng rộng rãi trong các ngành công nghiệp liên quan đến dự báo và quản lý thời tiết, như nông nghiệp, giao thông, và quản lý thiên tai. Việc sử dụng các công cụ phân tích thời tiết tự động có thể giúp các tổ chức và cơ quan quản lý đưa ra quyết định chính xác và kịp thời hơn, từ đó giảm thiểu tác động của các hiện tượng thời tiết cực đoan.

Tài liệu tham khảo

1. HCMUTE. (2024, November 17) . Giáo trình lập trình Python cơ bản.
<https://thuvienso.hcmute.edu.vn/doc/giao-trinh-lap-trinh-python-can-ban-831878.html>
2. UTEXLMS. (2024, November 17) . Slide bài giảng lập trình Python (GV: Trần Quang Khải)
https://hcmuteedu.vn0-my.sharepoint.com/personal/khaitq_hcmute_edu_vn/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fkhaitq%5Fhcmute%5Fedu%5Fvn%2FDocuments%2F10%2E%20UTE%2FM%C3%B4n%20h%E1%BB%8Dc%2F04%2E%20L%E1%BA%ADp%20tr%C3%ACnh%20Python%20%28IPPA233277%20%2D%20INPY131685%29%2FGi%E1%BA%A3ng%20d%E1%BA%A1y%2FGuiSV&ga=1