**BỘ GIÁO DỤC VÀ ĐÀO TẠO**



**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐỒ ÁN CUỐI KỲ**

**MÔN HỌC: LẬP TRÌNH PYTHON**

**PHÂN TÍCH DỮ LIỆU HẠNH PHÚC TOÀN CẦU**

Mã lớp học phần: 242IPPA233277\_06CLC

Học kỳ II – Năm học: 2024 - 2025

Giảng viên hướng dẫn: ThS. Trần Quang Khải

Danh sách sinh viên thực hiện:

**MSSV**

22110149

22146367

22110247

**Họ tên**

Phạm Gia Huy

Mai Công Phát

Nguyễn Thanh Tính

22110119 Nguyễn Khánh Duy

*Thành phố Hồ Chí Minh, tháng 05 năm 2025*

***Nhận xét của giảng viên***

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*...........................................................................................................................................*

*TP. Hồ Chí Minh, ngày … tháng… năm 2025*

*Giảng viên ký tên*

**MỤC LỤC**

[PHẦN 1: MỞ ĐẦU 6](#_Toc197749619)

[1. Lý do chọn đề tài 6](#_Toc197749620)

[2. Mục tiêu đề tài 6](#_Toc197749621)

[3. Đối tượng và phạm vi nghiên cứu 7](#_Toc197749622)

[PHẦN 2. NỘI DUNG 8](#_Toc197749623)

[CHƯƠNG 1: TÓM TẮT CƠ SỞ LÝ THUYẾT 8](#_Toc197749624)

[1.1. Giới thiệu về ngôn ngữ Python 8](#_Toc197749625)

[1.2. Tại sao Python thường được dùng để phân tích dữ liệu 8](#_Toc197749626)

[1.3. Giới thiệu về các thư viện thường được sử dụng trong Phân tích dữ liệu 9](#_Toc197749627)

[1.3.1. NumPy (Numerical Python) 9](#_Toc197749628)

[1.3.2. Pandas 9](#_Toc197749629)

[1.3.3. Matplotlib 10](#_Toc197749630)

[1.3.4. Seaborn 10](#_Toc197749631)

[1.3.5. Scikit-learn 11](#_Toc197749632)

[CHƯƠNG 2: MÔ TẢ TẬP DỮ LIỆU HẠNH PHÚC TOÀN CẦU 12](#_Toc197749633)

[2.1. Quá trình thu thập dữ liệu 12](#_Toc197749634)

[2.2. Tổng quan về dữ liệu 12](#_Toc197749635)

[2.3. Các hạn chế của dữ liệu 14](#_Toc197749636)

[2.3.1 Thiếu dữ liệu 14](#_Toc197749637)

[CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU HẠNH PHÚC TOÀN CẦU 15](#_Toc197749638)

[3.1. Xử lý dữ liệu 15](#_Toc197749639)

[3.1.1. Kiểm tra dữ liệu ban đầu 15](#_Toc197749640)

[3.1.2. Xử lý các giá trị bị thiếu 15](#_Toc197749641)

[3.1.3. Loại bỏ các giá trị trùng lặp 15](#_Toc197749642)

[3.1.4. Kiểm tra và xác thực dữ liệu hợp lệ 15](#_Toc197749643)

[3.1.5. Lưu dữ liệu đã làm sạch 15](#_Toc197749644)

[3.2. Trực quan hóa dữ liệu 16](#_Toc197749645)

[3.3. Thiết kế giao diện 21](#_Toc197749646)

[3.4. Các chức năng chính của chương trình 22](#_Toc197749647)

[3.4.1. Chức năng tìm kiếm 22](#_Toc197749648)

[3.4.2. Chức năng thêm record 23](#_Toc197749649)

[3.4.3. Chức năng sửa record 24](#_Toc197749650)

[3.4.4. Chức năng xóa dữ liệu 25](#_Toc197749651)

[PHẦN KẾT LUẬN 26](#_Toc197749652)

[1. Kết quả đạt được 26](#_Toc197749653)

[2. Hướng phát triển 27](#_Toc197749654)

[PHẦN TÀI LIỆU THAM KHẢO 28](#_Toc197749655)

**DANH MỤC HÌNH ẢNH**

[Hình 1. Biểu đồ đường : Xu hướng Thanh Hạnh Phúc của một số quốc gia 18](#_Toc197753626)

[Hình 2. Biểu đồ cột : Thang Hạnh Phúc trung bình theo năm 19](#_Toc197753627)

[Hình 3. Biểu đồ đường : Thang Hạnh Phúc và Log GDP bình quân đầu người của Việt Nam qua các năm 20](#_Toc197753628)

[Hình 4. Biểu đồ tròn Tỷ lệ Thang Hạnh Phúc của một số quốc gia 22](#_Toc197753629)

[Hình 5. Biểu đồ phân tán : So sánh Thang Hạnh Phúc của các nước năm 2019 và 2020 24](#_Toc197753630)

[Hình 6. Giao diện người dùng 27](#_Toc197753631)

[Hình 7. Chức năng tìm kiếm 28](#_Toc197753632)

[Hình 8. Chức năng thêm record 29](#_Toc197753633)

[Hình 9. Chức năng sửa record 30](#_Toc197753634)

[Hình 10. Chức năng xóa dữ liệu 31](#_Toc197753635)

**KẾ HOẠCH PHÂN CÔNG NHIỆM VỤ THỰC HIỆN ĐỀ TÀI**

**CUỐI KỲ MÔN LẬP TRÌNH PYTHON**

**HỌC KỲ I NĂM HỌC 2024-2025**

**1. Mã lớp môn học:** 242IPPA233277\_06CLC

**2. Giảng viên hướng dẫn:** ThS. Trần Quang Khải

**3. Tên đề tài: PHÂN TÍCH DỮ LIỆU HẠNH PHÚC TOÀN CẦU**

**4. Bảng phân công nhiệm vụ:**

|  |  |
| --- | --- |
| **Sinh viên thực hiện** | **Nội dung thực hiện** |
| Phạm Gia Huy | Giao diện, các chức năng CRUD |
| Mai Công Phát | Vẽ biểu đồ |
| Nguyễn Thanh Tính | Chỉnh sửa UI/UX, Tìm kiếm, phân tích, lọc dữ liệu, |
| Nguyễn Khánh Duy | Vẽ biểu đồ, Báo cáo, Slides thuyết trình |

**LỜI CẢM ƠN**

Để thực hiện được đề tài này, em xin gửi lời cảm ơn chân thành đến thầy Trần

Quang Khải - giảng viên dạy học và hướng dẫn bộ môn Lập trình Python của lớp 242IPPA233277\_06CLC vì đã chỉ dẫn và hỗ trợ chúng em hết mình trong việc học tập và trong quá trình thực hiện đồ án cuối kỳ.

Trong quá trình học tập, nghiên cứu và thực hiện đề tài chúng em đã nổ lực và cố gắng

rất nhiều, nhưng do kinh nghiệm và hiểu biết còn hạn hẹp nên có thể đồ án cuối kì của

chúng em có thể còn nhiều thiếu sót. Chúng em rất mong nhận được sự động viên, hỗ

trợ và góp ý từ thầy để nhóm em có thể hoàn thiện hơn không chỉ trong môn học này

mà còn có thể sẽ làm tốt hơn nữa trong những chặng đường sắp tới.

Một lần nữa, chúng em xin trân trọng cảm ơn sự quan tâm giúp đỡ của thầy trong suốt

hành trình vừa qua.

# **PHẦN 1: MỞ ĐẦU**

## **1. Lý do chọn đề tài**

Trong thời đại dữ liệu hiện nay, việc phân tích các chỉ số xã hội như chỉ số hạnh phúc đóng vai trò quan trọng trong việc hỗ trợ hoạch định chính sách và ra quyết định chiến lược ở cấp độ quốc gia và toàn cầu. Bộ dữ liệu "World Happiness Report" cung cấp cái nhìn sâu sắc về các yếu tố ảnh hưởng đến mức độ hạnh phúc của người dân tại nhiều quốc gia trên thế giới, từ đó giúp hiểu rõ hơn mối liên hệ giữa các yếu tố kinh tế, xã hội và tâm lý với chất lượng cuộc sống. Đề tài này được lựa chọn nhằm khai thác giá trị thực tiễn của việc phân tích dữ liệu hạnh phúc toàn cầu – như tác động của GDP, mức độ hỗ trợ xã hội, sự tự do cá nhân, và cảm nhận về tham nhũng đến mức độ hài lòng trong cuộc sống. Đồng thời, việc thực hiện tiểu luận giúp người học phát triển kỹ năng sử dụng Python và các thư viện như pandas, seaborn và matplotlib để xử lý và trực quan hóa dữ liệu một cách khoa học, góp phần rèn luyện tư duy phân tích và năng lực trình bày thông tin.

## **2. Mục tiêu đề tài**

Mục tiêu của đề tài là khai thác, xử lý và phân tích bộ dữ liệu "World Happiness Report" để tìm ra các yếu tố ảnh hưởng đến chỉ số hạnh phúc của các quốc gia, từ đó rút ra các kết luận có giá trị thực tiễn trong việc xây dựng chính sách phát triển xã hội. Cụ thể, đề tài tập trung vào:

* + - Làm sạch và chuẩn hóa dữ liệu, xây dựng bảng dữ liệu dễ khai thác.
    - Phân tích mối tương quan giữa các yếu tố như GDP bình quân đầu người, hỗ trợ xã hội, kỳ vọng sống khỏe mạnh, tự do cá nhân, lòng tin trong xã hội và mức độ hài lòng tổng thể.
    - Trực quan hóa các xu hướng và mẫu hình nổi bật bằng biểu đồ, heatmap, scatter plot,... để rút ra các nhận định rõ ràng.

Đưa ra các khuyến nghị hoặc giả thuyết về sự khác biệt trong chỉ số hạnh phúc giữa các khu vực, nhóm thu nhập hoặc mức độ phát triển của quốc gia.

Thông qua đó, đề tài không chỉ hướng đến việc tạo ra sản phẩm phân tích dữ liệu hữu ích mà còn giúp người thực hiện nâng cao kỹ năng lập trình, phân tích định lượng và trình bày thông tin trực quan.

## **3. Đối tượng và phạm vi nghiên cứu**

Đối tượng nghiên cứu

* Bộ dữ liệu "World Happiness Report": Bao gồm các trường dữ liệu như tên quốc gia, chỉ số hạnh phúc, GDP bình quân đầu người, hỗ trợ xã hội, kỳ vọng sống, tự do đưa ra quyết định trong cuộc sống, mức độ tham nhũng, và cảm xúc tích cực/tiêu cực của người dân.
* Các quốc gia trên thế giới: Là các đơn vị phân tích chính, với đặc điểm kinh tế – xã hội đa dạng.

Phạm vi nghiên cứu

* Phân tích dữ liệu : Tập trung vào mối quan hệ giữa các yếu tố ảnh hưởng đến chỉ số hạnh phúc, chẳng hạn như GDP và cảm xúc tích cực, hoặc mối liên hệ giữa mức độ tham nhũng và mức độ hài lòng.
* Trực quan hóa dữ liệu : Sử dụng các biểu đồ để minh họa sự khác biệt giữa các quốc gia, các khu vực địa lý, hoặc theo nhóm chỉ số
* Ứng dụng và đánh giá : Đề xuất các khuyến nghị chính sách dựa trên kết quả phân tích, đồng thời đánh giá những yếu tố có ảnh hưởng lớn nhất đến hạnh phúc toàn cầu

# **PHẦN 2. NỘI DUNG**

# **CHƯƠNG 1: TÓM TẮT CƠ SỞ LÝ THUYẾT**

## **1.1. Giới thiệu về ngôn ngữ Python**

Python là một ngôn ngữ lập trình bậc cao, được phát triển bởi Guido van Rossum và

ra mắt lần đầu vào năm 1991. Python được biết đến với cú pháp rõ ràng, dễ đọc, và dễ

học, là công cụ lý tưởng cho cả những người mới bắt đầu lẫn các chuyên gia lập trình.

Đây là một ngôn ngữ mã nguồn mở và đa nền tảng, được ứng dụng rộng rãi trong nhiều

lĩnh vực như phát triển ứng dụng web, học máy (machine learning), trí tuệ nhân tạo (AI), và đặc biệt là phân tích dữ liệu.

Python đã trải qua nhiều phiên bản phát triển và hiện nay phiên bản Python 3.x là

phiên bản được sử dụng phổ biến nhất. Với đặc điểm thiết kế đơn giản nhưng mạnh mẽ,

Python cung cấp khả năng xử lý và phân tích dữ liệu hiệu quả thông qua các thư viện

phong phú và đa dạng. Điều này giúp Python trở thành một trong những công cụ hàng

đầu trong lĩnh vực khoa học dữ liệu.

## **1.2. Tại sao Python thường được dùng để phân tích dữ liệu**

Python được sử dụng rộng rãi trong phân tích dữ liệu nhờ vào cú pháp đơn giản,

dễ hiểu, gần gũi với ngôn ngữ tự nhiên, giúp người dùng dễ dàng triển khai các ý tưởng.

Điểm mạnh lớn nhất của Python là hệ sinh thái thư viện phong phú và mạnh mẽ, bao gồm pandas để thao tác dữ liệu dạng bảng, NumPy cho xử lý dữ liệu số, matplotlib và seaborn để trực quan hóa dữ liệu, và scikit-learn cho các thuật toán học máy. Python cũng hỗ trợ tích hợp với nhiều định dạng dữ liệu phổ biến như CSV, Excel, SQL, JSON, và có khả năng làm việc với dữ liệu lớn từ các hệ thống phân tán. Sự linh hoạt của Python cho phép nó dễ dàng kết hợp với các công cụ và ngôn ngữ khác, đồng thời ứng dụng trong các lĩnh vực đa dạng như xây dựng ứng dụng web, học máy, và xử lý ngôn ngữ tự nhiên. Hơn nữa, cộng đồng lớn mạnh và tài liệu phong phú giúp Python trở thành công cụ lý tưởng cho cả người mới bắt đầu và các chuyên gia trong lĩnh vực phân tích dữ liệu. Những ưu điểm này đã biến Python thành một công cụ không thể thiếu, giúp các nhà phân tích nhanh chóng xử lý, trực quan hóa và đưa ra những nhận định chính xác dựa trên dữ liệu.

## **1.3. Giới thiệu về các thư viện thường được sử dụng trong Phân tích dữ liệu**

Python cung cấp một hệ sinh thái phong phú với nhiều thư viện mạnh mẽ, hỗ trợ

toàn diện cho việc xử lý, phân tích và trực quan hóa dữ liệu. Dưới đây là những thư viện

thường được sử dụng:

### **1.3.1. NumPy (Numerical Python)**

NumPy là một thư viện cốt lõi trong Python, cung cấp các công cụ hiệu suất cao

để thao tác với mảng đa chiều (ndarray) và thực hiện các phép toán số học.

*Xử lý mảng hiệu suất cao:* NumPy hỗ trợ các phép toán vector hóa, cho phép áp

dụng cùng một phép tính cho toàn bộ mảng, tăng tốc đáng kể so với việc sử dụng

vòng lặp.

*Phép toán tuyến tính:* Hỗ trợ các phép toán phức tạp như nhân ma trận, nghịch đảo ma trận, và tính định thức, giúp giải quyết các bài toán toán học và khoa học.

*Khả năng tương thích:* NumPy tích hợp tốt với các thư viện khác như pandas,

matplotlib, và scikit-learn, tạo điều kiện thuận lợi cho toàn bộ chuỗi phân tích dữ

liệu.

Nhờ vào hiệu suất cao và tính linh hoạt, NumPy là nền tảng cho nhiều thư viện

khác và là công cụ không thể thiếu trong xử lý dữ liệu lớn.

### **1.3.2. Pandas**

Pandas là một thư viện mạnh mẽ, được xây dựng trên NumPy, chuyên để thao tác

và phân tích dữ liệu dạng bảng. Thư viện này được sử dụng rộng rãi nhờ các đặc điểm

sau:

*Cấu trúc dữ liệu linh hoạt:* Pandas cung cấp hai cấu trúc chính: Series (dữ liệu một

chiều) và DataFrame (dữ liệu hai chiều, tương tự bảng tính Excel).

*Đọc và ghi dữ liệu:* Pandas hỗ trợ đọc và ghi dữ liệu từ nhiều định dạng phổ biến

như CSV, Excel, SQL, JSON, giúp dễ dàng làm việc với dữ liệu từ các nguồn khác

nhau.

*Làm sạch và chuẩn hóa dữ liệu:* Pandas có các công cụ xử lý dữ liệu mạnh mẽ,

giúp loại bỏ các giá trị không hợp lệ, xử lý dữ liệu bị thiếu, và chuẩn hóa dữ liệu một cách nhanh chóng.

*Thao tác dữ liệu hiệu quả:* Pandas hỗ trợ lọc, nhóm, sắp xếp, và tính toán trên dữ

liệu, giúp quá trình phân tích dữ liệu trở nên đơn giản và trực quan hơn.

Pandas là một công cụ không thể thiếu khi xử lý và phân tích dữ liệu dạng bảng,

đặc biệt trong các dự án khoa học dữ liệu.

### **1.3.3. Matplotlib**

Matplotlib là thư viện cơ bản và phổ biến nhất để trực quan hóa dữ liệu trong

Python.

*Đa dạng biểu đồ:* Matplotlib hỗ trợ tạo nhiều loại biểu đồ như biểu đồ cột,

biểu đồ đường, biểu đồ phân tán, và biểu đồ tròn.

*Khả năng tùy chỉnh cao:* Người dùng có thể tùy chỉnh màu sắc, kiểu đường, nhãn trục, và các yếu tố khác để tạo ra các biểu đồ chuyên nghiệp.

*Tích hợp với Pandas và NumPy:* Matplotlib có thể trực tiếp làm việc với các

DataFrame hoặc mảng, giúp việc trực quan hóa trở nên nhanh chóng.

Matplotlib là công cụ nền tảng cho việc trình bày dữ liệu trực quan, hỗ trợ tốt cho

các bài toán phân tích dữ liệu.

### **1.3.4. Seaborn**

Seaborn được xây dựng trên matplotlib, cung cấp các công cụ trực quan hóa dữ

liệu cấp cao, đặc biệt là các biểu đồ thống kê.

*Biểu đồ nâng cao:* Seaborn hỗ trợ các biểu đồ phức tạp như biểu đồ tương

quan, biểu đồ phân phối, và biểu đồ hộp, giúp khám phá dữ liệu một cách sâu sắc hơn.

*Tích hợp dễ dàng:* seaborn tích hợp chặt chẽ với pandas, cho phép vẽ biểu đồ

trực tiếp từ DataFrame mà không cần chuyển đổi dữ liệu.

*Thẩm mỹ cao:* Các biểu đồ của seaborn được thiết kế sẵn với màu sắc và bố

cục đẹp mắt, giúp tăng tính trực quan và chuyên nghiệp.

Seaborn thường được sử dụng cùng matplotlib để nâng cao khả năng trực quan hóa

dữ liệu.

### **1.3.5. Scikit-learn**

Scikit-learn là thư viện hàng đầu cho học máy (machine learning) trong Python,

hỗ trợ xây dựng và triển khai các thuật toán phân tích dữ liệu.

*Thuật toán đa dạng:* Thư viện cung cấp các thuật toán phổ biến như hồi quy,

phân loại, cụm, và giảm chiều dữ liệu.

*Công cụ tiền xử lý:* Scikit-learn hỗ trợ chuẩn hóa dữ liệu, xử lý dữ liệu bị

thiếu, và mã hóa dữ liệu dạng phân loại.

*Đánh giá và lựa chọn mô hình:* Các công cụ kiểm tra chéo và lựa chọn siêu

tham số giúp tối ưu hóa mô hình.

Scikit-learn là thư viện không thể thiếu trong việc xây dựng các mô hình dự đoán

và phân tích dữ liệu nâng cao.

# **CHƯƠNG 2: MÔ TẢ TẬP DỮ LIỆU HẠNH PHÚC TOÀN CẦU**

## **2.1. Quá trình thu thập dữ liệu**

Để làm đề tài này, nhóm chúng em quyết định lựa chọn tập dữ liệu:

[*WorldHappinessReport,*](https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023) từ kho dữ liệu của cộng đồng Học máyvà trí tuệ nhân tạo Kaggle,

một kho dữ liệu uy tín được cộng đồng học máy và trí tuệ nhân tạo đánh giá cao. Đây là

một bộ dữ liệu công khai, thường được sử dụng để nghiên cứu và phân tích các yếu tố ảnh hưởng đến kết quả học tập của học sinh.

Tập dữ liệu bao gồm các thông tin như tên quốc gia, điểm số hạnh phúc tổng thể (Happiness Score), GDP bình quân đầu người, mức độ hỗ trợ xã hội, kỳ vọng sống khỏe mạnh, tự do trong việc ra quyết định, cảm nhận về tham nhũng, cũng như các cảm xúc tích cực và tiêu cực của người dân. Các thông tin này cho phép nhóm chúng em phân tích mối quan hệ giữa các yếu tố xã hội – kinh tế với mức độ hài lòng trong cuộc sống.

Tập dữ liệu *WorldHappinessReport* được lựa chọn vì tính sẵn có, chất lượng cao với dữ liệu đầy đủ và có cấu trúc rõ ràng, không chứa giá trị thiếu nghiêm trọng, đồng thời có tính ứng dụng thực tiễn cao. Đây là nền tảng tốt để nhóm thực hiện các bước xử lý, phân tích, trực quan hóa, và triển khai mô hình dự đoán nhằm khám phá các yếu tố chính ảnh hưởng đến hạnh phúc toàn cầu

## **2.2. Tổng quan về dữ liệu**

1. Lịch sử hình thành: Tập dữ liệu *WorldHappinessReport* được tải lên nền tảng Kaggle

bởi cộng đồng Học máy vào 2 năm trước.

2. Nguồn gốc: Bộ dữ liệu thuộc chủ sở hữu là Ajaypal Singh.

3. Công dụng: Phân tích và dự đoán mức độ hạnh phúc của người dân tại các quốc gia trên thế giới bị ảnh hưởng như thế nào bởi các yếu tố khác nhau, bao gồm GDP bình quân đầu người, mức độ hỗ trợ xã hội, kỳ vọng sống khỏe mạnh, sự tự do cá nhân, mức độ tham nhũng được cảm nhận, cũng như các cảm xúc tích cực và tiêu cực. Qua đó, đề tài giúp xác định các yếu tố then chốt ảnh hưởng đến hạnh phúc và đề xuất các định hướng chính sách phù hợp nhằm nâng cao chất lượng sống của người dân.

4. Dữ liệu gồm 2363 dòng trước khi lọc được đánh số từ 1 đến 2363 và có 11 cột dữ liệu.   
5. Danh sách các đặc trưng:

+ Country name: Tên quốc gia.

+ Year: Năm.

+ Life Ladder: Chỉ số thanh đo hạnh phúc.   
+ Long GDP per capita: Tổng sản phẩm quốc nội (GDP) đầu người tính theo logarit tự nhiên.  
+ Social Support : Mức độ hỗ trợ xã hội   
+ Healthy life expectancy at birth : Tuổi thọ trung bình khỏe mạnh khi sinh.

+ Freedom to make life choices: Mức độ tự do cảm nhận được trong việc ra quyết định cuộc sống của cá nhân.

+ Generosity: Mức độ rộng lượng.

+ Perceptions of corruption: Mức độ cảm nhận về sự tham nhũng trong chính phủ và doanh nghiệp.

+ Positive affect: Tỷ lệ người cảm thấy tích cực vào ngày hôm trước.

+ Negative affect : Tỷ lệ người cảm thấy tiêu cực vào ngày hôm trước.

## **2.3. Các hạn chế của dữ liệu**

### **2.3.1 Thiếu dữ liệu**

 Thiếu trường id làm khóa chính

 Một số bản ghi thiếu dữ liệu

 Một số giá trị dữ liệu không hợp lệ hoặc bất thường, ví dụ như điểm số hạnh phúc âm hoặc vượt quá giới hạn hợp lý, cần được xử lý hoặc loại bỏ.

 Chưa có trường tổng hợp đánh giá mức sống tổng thể

# **CHƯƠNG 3: PHÂN TÍCH DỮ LIỆU HẠNH PHÚC TOÀN CẦU**

## **3.1. Xử lý dữ liệu**

Trong bước xử lý dữ liệu, chúng tôi tiến hành kiểm tra và làm sạch toàn bộ dữ liệu

bằng các thư viện Pandas và NumPy để đảm bảo rằng dữ liệu đầu vào có chất lượng tốt

nhất cho quá trình phân tích. Các bước xử lý bao gồm:

### **3.1.1. Kiểm tra dữ liệu ban đầu**

Đầu tiên, chúng tôi đọc file dữ liệu **WorldHappiness.csv** bằng thư viện

pandas để nạp dữ liệu vào DataFrame. Sau đó, chúng tôi duyệt qua dữ liệu để xác định những dòng bị thiếu (NaN), trùng lặp hoặc không hợp lệ trong các cột nhưđiểm số (math score, reading score, writing score).

### **3.1.2. Xử lý các giá trị bị thiếu**

Những dòng dữ liệu bị thiếu giá trị được loại bỏ để tránh ảnh hưởng đến quá trình

phân tích. Cụ thể, chúng tôi sử dụng hàm hỗ trợ dropna() của pandas để xóa các dòng

chứa giá trị NaN.

### **3.1.3. Loại bỏ các giá trị trùng lặp**

Dữ liệu trùng lặp được kiểm tra và xóa dựa trên cột khóa chính id, đảm bảo rằng

mỗi học sinh chỉ xuất hiện một lần trong tập dữ liệu. Việc này được thực hiện bằng hàm

drop\_duplicates() của pandas.

### **3.1.4. Kiểm tra và xác thực dữ liệu hợp lệ**

Để đảm bảo tính chính xác, chúng tôi kiểm tra các giá trị trong các cột điểm số

(math score, reading score, writing score). Các giá trị nằm ngoài phạm vi từ 0 đến 100

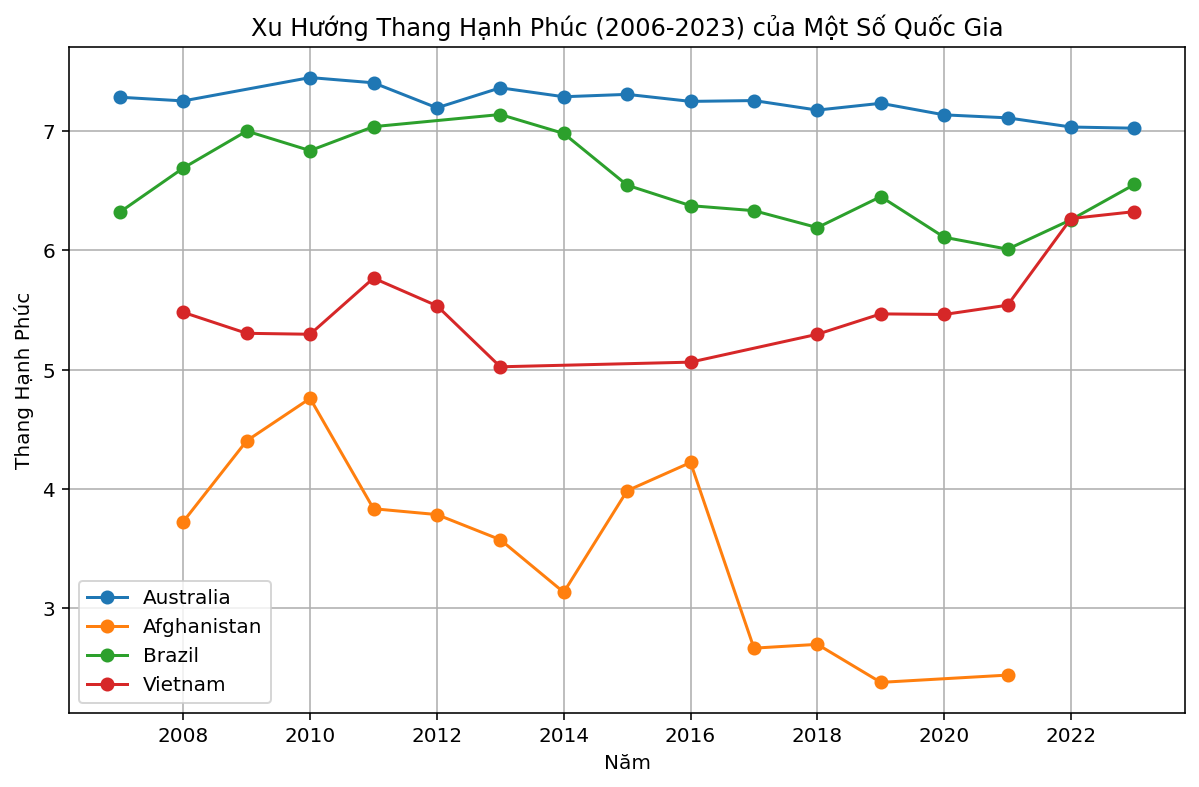
được coi là không hợp lệ và bị loại bỏ. Điều kiện này được thực hiện thông qua việc áp

dụng các phép lọc trực tiếp trên DataFrame.

### **3.1.5. Lưu dữ liệu đã làm sạch**

Sau khi dữ liệu được làm sạch, chúng tôi lưu lại dữ liệu vào một file CSV mới có

tên **WorldHappiness\_clean.csv**. Việc lưu dữ liệu sạch vào file mới giúp thuận tiện hơn cho các bước phân tích tiếp theo.

**3.2. Trực quan hóa dữ liệu** Chúng em tiến hành tạo các biểu đồ và đồ thị để trực quan hóa dữ liệu từ tệp dữ liệu đã làm sạch, nhằm minh họa rõ ràng các mỗi quan hệ giữa các yếu tố. Quá trình trực quan hóa này không chỉ giúp phân tích đặc điểm của từng nhóm đối tượng mà còn hỗ trợ nhận diện xu hướng, sự chênh lệch và mối tương quan giữa các yếu tố. Dưới đây là các biểu đồ :   


Hình 1. Biểu đồ đường : Xu hướng Thanh Hạnh Phúc của một số quốc gia

Biểu đồ đường này thể hiện xu hướng chỉ số hạnh phúc (Thang Hạnh Phúc - Life Ladder) từ năm 2006 đến năm 2023 cho bốn quốc gia: Australia, Afghanistan, Brazil và Việt Nam. Trục hoành (x) biểu thị các năm từ 2006 đến 2023, trong khi trục tung (y) biểu thị giá trị Thang Hạnh Phúc, dao động từ 0 đến 10. Mỗi quốc gia được đại diện bằng một đường riêng biệt, kèm theo các điểm đánh dấu hình tròn tại mỗi năm để dễ theo dõi.

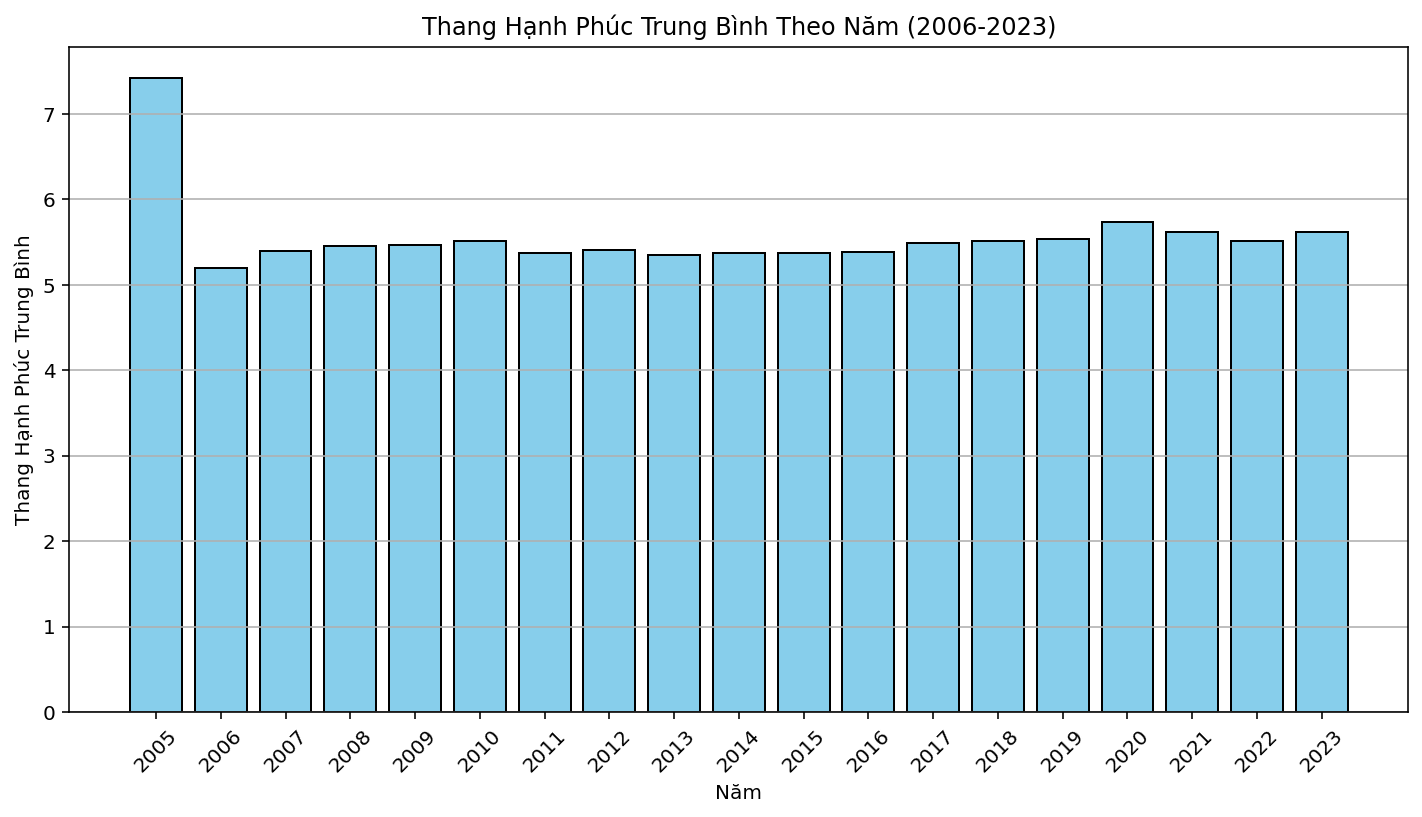
Australia: Đường biểu thị chỉ số hạnh phúc của Australia dao động ổn định ở mức cao, thường quanh 7.0 đến 7.5, cho thấy mức độ hạnh phúc tương đối ổn định và cao trong suốt giai đoạn.

Afghanistan: Đường của Afghanistan có xu hướng giảm dần, với chỉ số hạnh phúc bắt đầu từ khoảng 3.7 vào năm 2008, tăng nhẹ lên 4.8 vào năm 2010, nhưng sau đó giảm mạnh xuống dưới 2.5 vào các năm 2018-2021, phản ánh sự sụt giảm đáng kể về mức độ hạnh phúc.

Brazil: Chỉ số hạnh phúc của Brazil dao động từ khoảng 6.0 đến 7.0, với một số biến động nhẹ nhưng nhìn chung duy trì ở mức trung bình cao, cho thấy sự ổn định tương đối.

Việt Nam: Đường của Việt Nam cho thấy xu hướng tăng dần, từ khoảng 5.3 vào năm 2008 lên khoảng 6.3 vào năm 2023, phản ánh sự cải thiện dần về mức độ hạnh phúc qua thời gian.

Biểu đồ có lưới nền (grid) và chú thích (legend) để phân biệt các quốc gia, giúp dễ dàng so sánh xu hướng hạnh phúc giữa các nước.



Hình 2. Biểu đồ cột : Thang Hạnh Phúc trung bình theo năm

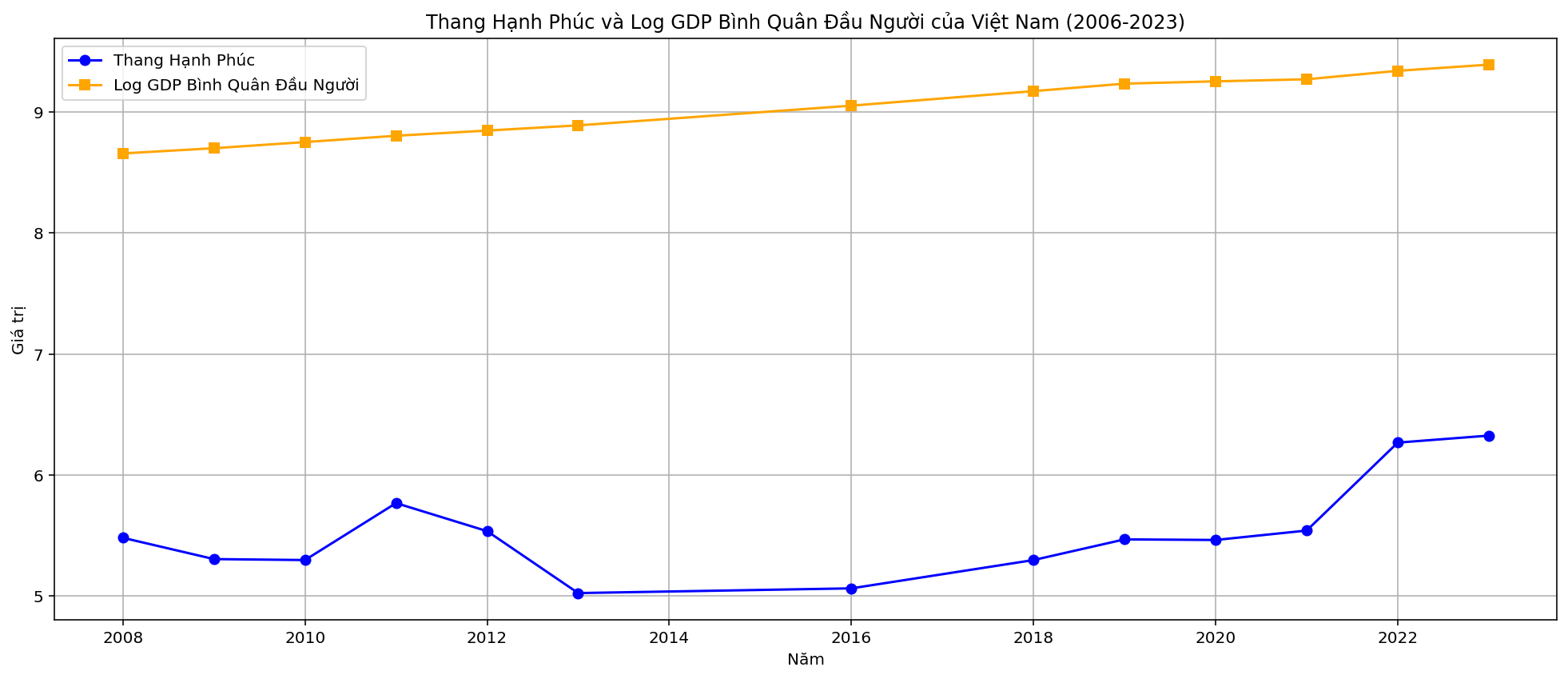
Biểu đồ cột này minh họa giá trị trung bình của Thang Hạnh Phúc (Life Ladder) trên toàn bộ các quốc gia trong dataset qua từng năm từ 2006 đến 2023. Trục hoành (x) liệt kê các năm, với các nhãn được xoay 45 độ để dễ đọc, trong khi trục tung (y) biểu thị giá trị trung bình của Thang Hạnh Phúc, dao động từ 0 đến 10.

Mỗi cột đại diện cho một năm, được tô màu xanh nhạt (skyblue) với viền đen để tăng độ tương phản. Chiều cao của cột phản ánh mức độ hạnh phúc trung bình của tất cả các quốc gia trong năm đó.

Quan sát biểu đồ, giá trị trung bình của Thang Hạnh Phúc dao động trong khoảng 5.0 đến 5.5 qua các năm, với một số biến động nhỏ. Ví dụ, một số năm như 2009 hoặc 2020 có thể cho thấy mức hạnh phúc trung bình thấp hơn nhẹ, trong khi các năm như 2018 hoặc 2023 có thể cao hơn một chút.

Biểu đồ có lưới nền theo trục tung để hỗ trợ việc ước lượng giá trị, giúp người xem dễ dàng nhận biết sự thay đổi của mức hạnh phúc trung bình qua thời gian.

Biểu đồ này cung cấp cái nhìn tổng quan về xu hướng hạnh phúc trung bình toàn cầu, nhấn mạnh sự ổn định tương đối nhưng cũng có những biến động nhỏ qua các năm.



Hình 3. Biểu đồ đường : Thang Hạnh Phúc và Log GDP bình quân đầu người của Việt Nam qua các năm

Biểu đồ đường này so sánh xu hướng của Thang Hạnh Phúc (Life Ladder) và Log GDP bình quân đầu người (Log GDP per capita) của Việt Nam qua các năm. Trục hoành (x) biểu thị các năm, trong khi trục tung (y) biểu thị giá trị của cả hai chỉ số (Thang Hạnh Phúc và Log GDP). Lưu ý rằng hai chỉ số này có thang đo khác nhau, nhưng được vẽ trên cùng một trục y để so sánh xu hướng.

Phân bố dữ liệu:

Thang Hạnh Phúc (đường màu xanh dương, điểm đánh dấu hình tròn): Bắt đầu từ khoảng 5.3 (2008) và tăng dần lên khoảng 6.3 (2023), cho thấy xu hướng cải thiện mức độ hạnh phúc.

Log GDP bình quân đầu người (đường màu cam, điểm đánh dấu hình vuông): Bắt đầu từ khoảng 7.5-8.0 (2008) và tăng lên khoảng 8.5-9.0 (2023), phản ánh sự phát triển kinh tế của Việt Nam.

Dữ liệu có thể bắt đầu từ 2008 thay vì 2006 (như tiêu đề) nếu dataset thiếu dữ liệu cho Việt Nam trong các năm 2006-2007, như đã thảo luận trước.

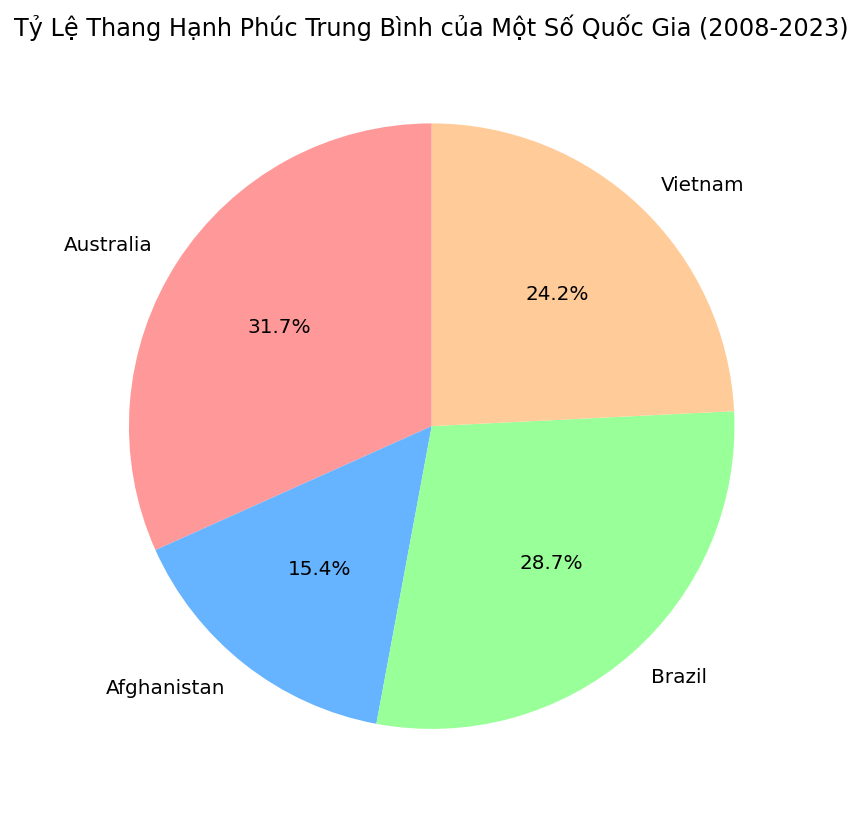
Đặc điểm trực quan:

Hai đường được phân biệt bằng màu sắc (xanh dương cho hạnh phúc, cam cho GDP) và kiểu điểm đánh dấu (hình tròn và hình vuông).

Biểu đồ có lưới nền và chú thích (legend) để phân biệt hai chỉ số. Tiêu đề "Thang Hạnh Phúc và Log GDP Bình Quân Đầu Người của Việt Nam (2006-2023)" có thể cần sửa thành "2008-2023" nếu dữ liệu chỉ bắt đầu từ 2008.

Do hai chỉ số có thang đo khác nhau (hạnh phúc: 0-10; Log GDP: khoảng 7-9), xu hướng tăng của cả hai được hiển thị, nhưng giá trị tuyệt đối không thể so sánh trực tiếp.

Ý nghĩa: Biểu đồ làm nổi bật mối quan hệ tích cực giữa sự phát triển kinh tế (Log GDP) và mức độ hạnh phúc tại Việt Nam, với cả hai chỉ số đều tăng qua thời gian. Nó cho thấy sự cải thiện đồng thời về thu nhập và chất lượng sống, dù cần lưu ý rằng thang đo khác nhau có thể làm khó so sánh trực tiếp.

\

Hình 4. Biểu đồ tròn Tỷ lệ Thang Hạnh Phúc của một số quốc gia

Biểu đồ tròn này minh họa tỷ lệ phần trăm của chỉ số hạnh phúc trung bình (Life Ladder) của bốn quốc gia—Australia, Afghanistan, Brazil, và Việt Nam—trong tổng chỉ số hạnh phúc trung bình của nhóm này, dựa trên dữ liệu từ năm 2008 đến 2023. Biểu đồ được hiển thị trong một tab riêng biệt trong giao diện tkinter, cung cấp một cách trực quan và dễ hiểu để so sánh mức độ hạnh phúc trung bình giữa các quốc gia.

Phân bố dữ liệu:

Mỗi quốc gia được biểu thị bằng một phần của hình tròn, với kích thước phần tỷ lệ với giá trị trung bình của chỉ số hạnh phúc (Life Ladder) của quốc gia đó trong giai đoạn 2008-2023.

Australia: Dự kiến chiếm phần lớn nhất, có thể khoảng 30-40% tổng tỷ lệ, do chỉ số hạnh phúc trung bình cao, khoảng 7.0-7.5. Phần này thể hiện mức độ hạnh phúc ổn định và cao của quốc gia này.

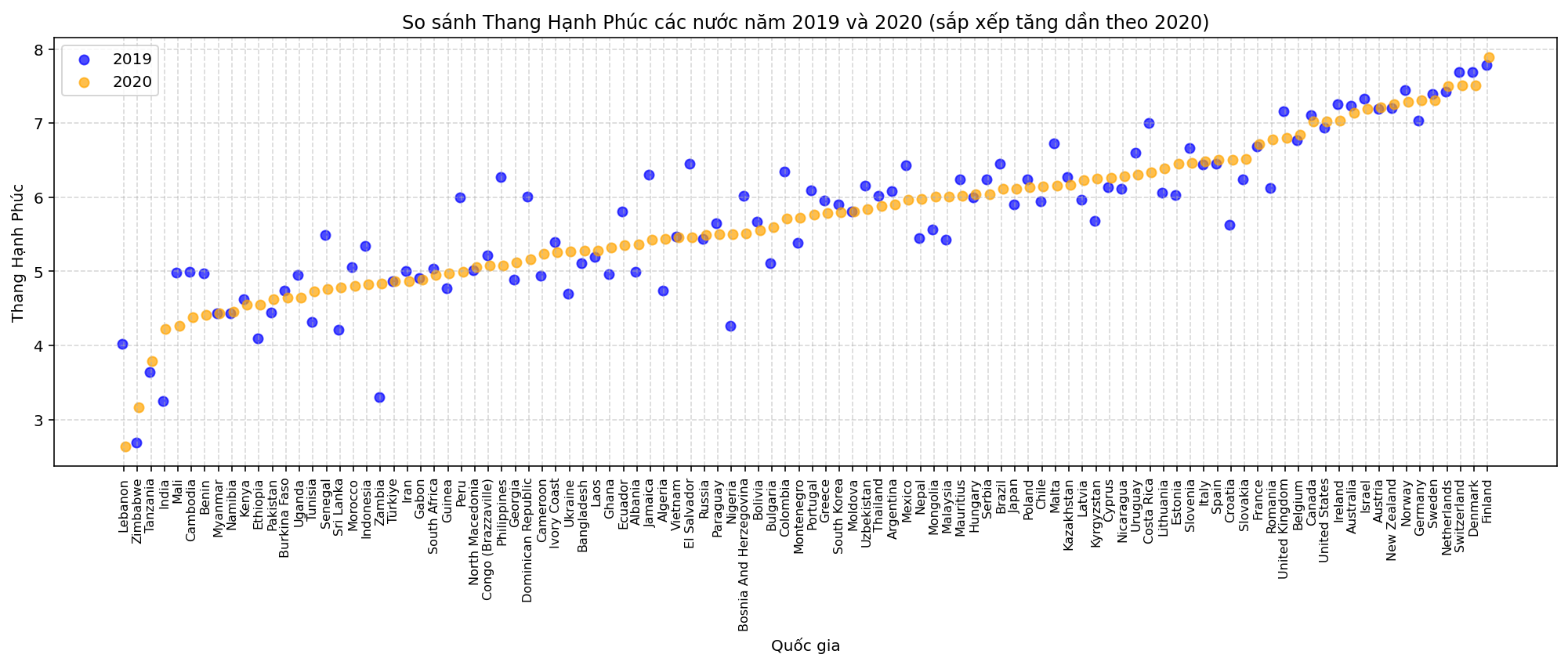
Brazil: Chiếm phần lớn thứ hai, khoảng 25-30%, với chỉ số hạnh phúc trung bình khoảng 6.0-6.5, phản ánh mức hạnh phúc trung bình cao và tương đối ổn định.

Việt Nam: Chiếm phần nhỏ hơn, khoảng 20-25%, với chỉ số hạnh phúc trung bình khoảng 5.5-6.0, cho thấy mức hạnh phúc đã cải thiện nhưng thấp hơn Australia và Brazil.

Afghanistan: Chiếm phần nhỏ nhất, khoảng 10-15%, do chỉ số hạnh phúc trung bình thấp, khoảng 3.0-4.0, phản ánh các điều kiện kinh tế và xã hội khó khăn. Tỷ lệ phần trăm được hiển thị trực tiếp trên mỗi phần của biểu đồ tròn (ví dụ: "35.2%"), với định dạng số thập phân một chữ số, giúp người xem dễ dàng nhận biết mức độ đóng góp của mỗi quốc gia.

Đặc điểm trực quan: Biểu đồ có dạng hình tròn, với mỗi quốc gia được phân biệt bằng một màu sắc riêng biệt để tăng tính trực quan làm nổi bật sự cải thiện của quốc gia này.

Ý nghĩa: Biểu đồ tròn cung cấp một cái nhìn tổng quan và trực quan về sự khác biệt trong mức độ hạnh phúc trung bình giữa bốn quốc gia, nhấn mạnh rằng các quốc gia có điều kiện kinh tế tốt hơn (như Australia) thường có chỉ số hạnh phúc cao hơn, trong khi các quốc gia gặp bất ổn (như Afghanistan) có chỉ số thấp hơn. Nó làm nổi bật mức độ đóng góp tương đối của mỗi quốc gia vào tổng chỉ số hạnh phúc trung bình của nhóm, với Australia dẫn đầu và Afghanistan thấp nhất. Bằng cách sử dụng tỷ lệ phần trăm, biểu đồ giúp người xem dễ dàng so sánh và hiểu ngay lập tức rằng Australia và Brazil có mức hạnh phúc vượt trội so với Việt Nam và Afghanistan. Trong bối cảnh bộ năm biểu đồ, biểu đồ tròn này bổ sung một góc nhìn mới, tập trung vào so sánh tĩnh giữa các quốc gia, khác với các biểu đồ khác nhấn mạnh xu hướng thời gian (biểu đồ 1, 3), trung bình toàn cầu (biểu đồ 2), hoặc so sánh năm cụ thể (biểu đồ 5).



Hình 5. Biểu đồ phân tán : So sánh Thang Hạnh Phúc của các nước năm 2019 và 2020

Biểu đồ phân tán này so sánh chỉ số hạnh phúc (Life Ladder) của các quốc gia trong hai năm 2019 và 2020, với các quốc gia được sắp xếp tăng dần theo chỉ số hạnh phúc năm 2020. Trục hoành (x) liệt kê tên các quốc gia, trong khi trục tung (y) biểu thị giá trị Thang Hạnh Phúc (0-10).

Phân bố dữ liệu:

Năm 2019 (điểm màu xanh dương): Các điểm phân tán biểu thị chỉ số hạnh phúc của từng quốc gia trong năm 2019, nằm hơi lệch sang trái so với vị trí quốc gia.

Năm 2020 (điểm màu cam): Các điểm biểu thị chỉ số hạnh phúc năm 2020, nằm hơi lệch sang phải, được sắp xếp theo thứ tự tăng dần từ quốc gia có hạnh phúc thấp nhất đến cao nhất.

Sự khác biệt giữa hai năm được thể hiện qua vị trí tương đối của các điểm. Ví dụ, một số quốc gia có thể có chỉ số hạnh phúc giảm trong năm 2020 (do đại dịch), trong khi một số khác tăng hoặc ổn định.

Dữ liệu được lấy từ các quốc gia có mặt trong cả hai năm 2019 và 2020, sử dụng pd.merge để đảm bảo so sánh chính xác.

Đặc điểm trực quan:

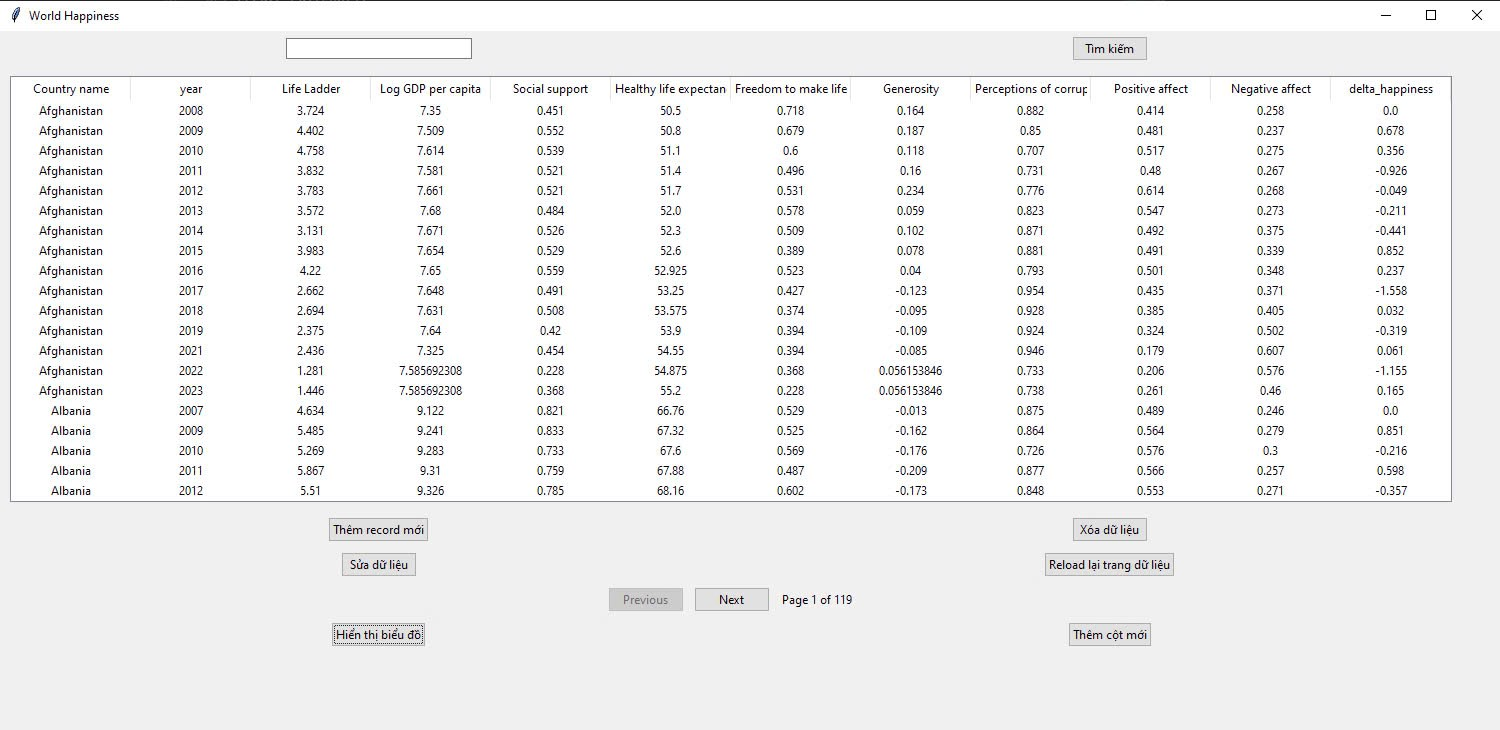
Các điểm xanh dương (2019) và cam (2020) được đặt cạnh nhau cho mỗi quốc gia, với độ trong suốt (alpha=0.7) để tránh chồng lấn.

Nhãn quốc gia trên trục x được xoay 90 độ và font chữ nhỏ (8) để hiển thị rõ ràng, dù có thể hơi khó đọc nếu số quốc gia lớn. Biểu đồ có lưới nền kiểu dấu gạch ngang (--) với độ trong suốt thấp (alpha=0.5), hỗ trợ đọc giá trị.

Chú thích (legend) phân biệt hai năm, và tiêu đề "So sánh Thang Hạnh Phúc các nước năm 2019 và 2020 (sắp xếp tăng dần theo 2020)" mô tả rõ mục đích.

Ý nghĩa: Biểu đồ giúp đánh giá tác động của các sự kiện toàn cầu (như đại dịch COVID-19) đến mức độ hạnh phúc giữa hai năm. Việc sắp xếp theo năm 2020 làm nổi bật các quốc gia có chỉ số hạnh phúc thấp nhất và cao nhất trong năm này, đồng thời cho phép so sánh với năm trước để nhận diện xu hướng tăng hoặc giảm.

**3.3. Thiết kế giao diện** Để cho người dùng dễ dàng tương tác với dữ liệu, có thể tương tác trực tiếp bằng các nút cơ bản và trực quan thì nhóm chúng em dùng thư viện Tkinter để tạo một giao diện người dùng đơn giản.

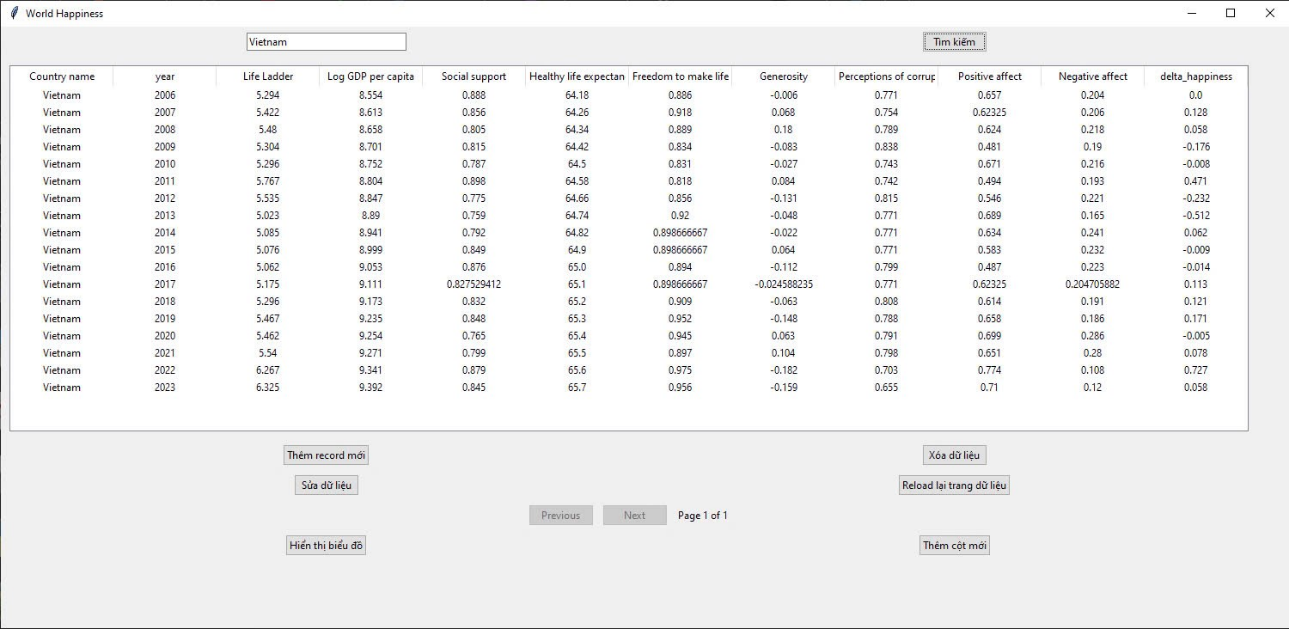
****

Hình 6. Giao diện người dùng

Giao diện được thiết kế bao gồm các nút cơ bản như “Thêm record mới”, “Sửa dữ liệu”, “Xóa dữ liệu”, 2 nút chuyển trang, “Reload lại trang dự liệu”, “ Hiển thị biểu đồ” và “Thêm cột mới”. Với giao diện đơn giản như này thì người dùng có thể dễ dàng quản lí và tổ chức dữ liệu mà không cần phải có nhiều kiến thức về lập trình

## **3.4. Các chức năng chính của chương trình**

### **3.4.1. Chức năng tìm kiếm**

****

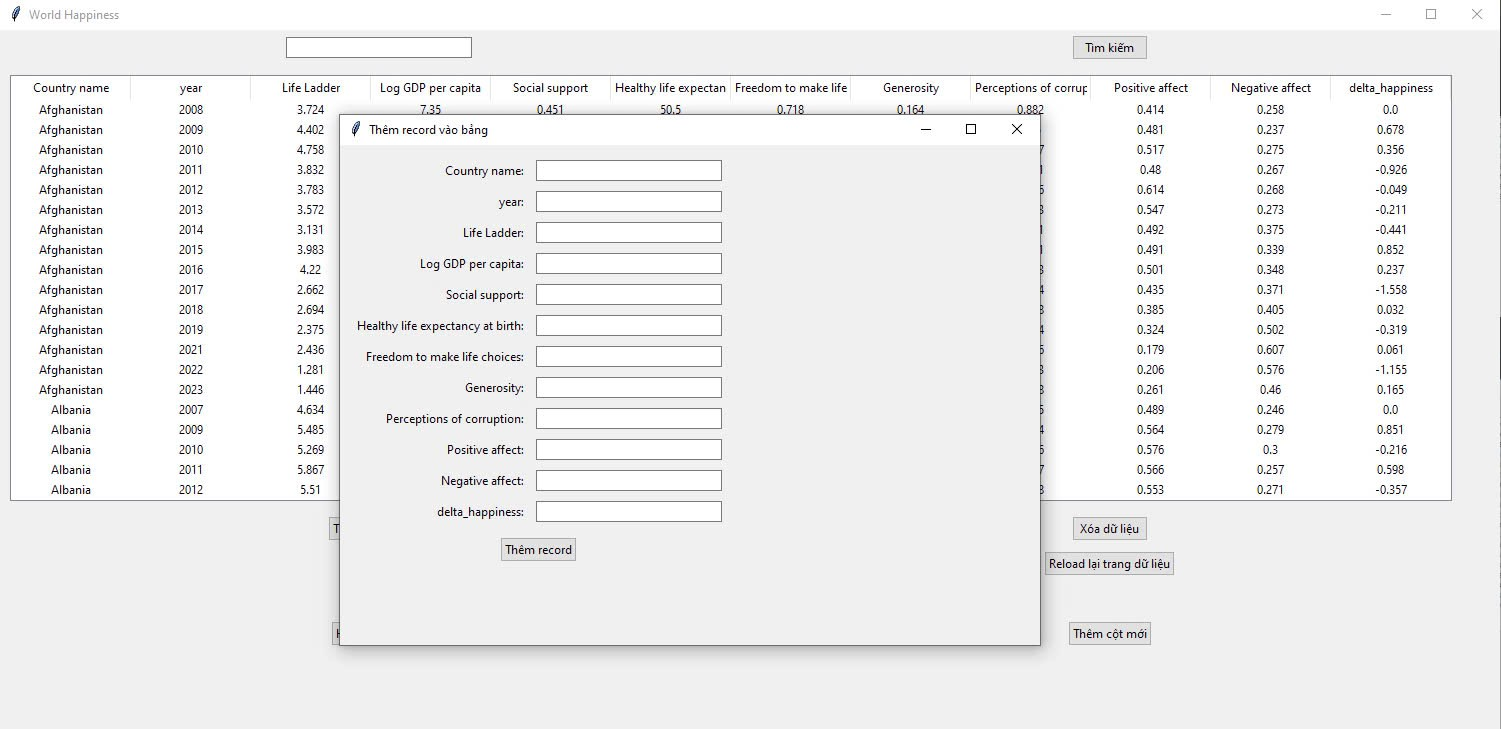
Hình 7. Chức năng tìm kiếm

Mục đích : Cho phép người dùng lọc và hiển thị dữ liệu chỉ liên quan đến quốc gia được nhập trong ô tìm kiếm  
 Cách sử dụng :  
 - Nhập tên quốc gia vào ô tìm kiếm

- Nhấn nút “Tìm kiếm”  
 - Ứng dụng sẽ hiển thị bảng dữ liệu chứa các thông tin của quốc gia đó.

Kết quả : Dữ liệu được lọc theo tên quốc gia nhập vào, giúp người dùng dễ dàng theo dõi và phân tích dữ liệu cụ thể cho từng quốc gia.

### **3.4.2. Chức năng thêm record**

  
Hình 8. Chức năng thêm record

Mục đích : Cho phép người dùng thêm một record mới vào bảng dữ liệu

Cách sử dụng :

- Khi ấn vào nút thêm record vào bảng trong giao diện chính một cửa sổ sẽ hiện ra

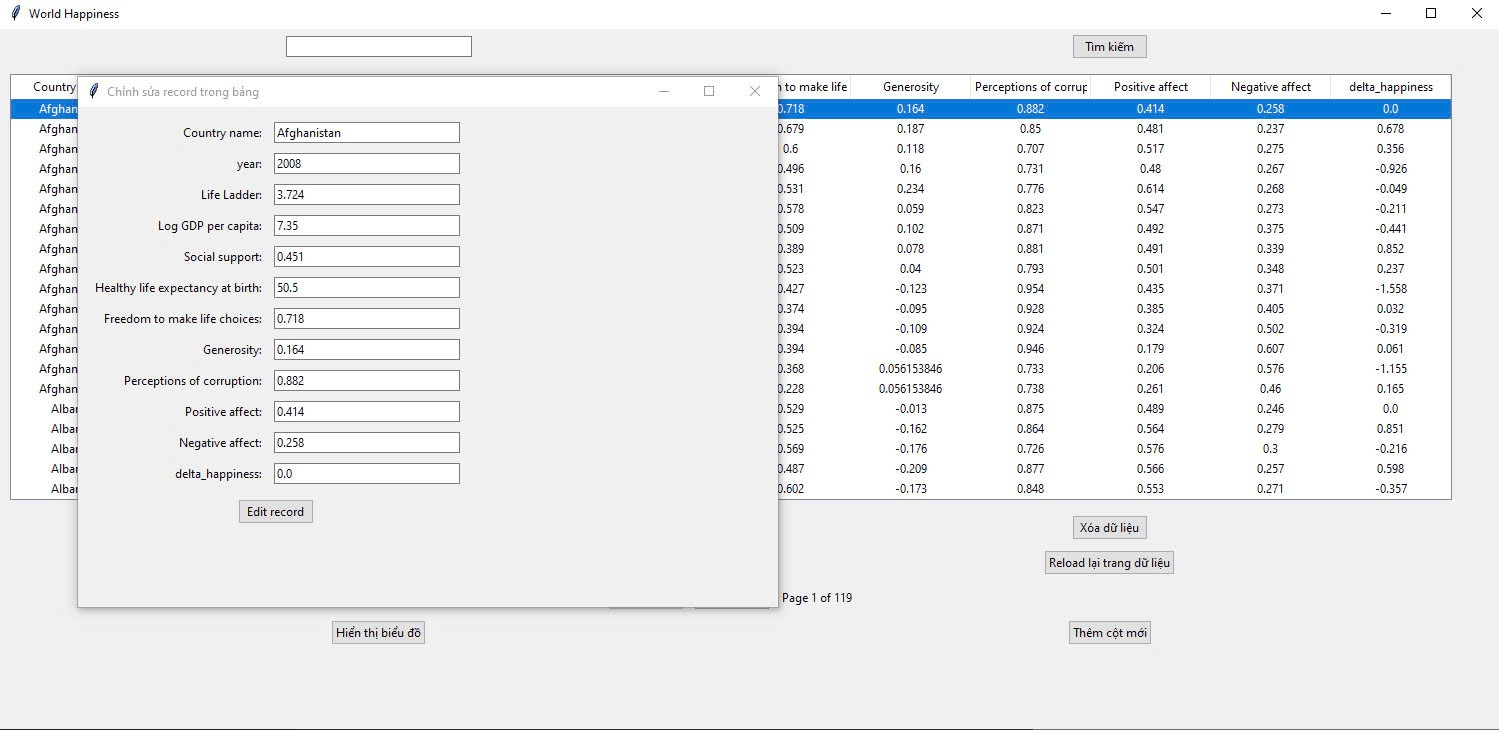
- Người dùng điền các thông tin vào các ô

- Sau khi điền đầy đủ, người dùng nhất nút “Thêm record”

- Dữ liệ vừa nhập sẽ được thêm vào bảng chính

Chức năng này giúp cập nhật, mở rộng dữ liệu khảo sát hạnh phúc thế giới bằng cách thêm thông tin cho các quốc gia hoặc năm chưa có trong bảng

### **3.4.3. Chức năng sửa record**

****

Hình 9. Chức năng sửa record

Chức năng cho phép người dùng cập nhật thông tin của một record đã có trong bảng dữ liệu

Cách sử dụng :  
 - Người dùng chọn một dòng dữ liệu trong bảng

- Nhấn nút “Sửa dữ liệu” trong giao diện chính.

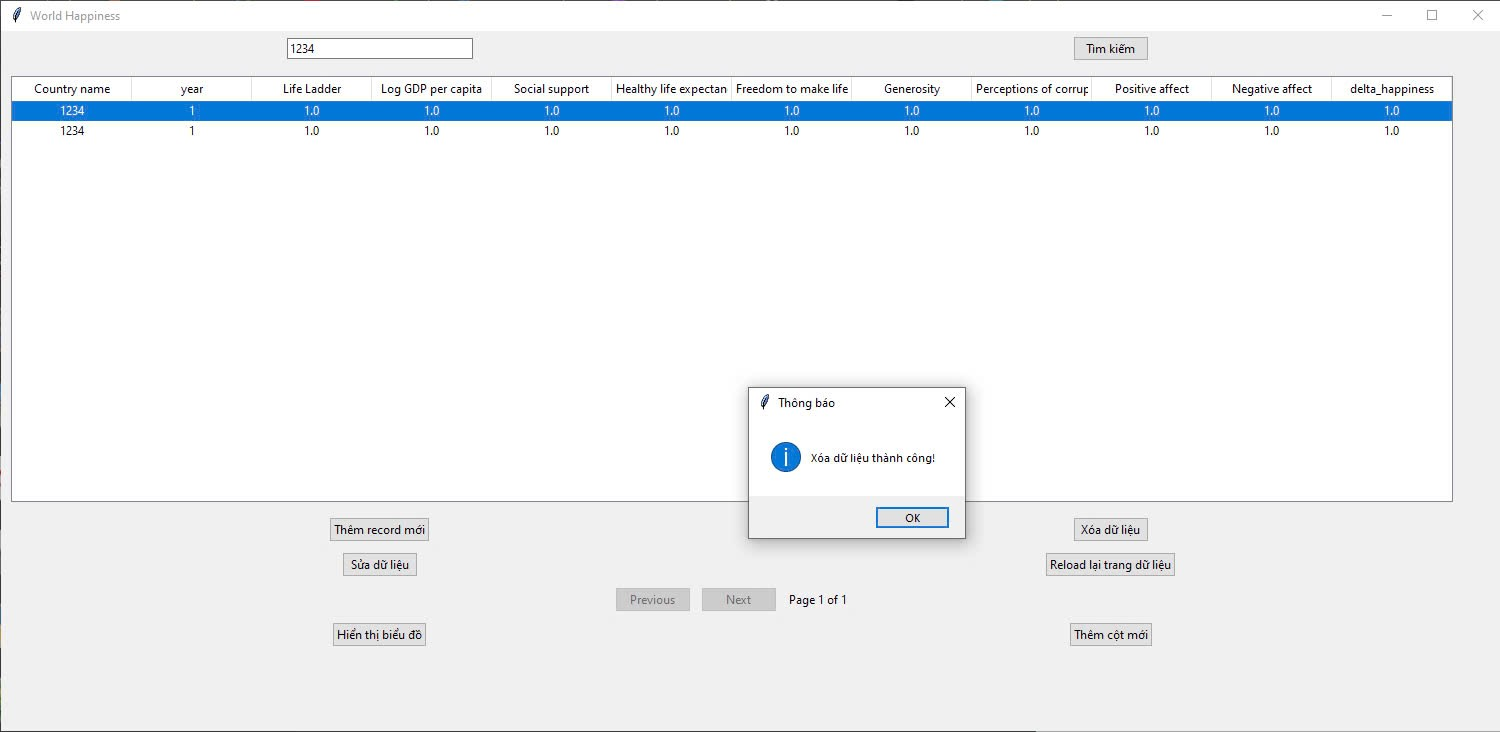
- Một cửa sổ mới xuất hiện, hiển thị thông tin chi tiết của bản ghi đã chọn

- Người dùng có thể chỉnh sửa các giá trị trong các ô nhập tương ứng

- Sau khi chỉnh sửa xong, nhấn nút “Edit record” để lưu thay đổi

Chức năng này hỗ trợ cập nhật dữ liệu bị sai hoặc cần thay đổi đảm bảo tính chính xác và linh hoạt của hệ thống dữ liệu

### **3.4.4. Chức năng xóa dữ liệu**



Hình 10. Chức năng xóa dữ liệu

Chức năng này cho phép người dùng xóa một record đã chọn khỏi bảng dữ liệu

Cách sử dụng :

- Người dùng chọn một dòng dữ liệu muốn xóa trong bảng

- Nhấn nút xóa dữ liệu

- Một hộp thoại thông báo hiện ra báo rằng “Xóa dữ liệu thành công”

- Dữ liệu được gỡ bỏ khỏi bảng và giao diện cập nhật lại

Ý nghĩa :

- Hỗ trợ quản lý dữ liệu dễ dàng hơn

- Loại bỏ các bản ghi không còn cần thiết, sai sót hoặc là nhập thử nghiệm

- Đảm bảo dữ liệu sạch sẽ, chính xác

# **PHẦN KẾT LUẬN**

## **1. Kết quả đạt được**

**Về kiến thức:**

Qua quá trình thực hiện đề tài, chúng tôi đã có cơ hội làm việc với tập dữ

liệu thực tế, giúp hiểu rõ hơn về cách phân tích và xử lý dữ liệu lớn. Chúng tôi tập

trung nhận diện các yếu tố chính ảnh hưởng đến kết quả học tập của học sinh, nắm bắt được cách các thành phần trong dữ liệu tương tác và ảnh hưởng lẫn nhau.

Các bước làm sạch và chuẩn hóa dữ liệu giúp chúng tôi nhận ra tầm quan

trọng của việc đảm bảo chất lượng dữ liệu trước khi phân tích. Những lỗi như dữ

liệu bị thiếu, trùng lặp, hay không hợp lệ cần được xử lý kỹ lưỡng để đảm bảo độ

chính xác và tính toàn vẹn của dữ liệu đầu vào.

Quá trình phân tích và trực quan hóa đã giúp nhóm chúng tôi nắm rõ hơn về việc sử dụng các công cụ như pandas, matplotlib, seaborn để khai thác thông tin từ dữ liệu. Chúng tôi nhận thấy rằng trực quan hóa là một bước quan trọng để nhận diện các xu hướng và mối quan hệ phức tạp mà dữ liệu thô không thể hiện rõ.

Bên cạnh đó, qua việc phân tích các biểu đồ, chúng tôi đã học cách trình bày

và diễn giải kết quả phân tích sao cho trực quan, dễ hiểu, từ đó hỗ trợ quá trình đưa

ra nhận định chính xác.

**Về sản phẩm:**

Sản phẩm đầu ra của đề tài là một tập dữ liệu đã được làm sạch và chuẩn hóa, sẵn sàng phục vụ cho các bước phân tích chuyên sâu hoặc ứng dụng trong các dự án khác. Quy trình làm sạch và lưu trữ dữ liệu đã góp phần nâng cao chất lượng dữ liệu, đảm bảo độ chính xác và hiệu quả trong xử lý.  
Ngoài ra, nhóm đã xây dựng một hệ thống giao diện quản lý dữ liệu với đầy đủ chức năng như: tải dữ liệu, làm sạch, chuẩn hóa, thêm/xóa/cập nhật dữ liệu và trực quan hóa bằng biểu đồ. Giao diện này không chỉ hỗ trợ hiệu quả cho nhóm nghiên cứu, mà còn có tiềm năng áp dụng trong nhiều đề tài phân tích dữ liệu khác.

Thông qua việc phân tích dữ liệu và biểu đồ, nhóm đã rút ra được các xu hướng và mối quan hệ đáng chú ý, có giá trị tham khảo trong lĩnh vực giáo dục. Các biểu đồ và báo cáo được tạo ra có thể làm nền tảng cho các nghiên cứu chuyên sâu hoặc xây dựng mô hình dự đoán trong tương lai.

Ưu điểm:

- Mô hình tự động hóa hiệu quả việc làm sạch và xử lý dữ liệu.

- Giao diện thân thiện, dễ sử dụng, phù hợp với cả người dùng không chuyên.

- Khả năng trực quan hóa mạnh mẽ với đa dạng biểu đồ, hỗ trợ minh họa rõ ràng các mối quan hệ trong dữ liệu.

Nhược điểm :  
 - Tốc độ xử lý còn hạn chế khi áp dụng với các tập dữ liệu lớn.

- Chưa tích hợp các thuật toán học máy để dự đoán hoặc phân tích loại dữ liệu.

- Giao diện còn sơ sài, chưa hỗ trợ các tính năng nâng cao như tải file hoặc xuất biểu đồ dưới dạng hình ảnh.

## **2. Hướng phát triển**

 Mở rộng chức năng phân tích.

 Nâng cấp UI và thêm tính năng

 Ứng dụng kết quả vào dự báo và khuyến nghị chính sách xã hội

# **PHẦN TÀI LIỆU THAM KHẢO**

**[1].** Kaggle (2024). *World Happiness Report*, đường dẫn:

<https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023>

**[2].** GeeksforGeeks. *Python Numpy*. Đường dẫn:

<https://www.geeksforgeeks.org/python-numpy/>

**[3].** GeeksforGeeks. *Pandas Tutorial*. Đường dẫn:

x

<https://www.geeksforgeeks.org/pandas-tutorial/>

**[4].** GeeksforGeeks. *Matplotlib Tutorial*. Đường dẫn: <https://www.geeksforgeeks.org/matplotlib-tutorial/>

**[5].** GeeksforGeeks. *Python Seaborn Tutorial*. Đường dẫn:

<https://www.geeksforgeeks.org/python-seaborn-tutorial/>

**[6].** GeeksforGeeks. *Learning Model Building in Scikit-learn*. Đường dẫn:

[https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-](https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/)

[learning-library/](https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/)

**[7].** GeeksforGeeks. *PythonTkinter*. Đường dẫn: <https://www.geeksforgeeks.org/python-gui-tkinter/>

**[8].** GeeksforGeeks. *Python Tkinter Tutorial*. Đường dẫn:

<https://www.geeksforgeeks.org/python-tkinter-tutorial/>