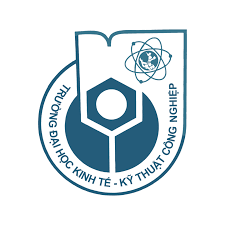
**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC KINH TẾ - KỸ THUẬT CÔNG NGHIỆP**

**KHOA KHOA HỌC ỨNG DỤNG**

****

**BÁO CÁO ĐỒ ÁN 1**

*(Nhóm 1)*

**THĂM DÒ VÀ TRỰC QUAN HÓA DỮ LIỆU CHỨNG KHOÁN NHẰM PHÂN TÍCH GIAO DỊCH KHOẢN VAY CỦA BA NGÂN HÀNG MBBANK, VIETCOMBANK VÀ SACOMBANK**

****

*( GV hướng dẫn : ThS. TRẦN CHÍ LÊ )*

**Hà Nội - 2024**

**BỘ CÔNG THƯƠNG**

**TRƯỜNG ĐẠI HỌC KINH TẾ KỸ THUẬT CÔNG NGHIỆP**

**KHOA KHOA HỌC ỨNG DỤNG**

**BÁO CÁO ĐỒ ÁN 1**

*(Nhóm 1)*

**THĂM DÒ VÀ TRỰC QUAN HÓA DỮ LIỆU CHỨNG KHOÁN NHẰM PHÂN TÍCH GIAO DỊCH KHOẢN VAY CỦA BA NGÂN HÀNG MBBANK, VIETCOMBANK VÀ SACOMBANK**

|  |  |
| --- | --- |
| Họ và tên các thành viên nhóm | Mã Sinh viên |
| Nguyễn Thị Na | 22174600084 |
| Lã Thị Minh Phương | 22174600075 |
| Phùng Thế Ngọc | 22174600101 |
| Phạm Thị Thu Trang | 22174600030 |

*(GV hướng dẫn : ThS. Trần Chí Lê)*

**Hà Nội – 2024**

# **LỜI CẢM ƠN**

Để hoàn thành báo cáo đồ án 1 này trước hết nhóm chúng em xin gửi lời cảm ơn chân thành đến quý Thầy/Cô trong khoa Khoa học ứng dụng trường Đại học Kinh Tế - Kĩ Thuật-Công Nghiệp đã trang bị cho các thành viên trong nhóm những kiến thức nền tảng và tạo điều kiện thuận lợi cho nhóm chúng em trong suốt quá trình học tập và nghiên cứu tại trường. Đặc biệt, nhóm chúng em xin gửi đến thầy Trần Chí Lê - giảng viên bộ môn Đồ án 1, người đã tận tình hướng dẫn, giúp đỡ các thành viên trong nhóm hoàn thành báo cáo đồ án 1 này lời cảm ơn sâu sắc nhất. Sự chỉ bảo tận tình và những góp ý quý báu của thầy đã giúp các thành viên trong nhóm hoàn thiện đồ án 1 và mở rộng kiến thức của mình. Giúp chúng em hiểu rõ hơn về các kiến thức chuyên môn, cũng như cách thức tiếp cận và giải quyết vấn đề trong đề tài của mình.

Cuối cùng, các thành viên trong nhóm chúng em xin gửi lời chúc sức khỏe và thành công đến Thầy/Cô trong khoa và mong rằng sẽ tiếp tục nhận được sự hướng dẫn và hỗ trợ từ Thầy/Cô trong những học kỳ tiếp theo.

Một lần nữa, chúng em xin chân thành cảm ơn !

Hà Nội, ngày 15 tháng 12 năm 2024

Nhóm thực hiện đồ án

# **TÊN ĐỒ ÁN VÀ YÊU CẦU CHI TIẾT**

1. **Thông tin thành viên nhóm**

|  |  |  |
| --- | --- | --- |
| Họ và tên các thành viên nhóm | Mã Sinh viên | Điện thoại |
| Nguyễn Thị Na | 22174600084 | 0763045385 |
| Lã Thị Minh Phương | 22174600075 | 0787006865 |
| Phùng Thế Ngọc | 22174600101 | 0365111161 |
| Phạm Thị Thu Trang | 22174600030 | 0328354259 |

1. **Thông tin đề tài**

Tên của đề tài: Thăm dò và trực quan hóa dữ liệu chứng khoán nhằm phân tích giao dịch khoản vay của ba ngân hàng MBBANK ,VIETCOMBANK và SACOMBANK.

Mục đích của đồ án: vì vấn đề về khoản vay của ngân hàng có ảnh hưởng tới sự ra quyết định chiến lược của ngân hàng và nhà đầu tư .Đặc biệt cả MBBANK, VIETCOMBANK và SACOMBANK đều là những cái tên lớn trong hệ thống tài chính Việt Nam, có sự ảnh hưởng lớn đến nền kinh tế. Nên chúng em chọn đồ án này để sử dụng dữ liệu đã thống kê từ đó giúp nhà đầu tư và ngân hàng đưa ra những chiến lược đúng đắn.

Đồ án môn học được thực hiện tại: Bộ Môn Toán & Khoa Học Dữ Liệu, Khoa Khoa Học Ứng Dụng, Trường Đại Học Kinh Tế - Kỹ Thuật Công Nghiệp.

Thời gian thực hiện: Từ ngày 25/03/2024 đến 26/05/2024.

1. **Các nhiệm vụ cụ thể của đề tài**

* Thu thập dữ liệu.
* Thống kê mô tả.
* Phân tích mối liên hệ giữa các biến.
* Trực quan hóa dữ liệu.
* Một số chủ đề nâng cao: Xây dựng mô hình dự đoán.
* Viết báo cáo.

# 

# **MỤC LỤC**

Contents

[**LỜI CẢM ƠN** 4](#_Toc185725516)

[**TÊN ĐỒ ÁN VÀ YÊU CẦU CHI TIẾT** 5](#_Toc185725517)

[**MỤC LỤC** 7](#_Toc185725518)

[**MỤC LỤC HÌNH VẼ** 9](#_Toc185725519)

[**MỤC LỤC BẢNG** 11](#_Toc185725520)

[**LỜI GIỚI THIỆU** 1](#_Toc185725521)

[**CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ THƯ VIỆN MATPLOTLIB VÀ PHÂN TÍCH THỐNG KÊ** 4](#_Toc185725522)

[**1.1. THĂM DÒ DỮ LIỆU BẰNG BIỂU ĐỒ** 4](#_Toc185725523)

[**1.2. TÓM TẮT KẾT QUẢ THEO SUY DIỄN THỐNG KÊ** 14](#_Toc185725524)

[**1.2.1. Thống kê mô tả** 15](#_Toc185725525)

[**1.2.2 Thống kê suy diễn trong các bài toán kiểm định** 15](#_Toc185725526)

[**1.2.3 Thống kê suy diễn trong các bài toán phân tích tương quan** 16](#_Toc185725527)

[**CHƯƠNG 2: PHÂN TÍCH THĂM DÒ DỮ LIỆU** 18](#_Toc185725528)

[**2.1. GIỚI THIỆU VỀ BÀI TOÁN THĂM DÒ VÀ TRỰC QUAN HÓA DỮ LI͏ỆU CHỨNG KHOÁN NHẰM PHÂN TÍCH GIAO DỊCH KHOẢN VAY CỦA BA NGÂN HÀNG MBBANK, VIETCOMBANK, SACOMBANK** 18](#_Toc185725529)

[**2.1.1. Giới thiệu** 18](#_Toc185725530)

[**2.1.2. Nguồn dữ liệu thu thập** 18](#_Toc185725531)

[**2.1.3. Bài toán** 19](#_Toc185725532)

[**2.1.4. Một số yếu tố cần quan tâm** 19](#_Toc185725533)

[**2.2. LÀM SẠCH DỮ LIỆU** 21](#_Toc185725534)

[**2.2.1. Nhập dữ liệu** 21](#_Toc185725535)

[**2.2.2. Kiểm tra dữ liệu tổng quan dữ liệu** 23](#_Toc185725536)

[**2.2.3. Chuyển đổi dữ liệu của cột** 24](#_Toc185725537)

[**2.3. THỐNG KÊ MÔ TẢ** 26](#_Toc185725538)

[**2.3.1. Trung bình ,trung vị và độ lệch chuẩn của Ngân hàng MBBank:** 28](#_Toc185725539)

[**2.3.2. Trung bình, trung vị và độ lệch chuẩn của Ngân hàng Vietcombank** 29](#_Toc185725540)

[**2.3.3.Trung bình, trung vị và độ lệch chuẩn của Ngân hàng Sacombank** 31](#_Toc185725541)

[**CHƯƠNG 3: TRỰC QUAN HÓA DỮ LIỆU** 33](#_Toc185725542)

[**3.1. BIỂU ĐỒ VỀ TRỰC QUAN DỮ LIỆU** 33](#_Toc185725543)

[**3.1.1. Biểu đồ đường so sánh giá cổ phiếu trung bình của ba ngân hàng MB Bank, Sacombank, Vietcombank** 33](#_Toc185725544)

[**3.1.2. Biểu đồ cột minh họa giá trị cho vay của MBBank, Sacombank, Vietcombank theo thời gian** 36](#_Toc185725545)

[**3.1.3. Biểu đồ hộp so sánh tỷ lệ nợ xấu giữa MBBank, Sacombank, Vietcombank** 38](#_Toc185725546)

[**3.1.4. Biểu đồ tròn thể hiện tỷ lệ đóng góp của từng ngân hàng vào tổng lợi nhuận thuần** 41](#_Toc185725547)

[**3.1.5. So sánh trung bình, trung vị đọ lệch chuẩn của MBBank với Vietcombank** 44](#_Toc185725548)

[**3.1.6. So sánh trung bình, trung vị độ lệch chuẩn của MBBank với Sacombank** 51](#_Toc185725549)

[**3.1.7. So sánh trung bình, trung vị độ lệch chuẩn của Vietcombank với Sacombank** 59](#_Toc185725550)

[**3.2. HỆ SỐ TƯƠNG QUAN VÀ BIỂU ĐỒ TƯƠNG QUAN** 66](#_Toc185725551)

[**3.2.1. Hệ số tương quan** 66](#_Toc185725552)

[**3.2.2. Biểu đồ tương quan** 67](#_Toc185725553)

[**3.3 PHÂN TÍCH PCA** 72](#_Toc185725554)

[**CHƯƠNG 4: TỔNG KẾT BÁO CÁO** 80](#_Toc185725555)

[**4.1. MỤC TIÊU VÀ PHẠM VI CỦA ĐỒ ÁN** 80](#_Toc185725556)

[**4.2. QUY TRÌNH VÀ PHƯƠNG PHÁP SỬ DỤNG** 80](#_Toc185725557)

[**4.3. ĐƯA RA KẾT QUẢ CHÍNH VÀ PHÁT HIỆN** 81](#_Toc185725558)

[**4.4. THẢO LUẬN VỀ Ý NGHĨA VÀ ỨNG DỤNG KẾT QUẢ** 82](#_Toc185725559)

[**4.5. NÊU RA CÁC HẠN CHẾ VÀ ĐỀ XUẤT HƯỚNG PHÁT TRIỂN** 82](#_Toc185725560)

[**TÀI LIỆU THAM KHẢO** 84](#_Toc185725561)

# 

# **MỤC LỤC HÌNH VẼ**

[Hình 1.1: Minh họa quy trình thêm các lớp hàm phân tích trong seaborn. 4](#_Toc185726215)

[Hình 1.2: Cách gán biến số vào các trục. 7](#_Toc185726218)

[Hình 1.3: Phối màu sắc trong biểu đồ. 11](#_Toc185726219)

[Hình 1.4 : Một số tùy chỉnh thang màu qua các hàm gradian. 13](#_Toc185726220)

[Hình 1.5: Chỉnh sửa biểu đồ đã lưu trữ. 14](#_Toc185726221)

[Hình 3.1  Biểu đồ đường so sánh giá cổ phiếu trung bình của ngân hàng MB Bank, Sacombank, Vietcombank 34](#_Toc185726244)

[Hình 3.2 Biểu đồ cột minh họa giá trị cho vay của MBbank, Sacombank, Vietcombank theo thời gian 37](#_Toc185726246)

[Hình 3.3 Biểu đồ hộp so sánh tỷ lệ nợ xấu giữa MBbank, Sacombank, Vietcombank 39](#_Toc185726248)

[Hình 3.4 Biểu đồ tròn thể hiện tỷ lệ đóng góp của từng ngân hàng vào tổng lợi nhuận thuần 42](#_Toc185726250)

[Hình 3.5. Biểu đồ so sánh trung bình các chỉ số tài chính của MBBank với Vietcombank 46](#_Toc185726253)

[Hình 3.6 Biểu đồ so sánh trung vị các chỉ số tài chính của MBBank với Vietcombank 47](#_Toc185726255)

[Hình 3.7 Biểu đồ so sánh Độ lệch chuẩn các chỉ số tài chính của MBbank Vietcombank 49](#_Toc185726257)

[Hình 3.8 Biểu đồ so sánh trung bình chỉ số tài chính của MBBank và Sacombank 53](#_Toc185726261)

[Hình 3.8 Biểu đồ so sánh trung vị chỉ số tài chính của MBBank và Sacombank 54](#_Toc185726263)

[Hình 3.9 Biểu đồ so sánh độ lệch chuẩn chỉ số tài chính của MBBank và Sacombank 55](#_Toc185726265)

[Hình 3.10 Biểu đồ so sánh trung bình chỉ số tài chính của Vietcombank và Sacombank 60](#_Toc185726269)

[Hình 3.11 Biểu đồ so sánh trung vị chỉ số tài chính của Vietcombank và Sacombank 61](#_Toc185726271)

[Hình 3.12 Biểu đồ so sánh độ lệch chuẩn chỉ số tài chính của Vietcombank và Sacombank 62](#_Toc185726273)

[Hình 3.13 Biểu đồ ma trận 69](#_Toc185726280)

[Hình 3.14 biểu đồ ma trận tương quan 71](#_Toc185726282)

[Hình 3.15 Biểu đồ trung bình trượt 30 ngày của 3 ngân hàng 76](#_Toc185726284)

[Hình 3.16. Biểu đồ phân tích PCA 78](#_Toc185726285)

# 

# **MỤC LỤC BẢNG**

[*Bảng 1.1 : Dữ liệu quan sát về các mẫu ô tô ( nguồn : ggplot2).* 4](#_heading=h.35nkun2)

[*Bảng 1.2 Những biểu đồ thường dùng trong Seaborn* 6](#_heading=h.44sinio)

[*Bảng 1.3: Các hàm tính thống kê mô tả cơ bản* 15](#_heading=h.147n2zr)

# **LỜI GIỚI THIỆU**

Trực quan hóa dữ liệu là một phương thức biểu diễn trực quan của dữ liệu. Trực quan hóa dữ liệu có thể cho biết tổng quan về dữ liệu nghiên cứu, các xu hướng, các giá trị ngoại lệ và mối tương quan trong dữ liệu có thể cho biết tổng quan về dữ liệu. Trực quan hóa dữ liệu giúp những người phân tích dễ dàng hiểu được ý nghĩa của dữ liệu, từ đó đưa ra các quyết định phù hợp trong lĩnh vực mà mình đang phân tích, quản lý. Phương thức phổ biến nhất làm cho dữ liệu trở nên trực quan là thông qua phân tích thống kê và phân tích biểu đồ.

Mục đích của trực quan hóa dữ liệu: Trực quan hóa dữ liệu đóng vai trò quan trọng trong quá trình phân tích và truyền tải thông tin. Một số mục đích chính của trực quan hóa dữ liệu đó là: hiểu dữ liệu, trình bày kết quả, ra quyết định, giao tiếp và thảo luận, tăng tương tác, giáo dục và đào tạo.

*Thứ nhất*, trực quan hóa dữ liệu giúp chúng ta hiểu sâu hơn về các mẫu, xu hướng và mối quan hệ tiềm ẩn trong dữ liệu. Khi xem xét các biểu đồ và đồ thị, chúng ta có thể nhận diện ngay lập tức các đặc điểm nổi bật mà có thể không rõ ràng khi chỉ xem xét các bảng số liệu đơn thuần. Việc khám phá này cho phép chúng ta phát hiện các ngoại lệ hoặc bất thường trong dữ liệu, kiểm tra các giả thuyết và thu thập thông tin chi tiết để làm rõ các hiện tượng phức tạp.

*Thứ hai,* một trong những mục đích chính của trực quan hóa dữ liệu là truyền tải thông tin một cách rõ ràng và ngắn gọn. Các biểu đồ và đồ thị giúp chúng ta trình bày kết quả phân tích một cách dễ hiểu, đặc biệt là khi đối tượng khán giả không có nền tảng kỹ thuật. Hình ảnh trực quan hấp dẫn không chỉ gây ấn tượng mạnh mẽ mà còn giúp thông tin dễ dàng được nhớ lâu hơn so với việc trình bày bằng các bảng số liệu khô khan.

*Thứ ba,* trực quan hóa dữ liệu đóng vai trò quan trọng trong việc hỗ trợ ra quyết định. Các biểu đồ và đồ thị cung cấp cái nhìn tổng quan và chi tiết về dữ liệu, giúp người quản lý và nhà phân tích đưa ra quyết định dựa trên dữ liệu thực tế. Khi các thông tin được trình bày một cách rõ ràng và dễ hiểu, nguy cơ sai sót trong quá trình ra quyết định được giảm thiểu, đồng thời tăng cường độ chính xác và hiệu quả của các quyết định chiến lược.

*Thứ tư,* trực quan hóa dữ liệu giúp giao tiếp thông tin một cách hiệu quả và dễ hiểu, tạo điều kiện cho các cuộc thảo luận và trao đổi thông tin. Các biểu đồ và đồ thị minh họa rõ ràng các điểm chính, giúp các bên liên quan dễ dàng hiểu và đồng thuận với các thông tin được trình bày. Điều này không chỉ giúp tăng cường sự minh bạch mà còn thúc đẩy quá trình hợp tác và thảo luận sâu hơn.

*Thứ năm,* với các công cụ trực quan hóa dữ liệu tương tác, người dùng có thể khám phá dữ liệu theo nhiều cách khác nhau, chẳng hạn như phóng to, thu nhỏ, và lọc dữ liệu. Khả năng tương tác này cho phép người dùng thực hiện các phân tích động, xem xét dữ liệu từ nhiều góc độ và tìm hiểu các chi tiết cụ thể. Trải nghiệm tương tác này không chỉ làm cho quá trình khám phá dữ liệu trở nên thú vị hơn mà còn tăng cường khả năng phát hiện những thông tin giá trị.

*Thứ sáu,* trực quan hóa dữ liệu là một công cụ mạnh mẽ trong giáo dục và đào tạo, giúp học sinh và sinh viên hiểu rõ các khái niệm phức tạp và ứng dụng thực tế của dữ liệu. Các biểu đồ và đồ thị giúp minh họa các quy trình và hệ thống phức tạp, làm cho việc giảng dạy và học tập trở nên trực quan và sinh động hơn. Điều này không chỉ giúp nâng cao hiệu quả học tập mà còn khuyến khích sự hứng thú và đam mê trong việc khám phá dữ liệu.

Để hiểu rõ hơn và thành thạo hơn trong việc trực quan hóa dữ liệu nhóm em đã quyết định lựa chọn đề tài liên quan tới khoản vay của ba ngân hàng MBBANK ,VIETCOMBANK và SACOMBANK. Trực quan hóa dữ liệu ba ngân hàng tạo nên sự minh bạch, dễ hiểu, giúp lãnh đạo so sánh hiệu quả hoạt động với đối thủ và ra quyết định nhanh chóng. Phân tích khoản vay của ba ngân hàng đóng vai trò quan trọng đối với cả nhà đầu tư lẫn các nhà lãnh đạo ngân hàng. Đối với nhà đầu tư, việc này giúp họ hiểu rõ sức khỏe tài chính của từng ngân hàng thông qua các chỉ số về chất lượng tín dụng, khả năng quản lý nợ xấu, và hiệu quả sử dụng vốn. Ngoài ra, trực quan hóa dữ liệu hỗ trợ nhà đầu tư nhận diện xu hướng vay vốn, dự đoán các cơ hội và rủi ro tiềm ẩn trên thị trường, từ đó tối ưu hóa danh mục đầu tư bằng các quyết định dựa trên thông tin trực quan và chính xác. Đối với các nhà lãnh đạo ngân hàng, phân tích này cung cấp góc nhìn toàn diện về hiệu suất danh mục cho vay, giúp phát hiện các khoản vay rủi ro và tối ưu hóa chiến lược tín dụng. Bên cạnh đó, dữ liệu thăm dò còn hỗ trợ dự đoán nhu cầu tài chính của thị trường, điều chỉnh chiến lược kinh doanh, và tăng cường quản trị rủi ro thông qua việc quản lý nợ xấu hiệu quả.

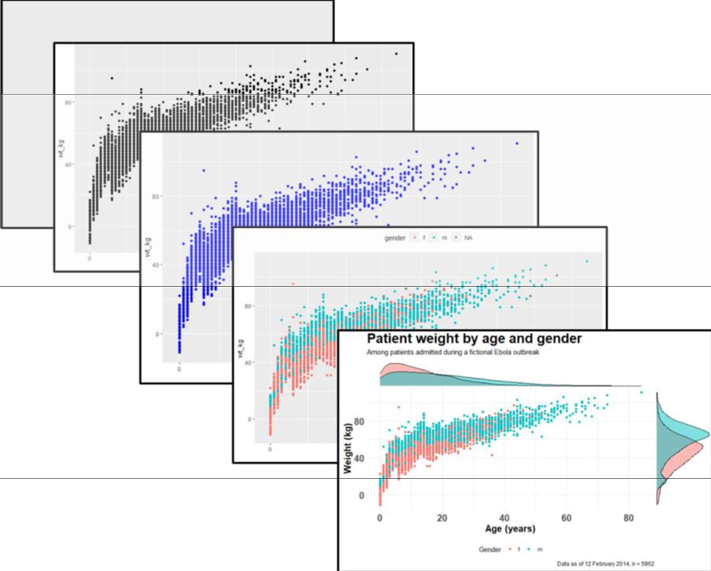
      Biết được lợi ích trong việc trực quan hóa dữ liệu thu thập được cùng với sự hướng dẫn của Giảng viên môn học “ Đồ án 1: Trực quan hóa dữ liệu bằng R” Thầy *Trần Chí Lê* để xây dựng nên đồ án “*Thăm dò và trực quan hóa dữ liệu chứng khoán nhằm phân tích giao dịch khoản vay của ba ngân hàng MBBANK ,VIETCOMBANK và SACOMBANK.”.* Để có thể thấy rõ được tỷ lệ nợ xấu của từng ngân hàng qua các quý biến động ra sao? Ngân hàng nào có tốc độ tăng trưởng cho vay tốt nhất trong cùng một khoảng thời gian?

      Để đi đến chủ đề trực quan hóa dữ liệu thu thập được chúng ta sẽ đi tìm hiểu về các bước để trực quan hóa dữ liệu thông qua tóm tắt bốn chương sau : *“Chương 1: Giới thiệu tổng quan về thư viện matplotlib và phân tích thống kê , Chương 2: Phân tích thăm dò dữ liệu , Chương 3: Trực quan hóa dữ liệu và cuối cùng được tổng hợp vào Chương 4: Tổng kết báo cáo”.*

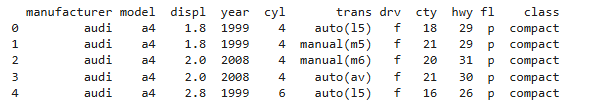
# **CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ THƯ VIỆN MATPLOTLIB VÀ PHÂN TÍCH THỐNG KÊ**

## **1.1. THĂM DÒ DỮ LIỆU BẰNG BIỂU ĐỒ**

Biểu đồ với seaborn:



#### Hình 1.1: Minh họa quy trình thêm các lớp hàm phân tích trong seaborn.

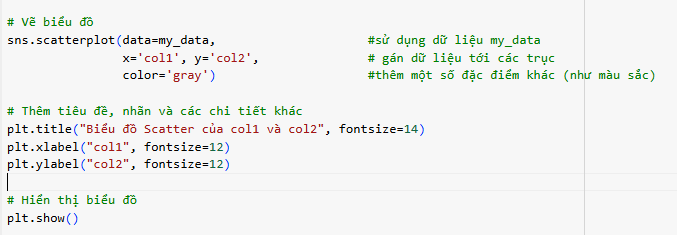
Để minh họa cho việc sử dụng searborn chúng ta sẽ làm việc trên một dữ liệu, đó là dữ liệu mpg chứa các quan sát được Cơ quan Bảo vệ Môi trường Hoa Kỳ thu thập trên 38 mẫu ô tô với 233 quan sát và 11 biến. (trong thư viện ggplot2 thuộc R gõ : ?mpg để biết chi tiết về nguồn gốc dữ liệu)

##### *Bảng 1.1 : Dữ liệu quan sát về các mẫu ô tô ( nguồn : ggplot2).*

[*Bảng 1.1 : Dữ liệu quan sát về các mẫu ô tô ( nguồn : ggplot2).*](#_heading=h.35nkun2)

*a) Cú pháp cơ bản*

Chúng ta có thể minh họa cú pháp cơ bản như sau:



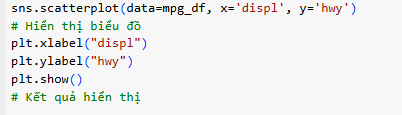
*b) Gán các biến dữ liệu cho biểu đồ*

Những biểu đồ thường dùng trong Seaborn

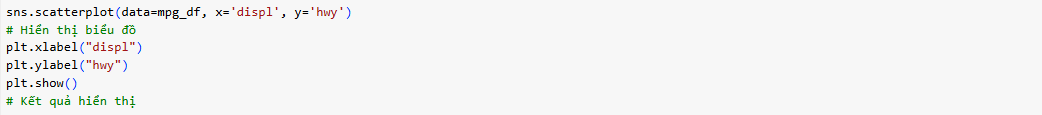
|  |  |  |
| --- | --- | --- |
| **Biểu Đồ** | **Hàm Seaborn** | **Chức năng** |
| Scatter plot | sns.scatterplot() | Biểu diễn mối quan hệ giữa hai biến số. |
| Scatter plot | sns.lineplot() | Hiển thị dữ liệu theo thời gian hoặc xu hướng. |
| Histogram | sns.histplot() | Biểu diễn phân phối dữ liệu dạng histogram. |
| Box plot | sns.boxplot() | Tóm tắt dữ liệu thông qua phân vị, xác định giá trị ngoại lai. |
| Violin plot | sns.violinplot() | Kết hợp box plot và phân phối KDE. |
| Heatmap | sns.heatmap() | Hiển thị ma trận dữ liệu (thường là tương quan) dưới dạng bản đồ nhiệt. |
| Pair plot | sns.pairplot() | Vẽ tất cả các biểu đồ scatter giữa các cặp biến trong tập dữ liệu. |
| Cat plot | sns.catplot() | Biểu đồ chuyên hiển thị dữ liệu phân loại như bar plot, box plot. |

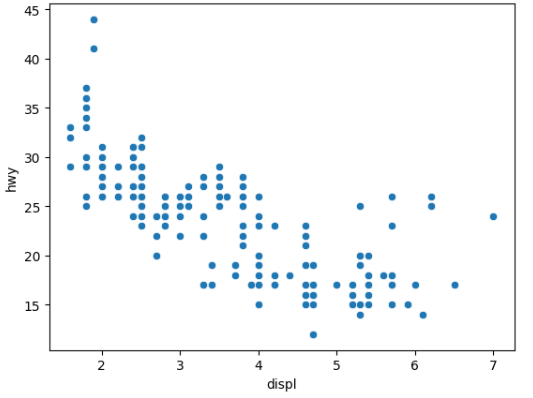
##### *Bảng 1.2 Những biểu đồ thường dùng trong Seaborn*

Hầu hết các hàm-hình để tạo biểu đồ trong thư viện seaborn phải được cho biết cái gì được sử dụng để vẽ biểu đồ, vì vậy chúng ta phải cung cấp map tránh các biến số trong dữ liệu từ các thành phần của biểu đồ như là các trục, màu đối tượng, kích thước đối tượng, v.v. Đối với hầu hết các biểu đồ, các thành phần thiết yếu phải được gán tới các cột trong dữ liệu là trục x, và (nếu cần) là trục y.



Ví dụ 1.1 Trong lệnh dưới đây, dữ liệu đc thiết lập là bộ dữ liệu mpg. Trong đó cột displ được gán cho trục x,và cột hwy được gán cho trục y.





#### Hình 1.2: Cách gán biến số vào các trục.

*c) Tính thẩm mỹ trong biểu đồ*

Tính thẩm mỹ trong biểu đồ có thể là màu sắc, kích thước, độ trong suốt, vị trí, v.v. của dữ liệu được vẽ. Không phải tất cả các biểu đồ sẽ có các tùy chọn về tính thẩm mỹ, trang trí giống nhau, nhưng một số tùy chọn được áp dụng với phần lớn các biểu đồ. Dưới đây là một số trang trí hay gặp:

* marker = Xác định hình dạng của các điểm trong biểu đồ phân tán (scatter). Có nhiều kiểu marker như o (hình tròn), ^ (tam giác), s (hình vuông), v.v.
* color = Xác định màu sắc của các đối tượng trong biểu đồ, như điểm, cột, hoặc đường. Hỗ trợ màu sắc cố định hoặc thay đổi theo biến.
* edgecolor =Xác định màu sắc của đường viền xung quanh các đối tượng như cột, điểm trong biểu đồ. Phù hợp để làm nổi bật đường viền hoặc tạo phân biệt rõ ràng hơn.
* s = Điều chỉnh kích thước của các điểm trong biểu đồ phân tán. Có thể cố định kích thước hoặc liên kết với biến dữ liệu.
* alpha = Điều chỉnh độ trong suốt của các đối tượng trong biểu đồ (giá trị từ 0 đến 1, với 1 là không trong suốt).
* binwidth = Xác định độ rộng của các bins trong biểu đồ histogram.
* width = Điều chỉnh độ rộng của các cột trong biểu đồ cột (bar chart).
* linestyle = Quy định kiểu đường (nét liền, nét đứt, chấm chấm) trong biểu đồ đường.

Trang trí của đối tượng biểu đồ cụ thể được gán giá trị theo hai cách: Gán một giá trị tĩnh (vd: color = "blue") để áp dụng cho tất cả các quan sát được vẽ biểu đồ hoặc gán cho từng biến của dữ liệu (vd: color = hospital) để hiển thị từng quan sát phụ thuộc vào giá trị của nó trong biến đó

* ***Trang trí theo giá trị của từng biến***

 Để thực hiện được điều này, chúng ta gán yếu tố trang trí của biểu đồ với một biến (không trong dấu ngoặc kép). Điều này phải được thực hiện bên trong phần vẽ biểu đồ. Việc gán được thực hiện trực tiếp thông qua các tham số như hue, size, và style.Và gán các yếu tố thẩm mỹ như hue hoặc size có thể được thực hiện một lần trong hàm chính hoặc tách ra theo từng lớp

* ***Trang trí theo nhóm đối tượng***

Trong Python, khi sử dụng thư viện Seaborn để trực quan hóa dữ liệu, cách trang trí biểu đồ phụ thuộc vào loại biểu đồ được sử dụng và mục tiêu phân nhóm dữ liệu. Chẳng hạn, với biểu đồ sns.scatterplot() (tương đương với geom\_point() trong ggplot2), các tham số như hue, style, hoặc size thường được sử dụng để phân nhóm và trang trí dữ liệu, giúp thể hiện các đặc điểm khác nhau giữa các nhóm trên cùng một biểu đồ. Trong khi đó, đối với biểu đồ cột như sns.barplot() (tương đương với geom\_bar()), tham số hue thường được dùng để phân nhóm, làm nổi bật sự khác biệt giữa các nhóm trong các thanh biểu diễn dữ liệu. Sự linh hoạt này giúp Seaborn phù hợp với nhiều nhu cầu trực quan hóa khác nhau trong phân tích dữ liệu.

*d) Gán nhãn cho biểu đồ*

Việc gán nhãn cho biểu đồ trong Seaborn (như tiêu đề, nhãn trục x/y, chú thích) thường được thực hiện bằng cách sử dụng các hàm của Matplotlib như plt.title(), plt.xlabel(), plt.ylabel(), và plt.figtext(). Sau khi tạo biểu đồ với Seaborn, bạn có thể sử dụng các hàm này để gán nhãn:

* plt.xlabel() và plt.ylabel(): Đặt nhãn cho trục x và trục y.
* plt.title(): Đặt tiêu đề chính cho biểu đồ.
* plt.suptitle(): Thêm tiêu đề phụ (subtitle) nhỏ hơn và thường nằm bên dưới tiêu đề chính.
* plt.figtext(): Thêm chú thích (caption) cho biểu đồ, thường đặt ở góc dưới cùng của hình vẽ.

*e) Căn chỉnh trong biểu đồ*

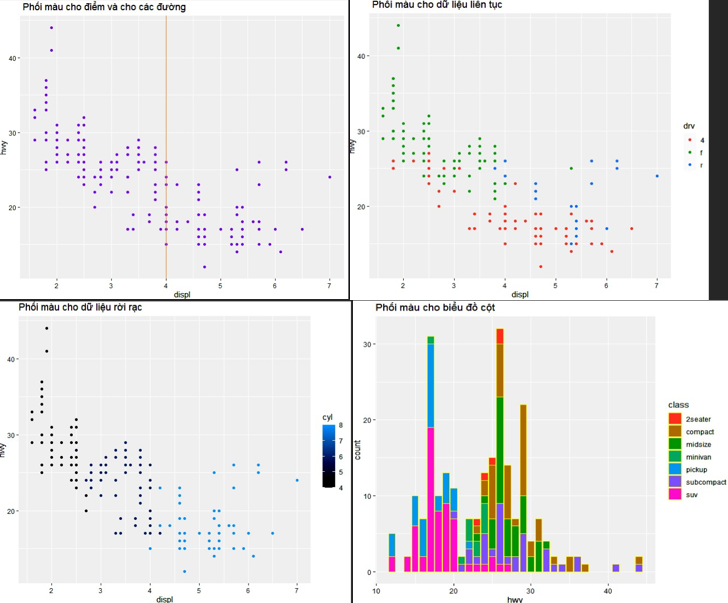
Trong Python, khi trực quan hóa dữ liệu với các thư viện như Matplotlib và Seaborn, việc căn chỉnh biểu đồ có thể thực hiện theo nhiều cách khác nhau, tùy thuộc vào mục đích trình bày và bố cục. Dưới đây là các dạng căn chỉnh phổ biến và cách thực hiện:

* **Căn chỉnh tiêu đề và nhãn**Trong Python, việc căn chỉnh tiêu đề và nhãn của biểu đồ được thực hiện thông qua các hàm như plt.title() để điều chỉnh tiêu đề chính, và plt.xlabel() hoặc plt.ylabel() để gán và căn chỉnh nhãn cho các trục. Bạn có thể tùy chỉnh vị trí của tiêu đề bằng tham số loc với các giá trị như "left", "center", hoặc "right". Ngoài ra, tham số pad cho phép thay đổi khoảng cách giữa tiêu đề hoặc nhãn và biểu đồ.
* **Căn chỉnh chú thích (Legend)**Chú thích (legend) giúp phân biệt các nhóm dữ liệu trên biểu đồ, và bạn có thể căn chỉnh vị trí của chú thích bằng hàm plt.legend() hoặc ax.legend(). Tham số loc giúp định vị chú thích tại các vị trí cụ thể như "best", "upper left", hoặc "lower right". Nếu cần, bạn có thể sử dụng bbox\_to\_anchor để điều chỉnh chi tiết vị trí của chú thích, hoặc dùng ncol để thiết lập số cột trong chú thích.
* **Căn chỉnh lề và không gian giữa các biểu đồ**Để tạo bố cục biểu đồ cân đối và gọn gàng, bạn có thể sử dụng plt.tight\_layout() để tự động căn chỉnh các khoảng trống giữa các phần tử trên biểu đồ. Đối với biểu đồ con (subplots), hàm fig.subplots\_adjust() cho phép tùy chỉnh khoảng cách dọc (hspace) và ngang (wspace) giữa các biểu đồ. Ngoài ra, bạn có thể điều chỉnh lề của toàn bộ hình vẽ bằng các tham số left, right, top, hoặc bottom.
* **Căn chỉnh kích thước và tỷ lệ biểu đồ**Kích thước và tỷ lệ biểu đồ có thể được điều chỉnh bằng cách sử dụng plt.figure() với tham số figsize, hoặc sử dụng hàm sns.set\_context() của Seaborn để thay đổi tỷ lệ theo các ngữ cảnh khác nhau như "paper", "notebook", "talk", và "poster". Việc điều chỉnh này giúp biểu đồ dễ nhìn hơn trong các tài liệu hoặc bài thuyết trình.
* **Căn chỉnh lưới và trục**Lưới (grid) giúp biểu đồ dễ đọc hơn, và bạn có thể bật hoặc tùy chỉnh lưới bằng plt.grid(). Hàm này hỗ trợ các tham số như axis để chọn trục lưới ("x", "y", "both") hoặc alpha để điều chỉnh độ trong suốt. Ngoài ra, bạn có thể dùng plt.xlim() và plt.ylim() để giới hạn phạm vi dữ liệu hiển thị trên các trục.
* **Căn chỉnh nhiều biểu đồ (Subplots)**Khi cần hiển thị nhiều biểu đồ trên cùng một hình, bạn có thể sử dụng plt.subplots() để tạo bố cục lưới các biểu đồ con. Tham số gridspec\_kw cho phép điều chỉnh tỷ lệ giữa các hàng và cột, trong khi các tham số hspace và wspace giúp kiểm soát khoảng cách dọc và ngang giữa các biểu đồ con. Bạn cũng có thể sử dụng sharex hoặc sharey để chia sẻ cùng trục x hoặc y giữa các biểu đồ.
* **Căn chỉnh tỉ lệ khung hình**Để điều chỉnh tỷ lệ khung hình của biểu đồ, bạn có thể dùng ax.set\_aspect(). Giá trị "equal" giúp đảm bảo tỷ lệ giữa trục x và trục y là bằng nhau, hoặc bạn có thể cung cấp một giá trị số cụ thể để tùy chỉnh tỷ lệ.

*f) Phối màu sắc, tô màu, thang đo*

* ***Phối màu***

Trong Python, khi sử dụng thư viện Seaborn hoặc Matplotlib để trực quan hóa dữ liệu, màu sắc của các đối tượng biểu đồ (như điểm, cột, đường, ô,...) có thể được điều chỉnh thông qua các tham số color (màu bên ngoài) hoặc fill (màu bên trong). Với biểu đồ dạng điểm, chẳng hạn như sns.scatterplot(), bạn có thể dùng tham số hue để thiết lập màu dựa trên một biến phân nhóm. Đối với các biểu đồ khác như sns.barplot() hay sns.heatmap(), tham số palette cho phép sử dụng bảng màu tùy chỉnh. Các bảng màu có thể được đặt bằng tên (ví dụ: "Blues", "coolwarm", hoặc "viridis"), mã màu HEX, hoặc danh sách màu cụ thể.



#### Hình 1.3: Phối màu sắc trong biểu đồ.

* ***Thang đo cho yếu tố trang trí (thẩm mỹ)***

Khi gán một biến cho yếu tố thẩm mỹ trên biểu đồ (như x, y, hue, hoặc size), biểu đồ sẽ tự động tạo thang đo hoặc chú giải phù hợp, hiển thị các giá trị liên tục, rời rạc hoặc ngày tháng tùy thuộc vào kiểu dữ liệu. Bạn có thể kiểm soát thang đo bằng cách sử dụng các hàm thích hợp. Trong Seaborn, các yếu tố thẩm mỹ như màu sắc và kích thước được điều chỉnh qua các tham số như hue, size, và các bảng màu thông qua palette hoặc cmap (cho gradient liên tục).

* ***Điều chỉnh thủ công***

Trong Python, chúng ta có thể gán màu thủ công cho các giá trị cụ thể bằng cách tạo một từ điển ánh xạ và sử dụng tham số hue hoặc palette. Để điều chỉnh nhãn hiển thị, bạn có thể sử dụng các tham số legend hoặc chỉnh sửa thủ công chú thích bằng Matplotlib.

* ***Thang đo trên các trục***

Đối với dữ liệu liên tục, bạn có thể điều chỉnh các ngắt giá trị trên trục (breaks) bằng Matplotlib với plt.xticks() hoặc plt.yticks().

* ***Hiển thị phần trăm trên trục***

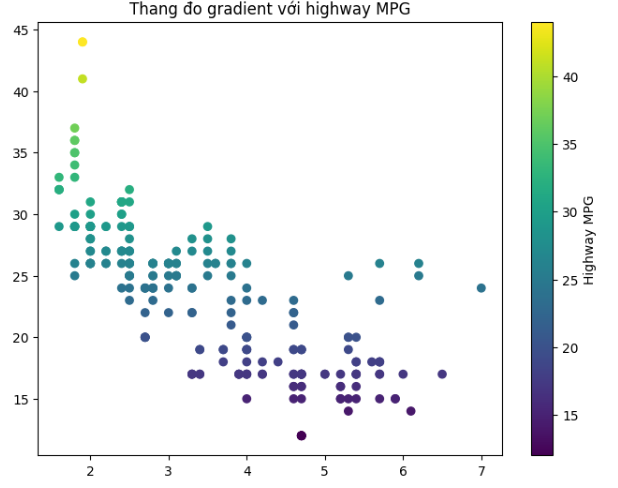
Đối với dữ liệu tỷ lệ phần trăm, bạn có thể chuyển đổi giá trị bằng cách sử dụng định dạng hoặc thêm ký tự %.

* ***Thang đo log***

Khi dữ liệu có khoảng giá trị rộng, bạn có thể chuyển đổi thang đo của trục sang log bằng plt.yscale() hoặc plt.xscale() với tham số "log". Điều này giúp biểu đồ hiển thị dữ liệu rõ ràng hơn.

* ***Thang đo Gradient***

Để tô màu theo thang đo gradient, bạn có thể sử dụng các tham số như cmap trong Matplotlib hoặc Seaborn. Hàm sns.heatmap() hỗ trợ tùy chỉnh gradient với tham số cmap, trong khi Matplotlib cung cấp hàm plt.scatter() với tham số c. Bạn cũng có thể điều chỉnh điểm ngắt màu bằng Normalize() từ thư viện matplotlib.colors hoặc sử dụng hàm np.interp() để nội suy giá trị.



#### Hình 1.4 : Một số tùy chỉnh thang màu qua các hàm gradian.

*g)Lưu trữ, chỉnh sửa và xuất biểu đồ*

* ***Lưu biểu đồ***

Trong Python, biểu đồ được tạo bằng thư viện Matplotlib hoặc Seaborn có thể được lưu trữ dưới dạng đối tượng để chỉnh sửa hoặc tái sử dụng sau này. Chúng ta gán biểu đồ vào một biến bằng cách tạo đối tượng Figure hoặc Axes thông qua hàm plt.subplots() hoặc trực tiếp từ biểu đồ Seaborn.

*Ví dụ 1.16.* lệnh lưu biểu đồ hình 1.1 vào biến có tên ax

# Tạo biểu đồ và lưu vào biến

fig, ax = plt.subplots()

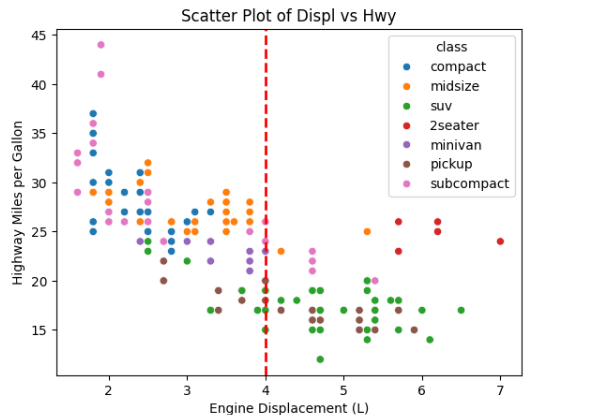
sns.scatterplot(data=mpg, x='displ', y='hwy', hue='class', ax=ax)

# Hiển thị biểu đồ

plt.show()

* ***Chỉnh sửa biều đồ đã lưu***

***Ví dụ 1.17.*** Ta chỉnh sửa đổi biểu đồ ax đã được lưu ở Ví dí 1.16 ở bên trên, thêm một trục dọc tại dung tích động cơ bằng 4, chúng ta chỉ cần thêm hoặc chỉnh sửa các thành phần của biểu đồ bằng cách gọi biến đã lưu trữ.



#### Hình 1.5: Chỉnh sửa biểu đồ đã lưu trữ.

* ***Xuất bản biểu đồ***

Việc xuất bản biểu đồ được thực hiện dễ dàng với hàm plt.show()

Chúng ta có thể xuất biểu đồ ra dưới nhiều định dạng như PNG, JPEG, PDF, hoặc SVG bằng cách sử dụng hàm savefig() của Matplotlib. Các tham số như kích thước, độ phân giải (dpi), và đường dẫn lưu trữ đều có thể được tùy chỉnh.

## **1.2. TÓM TẮT KẾT QUẢ THEO SUY DIỄN THỐNG KÊ**

Tóm tắt các kết quả theo suy diễn thống kê như các tính toán về đặc trưng của dữ liệu mẫu, các bài toán ước lượng, bài toán kiểm định tham số, bài toán phân tích hệ số tương quan để tạo thành các module phân tích thống kê. Những module này kết với phân tích dữ liệu qua biểu đồ sẽ cho kết quả trực quan dữ liệu chính xác hơn. Ngoài ra, trong phần này các ví dụ minh họa sử dụng file dữ liệu Diem\_TN, xem [1]

### **1.2.1. Thống kê mô tả**

Cho một biến số , , ,…., chúng ta có thể tính toán một số chỉ số thống kê mô tả như sau:

|  |  |
| --- | --- |
| Lý thuyết | Hàm |
|  | mean(x) |
|  | *var(x)* |
| *Độ lệch chuẩn: s=* | sqrt (x) |
| *Trị số thấp nhất* | min(x) |
| *Trị số cao nhất* | max(x) |

##### *Bảng 1.3: Các hàm tính thống kê mô tả cơ bản*

### **1.2.2 Thống kê suy diễn trong các bài toán kiểm định**

*a)  Trị số P-value*

Trong nghiên cứu khoa học, ngoài những dữ kiện bằng số, biểu đồ và hình ảnh, con số mà chúng ta thường hay gặp nhất là trị số P (P-value). Do đó, trước khi nói đến các phương pháp phân tích thống kê, chúng ta cùng tìm hiểu về ý nghĩa của trị số này. Một giả thiết được xem là mang tính “khoa học” nếu giả thiết đó có khả năng “phản nghiệm”. Theo Karl Popper, nhà triết học khoa học, đặc điểm duy nhất để có thể phân biệt giữa một lý thuyết khoa học thực thụ với ngụy khoa học (pseudoscience) là thuyết khoa học luôn có đặc tính có thể “bị bác bỏ” (hay bị phản bác – falsified) bằng những thực nghiệm đơn giản. Ông gọi đó là “khả năng phản nghiệm”. Phép phản nghiệm là phương cách tiến hành những thực nghiệm không phải để xác minh mà để phê phán các lý thuyết khoa học và có thể coi đây như là một nền tảng cho khoa học thực thụ.

Vì thế, giá trị P có nghĩa là xác suất của dữ kiện D xảy ra nếu giả thuyết đảo H0 là sự thật. Như vậy, giá trị P không trực tiếp cho chúng ta một ý niệm gì về sự thật của giả thuyết chính ; nó chỉ gián tiếp cung cấp bằng chứng để chúng ta chấp nhận giả thuyết chính và bác bỏ giả thuyết đảo.

*b)  Các loại sai lầm trong kiểm định giả thuyết*

Sai lầm loại I: Nếu ta bác bỏ H0 khi H0 đúng thì sai lầm đó gọi là sai lầm loại I.

Sai lầm loại II: Nếu H0 sai mà ta không bác bỏ H0 thì sai lầm đó gọi là sai lầm loại II.

### **1.2.3 Thống kê suy diễn trong các bài toán phân tích tương quan**

Hệ số tương quan (r) là một chỉ số thống kê đo lường mối liên hệ tương quan giữa hai biến số. Hệ số tương quan có giá trị từ -1 đến 1. Hệ số tương quan bằng 0 (hay gần 0) có nghĩa là hai biến số không có liên hệ gì với nhau; ngược lại nếu hệ số bằng -1 hay 1 có nghĩa là hai biến số có một mối liên hệ tuyệt đối. Nếu giá trị của hệ số tương quan là âm (r < 0) có nghĩa là hai biến tương quan nghịch (biến này tăng thì biến kia giảm và ngược lại); nếu giá trị hệ số tương quan là dương (r > 0) có nghĩa là hai biến tương quan thuận (hai biến cùng tăng hoặc cùng giảm).

Có nhiều hệ số tương quan trong thống kê, nhưng ở đây chúng ta sẽ trình bày 3 hệ số tương quan thông dụng nhất: hệ số tương quan Pearson r, Spearman ρ, và Kendall τ. Trong tiểu mục này dữ liệu dùng để minh họa là file dữ liệu markettimng.csv tham khảo từ link:

https://drive.google.com/drive/folders/1maNUAWyCcjXrU0m6hMgZNhjEI0jUI9Gu

*a)   H͏ số tương quan mẫu*

* Hệ số tương quan Pearson

Cho hai biến số x và y từ n mẫu, hệ số tương quan Pearson được tính bằng công thức sau đây: Trong đó, x và y là giá trị trung bình của biến số x và y. Để tính hệ số tương quan Pearson trong R, cú pháp như sau:

cor(data, method = "pearson")

* Hệ số tương quan Spearman ρ

Hệ số tương quan Pearson chỉ hợp lý nếu biến số x và y tuân theo luật phân phối chuẩn. Nếu x và y không tuân theo luật phân phối chuẩn, chúng ta phải sử dụng một hệ số tương quan khác tên là Spearman, một phương pháp phân tích phi tham số. Hệ số này được ước tính bằng cách biến đổi hai biến số x và y thành thứ bậc (rank), và xem độ tương quan giữa hai dãy số bậc. Do đó, hệ số còn có tên tiếng Anh là Spearman’s Rank correlation.

Để tính hệ số tương quan spearman trong R, cú pháp như sau:

​​cor(data, method = " spearman ")

* Hệ số tương quan Kendall τ

Hệ số tương quan Kendall (cũng là một phương pháp phân tích phi tham số) được ước tính bằng cách tìm các cặp số (x, y) ”song hành” với nhau. Một cặp (x, y) song hành ở đây được định nghĩa là hiệu (độ khác biệt) trên trục hoành có cùng dấu hiệu (dương hay âm) với hiệu trên trục tung. Nếu hai biến số x và y không có liên hệ với nhau, thì cặp số song hành bằng hay tương đương với cặp số không song hành.

Vì có nhiều cặp phải kiểm định, phương pháp tính toán hệ số tương quan Kendall đòi hỏi thời gian của máy tính khá cao. Tuy nhiên, nếu một dữ liệu dưới 5000 đối tượng thì một máy vi tính có thể tính toán khá dễ dàng.

      Để tính hệ số tương quan Kendall trong R, cú pháp như sau:

cor(data, method = " kendall ")

*b)   Kiểm định h͏ệ số tương quan*

Bên cạnh việc tính các giá trị tương quan mẫu, chúng ta cũng có thể kiểm định hệ số tương quan lý thuyết với giả thuyết kiểm định:

* : Không có tương quan (hệ số tương quan = 0).
* 1: Có tương quan.

# **CHƯƠNG 2: PHÂN TÍCH THĂM DÒ DỮ LIỆU**

## **2.1. GIỚI THIỆU VỀ BÀI TOÁN THĂM DÒ VÀ TRỰC QUAN HÓA DỮ LI͏ỆU CHỨNG KHOÁN NHẰM PHÂN TÍCH GIAO DỊCH KHOẢN VAY CỦA BA NGÂN HÀNG MBBANK, VIETCOMBANK, SACOMBANK**

### **2.1.1. Giới thiệu**

**Chứng khoán giao dịch** (còn gọi là giao dịch chứng khoán) là một hình thức đầu tư tài chính, nơi các nhà đầu tư mua và bán các sản phẩm chứng khoán (như cổ phiếu, trái phiếu, quỹ, hoặc các công cụ tài chính khác) trên thị trường chứng khoán. Đây là một phần quan trọng trong hệ thống tài chính toàn cầu, giúp huy động vốn và tăng tính thanh khoản cho các doanh nghiệp. Từ chứng khoán có thể phản ánh hiệu quả và mức độ tín nhiệm của các ngân hàng như **MBBank**, **Vietcombank**, và **Sacombank.**

Trên nền tảng của sự hấp dẫn và sự đa dạng của dữ liệu, Nhóm 1 chúng tôi đã khám phá bộ dữ liệu về khoản vay của ba ngân hàng MBBANK, VIETCOMBANK và SACOMBANK mà chúng tôi thu thập đc. Bộ dữ liệu này bao gồm các thông tin đa dạng về Giá cổ phiếu trung bình(VND), Cho vay (Tỷ đồng), Nợ xấu (%), Lợi nhuận thuần (Tỷ đồng), Tổng tài sản (Tỷ đồng), Quý, Ngân hàng.

### **2.1.2. Nguồn dữ liệu thu thập**

       o   Tên bộ dữ liệu : “bang\_thong\_ke,csv”.

       o  Kỹ thuật thu thập dữ liệu: Tìm các giá trị chứng khoán trên các trang web. Tải các dữ liệu chứng khoán về tính giá trị trung bình và tạo thành cột giá cổ phiếu trung bình. Tìm các giá trị nợ xấu trên các trang các trang web chính thức của 3 ngân hàng. Các giá trị khác như cho vay, lợi nhuận thuần và tổng tài sản thu thập bằng thiết lập thời gian theo quý trên web rồi tải dữ liệu về tạo thành các cột cho vay, lợi nhuận thuần và tổng tài sản thu thập.

       o   Nguồn dữ liệu :

Trên các trang web của các sàn giao dịch:

Sở Giao dịch Chứng khoán TP.HCM (HOSE): [hsx.vn](https://hsx.vn)

Sở Giao dịch Chứng khoán Hà Nội (HNX): [hnx.vn](https://hnx.vn)

Trên các cổng thông tin tài chính:CafeF và Vietstock,TradingView

Trên API tài chính:Yahoo Finance API

Trên trang web chính thức của 3 ngân hàng MBB,VCB, SCB

### **2.1.3. Bài toán**

     Kết quả về sự tác động của các tác nhân như : Cho vay (Tỷ đồng), Nợ xấu (%), Lợi nhuận thuần (Tỷ đồng), Tổng tài sản (Tỷ đồng) … có ảnh hưởng tới “Giá cổ phiếu trung bình ” như thế nào ?

### **2.1.4. Một số yếu tố cần quan tâm**

*a)   Tập dữ liệu này gồm các biến:*

* Ngân hàng : Biểu thị tên ba ngân hàng.
* Quý: Biểu thị thời gian.
* Giá cổ phiếu trung bình: Biểu thị giá trị của cổ phiếu trên thị trường của 3 ngân hàng theo quý cho biết tình hình hoạt động của các ngân hàng.
* Cho vay (Tỷ đồng): Tổng số tiền cho vay của từng ngân hàng ngân hàng theo quý cho biết tình hình nguồn vốn của các ngân hàng.
* Nợ xấu (%): Tỷ lệ nợ xấu trên tổng khoản vay cho biết tình hình hoạt động cho vay của các ngân hàng.
* Lợi nhuận thuần (Tỷ đồng): Lợi nhuận sau khi đã trừ các chi phí hoạt động cho biết lợi nhuận của các ngân hàng.
* Tổng tài sản (Tỷ đồng): Tổng giá trị tài sản mà ngân hàng nắm giữ cho biết giá trị của các ngân hàng.

## 

* 1. *Tìm ra mối liên hệ giữa các biến:*

-Mối liên hệ giữa giá cổ phiếu trung bình và các biến khác

Giá cổ phiếu trung bình phản ánh kỳ vọng của thị trường về hiệu quả hoạt động và giá trị của ngân hàng. Khi tổng số tiền cho vay tăng trưởng (với điều kiện tỷ lệ nợ xấu được kiểm soát tốt), giá cổ phiếu thường có xu hướng tăng nhờ doanh thu từ lãi vay tăng lên. Đồng thời, tỷ lệ nợ xấu thấp làm tăng niềm tin của nhà đầu tư, từ đó thúc đẩy giá cổ phiếu. Lợi nhuận thuần cao cũng là một yếu tố quan trọng, vì nó cho thấy khả năng tạo ra giá trị cho cổ đông, góp phần nâng cao giá cổ phiếu. Bên cạnh đó, tổng tài sản lớn hơn cho thấy ngân hàng có nền tảng tài chính vững chắc, điều này cũng giúp củng cố giá trị cổ phiếu trên thị trường.

-Mối liên hệ giữa cho vay và các biến khác

Cho vay là yếu tố quyết định nguồn thu chính của ngân hàng và ảnh hưởng lớn đến các chỉ số tài chính. Khi tỷ lệ nợ xấu tăng cao, hiệu quả từ hoạt động cho vay bị suy giảm, đồng thời chi phí dự phòng rủi ro tăng lên, dẫn đến lợi nhuận giảm. Ngược lại, khi tổng số tiền cho vay tăng trưởng với tỷ lệ nợ xấu thấp, doanh thu từ lãi vay sẽ tăng lên, cải thiện lợi nhuận thuần của ngân hàng. Hơn nữa, cho vay chiếm tỷ trọng lớn trong tài sản ngân hàng, vì vậy khi tổng số tiền cho vay tăng, tổng tài sản của ngân hàng cũng tăng theo.

**-**Mối liên hệ giữa nợ xấu và các biến khác

Nợ xấu là một chỉ số quan trọng để đánh giá chất lượng tín dụng và rủi ro của ngân hàng. Khi tỷ lệ nợ xấu tăng cao, ngân hàng phải dự phòng rủi ro tín dụng nhiều hơn, làm giảm lợi nhuận thuần. Điều này cũng kéo theo sự sụt giảm niềm tin từ nhà đầu tư, dẫn đến giá cổ phiếu giảm. Ngoài ra, tỷ lệ nợ xấu cao có thể ảnh hưởng tiêu cực đến tổng tài sản, bởi các khoản vay khó thu hồi làm giảm giá trị thực của tài sản.

-Mối liên hệ giữa lợi nhuận thuần và các biến khác

Lợi nhuận thuần là thước đo cuối cùng phản ánh hiệu quả hoạt động của ngân hàng. Lợi nhuận cao thường làm tăng giá cổ phiếu, bởi nó cho thấy tiềm năng chia cổ tức và khả năng tăng trưởng trong tương lai. Phần lớn lợi nhuận của ngân hàng đến từ hoạt động cho vay, do đó tăng trưởng tín dụng với tỷ lệ nợ xấu thấp sẽ kéo theo lợi nhuận tăng. Ngoài ra, ngân hàng với quy mô tài sản lớn thường có khả năng tạo ra lợi nhuận cao hơn, mặc dù điều này cũng phụ thuộc vào hiệu quả quản lý tài sản.

-Mối liên hệ giữa tổng tài sản và các biến khác

Tổng tài sản thể hiện quy mô và tiềm lực tài chính của ngân hàng, ảnh hưởng đến nhiều biến số khác. Cho vay thường là thành phần chính của tài sản, vì vậy tổng tài sản và tổng số tiền cho vay thường tăng song hành. Một ngân hàng có quy mô tài sản lớn và được quản trị tốt sẽ thu hút nhà đầu tư, từ đó đẩy giá cổ phiếu tăng. Quy mô lớn cũng giúp ngân hàng tận dụng lợi thế về kinh tế, nâng cao lợi nhuận thuần. Tuy nhiên, nếu tổng tài sản lớn nhưng đi kèm với tỷ lệ nợ xấu cao, giá trị thực của tài sản có thể bị suy giảm, dẫn đến tác động tiêu cực đến các chỉ số tài chính khác.

## **2.2. LÀM SẠCH DỮ LIỆU**

Trước khi có thể tiến hành trực quan hóa với dữ liệu thì bước quan trọng là làm sạch dữ liệu. Làm sạch dữ liệu là quá trình điều chỉnh hoặc loại bỏ dữ liệu không chính xác, sai định dạng, trùng lặp\, không đầy đủ hoặc có vấn đề trong bộ dữ liệu.

Quy trình làm sạch dữ liệu bao gồm:

* Nhập dữ liệu.
* Làm sạch hoặc đổi tên cột.
* Kiểm tra tổng quan dữ liệu
* Chuyển đổi dữ liệu cột

### **2.2.1. Nhập dữ liệu**

Bước đầu tiên của quy trình làm sạch dữ liệu là nhập dữ liệu. Để đọc được dữ liệu ta dùng hàm “read\_csv()”.

# Đoc file csv

df = pd.read\_csv('bang\_thong\_ke.csv')

df['Ngân Hàng'] = df['Ngân Hàng'].fillna(method='ffill')

# Hiển thị file

print(df)

#kết quả hiển thị:

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

0 Mbank 3\_2023 $17,950 $523,969,217

1 Mbank 4\_2023 $18,052 $599,579,267

2 Mbank 1\_2024 $22,950 $603,072,394

3 Mbank 2\_2024 $22,970 $622,584,515

4 Sacombank 3\_2023 $28,500 $165,403,139

5 Sacombank 4\_2023 $29,000 $475,169,108

6 Sacombank 1\_2024 $30,510 $492,140,463

7 Sacombank 2\_2024 $29,000 $507,883,143

8 Vietcombank 3\_2023 $80,000 $1,150,958,460

9 Vietcombank 4\_2023 $80,300 $1,241,675,333

10 Vietcombank 1\_2024 $91,830 $1,235,587,240

11 Vietcombank 2\_2024 $90,370 $1,334,761,763

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

0 1,09% $9,811,753 $815,880,535

1 1,6% $9,163,492 $944,953,640

2 2,49% $9,162,293 $900,697,141

3 1,5% $10,531,113 $988,604,894

4 1,22% $4,851,108 $651,288,299

5 2,88% $5,633,088 $674,389,660

6 2,15% $5,950,840 $693,534,553

7 2,43% $6,116,566 $717,312,771

8 1,21% $12,596,118 $1,731,305,572

9 0,98% $12,801,224 $1,839,223,339

10 1,22% $14,078,056 $1,773,713,584

11 1,2% $13,907,711 $1,905,638,845

<ipython-input-4-d441c1e9eb62>:7: FutureWarning: Series.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.

Tập dữ liệu bang\_thong\_ke.csv cung cấp cái nhìn tổng quan về các yếu tố về khoản vay ảnh hưởng đến cổ phiếu với 12 quan sát và 7 biến. Bao gồm các biến: Cho vay (Tỷ đồng), Nợ xấu (%), Lợi nhuận thuần (Tỷ đồng), Tổng tài sản (Tỷ đồng), Ngân hàng, Quý, Giá cổ phiếu trung bình(VND).

### **2.2.2. Kiểm tra dữ liệu tổng quan dữ liệu**

*a)Kiểm tra trùng lặp của dữ liệu*

# Kiểm tra các dòng trùng lặp

print('Các dòng trùng lặp:')

print(df.duplicated().sum())

#kết quả hiển thị

Các dòng trùng lặp:

0

Trong Python, hàm 'Duplicated().sum()' được sử dụng để tìm phần tử trùng lặp đầu tiên trong một đối tượng (vector, list, data frame, v.v.). Kết quả trả về của hàm là vị trí của phần tử trùng lặp đầu tiên nếu có, hoặc giá trị 0 nếu không có phần tử trùng lặp nào. Kết quả cho thấy không có dữ liệu bị trùng lặp.

*b)Kiểm tra dữ liệu null,na*

# Kiểm tra các giá trị bị thiếu

print(df.isnull().sum())

#kết quả hiển thị

Ngân Hàng 0

Quý 0

Giá cổ phiếu trung bình(VNĐ) 0

Cho vay (Tỷ đồng) 0

Nợ xấu (%) 0

Lợi nhuận thuần (Tỷ đồng) 0

Tổng tài sản (Tỷ đồng) 0

dtype: int64

Tất cả các cột đều không có giá trị NULL. Có 11 dòng trong data frame của bạn không có bất kỳ giá trị NA nào, tức là chúng có đầy đủ dữ liệu cho tất cả các cột.

*c) Cấu trúc của dữ liệu*

Hàm “str()” hiển thị cấu trúc của một đối tượng , bao gồm kiểu dữ liệu của từng thành phần và một bản tóm tắt ngắn gọn về nội dung của từng thành phần.

str(df)

#kết quả hiển thị

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

0 Mbank 3\_2023 $17,950 $523,969,217

1 Mbank 4\_2023 $18,052 $599,579,267

2 Mbank 1\_2024 $22,950 $603,072,394

3 Mbank 2\_2024 $22,970 $622,584,515

4 Sacombank 3\_2023 $28,500 $165,403,139

5 Sacombank 4\_2023 $29,000 $475,169,108

6 Sacombank 1\_2024 $30,510 $492,140,463

7 Sacombank 2\_2024 $29,000 $507,883,143

8 Vietcombank 3\_2023 $80,000 $1,150,958,460

9 Vietcombank 4\_2023 $80,300 $1,241,675,333

10 Vietcombank 1\_2024 $91,830 $1,235,587,240

11 Vietcombank 2\_2024 $90,370 $1,334,761,763

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

0 1,09% $9,811,753 $815,880,535

1 1,6% $9,163,492 $944,953,640

2 2,49% $9,162,293 $900,697,141

3 1,5% $10,531,113 $988,604,894

4 1,22% $4,851,108 $651,288,299

5 2,88% $5,633,088 $674,389,660

6 2,15% $5,950,840 $693,534,553

7 2,43% $6,116,566 $717,312,771

8 1,21% $12,596,118 $1,731,305,572

9 0,98% $12,801,224 $1,839,223,339

10 1,22% $14,078,056 $1,773,713,584

11 1,2% $13,907,711 $1,905,638,845

Kiểu dữ liệu không nhất quán: các cột đang còn đang hiện đơn vị (như Cho vay (Tỷ đồng), Nợ xấu (%), Lợi nhuận thuần (Tỷ đồng), Tổng tài sản (Tỷ đồng), Giá cổ phiếu trung bình(VND). Vì vậy cần chuyển đổi dữ liệu các cột cho phù hợp.

### **2.2.3. Chuyển đổi dữ liệu của cột**

# Loại bỏ khoảng trắng dư thừa ở đầu và cuối tên cột (nếu có)

df.columns = df.columns.str.strip()

# Chuyển đổi các cột số thành kiểu dữ liệu phù hợp

# Chuyển đổi cột 'Giá cổ phiếu trung bình(VNĐ)'

df['Giá cổ phiếu trung bình(VNĐ)'] = df['Giá cổ phiếu trung bình(VNĐ)'].replace('[\$,]', '', regex=True).astype(float)

# Chuyển đổi cột 'Cho vay (Tỷ đồng)'

df['Cho vay (Tỷ đồng)'] = df['Cho vay (Tỷ đồng)'].replace('[\$,]', '', regex=True).astype(float)

# Chuyển đổi cột 'Nợ xấu (%)'

df['Nợ xấu (%)'] = df['Nợ xấu (%)'].fillna('').astype(str).str.replace('%', '', regex=True).str.replace(',', '.', regex=True).astype(float)

#Chuyển đổi cột' Lợi nhuận thuần (Tỷ đồng)'

df['Lợi nhuận thuần (Tỷ đồng)'] = df['Lợi nhuận thuần (Tỷ đồng)'].replace('[\$,]', '', regex=True).astype(float)

#Chuyển đổi cột' Tổng tài sản (Tỷ đồng)'

df['Tổng tài sản (Tỷ đồng)'] = df['Tổng tài sản (Tỷ đồng)'].replace('[\$,]', '', regex=True).astype(float)

# Kết quả hiển thị

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

0 Mbank 3\_2023 17950.0 5.239692e+08

1 Mbank 4\_2023 18052.0 5.995793e+08

2 Mbank 1\_2024 22950.0 6.030724e+08

3 Mbank 2\_2024 22970.0 6.225845e+08

4 Sacombank 3\_2023 28500.0 1.654031e+08

5 Sacombank 4\_2023 29000.0 4.751691e+08

6 Sacombank 1\_2024 30510.0 4.921405e+08

7 Sacombank 2\_2024 29000.0 5.078831e+08

8 Vietcombank 3\_2023 80000.0 1.150958e+09

9 Vietcombank 4\_2023 80300.0 1.241675e+09

10 Vietcombank 1\_2024 91830.0 1.235587e+09

11 Vietcombank 2\_2024 90370.0 1.334762e+09

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

0 1.09 9811753.0 8.158805e+08

1 1.60 9163492.0 9.449536e+08

2 2.49 9162293.0 9.006971e+08

3 1.50 10531113.0 9.886049e+08

4 1.22 4851108.0 6.512883e+08

5 2.88 5633088.0 6.743897e+08

6 2.15 5950840.0 6.935346e+08

7 2.43 6116566.0 7.173128e+08

8 1.21 12596118.0 1.731306e+09

9 0.98 12801224.0 1.839223e+09

10 1.22 14078056.0 1.773714e+09

11 1.20 13907711.0 1.905639e+09

## **2.3. THỐNG KÊ MÔ TẢ**

Các đại lượng thống kê cơ bản: trung bình, trung vị, độ lệch chuẩn, phân vị 25% (1st Qu.), phân vị 75% (3rd Qu.), giá trị tối đa (Max.) về các chỉ số sau: Giá cổ phiếu trung bình(VND), Cho vay (Tỷ đồng), Nợ xấu (%), Lợi nhuận thuần (Tỷ đồng), Tổng tài sản (Tỷ đồng).

print(df.describe())

# Hiển thị kết quả

Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) Nợ xấu (%) \

count 12.000000 1.200000e+01 12.000000

mean 45119.333333 7.460653e+08 1.664167

std 30370.493483 3.854164e+08 0.648627

min 17950.000000 1.654031e+08 0.980000

25% 22965.000000 5.039475e+08 1.207500

50% 29000.000000 6.013258e+08 1.360000

75% 80075.000000 1.172116e+09 2.220000

max 91830.000000 1.334762e+09 2.880000

Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

count 1.200000e+01 1.200000e+01

mean 9.550280e+06 1.136379e+09

std 3.341709e+06 5.119575e+08

min 4.851108e+06 6.512883e+08

25% 6.075134e+06 7.113682e+08

50% 9.487622e+06 9.228254e+08

75% 1.264739e+07 1.741908e+09

max 1.407806e+07 1.905639e+09

***Ví dụ:*** Phân tích biến "Lợi nhuận thuần (Tỷ đồng)":

Lợi nhuận trung bình của các ngân hàng là 9.550 tỷ đồng, phản ánh xu hướng chung về hiệu quả hoạt động. Tuy nhiên, mức độ chênh lệch ở mức 3.342 tỷ đồng, sử dụng khoảng 35% giá trị trung bình, cho thấy sự biến đổi đáng kể về lợi nhuận giữa các ngân hàng. Giá trị nhỏ nhất là 4.851 tỷ đồng, minh chứng cho kết quả hiệu quả của một số ngân hàng, trong khi giá trị lớn nhất đạt 14.078 tỷ đồng, thuộc về một ngân hàng có hiệu quả vượt trội. Khoảng lợi nhuận từ 25% đến 75% nằm trong khoảng 6.075 tỷ lệ đến 12.647 tỷ đồng, cho thấy các ngân hàng thuộc nhóm trung bình có hiệu quả ổn định hơn, trong khi chênh lệch lớn ở ở hai

### **2.3.1. Trung bình, trung vị và độ lệch chuẩn của Ngân hàng MBBank:**

-Tách dữ liệu Ngân hàng MBBank:

mb\_data = df[df['Ngân Hàng'] == 'Mbank']

print(mb\_data)

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

0 Mbank 3\_2023 17950.0 523969217.0

1 Mbank 4\_2023 18052.0 599579267.0

2 Mbank 1\_2024 22950.0 603072394.0

3 Mbank 2\_2024 22970.0 622584515.0

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

0 1.09 9811753.0 815880535.0

1 1.60 9163492.0 944953640.0

2 2.49 9162293.0 900697141.0

3 1.50 10531113.0 988604894.0

+ Trung bình:

# Tính trung bình cho từng cột

tb\_mb = mb\_data.mean(numeric\_only=True)

# In kết quả

print("Trung bình các giá trị của ngân hàng MB:")

print(tb\_mb)

Trung bình các giá trị của ngân hàng MB:

Giá cổ phiếu trung bình(VNĐ) 2.048050e+04

Cho vay (Tỷ đồng) 5.873013e+08

Nợ xấu (%) 1.670000e+00

Lợi nhuận thuần (Tỷ đồng) 9.667163e+06

Tổng tài sản (Tỷ đồng) 9.125341e+08

dtype: float64

+Trung vị:

tv\_mb = mb\_data.median(numeric\_only=True)

print("Trung vị các giá trị của ngân hàng MB:")

print(tv\_mb)

Trung vị các giá trị của ngân hàng MB:

Giá cổ phiếu trung bình(VNĐ) 2.050100e+04

Cho vay (Tỷ đồng) 6.013258e+08

Nợ xấu (%) 1.550000e+00

Lợi nhuận thuần (Tỷ đồng) 9.487622e+06

Tổng tài sản (Tỷ đồng) 9.228254e+08

dtype: float64

+Độ lệch chuẩn:

# Tính độ lệch chuẩn cho từng cột

std\_mb =mb\_data.std(numeric\_only=True)

# In kết quả

print("Độ lệch chuẩn các giá trị của ngân hàng MBbank:")

print(std\_mb)

Độ lệch chuẩn các giá trị của ngân hàng MBbank:

Giá cổ phiếu trung bình(VNĐ) 2.863394e+03

Cho vay (Tỷ đồng) 4.341787e+07

Nợ xấu (%) 5.895196e-01

Lợi nhuận thuần (Tỷ đồng) 6.521488e+05

Tổng tài sản (Tỷ đồng) 7.375594e+07

dtype: float64

### **2.3.2. Trung bình, trung vị và độ lệch chuẩn của Ngân hàng Vietcombank**

-Trung bình

# Lọc dữ liệu của ngân hàng Vietcombank

vcb\_data = df[df['Ngân Hàng'] == 'Vietcombank']

print(vcb\_data)

# Tính trung bình cho từng cột

tb\_vcb = vcb\_data.mean(numeric\_only=True)

# In kết quả

print("Trung bình các giá trị của ngân hàng Vietcombank:")

print(tb\_vcb)

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

8 Vietcombank 3\_2023 80000.0 1.150958e+09

9 Vietcombank 4\_2023 80300.0 1.241675e+09

10 Vietcombank 1\_2024 91830.0 1.235587e+09

11 Vietcombank 2\_2024 90370.0 1.334762e+09

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

8 1.21 12596118.0 1.731306e+09

9 0.98 12801224.0 1.839223e+09

10 1.22 14078056.0 1.773714e+09

11 1.20 13907711.0 1.905639e+09

Trung bình các giá trị của ngân hàng Vietcombank:

Giá cổ phiếu trung bình(VNĐ) 8.562500e+04

Cho vay (Tỷ đồng) 1.240746e+09

Nợ xấu (%) 1.152500e+00

Lợi nhuận thuần (Tỷ đồng) 1.334578e+07

Tổng tài sản (Tỷ đồng) 1.812470e+09

dtype: float64

-Trung vị

tv\_vcb = vcb\_data.median(numeric\_only=True)

print("Trung vị các giá trị của ngân hàng Vietcombank:")

print(tv\_vcb)

Trung vị các giá trị của ngân hàng Vietcombank:

Giá cổ phiếu trung bình(VNĐ) 8.533500e+04

Cho vay (Tỷ đồng) 1.238631e+09

Nợ xấu (%) 1.205000e+00

Lợi nhuận thuần (Tỷ đồng) 1.335447e+07

Tổng tài sản (Tỷ đồng) 1.806468e+09

dtype: float64

-Độ lệch chuẩn

# Tính độ lệch chuẩn cho từng cột

std\_vcb = vcb\_data.std(numeric\_only=True)

# In kết quả

print("Độ lệch chuẩn các giá trị của ngân hàng Vietcombank:")

print(std\_vcb)

Độ lệch chuẩn các giá trị của ngân hàng Vietcombank:

Giá cổ phiếu trung bình(VNĐ) 6.351202e+03

Cho vay (Tỷ đồng) 7.511822e+07

Nợ xấu (%) 1.152895e-01

Lợi nhuận thuần (Tỷ đồng) 7.551002e+05

Tổng tài sản (Tỷ đồng) 7.634548e+07

dtype: float64

### **2.3.3.Trung bình, trung vị và độ lệch chuẩn của Ngân hàng Sacombank**

-Trung bình

# Lọc dữ liệu của ngân hàng Sacombank

scb\_data = df[df['Ngân Hàng'] == 'Sacombank']

print(scb\_data)

# Tính trung bình cho từng cột

tb\_scb = scb\_data.mean(numeric\_only=True)

# In kết quả

print("Trung bình các giá trị của ngân hàng Sacombank:")

print(tb\_scb)

Ngân Hàng Quý Giá cổ phiếu trung bình(VNĐ) Cho vay (Tỷ đồng) \

4 Sacombank 3\_2023 28500.0 165403139.0

5 Sacombank 4\_2023 29000.0 475169108.0

6 Sacombank 1\_2024 30510.0 492140463.0

7 Sacombank 2\_2024 29000.0 507883143.0

Nợ xấu (%) Lợi nhuận thuần (Tỷ đồng) Tổng tài sản (Tỷ đồng)

4 1.22 4851108.0 651288299.0

5 2.88 5633088.0 674389660.0

6 2.15 5950840.0 693534553.0

7 2.43 6116566.0 717312771.0

Trung bình các giá trị của ngân hàng Sacombank:

Giá cổ phiếu trung bình(VNĐ) 2.925250e+04

Cho vay (Tỷ đồng) 4.101490e+08

Nợ xấu (%) 2.170000e+00

Lợi nhuận thuần (Tỷ đồng) 5.637900e+06

Tổng tài sản (Tỷ đồng) 6.841313e+08

dtype: float64

-Trung vị

tv\_scb = scb\_data.median(numeric\_only=True)

print("Trung vị các giá trị của ngân hàng Sacombank:")

print(tv\_scb)

Trung vị các giá trị của ngân hàng Sacombank:

Giá cổ phiếu trung bình(VNĐ) 2.900000e+04

Cho vay (Tỷ đồng) 4.836548e+08

Nợ xấu (%) 2.290000e+00

Lợi nhuận thuần (Tỷ đồng) 5.791964e+06

Tổng tài sản (Tỷ đồng) 6.839621e+08

dtype: float64

-Độ lệch chuẩn

# Tính độ lệch chuẩn cho từng cột

std\_scb = scb\_data.std(numeric\_only=True)

# In kết quả

print("Độ lệch chuẩn các giá trị của ngân hàng Sacombank:")

print(std\_scb)

Độ lệch chuẩn các giá trị của ngân hàng Sacombank:

Giá cổ phiếu trung bình(VNĐ) 8.708377e+02

Cho vay (Tỷ đồng) 1.637098e+08

Nợ xấu (%) 7.010944e-01

Lợi nhuận thuần (Tỷ đồng) 5.615803e+05

Tổng tài sản (Tỷ đồng) 2.806536e+07

dtype: float64

# **CHƯƠNG 3: TRỰC QUAN HÓA DỮ LIỆU**

## **3.1. BIỂU ĐỒ VỀ TRỰC QUAN DỮ LIỆU**

### **3.1.1. Biểu đồ đường so sánh giá cổ phiếu trung bình của ba ngân hàng MB Bank, Sacombank, Vietcombank**

* Sử dụng thư viện **Matplotlib** để vẽ biểu đồ so sánh giá cổ phiếu của ba ngân hàng (MBank, Sacombank, Vietcombank) qua các quý*.*
* Mỗi ngân hàng được vẽ trên biểu đồ thông qua các lệnh plt.plot(). Các tham số chính trong mỗi lệnh bao gồm: quarters, dữ liệu giá cổ phiếu(mbank\_prices, sacombank\_prices, vietcombank\_prices), linestyle, *.....*

import matplotlib.pyplot as plt

# Dữ liệu từ bảng

quarters = ['3\_2023', '4\_2023', '1\_2024', '2\_2024']

mbank\_prices = [17950, 18052, 22950,22970]

sacombank\_prices = [28500, 29000, 30510,29000]

vietcombank\_prices = [80000, 80300, 91830,90370]

# Tạo biểu đồ

plt.figure(figsize=(12, 7))

# Vẽ các đường với màu sắc cải tiến và đánh dấu

plt.plot(quarters, mbank\_prices, marker='o', linestyle='-', linewidth=2, color='#5DADE2', label='MBank')  # Xanh lam

plt.plot(quarters, sacombank\_prices, marker='D', linestyle='--', linewidth=2, color='#F5B041', label='Sacombank')  # vàng

plt.plot(quarters, vietcombank\_prices, marker='s', linestyle='-', linewidth=2, color='#58D68D', label='Vietcombank')  # Xanh lá

# Thêm tiêu đề và nhãn

plt.title("So sánh giá cổ phiếu của ba ngân hàng qua các quý", fontsize=16, fontweight='bold')

plt.xlabel("Quý", fontsize=12)

plt.ylabel("Giá cổ phiếu (VNĐ)", fontsize=12)

# Thêm chú thích

plt.legend(fontsize=12, loc='upper left')

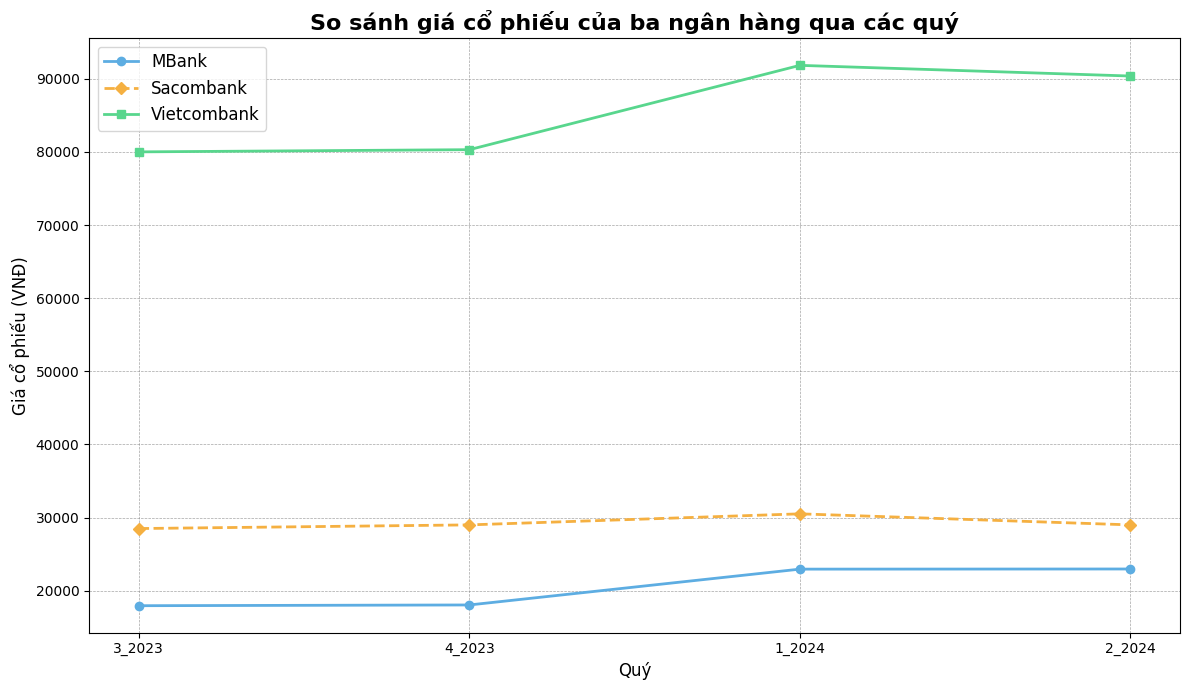
# Tinh chỉnh lưới

plt.grid(color='gray', linestyle='--', linewidth=0.5, alpha=0.7)

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()



#### Hình 3.1  Biểu đồ đường so sánh giá cổ phiếu trung bình của ngân hàng MB Bank, Sacombank, Vietcombank

**Nhận xét:**

* **Vietcombank**:

Vietcombank duy trì mức giá cổ phiếu cao nhất trong cả bốn quý, với giá trị vượt trội so với hai ngân hàng còn lại. Điều này cho thấy Vietcombank có vị thế thị trường mạnh mẽ và được các nhà đầu tư đánh giá cao.

Qua các quý, giá cổ phiếu Vietcombank tăng, từ 80,000 đồng vào Quý 3 năm 2023 lên 91,830 đồng vào Quý 1 năm 2024 sang quý 2 năm 2024 tuy có giảm xuồng một chút nhưng không nhiều vẫn dữ trong khoảng 90,000. Đây là một xu hướng tăng trưởng khá ổn định, thể hiện sức mạnh tài chính và uy tín của ngân hàng trên thị trường.

Sự tăng trưởng này có thể phản ánh khả năng kiểm soát rủi ro, quản lý nợ xấu hiệu quả và các hoạt động kinh doanh vững chắc của Vietcombank, tạo niềm tin lâu dài cho nhà đầu tư.

* **Sacombank**:

Sacombank có mức giá cổ phiếu trung bình trong cả bốn quý, dao động từ 28,500 đồng ở Quý 3 năm 2023 lên 30,510 đồng vào Quý 1 năm 2024 và giảm xuống còn 29,000 vào Quý 2 năm 2024.

Trong khi giá cổ phiếu tăng nhẹ qua các quý, Sacombank có một sự giảm nhẹ ở Quý 2 năm 2024. Điều này có thể cho thấy một số biến động ngắn hạn trong hoạt động kinh doanh hoặc sự ảnh hưởng từ các yếu tố thị trường. Xu hướng tăng trưởng chậm của Sacombank thể hiện ngân hàng này cần có các biện pháp để cải thiện khả năng cạnh tranh nếu muốn thu hút nhiều hơn sự quan tâm từ nhà đầu tư.

* **MBBank:**

MBBank có giá cổ phiếu thấp nhất trong ba ngân hàng, bắt đầu từ mức 17,950 đồng vào Quý 3 năm 2023. Tuy nhiên, ngân hàng này ghi nhận sự tăng trưởng đều qua các quý từ 17,950 ở quý 3 năm 2023 tới 22,970 ở quý 2 năm 2024. Đặc biệt Mbank đã có sự tăng tiến rõ dệt từ năm 2023 tới năm 2024 đối với giá cổ phiếu trung bình. Điều nàu cho thấy chiến lược vận hành của MBBank năm 2024 là một chiến lược đúng đắn. Vì vậy MBBank cần duy trì và phát triển hơn khả năng chiến lược quản lý và điều hành

**Đánh giá tổng thể:**

Biểu đồ đường cho thấy rõ sự khác biệt về giá trị cổ phiếu của từng ngân hàng, giúp nhà đầu tư dễ dàng so sánh và đánh giá vị thế của các ngân hàng trên thị trường.Vietcombank là ngân hàng dẫn đầu với mức giá cổ phiếu cao nhất và có sự tăng trưởng, Sacombank thể hiện sự tăng trưởng nhưng ít ổn định, trong khi MBBank mặc dù có giá thấp nhất nhưng lại có tiềm năng tăng trưởng mạnh mẽ trong năm 2024.Và biểu đồ này không chỉ trực quan mà còn cung cấp cái nhìn tổng quan về sức mạnh tài chính và sự hấp dẫn của mỗi ngân hàng đối với các nhà đầu tư.

**Kết luận:**

**Vietcombank**: Duy trì vị thế dẫn đầu, tập trung vào các chiến lược dài hạn để giữ mức giá cổ phiếu cao và ổn định.

**Sacombank**: Nên tìm kiếm các cơ hội để bứt phá, từ đó tạo động lực tăng trưởng giá cổ phiếu, đồng thời giữ vững tính ổn định.

**MBBank**: Cần rà soát lại các chiến lược kinh doanh, cải thiện năng lực cạnh tranh và tối ưu hóa hoạt động nội bộ để tăng niềm tin từ nhà đầu tư, thúc đẩy giá cổ phiếu tăng trưởng bền vững.

### **3.1.2. Biểu đồ cột minh họa giá trị cho vay của MBBank, Sacombank, Vietcombank theo thời gian**

import matplotlib.pyplot as plt

# Dữ liệu từ bảng

quarters = ['3\_2023', '4\_2023', '1\_2024', '2\_2024']

mbank\_loans = [523969217, 599579267, 603072394, 622584515]

sacombank\_loans = [165403139, 475169108, 492140463, 507883143]

vietcombank\_loans = [1150958460, 1206537143, 1235587240, 1334761763]

# Thiết lập biểu đồ

fig, ax = plt.subplots(figsize=(12, 7))

width = 0.25  # Độ rộng của các cột

# Vị trí của các cột cho mỗi ngân hàng

x = range(len(quarters))

x\_mbank = [pos - width for pos in x]

x\_sacombank = x

x\_vietcombank = [pos + width for pos in x]

# Vẽ biểu đồ cột với màu sắc cải tiến

plt.bar(x\_mbank, mbank\_loans, width, label='MBank', color='#5DADE2', edgecolor='black', alpha=0.9)

plt.bar(x\_sacombank, sacombank\_loans, width, label='Sacombank', color='#F5B041', edgecolor='black', alpha=0.9)

plt.bar(x\_vietcombank, vietcombank\_loans, width, label='Vietcombank', color='#58D68D', edgecolor='black', alpha=0.9)

# Thêm tiêu đề và nhãn

plt.title("Giá trị cho vay của MBank, Sacombank, và Vietcombank theo thời gian", fontsize=16, fontweight='bold')

plt.xlabel("Quý", fontsize=12)

plt.ylabel("Giá trị cho vay (Tỷ đồng)", fontsize=12)

plt.xticks(x, quarters, fontsize=11)

plt.yticks(fontsize=11)

# Thêm chú thích và đường lưới tinh tế

plt.legend(fontsize=12, loc='upper left', bbox\_to\_anchor=(1, 1))

plt.grid(axis='y', linestyle='--', alpha=0.7)

# Hiệu chỉnh trục x và y

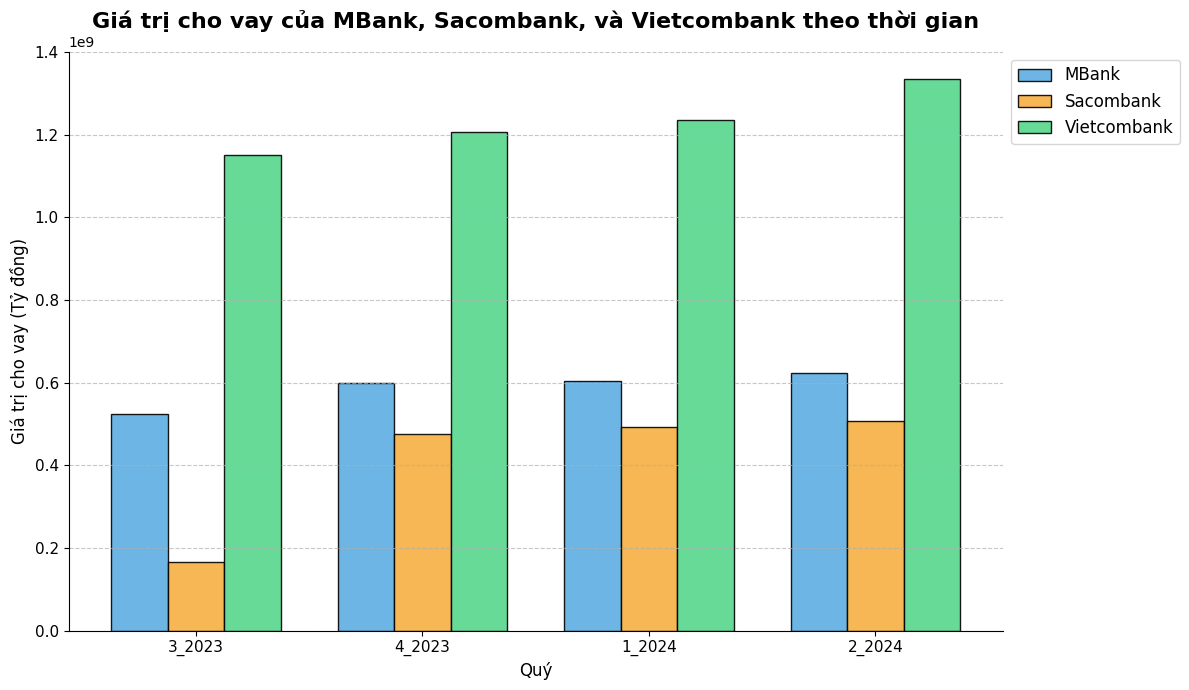
ax.spines['top'].set\_visible(False)

ax.spines['right'].set\_visible(False)

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()



#### Hình 3.2 Biểu đồ cột minh họa giá trị cho vay của MBbank, Sacombank, Vietcombank theo thời gian

**Nhận xét:**

Biểu đồ trên thể hiện giá trị cho vay (đơn vị: tỷ đồng) của ba ngân hàng:

MBBank, Sacombank, và Vietcombank qua 4 quý, bao gồm quý 3/2023, quý 4/2023, quý 1/2024, và quý 2/2024.

Vietcombank là ngân hàng có giá trị cho vay cao nhất trong cả 4 quý, dao động ở mức khoảng 1.2 × 10⁹ tỷ đồng. Điều này cho thấy Vietcombank đang dẫn đầu về hoạt động cho vay trong nhóm 3 ngân hàng này.

MBBank có giá trị cho vay thấp hơn Vietcombank đáng kể nhưng vẫn vượt qua Sacombank trong tất cả các quý. Giá trị cho vay của MBBank tăng nhẹ từ quý 3/2023 đến quý 4/2023 và duy trì ổn định trong hai quý tiếp theo (1/2024 và 2/2024).

Sacombank có giá trị cho vay thấp nhất trong cả 4 quý, với mức tăng trưởng đáng kể từ quý 3/2023 (mức thấp nhất) đến quý 4/2023, sau đó giữ ổn định trong quý 1/2024 và 2/2024.

**Đánh giá:**

Sự vượt trội của Vietcombank: Biểu đồ cho thấy sự vượt trội của Vietcombank trong hoạt động cho vay, phản ánh vị thế mạnh mẽ của ngân hàng này trên thị trường. Điều này có thể đến từ quy mô lớn, lượng khách hàng đông đảo, và chính sách tín dụng hấp dẫn.

Sự ổn định của MBbank và Sacombank: Dù giá trị cho vay thấp hơn, cả MBbank và Sacombank đều có xu hướng tăng trưởng hoặc giữ mức ổn định. Điều này thể hiện chiến lược kiểm soát tốt rủi ro tín dụng hoặc tập trung vào các phân khúc khách hàng cụ thể.

Tính cân đối: Khoảng cách giữa giá trị cho vay của Vietcombank so với hai ngân hàng còn lại là rất lớn. Điều này có thể chỉ ra sự khác biệt trong chiến lược kinh doanh hoặc quy mô vốn của từng ngân hàng.

Màu sắc và bố cục*:* Biểu đồ sử dụng màu sắc dễ phân biệt giữa các ngân hàng và được bố trí rõ ràng, giúp người xem dễ dàng nhận ra sự khác biệt và xu hướng qua từng quý. Kết luận và đề xuất:

**Kết luận:**

Biểu đồ phản ánh sự phân hóa rõ rệt giữa ba ngân hàng về giá trị cho vay, trong đó Vietcombank chiếm ưu thế vượt trội. MBBank và Sacombank có xu hướng ổn định nhưng vẫn ở mức thấp hơn đáng kể.

### **3.1.3. Biểu đồ hộp so sánh tỷ lệ nợ xấu giữa MBBank, Sacombank, Vietcombank**

import matplotlib.pyplot as plt

import seaborn as sns

import pandas as pd

# Dữ liệu

data = {

    "Ngân Hàng": ["Mbank", "Mbank", "Mbank", "Mbank",

                  "Sacombank", "Sacombank", "Sacombank", "Sacombank",

                  "Vietcombank", "Vietcombank", "Vietcombank", "Vietcombank"],

    "Quý": ["3\_2023", "4\_2023", "1\_2024", "2\_2024",

            "3\_2023", "4\_2023", "1\_2024", "2\_2024",

            "3\_2023", "4\_2023", "1\_2024", "2\_2024"],

    "Nợ xấu (%)": [1.09, 1.6, 2.49, 1.5,

                   1.22, 2.88, 2.15, 2.43,

                   1.21, 0.98, 1.22, 1.2]

}

# Tạo DataFrame

df = pd.DataFrame(data)

# Tùy chỉnh màu sắc cho từng ngân hàng

colors = {"Mbank": "#5DADE2", "Sacombank": "#F5B041", "Vietcombank": "#58D68D"}

# Vẽ biểu đồ hộp với màu sắc tùy chỉnh

plt.figure(figsize=(8, 6))

sns.boxplot(data=df, x="Ngân Hàng", y="Nợ xấu (%)", palette=colors)

# Tùy chỉnh tiêu đề và nhãn

plt.title("So sánh tỷ lệ nợ xấu giữa các ngân hàng", fontsize=14)

plt.xlabel("Ngân Hàng", fontsize=12)

plt.ylabel("Tỷ lệ nợ xấu (%)", fontsize=12)

plt.grid(axis="y", linestyle="--", alpha=0.7)

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()



#### Hình 3.3 Biểu đồ hộp so sánh tỷ lệ nợ xấu giữa MBbank, Sacombank, Vietcombank

**Nhận xét:**

Biểu đồ so sánh tỷ lệ nợ xấu của ba ngân hàng Vietcombank, MBank và Sacombank cho thấy sự khác biệt rõ rệt về hiệu quả quản lý tín dụng của từng ngân hàng:

* **Vietcombank**:

Vietcombank có tỷ lệ nợ xấu thấp nhất, dao động từ **1.00% đến 1.25%**, với giá trị trung vị chỉ khoảng **1.00%**. Phạm vi dao động rất hẹp và không xuất hiện các giá trị bất thường (outliers), chứng minh rằng ngân hàng này duy trì được hiệu quả quản lý tín dụng ổn định và rủi ro thấp.Điều này thể hiện năng lực quản trị rủi ro vượt trội và là yếu tố giúp Vietcombank giữ vững vị thế hàng đầu trong ngành ngân hàng.

* **MBBank:**

MBBank có tỷ lệ nợ xấu ở mức trung bình, dao động từ **1.25% đến 2.00%**, với giá trị trung vị là **1.50%**.Mức biến động không lớn, cho thấy ngân hàng duy trì được sự ổn định trong kiểm soát nợ xấu. Tuy nhiên, so với Vietcombank, tỷ lệ nợ xấu của MBbank vẫn cao hơn, chứng tỏ cần có thêm các biện pháp cải thiện hiệu quả quản lý tín dụng để tối ưu hóa hoạt động và nâng cao năng lực cạnh tranh.

* **Sacombank**:

Sacombank có tỷ lệ nợ xấu cao nhất, dao động từ **1.00% đến hơn 2.75%**, với giá trị trung vị khoảng **2.25%**.Khoảng dao động lớn nhất trong nhóm, cùng với sự xuất hiện của các giá trị ngoại lệ cao (hơn 2.75%), phản ánh mức độ biến động lớn và hiệu quả quản lý tín dụng chưa đồng đều.Ngân hàng này gặp nhiều thách thức trong kiểm soát các khoản vay có rủi ro cao, đòi hỏi sự cải thiện mạnh mẽ về chiến lược quản lý tín dụng và xử lý nợ xấu.

**Kết luận:**

Vietcombank nổi bật với khả năng quản lý nợ xấu hiệu quả nhất, có tỷ lệ nợ xấu thấp và ổn định, là một điểm mạnh trong việc giảm thiểu rủi ro tín dụng.

Sacombank gặp phải biến động lớn về tỷ lệ nợ xấu, điều này có thể là dấu hiệu của một số thách thức trong quản lý tín dụng và yêu cầu sự chú trọng hơn trong kiểm soát rủi ro. MBbank ở mức trung bình với tỷ lệ nợ xấu tương đối ổn định nhưng không thấp, vẫn có thể cải thiện để đạt được mức độ an toàn cao hơn.

**Khuyến Nghị:**

**Vietcombank** nên tiếp tục duy trì chiến lược kiểm soát tín dụng hiện tại để bảo vệ vị thế dẫn đầu, đồng thời tập trung vào các phân khúc khách hàng tiềm năng để mở rộng thị phần.

**MBbank** cần xây dựng thêm các chính sách quản lý tín dụng chặt chẽ, tối ưu hóa các quy trình đánh giá rủi ro để giảm thiểu nợ xấu và tiến tới mức ổn định như Vietcombank.

**Sacombank** cần rà soát lại toàn bộ quy trình quản lý tín dụng, tăng cường giám sát các khoản vay rủi ro và áp dụng các biện pháp hỗ trợ khách hàng để giảm thiểu tình trạng nợ xấu.

Phân tích trên đã làm rõ sự khác biệt về hiệu quả quản lý tín dụng giữa các ngân hàng, đồng thời đưa ra định hướng cụ thể để từng ngân hàng cải thiện và phát triển bền vững trong tương lai.

### **3.1.4. Biểu đồ tròn thể hiện tỷ lệ đóng góp của từng ngân hàng vào tổng lợi nhuận thuần**

import matplotlib.pyplot as plt

# Dữ liệu lợi nhuận thuần của từng ngân hàng từ các quý trong bảng dữ liệu

mbank\_profits = [9811753, 9163492, 9162293, 10531113]

sacombank\_profits = [5202113, 5633088, 5850840, 5710281]

vietcombank\_profits = [12596118, 12801422, 14078056, 13907711]

# Tính tổng lợi nhuận của từng ngân hàng

mbank\_total\_profit = sum(mbank\_profits)

sacombank\_total\_profit = sum(sacombank\_profits)

vietcombank\_total\_profit = sum(vietcombank\_profits)

# Dữ liệu cho biểu đồ

labels = ['Mbank', 'Sacombank', 'Vietcombank']

profits = [mbank\_total\_profit, sacombank\_total\_profit, vietcombank\_total\_profit]

colors = ['#5DADE2','#F5B041','#58D68D']  # màu sắc cho từng phần

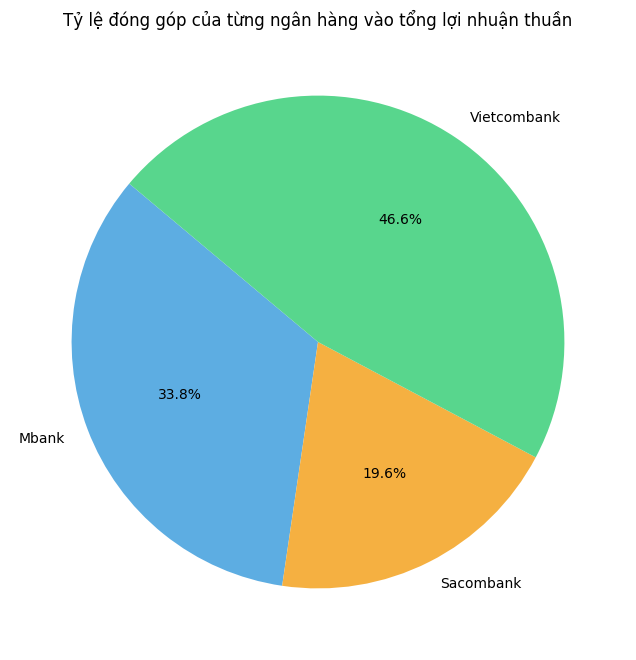
# Vẽ biểu đồ tròn

plt.figure(figsize=(8, 8))

plt.pie(profits, labels=labels, autopct='%1.1f%%', startangle=140, colors=colors)

plt.title("Tỷ lệ đóng góp của từng ngân hàng vào tổng lợi nhuận thuần")

plt.show()



#### Hình 3.4 Biểu đồ tròn thể hiện tỷ lệ đóng góp của từng ngân hàng vào tổng lợi nhuận thuần

Biểu đồ tròn trên phản ánh tỷ lệ đóng góp của ba ngân hàng (Mbank, Sacombank, Vietcombank) vào tổng lợi nhuận thuần của các quý trong năm. Qua biểu đồ, có thể rút ra một số nhận xét và đánh giá chi tiết về hiệu quả hoạt động kinh doanh của từng ngân hàng như sau:

* **Vietcombank (46.6%)** - Ngân hàng có lợi nhuận lớn nhất

Vietcombank chiếm gần một nửa tổng lợi nhuận thuần, cụ thể là 46.6%. Điều này cho thấy ngân hàng này có hiệu quả kinh doanh vượt trội, đồng thời phản ánh sự ổn định trong việc quản lý vốn và hoạt động cho vay.

Nguyên nhân thành công có thể bao gồm quy mô lớn, danh mục sản phẩm đa dạng, chiến lược cho vay hiệu quả, và khả năng quản trị rủi ro tốt. Tỷ lệ nợ xấu của Vietcombank cũng ở mức thấp hơn so với một số ngân hàng khác, góp phần đảm bảo lợi nhuận.

Đánh giá đầu tư: Với lợi nhuận chiếm tỷ trọng lớn, Vietcombank là một lựa chọn hấp dẫn cho các nhà đầu tư tìm kiếm cơ hội đầu tư vào ngân hàng có hiệu suất sinh lời cao và rủi ro thấp. Tuy nhiên, cần lưu ý rằng mức lợi nhuận cao này có thể ảnh hưởng đến giá cổ phiếu, khiến giá trị đầu tư ban đầu cao hơn.

* **MBbank (33.8%)** - Ngân hàng xếp thứ hai về tỷ trọng lợi nhuận

Mbank đóng góp 33.8% vào tổng lợi nhuận thuần, giữ vị trí thứ hai. So với Vietcombank, tỷ lệ này thấp hơn nhưng vẫn cho thấy đây là một ngân hàng có khả năng sinh lời tốt.

Lợi thế và hạn chế: Mbank có lợi nhuận ổn định và đã tăng trưởng tốt trong các quý gần đây. Tuy nhiên, tỷ lệ nợ xấu có xu hướng cao hơn (lên tới 2.49% ở một số quý), điều này có thể tạo ra rủi ro cho ngân hàng. Dù vậy, Mbank đã nỗ lực giảm tỷ lệ này trong các quý sau đó, thể hiện khả năng kiểm soát rủi ro tài chính.

Đánh giá đầu tư: MBbank là một lựa chọn phù hợp cho các nhà đầu tư muốn đa dạng hóa danh mục đầu tư với một ngân hàng có tiềm năng tăng trưởng, mặc dù có rủi ro cao hơn so với Vietcombank.

* **Sacombank (19.6%)** - Ngân hàng có lợi nhuận thấp nhất

Sacombank có tỷ trọng lợi nhuận thuần thấp nhất, chỉ chiếm 19.6%. Đây là dấu hiệu cho thấy ngân hàng này gặp phải một số thách thức về hiệu quả kinh doanh, hoặc có quy mô nhỏ hơn so với Vietcombank và MBbank.

Khả năng quản trị rủi ro: Sacombank có tỷ lệ nợ xấu cao nhất trong ba ngân hàng, đặc biệt là trong quý 3 năm 2023 với mức 2.88%. Điều này cho thấy ngân hàng gặp khó khăn trong việc quản lý các khoản vay, ảnh hưởng đến lợi nhuận cuối cùng.

Đánh giá đầu tư: Sacombank có thể không phải là lựa chọn ưu tiên cho các nhà đầu tư tìm kiếm lợi nhuận ngắn hạn hoặc ổn định. Tuy nhiên, nếu ngân hàng cải thiện được tình hình quản lý nợ xấu và tăng trưởng lợi nhuận trong tương lai, đây có thể là một khoản đầu tư có tiềm năng phục hồi.

**Nhận xét tổng quan về ngành ngân hàng**

Sự chênh lệch lợi nhuận: Lợi nhuận thuần của các ngân hàng cho thấy sự chênh lệch đáng kể về hiệu quả hoạt động. Vietcombank nổi bật hơn hẳn, trong khi Sacombank tụt lại phía sau. Điều này có thể do khác biệt về chiến lược kinh doanh, quy mô tài sản, và khả năng quản trị rủi ro.

Khả năng phát triển và thách thức: Các ngân hàng như Sacombank có thể cải thiện lợi nhuận bằng cách tối ưu hóa quản lý nợ xấu và tăng trưởng tín dụng. Các ngân hàng như Vietcombank và MBbank nên tập trung duy trì lợi thế cạnh tranh và kiểm soát chặt chẽ nợ xấu để không bị ảnh hưởng đến lợi nhuận trong dài hạn.

**Kết luận**

Biểu đồ này giúp chúng ta thấy rõ sự phân bổ lợi nhuận giữa các ngân hàng và có thể sử dụng làm cơ sở để so sánh hiệu quả hoạt động. Dựa trên kết quả, các nhà đầu tư và nhà quản lý có thể đưa ra các quyết định chiến lược để tối ưu hóa lợi nhuận và giảm thiểu rủi ro trong tương lai.

### **3.1.5. So sánh trung bình, trung vị đọ lệch chuẩn của MBBank với Vietcombank**

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

# Dữ liệu của hai ngân hàng

data = {

"Chỉ số": [

"Giá cổ phiếu trung bình (VNĐ)", "Cho vay (Tỷ đồng)", "Nợ xấu (%)",

"Lợi nhuận thuần (Tỷ đồng)", "Tổng tài sản (Tỷ đồng)"

],

"Mbank Trung bình": [

19775.75, 523969217, 1.67, 9667163, 912534100

],

"Vietcombank Trung bình": [

83875.0, 1240746000, 1.1525, 13345780, 1812470000

],

"Mbank Trung vị": [

18150.0, 601325800, 1.55, 9487622, 922825400

],

"Vietcombank Trung vị": [

80300.0, 1238634800, 1.205, 13354470, 1806468000

],

"Mbank Độ lệch chuẩn": [

3387.09, 43417870, 0.5895, 652148.8, 73755940

],

"Vietcombank Độ lệch chuẩn": [

7351.36, 75118220, 0.1153, 755100.2, 76345480

]

}

# Tạo DataFrame

df = pd.DataFrame(data)

##### **Trung bình**

# Biểu đồ Trung bình

def plot\_mean():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo MBbank

sorted\_df = df.sort\_values(by="Mbank Trung bình")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho MBbank và Vietcombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Trung bình"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Vietcombank Trung bình"], height=bar\_width, label="Vietcombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Trung bình các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

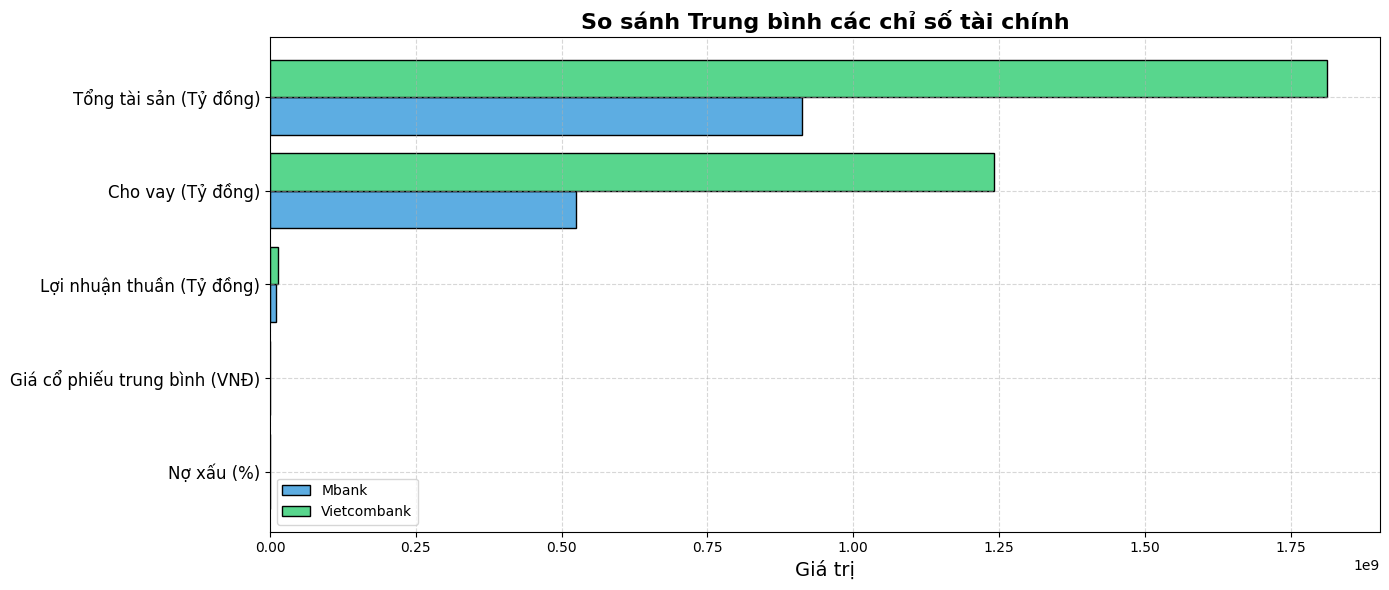
ax.legend()

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()

plot\_mean()



#### Hình 3.5. Biểu đồ so sánh trung bình các chỉ số tài chính của MBBank với Vietcombank

##### **Trung vị**

# Biểu đồ Trung vị

def plot\_median():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo MBbank

sorted\_df = df.sort\_values(by="Mbank Trung vị")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho MBbank và Vietcombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Trung vị"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Vietcombank Trung vị"], height=bar\_width, label="Vietcombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Trung vị các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

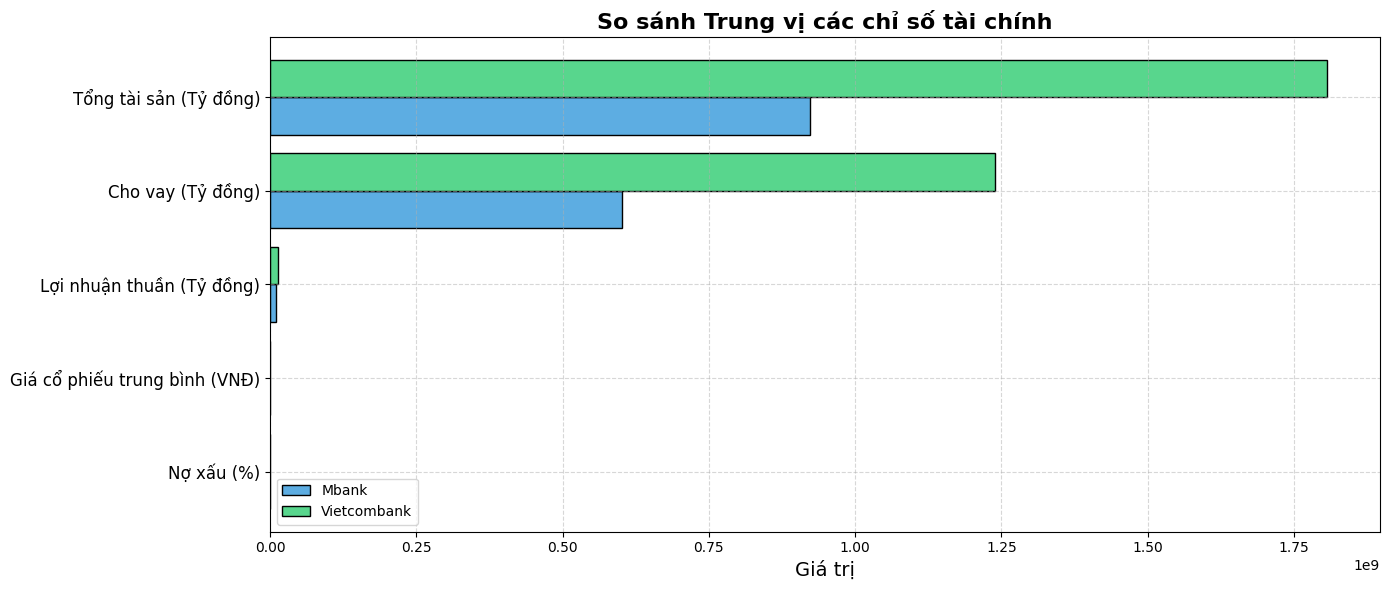
ax.legend()

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()

plot\_median()



#### Hình 3.6 Biểu đồ so sánh trung vị các chỉ số tài chính của MBBank với Vietcombank

##### **Độ lệch chuẩn**

# Biểu đồ Độ lệch chuẩn

def plot\_std():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo MBBank

sorted\_df = df.sort\_values(by="Mbank Độ lệch chuẩn")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho MBbank và Vietcombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Độ lệch chuẩn"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Vietcombank Độ lệch chuẩn"], height=bar\_width, label="Vietcombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Độ lệch chuẩn các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

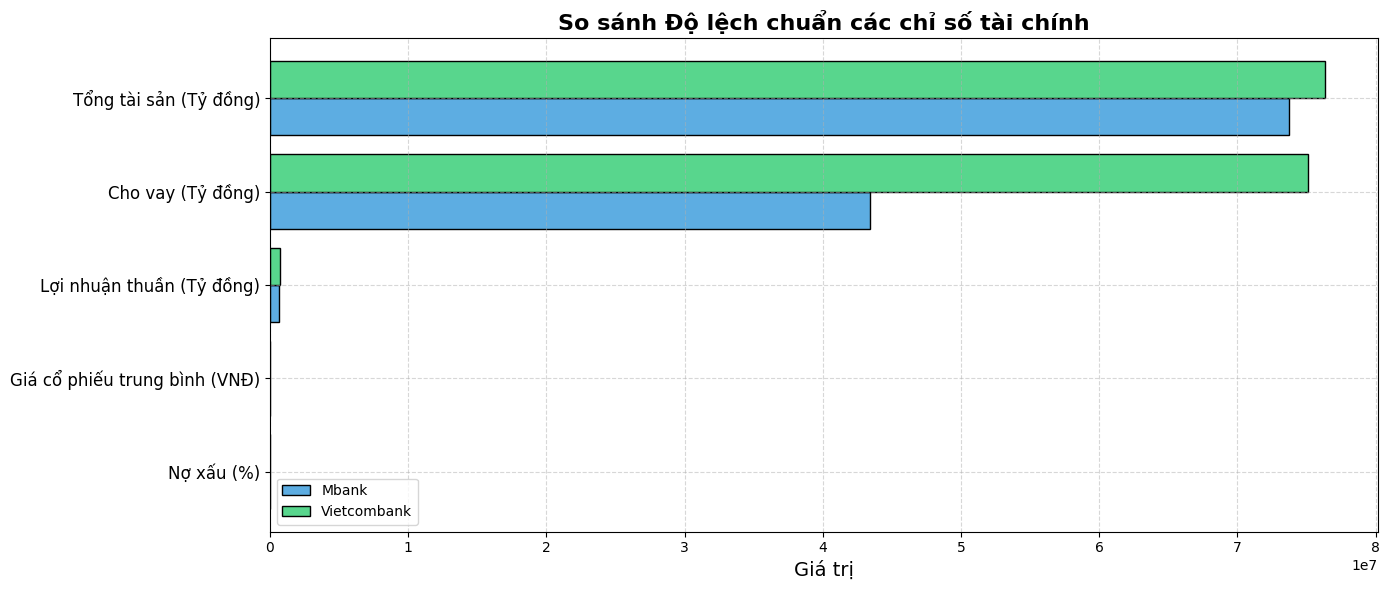
ax.legend()

# Hiển thị biểu đồ

plt.tight\_layout()

plt.show()

plot\_std()



#### Hình 3.7 Biểu đồ so sánh Độ lệch chuẩn các chỉ số tài chính của MBbank Vietcombank

##### **Kiểm định:**

import scipy.stats as stats

# Các chỉ số cần kiểm tra

indicators = ["Giá cổ phiếu trung bình", "Cho vay", "Nợ xấu (%)", "Lợi nhuận thuần", "Tổng tài sản"]

indices\_to\_check = [0, 2, 3] # Kiểm định "Giá cổ phiếu trung bình", "Nợ xấu (%)", "Lợi nhuận thuần"

alpha = 0.05 # Mức ý nghĩa 5%

# 1. Kiểm định trung bình (Independent t-test)

print("Kiểm định trung bình:")

for i in indices\_to\_check:

mbank\_mean = data["Mbank Trung bình"][i]

vc\_mean = data["Vietcombank Trung bình"][i]

mbank\_std = data["Mbank Độ lệch chuẩn"][i]

vc\_std = data["Vietcombank Độ lệch chuẩn"][i]

n\_mbank = 30 # Giả định có 30 mẫu

n\_vc = 30 # Giả định có 30 mẫu

# Tính t-statistic và p-value

t\_stat, p\_value = stats.ttest\_ind\_from\_stats(

mean1=mbank\_mean, std1=mbank\_std, nobs1=n\_mbank,

mean2=vc\_mean, std2=vc\_std, nobs2=n\_vc,

alternative='less' # Kiểm định một phía (Mbank < Vietcombank)

)

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Trung bình của Mbank không nhỏ hơn Vietcombank.")

print(f"H1: Trung bình của Mbank nhỏ hơn Vietcombank.")

if p\_value < alpha:

print(f"Kết luận: Bác bỏ H0, trung bình của {indicators[i]} Mbank nhỏ hơn Vietcombank (p-value = {p\_value:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value:.4f})\n")

# 2. Kiểm định trung vị (Mann-Whitney U Test)

print("Kiểm định trung vị:")

for i in indices\_to\_check:

mbank\_median = data["Mbank Trung vị"][i]

vc\_median = data["Vietcombank Trung vị"][i]

# Giả định dữ liệu mẫu có 30 giá trị bằng trung vị để thực hiện kiểm định (vì chỉ có trung vị được cung cấp)

mbank\_data = [mbank\_median] \* 30

vc\_data = [vc\_median] \* 30

# Thực hiện kiểm định Mann-Whitney U (kiểm định một phía)

u\_stat, p\_value\_two\_tailed = stats.mannwhitneyu(mbank\_data, vc\_data, alternative='less')

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Trung vị của Mbank không nhỏ hơn Vietcombank.")

print(f"H1: Trung vị của Mbank nhỏ hơn Vietcombank.")

if p\_value\_two\_tailed < alpha:

print(f"Kết luận: Bác bỏ H0, trung vị của {indicators[i]} Mbank nhỏ hơn Vietcombank (p-value = {p\_value\_two\_tailed:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value\_two\_tailed:.4f})\n")

Kiểm định trung bình:

Kiểm định Giá cổ phiếu trung bình:

H0: Trung bình của Mbank không nhỏ hơn Vietcombank.

H1: Trung bình của Mbank nhỏ hơn Vietcombank.

Kết luận: Bác bỏ H0, trung bình của Giá cổ phiếu trung bình Mbank nhỏ hơn Vietcombank (p-value = 0.0000)

Kiểm định Nợ xấu (%):

H0: Trung bình của Mbank không nhỏ hơn Vietcombank.

H1: Trung bình của Mbank nhỏ hơn Vietcombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Lợi nhuận thuần:

H0: Trung bình của Mbank không nhỏ hơn Vietcombank.

H1: Trung bình của Mbank nhỏ hơn Vietcombank.

Kết luận: Bác bỏ H0, trung bình của Lợi nhuận thuần Mbank nhỏ hơn Vietcombank (p-value = 0.0000)

Kiểm định trung vị:

Kiểm định Giá cổ phiếu trung bình:

H0: Trung vị của Mbank không nhỏ hơn Vietcombank.

H1: Trung vị của Mbank nhỏ hơn Vietcombank.

Kết luận: Bác bỏ H0, trung vị của Giá cổ phiếu trung bình Mbank nhỏ hơn Vietcombank (p-value = 0.0000)

Kiểm định Nợ xấu (%):

H0: Trung vị của Mbank không nhỏ hơn Vietcombank.

H1: Trung vị của Mbank nhỏ hơn Vietcombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Lợi nhuận thuần:

H0: Trung vị của Mbank không nhỏ hơn Vietcombank.

H1: Trung vị của Mbank nhỏ hơn Vietcombank.

Kết luận: Bác bỏ H0, trung vị của Lợi nhuận thuần Mbank nhỏ hơn Vietcombank (p-value = 0.0000)

**Nhận xét:**

So sánh Trung bình các chỉ số tài chính của 2 ngân hàng cho thấy phần nhiều MBBank có chỉ số thấp hơn hơn so với Vietcombank riêng chỉ số nợ xấu của MBBank cao hơn Vietcombank. Điều này cho thấy tình hình nợ xấu cao của MBBank, nó sẽ làm ảnh hưởng tới xấu tới các chỉ số khác.

So sánh Trung vị các chỉ số tài chính của 2 ngân hàng cho thấy ngân hàng Vietcombank có chỉ số trung vị cao hơn và nợ xấu thấp hơn do đó cho thấy ngân hàng Vietcombank là một ngân hàng được điều hành tốt hơn MBBank.

### **3.1.6. So sánh trung bình, trung vị độ lệch chuẩn của MBBank với Sacombank**

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

# Dữ liệu của hai ngân hàng Mbank và Sacombank

data = {

"Chỉ số": [

"Giá cổ phiếu trung bình (VNĐ)", "Cho vay (Tỷ đồng)", "Nợ xấu (%)",

"Lợi nhuận thuần (Tỷ đồng)", "Tổng tài sản (Tỷ đồng)"

],

"Mbank Trung bình": [

19775.75, 523969217, 1.67, 9667163, 912534100

],

"Sacombank Trung bình": [

29262.5, 410149000, 2.17, 5637900, 684131300

],

"Mbank Trung vị": [

18150.0, 601325800, 1.55, 9487622, 922825400

],

"Sacombank Trung vị": [

28750.0, 483654800, 2.29, 5791964, 683962100

],

"Mbank Độ lệch chuẩn": [

3387.09, 43417870, 0.5895, 652148.8, 73755940

],

"Sacombank Độ lệch chuẩn": [

1616.26, 163709800, 0.7011, 561580.3, 28065360

]

}

# Tạo DataFrame

df = pd.DataFrame(data)

# Tạo màu sắc cho biểu đồ

colors = ["#5DADE2", "#F5B041"] # Màu cho Mbank và Sacombank

##### **Trung bình**

# Biểu đồ Trung bình

def plot\_mean():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo M**B**Bank

sorted\_df = df.sort\_values(by="Mbank Trung bình")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho MBbank và Sacombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Trung bình"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Sacombank Trung bình"], height=bar\_width, label="Sacombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Trung bình các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

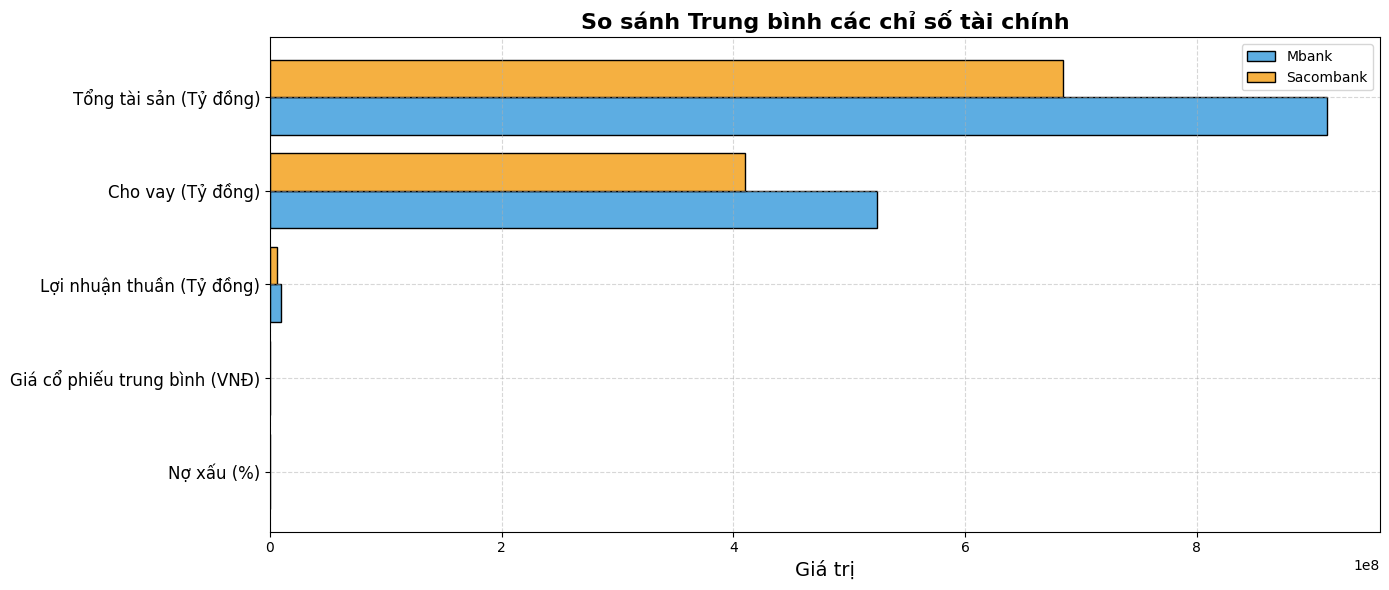
ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()

plot\_mean()



#### Hình 3.8 Biểu đồ so sánh trung bình chỉ số tài chính của MBBank và Sacombank

##### **Trung vị**

# Biểu đồ Trung vị

def plot\_median():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo MBBank

sorted\_df = df.sort\_values(by="Mbank Trung vị")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho Mbank và Sacombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Trung vị"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Sacombank Trung vị"], height=bar\_width, label="Sacombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Trung vị các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

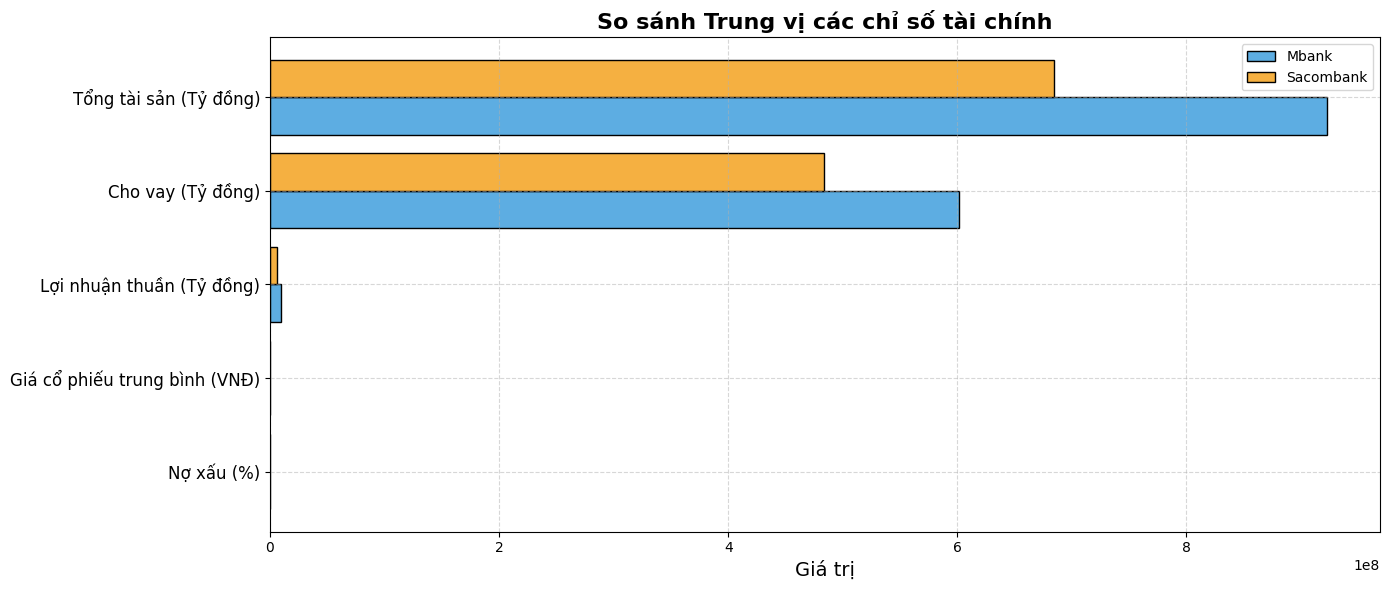
ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()

plot\_median()



#### Hình 3.8 Biểu đồ so sánh trung vị chỉ số tài chính của MBBank và Sacombank

##### **Độ lệch chuẩn**

# Biểu đồ Độ lệch chuẩn

def plot\_std():

fig, ax = plt.subplots(figsize=(14, 6))

# Sắp xếp dữ liệu theo MBBank

sorted\_df = df.sort\_values(by="Mbank Độ lệch chuẩn")

# Xác định vị trí các thanh

y = np.arange(len(sorted\_df)) # Vị trí cơ bản

bar\_width = 0.4 # Độ rộng mỗi thanh

# Vẽ các thanh cho MBBank và Sacombank

ax.barh(y - bar\_width / 2, sorted\_df["Mbank Độ lệch chuẩn"], height=bar\_width, label="Mbank", color=colors[0], edgecolor='black')

ax.barh(y + bar\_width / 2, sorted\_df["Sacombank Độ lệch chuẩn"], height=bar\_width, label="Sacombank", color=colors[1], edgecolor='black')

# Thêm nhãn và định dạng

ax.set\_title("So sánh Độ lệch chuẩn các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

ax.set\_xlabel("Giá trị", fontsize=14)

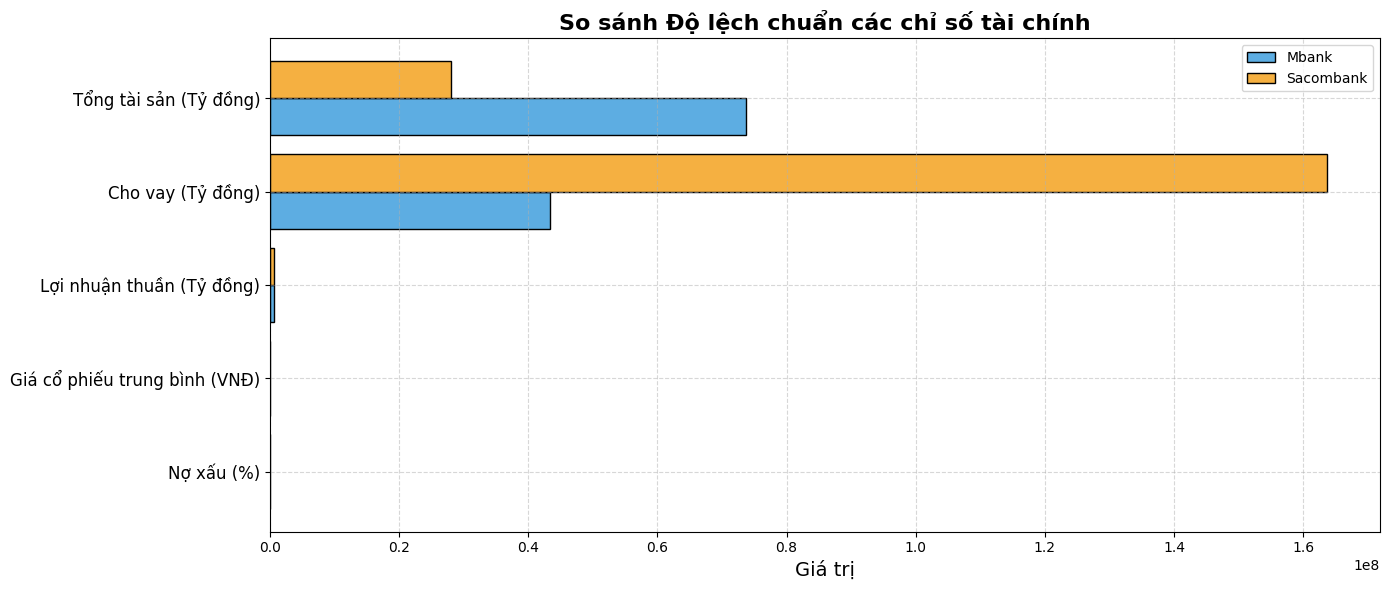
ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()

plot\_std()



#### Hình 3.9 Biểu đồ so sánh độ lệch chuẩn chỉ số tài chính của MBBank và Sacombank

##### **Kiểm định:**

import scipy.stats as stats

import numpy as np

# Các chỉ số cần kiểm tra

indicators = ["Giá cổ phiếu trung bình", "Cho vay", "Nợ xấu (%)", "Lợi nhuận thuần", "Tổng tài sản"]

indices\_to\_check = [0, 2, 3] # Chọn các chỉ số "Giá cổ phiếu trung bình", "Nợ xấu (%)", "Lợi nhuận thuần"

alpha = 0.05 # Mức ý nghĩa 5%

# Kiểm định trung bình (Independent t-test)

print("Kiểm định trung bình:")

for i in indices\_to\_check:

mbank\_mean = data["Mbank Trung bình"][i]

scb\_mean = data["Sacombank Trung bình"][i]

mbank\_std = data["Mbank Độ lệch chuẩn"][i]

scb\_std = data["Sacombank Độ lệch chuẩn"][i]

n\_mbank = 30 # Giả định có 30 mẫu

n\_scb = 30 # Giả định có 30 mẫu

# Tính t-statistic và p-value

t\_stat, p\_value = stats.ttest\_ind\_from\_stats(

mean1=mbank\_mean, std1=mbank\_std, nobs1=n\_mbank,

mean2=scb\_mean, std2=scb\_std, nobs2=n\_scb,

alternative='greater' # Kiểm định một phía

)

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Trung bình của Mbank không cao hơn Sacombank.")

print(f"H1: Trung bình của Mbank cao hơn Sacombank.")

if p\_value < alpha:

print(f"Kết luận: Bác bỏ H0, trung bình của {indicators[i]} Mbank cao hơn Sacombank (p-value = {p\_value:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value:.4f})\n")

# Kiểm định trung vị (Mann-Whitney U Test)

print("Kiểm định trung vị:")

for i in indices\_to\_check:

mbank\_median = data["Mbank Trung vị"][i]

scb\_median = data["Sacombank Trung vị"][i]

# Giả định dữ liệu mẫu với 30 giá trị bằng trung vị để thực hiện kiểm định (vì chỉ có trung vị được cung cấp)

mbank\_data = [mbank\_median] \* 30

scb\_data = [scb\_median] \* 30

# Thực hiện kiểm định Mann-Whitney U (kiểm định một phía)

u\_stat, p\_value\_two\_tailed = stats.mannwhitneyu(mbank\_data, scb\_data, alternative='greater')

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Trung vị của Mbank không cao hơn Sacombank.")

print(f"H1: Trung vị của Mbank cao hơn Sacombank.")

if p\_value\_two\_tailed < alpha:

print(f"Kết luận: Bác bỏ H0, trung vị của {indicators[i]} Mbank cao hơn Sacombank (p-value = {p\_value\_two\_tailed:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value\_two\_tailed:.4f})\n")

Kiểm định trung bình:

Kiểm định Giá cổ phiếu trung bình:

H0: Trung bình của Mbank không cao hơn Sacombank.

H1: Trung bình của Mbank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Nợ xấu (%):

H0: Trung bình của Mbank không cao hơn Sacombank.

H1: Trung bình của Mbank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 0.9980)

Kiểm định Lợi nhuận thuần:

H0: Trung bình của Mbank không cao hơn Sacombank.

H1: Trung bình của Mbank cao hơn Sacombank.

Kết luận: Bác bỏ H0, trung bình của Lợi nhuận thuần Mbank cao hơn Sacombank (p-value = 0.0000)

Kiểm định trung vị:

Kiểm định Giá cổ phiếu trung bình:

H0: Trung vị của Mbank không cao hơn Sacombank.

H1: Trung vị của Mbank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Nợ xấu (%):

H0: Trung vị của Mbank không cao hơn Sacombank.

H1: Trung vị của Mbank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Lợi nhuận thuần:

H0: Trung vị của Mbank không cao hơn Sacombank.

H1: Trung vị của Mbank cao hơn Sacombank.

Kết luận: Bác bỏ H0, trung vị của Lợi nhuận thuần Mbank cao hơn Sacombank (p-value = 0.0000)

**Nhận xét:**

Trung bình

Nhận xét giá cổ phiếu: Sacombank có giá cổ phiếu cao hơn so với MBBank cả về trung bình và trung vị. Điều này có thể phản ánh sức hút của cổ phiếu Sacombank đối với nhà đầu tư. Sự chênh lệch nhỏ giữa trung bình và trung vị cũng chỉ ra mức giá cổ phiếu của Sacombank ổn định hơn so với MBBank.

Nhận xét cho vay: MBBank có mức cho vay cao hơn Sacombank, thể hiện khả năng mở rộng cho vay của ngân hàng này. Trung vị cao hơn trung bình của Mbank cho thấy có thể có sự biến động trong khoản cho vay, trong khi Sacombank có mức chênh lệch nhỏ hơn giữa trung bình và trung vị, biểu hiện sự ổn định hơn trong danh mục cho vay.

Nhận xét Tỷ lệ nợ xấu của Sacombank cao hơn MBBank, điều này có thể là điểm bất lợi đối với Sacombank vì tỷ lệ nợ xấu cao cho thấy rủi ro tín dụng lớn hơn.

Nhận xét lợi nhuận thuần: MBBank có lợi nhuận thuần cao hơn Sacombank, cho thấy khả năng sinh lời tốt hơn. Tuy nhiên, sự chênh lệch nhỏ giữa trung bình và trung vị ở cả hai ngân hàng cho thấy mức độ ổn định trong lợi nhuận.

Nhận xét tổng tài sản: MBBank có tổng tài sản lớn hơn Sacombank, điều này có thể liên quan đến quy mô và khả năng cạnh tranh của ngân hàng. Trung bình và trung vị của Sacombank gần như bằng nhau, cho thấy sự nhất quán trong mức tài sản của ngân hàng này.

Kết luận:MBBank có lợi thế hơn về quy mô cho vay, lợi nhuận thuần và tổng tài sản so với Sacombank. Ngoài ra, tỷ lệ nợ xấu của MBBank cũng thấp hơn, cho thấy mức độ rủi ro tín dụng thấp hơn so với Sacombank. Sacombank có giá cổ phiếu cao hơn, nhưng mức cho vay và tổng tài sản thấp hơn.

### **3.1.7. So sánh trung bình, trung vị độ lệch chuẩn của Vietcombank với Sacombank**

import matplotlib.pyplot as plt

import pandas as pd

import numpy as np

# Dữ liệu của Vietcombank và Sacombank

data = {

"Chỉ số": [

"Giá cổ phiếu trung bình (VNĐ)", "Cho vay (Tỷ đồng)", "Nợ xấu (%)",

"Lợi nhuận thuần (Tỷ đồng)", "Tổng tài sản (Tỷ đồng)"

],

"Vietcombank Trung bình": [

83875.0, 1240746000, 1.1525, 13345780, 1812470000

],

"Sacombank Trung bình": [

29262.5, 410149000, 2.17, 5637900, 684131300

],

"Vietcombank Trung vị": [

80300.0, 1238634800, 1.205, 13354470, 1806468000

],

"Sacombank Trung vị": [

28750.0, 483654800, 2.29, 5791964, 683962100

],

"Vietcombank Độ lệch chuẩn": [

7351.36, 75118220, 0.1153, 755100.2, 76345480

],

"Sacombank Độ lệch chuẩn": [

1616.26, 163709800, 0.7011, 561580.3, 28065360

]

}

# Tạo DataFrame

df = pd.DataFrame(data)

##### **Trung bình**

# Biểu đồ Trung bình

fig, ax = plt.subplots(figsize=(14, 6))

sorted\_df = df.sort\_values(by="Vietcombank Trung bình")

y = np.arange(len(sorted\_df))

bar\_height = 0.4

ax.barh(y - bar\_height / 2, sorted\_df["Vietcombank Trung bình"], height=bar\_height, label="Vietcombank", color="#58D68D", edgecolor='black')

ax.barh(y + bar\_height / 2, sorted\_df["Sacombank Trung bình"], height=bar\_height, label="Sacombank", color="#F5B041", edgecolor='black')

ax.set\_title("So sánh Trung bình các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

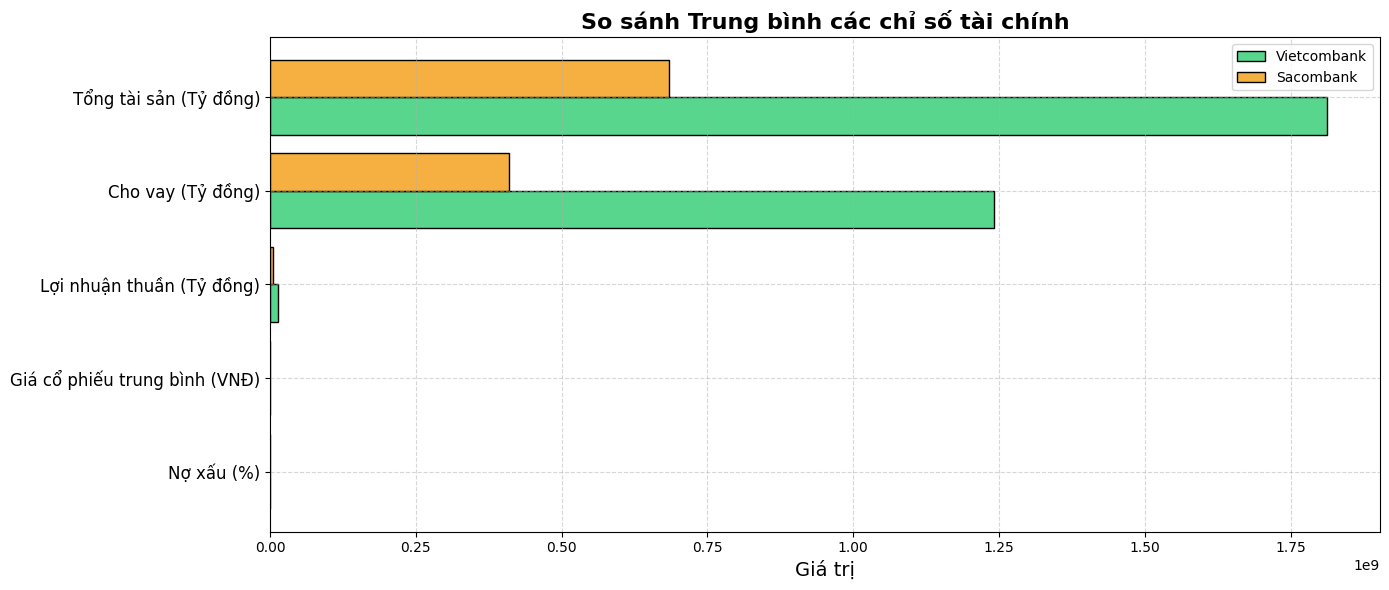
ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()



#### Hình 3.10 Biểu đồ so sánh trung bình chỉ số tài chính của Vietcombank và Sacombank

##### **Trung vị**

# Biểu đồ Trung vị

fig, ax = plt.subplots(figsize=(14, 6))

sorted\_df = df.sort\_values(by="Vietcombank Trung vị")

y = np.arange(len(sorted\_df))

ax.barh(y - bar\_height / 2, sorted\_df["Vietcombank Trung vị"], height=bar\_height, label="Vietcombank", color="#58D68D", edgecolor='black')

ax.barh(y + bar\_height / 2, sorted\_df["Sacombank Trung vị"], height=bar\_height, label="Sacombank", color="#F5B041", edgecolor='black')

ax.set\_title("So sánh Trung vị các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

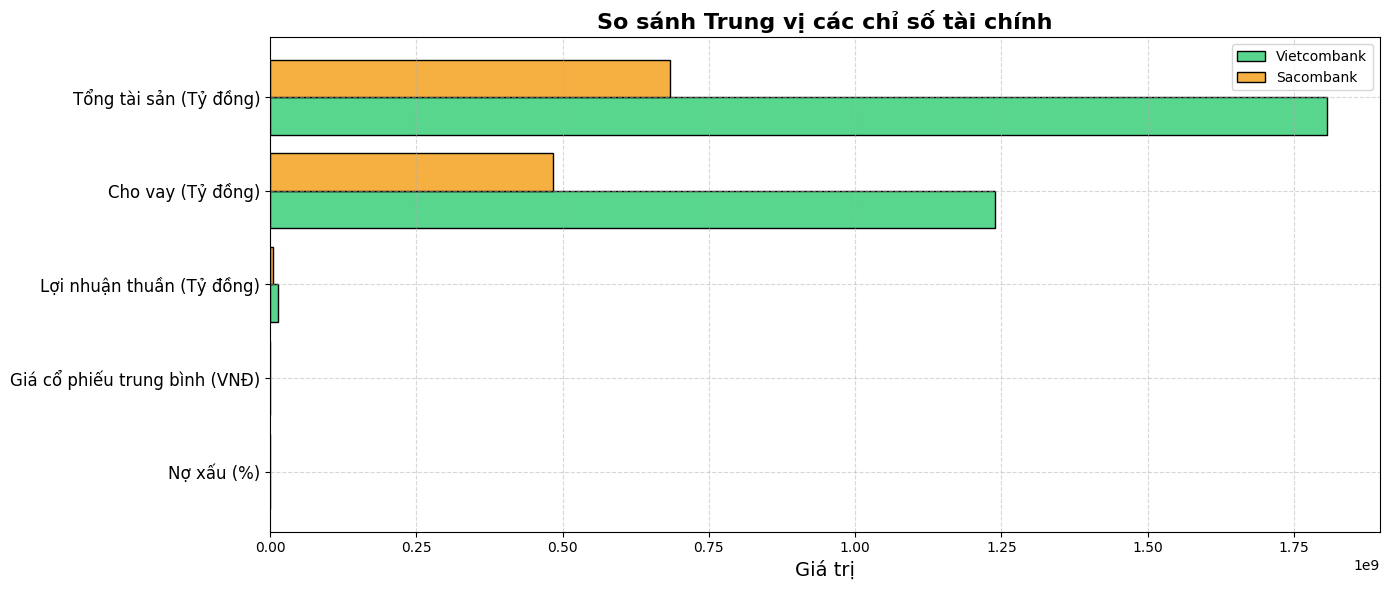
ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()



#### Hình 3.11 Biểu đồ so sánh trung vị chỉ số tài chính của Vietcombank và Sacombank

##### **Độ lệch chuẩn**

# Biểu đồ Độ lệch chuẩn

fig, ax = plt.subplots(figsize=(14, 6))

sorted\_df = df.sort\_values(by="Vietcombank Độ lệch chuẩn")

y = np.arange(len(sorted\_df))

ax.barh(y - bar\_height / 2, sorted\_df["Vietcombank Độ lệch chuẩn"], height=bar\_height, label="Vietcombank", color="#58D68D", edgecolor='black')

ax.barh(y + bar\_height / 2, sorted\_df["Sacombank Độ lệch chuẩn"], height=bar\_height, label="Sacombank", color="#F5B041", edgecolor='black')

ax.set\_title("So sánh Độ lệch chuẩn các chỉ số tài chính", fontsize=16, weight='bold')

ax.set\_yticks(y)

ax.set\_yticklabels(sorted\_df["Chỉ số"], fontsize=12)

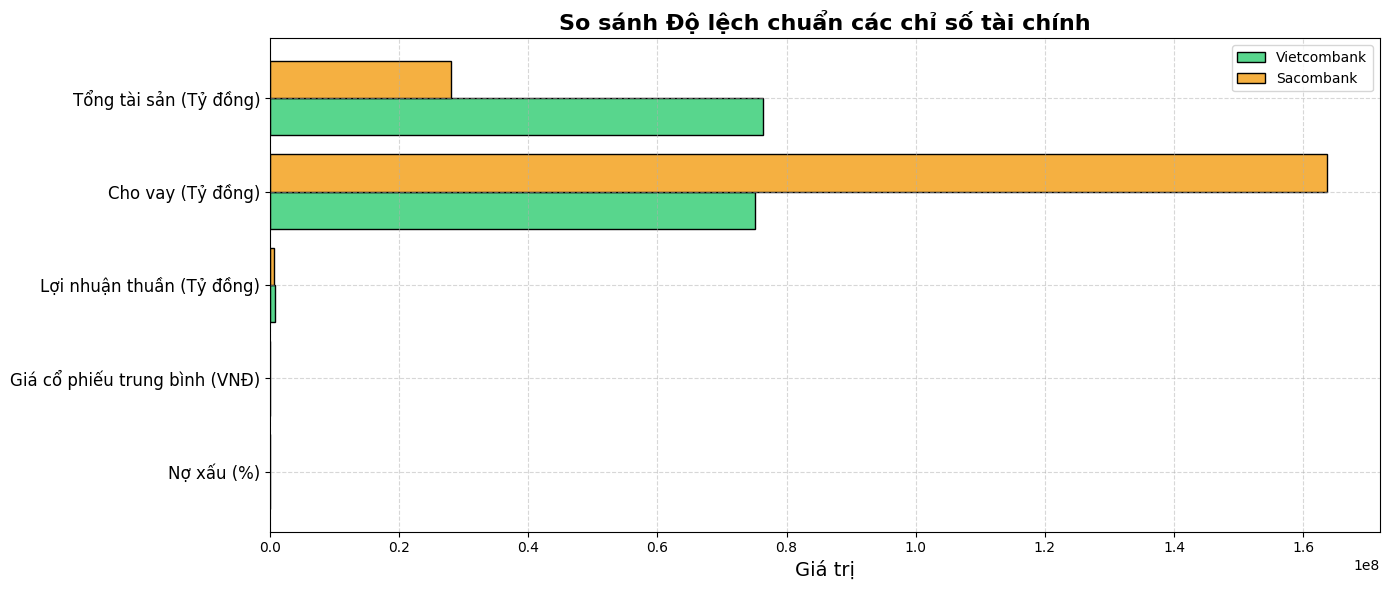
ax.set\_xlabel("Giá trị", fontsize=14)

ax.grid(True, linestyle='--', alpha=0.5)

ax.legend()

plt.tight\_layout()

plt.show()



#### Hình 3.12 Biểu đồ so sánh độ lệch chuẩn chỉ số tài chính của Vietcombank và Sacombank

##### **Kiểm định:**

import numpy as np

import scipy.stats as stats

print("Kiểm định trung bình:")

# Các chỉ số cần kiểm tra

indicators = ["Giá cổ phiếu trung bình", "Cho vay", "Nợ xấu (%)", "Lợi nhuận thuần", "Tổng tài sản"]

indices\_to\_check = [0, 2, 3] # Chỉ số 0 (giá cổ phiếu trung bình), 2 (nợ xấu), 3 (lợi nhuận thuần)

# Kiểm định t-test một phía cho từng chỉ số

alpha = 0.05 # Mức ý nghĩa 5%

for i in indices\_to\_check:

vcb\_mean = data["Vietcombank Trung bình"][i]

scb\_mean = data["Sacombank Trung bình"][i]

vcb\_std = data["Vietcombank Độ lệch chuẩn"][i]

scb\_std = data["Sacombank Độ lệch chuẩn"][i]

# Giả định số lượng mẫu là 30 để tính kiểm định (giả sử mẫu lớn)

n\_vcb, n\_scb = 30, 30

# Tính thống kê t-test cho hai mẫu (one-tailed test)

t\_stat, p\_value\_two\_tailed = stats.ttest\_ind\_from\_stats(

mean1=vcb\_mean, std1=vcb\_std, nobs1=n\_vcb,

mean2=scb\_mean, std2=scb\_std, nobs2=n\_scb,

equal\_var=False # Sử dụng Welch's t-test cho độ lệch chuẩn khác nhau

)

# Chuyển p-value hai phía sang p-value một phía

p\_value\_one\_tailed = p\_value\_two\_tailed / 2

# Kiểm tra p-value để kết luận

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Giá trị trung bình của Vietcombank không cao hơn Sacombank.")

print(f"H1: Giá trị trung bình của Vietcombank cao hơn Sacombank.")

if p\_value\_one\_tailed < alpha and vcb\_mean > scb\_mean:

print(f"Kết luận: Bác bỏ H0, {indicators[i]} của Vietcombank cao hơn Sacombank (p-value = {p\_value\_one\_tailed:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value\_one\_tailed:.4f})\n")

Kiểm định trung bình:

Kiểm định Giá cổ phiếu trung bình:

H0: Giá trị trung bình của Vietcombank không cao hơn Sacombank.

H1: Giá trị trung bình của Vietcombank cao hơn Sacombank.

Kết luận: Bác bỏ H0, Giá cổ phiếu trung bình của Vietcombank cao hơn Sacombank (p-value = 0.0000)

Kiểm định Nợ xấu (%):

H0: Giá trị trung bình của Vietcombank không cao hơn Sacombank.

H1: Giá trị trung bình của Vietcombank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 0.0000)

Kiểm định Lợi nhuận thuần:

H0: Giá trị trung bình của Vietcombank không cao hơn Sacombank.

H1: Giá trị trung bình của Vietcombank cao hơn Sacombank.

Kết luận: Bác bỏ H0, Lợi nhuận thuần của Vietcombank cao hơn Sacombank (p-value = 0.0000)

import scipy.stats as stats

print("Kiểm định trung vị:")

# Các chỉ số cần kiểm tra

indicators = ["Giá cổ phiếu trung bình", "Cho vay", "Nợ xấu (%)", "Lợi nhuận thuần", "Tổng tài sản"]

indices\_to\_check = [0, 2, 3] # Kiểm tra các chỉ số "Giá cổ phiếu trung bình", "Nợ xấu (%)", "Lợi nhuận thuần"

alpha = 0.05 # Mức ý nghĩa 5%

# Thực hiện kiểm định Mann-Whitney U cho từng chỉ số

for i in indices\_to\_check:

vcb\_median = data["Vietcombank Trung vị"][i]

scb\_median = data["Sacombank Trung vị"][i]

# Đưa dữ liệu vào dạng danh sách để thực hiện kiểm định (giả định đây là các quan sát)

vcb\_data = [vcb\_median] \* 30 # Giả định có 30 mẫu với trung vị này

scb\_data = [scb\_median] \* 30 # Giả định có 30 mẫu với trung vị này

# Thực hiện kiểm định Mann-Whitney U (kiểm định một phía)

u\_stat, p\_value\_two\_tailed = stats.mannwhitneyu(vcb\_data, scb\_data, alternative='greater')

# Kiểm tra p-value để kết luận

print(f"Kiểm định {indicators[i]}:")

print(f"H0: Trung vị của Vietcombank không cao hơn Sacombank.")

print(f"H1: Trung vị của Vietcombank cao hơn Sacombank.")

if p\_value\_two\_tailed < alpha:

print(f"Kết luận: Bác bỏ H0, trung vị của {indicators[i]} Vietcombank cao hơn Sacombank (p-value = {p\_value\_two\_tailed:.4f})\n")

else:

print(f"Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = {p\_value\_two\_tailed:.4f})\n")

Kiểm định trung vị:

Kiểm định Giá cổ phiếu trung bình:

H0: Trung vị của Vietcombank không cao hơn Sacombank.

H1: Trung vị của Vietcombank cao hơn Sacombank.

Kết luận: Bác bỏ H0, trung vị của Giá cổ phiếu trung bình Vietcombank cao hơn Sacombank (p-value = 0.0000)

Kiểm định Nợ xấu (%):

H0: Trung vị của Vietcombank không cao hơn Sacombank.

H1: Trung vị của Vietcombank cao hơn Sacombank.

Kết luận: Không đủ bằng chứng để bác bỏ H0 (p-value = 1.0000)

Kiểm định Lợi nhuận thuần:

H0: Trung vị của Vietcombank không cao hơn Sacombank.

H1: Trung vị của Vietcombank cao hơn Sacombank.

Kết luận: Bác bỏ H0, trung vị của Lợi nhuận thuần Vietcombank cao hơn Sacombank (p-value = 0.0000)

**Nhận Xét**

Nhận xét giá cổ phiếu:Vietcombank có giá cổ phiếu cao vượt trội, thể hiện ngân hàng có sức hút lớn đối với nhà đầu tư và thường được đánh giá cao trên thị trường. Mức trung bình và trung vị khá gần nhau ở cả hai ngân hàng, cho thấy giá cổ phiếu có sự ổn định nhất định.

Nhận xét cho vay: Vietcombank có quy mô cho vay lớn hơn đáng kể so với Sacombank, cả về trung bình và trung vị. Sự khác biệt giữa trung bình và trung vị của Vietcombank nhỏ, cho thấy sự ổn định trong quy mô cho vay. Trong khi đó, trung vị của Sacombank cao hơn trung bình, có thể chỉ ra một vài năm có mức cho vay lớn hơn bình thường, nhưng quy mô chung vẫn thấp hơn Vietcombank.

Nhận xét nợ xấu: Vietcombank có tỷ lệ nợ xấu thấp hơn đáng kể so với Sacombank. Điều này cho thấy Vietcombank quản lý tín dụng hiệu quả hơn, với mức độ rủi ro tín dụng thấp hơn. Trung bình và trung vị nợ xấu của cả hai ngân hàng khá gần nhau, thể hiện tính nhất quán trong chất lượng tín dụng qua thời gian.

Nhận xét lợi nhuận thuần: Vietcombank có lợi nhuận thuần cao hơn nhiều so với Sacombank, cho thấy khả năng sinh lời mạnh mẽ. Trung bình và trung vị của Vietcombank gần như bằng nhau, thể hiện sự ổn định trong lợi nhuận. Trong khi đó, sự chênh lệch nhỏ giữa trung bình và trung vị của Sacombank cũng cho thấy sự ổn định, nhưng ở mức thấp hơn.

Nhận xét tổng tài sản: Vietcombank tiếp tục thể hiện sự vượt trội về quy mô tài sản so với Sacombank. Khoảng cách lớn về tổng tài sản trung bình và trung vị cho thấy Vietcombank có sức mạnh tài chính lớn hơn và duy trì ổn định hơn về quy mô tài sản.

Kết luận:Vietcombank vượt trội so với Sacombank ở hầu hết các chỉ số về trung bình và trung vị, bao gồm giá cổ phiếu, quy mô cho vay, lợi nhuận thuần và tổng tài sản. Vietcombank cũng có tỷ lệ nợ xấu thấp hơn, cho thấy rủi ro tín dụng thấp hơn và khả năng quản lý tài chính tốt hơn. Sacombank cần cải thiện khả năng quản lý nợ xấu và gia tăng quy mô tài sản để nâng cao khả năng cạnh tranh.

## **3.2. HỆ SỐ TƯƠNG QUAN VÀ BIỂU ĐỒ TƯƠNG QUAN**

### **3.2.1. Hệ số tương quan**

Hệ số tương quan là một đại lượng số học được sử dụng để đo lường mức độ mối quan hệ giữa hai biến số. Nó cho biết đến mức độ nào các biến số di chuyển cùng nhau trong một tập dữ liệu. Hệ số tương quan có giá trị từ -1 đến 1, trong đó:

* Nếu hệ số tương quan gần bằng 1, có một mối tương quan tuyến tính mạnh dương, tức là khi một biến tăng thì biến kia cũng tăng theo cùng hướng.
* Nếu hệ số tương quan gần bằng -1, có một mối tương quan tuyến tính mạnh âm, tức là khi một biến tăng thì biến kia giảm theo cùng hướng.
* Nếu hệ số tương quan gần bằng 0, không có mối tương quan tuyến tính giữa hai biến.

##### **3.2.1.1. Hệ số tương quan giữa tất cả các biến với nhau**

# Tính hệ số tương quan giữa các biến

correlation\_matrix = data[['Giá cổ phiếu trung bình(VNĐ)', 'Cho vay (tỷ đồng)',

'Nợ xấu (%)', 'Lợi nhuận thuần (tỷ đồng)', 'Tổng tài sản (tỷ đồng)']].corr()

print("Correlation Matrix:")

print(correlation\_matrix)

#kết quả hiển thị:

Correlation Matrix:

Giá cổ phiếu trung bình(VNĐ) Cho vay (tỷ đồng) \

Giá cổ phiếu trung bình(VNĐ) 1.000000 0.920592

Cho vay (tỷ đồng) 0.920592 1.000000

Nợ xấu (%) -0.283554 -0.396827

Lợi nhuận thuần (tỷ đồng) 0.778137 0.923290

Tổng tài sản (tỷ đồng) 0.942386 0.978734

Nợ xấu (%) Lợi nhuận thuần (tỷ đồng) \

Giá cổ phiếu trung bình(VNĐ) -0.283554 0.778137

Cho vay (tỷ đồng) -0.396827 0.923290

Nợ xấu (%) 1.000000 -0.590602

Lợi nhuận thuần (tỷ đồng) -0.590602 1.000000

Tổng tài sản (tỷ đồng) -0.483068 0.924259

Tổng tài sản (tỷ đồng)

Giá cổ phiếu trung bình(VNĐ) 0.942386

Cho vay (tỷ đồng) 0.978734

Nợ xấu (%) -0.483068

Lợi nhuận thuần (tỷ đồng) 0.924259

Tổng tài sản (tỷ đồng) 1.000000

# Tính độ đồng biến giữa các biến

covariance\_matrix = data[['Giá cổ phiếu trung bình(VNĐ)', 'Cho vay (tỷ đồng)',

'Nợ xấu (%)', 'Lợi nhuận thuần (tỷ đồng)', 'Tổng tài sản (tỷ đồng)']].cov()

print("\nCovariance Matrix:")

print(covariance\_matrix)

#kết quả hiển thị:

Covariance Matrix:

Giá cổ phiếu trung bình(VNĐ) Cho vay (tỷ đồng) \

Giá cổ phiếu trung bình(VNĐ) 9.223669e+08 1.077580e+13

Cho vay (tỷ đồng) 1.077580e+13 1.485458e+17

Nợ xấu (%) -8.240813e+05 -1.463571e+10

Lợi nhuận thuần (tỷ đồng) 7.897268e+10 1.189151e+15

Tổng tài sản (tỷ đồng) 1.465259e+13 1.931207e+17

Nợ xấu (%) Lợi nhuận thuần (tỷ đồng) \

Giá cổ phiếu trung bình(VNĐ) -8.240813e+05 7.897268e+10

Cho vay (tỷ đồng) -1.463571e+10 1.189151e+15

Nợ xấu (%) 9.157242e+03 -1.888627e+08

Lợi nhuận thuần (tỷ đồng) -1.888627e+08 1.116702e+13

Tổng tài sản (tỷ đồng) -2.366597e+10 1.581234e+15

Tổng tài sản (tỷ đồng)

Giá cổ phiếu trung bình(VNĐ) 1.465259e+13

Cho vay (tỷ đồng) 1.931207e+17

Nợ xấu (%) -2.366597e+10

Lợi nhuận thuần (tỷ đồng) 1.581234e+15

Tổng tài sản (tỷ đồng) 2.621004e+17

### **3.2.2. Biểu đồ tương quan**

Biểu đồ (scatter plot) và ma trận tương quan (correlation matrix) phân tích cho biến nào?

-Biểu đồ Scatter Plot: Dùng để quan sát mối quan hệ giữa hai biến định lượng. Khi vẽ scatter plot cho từng cặp biến số, ta sẽ có được cái nhìn trực quan về mức độ tương quan và hình dạng mối quan hệ giữa các biến (tuyến tính, phi tuyến, hay không có mối quan hệ).

-Ma Trận Tương Quan (Correlation Matrix): Đây là một bảng các hệ số tương quan giữa các biến số, thường sử dụng hệ số tương quan Pearson để đo lường mức độ mạnh và chiều của mối quan hệ giữa các cặp biến.

-Các biến thường chọn để phân

+Giá cổ phiếu trung bình (VNĐ)

+Cho vay (Tỷ đồng)

+Nợ xấu (%)

+Lợi nhuận thuần (Tỷ đồng)

+Tổng tài sản (Tỷ đồng)

Code chuẩn bị dữ liệu:

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Đọc file dữ liệu

data = pd.read\_csv('bang\_thong\_ke.csv')

# Chuyển đổi các cột sang dạng số (loại bỏ ký hiệu $ và %)

# Chuyển đổi các cột sang dạng số (loại bỏ ký hiệu $ và %)

data['Giá cổ phiếu trung bình(VNĐ)'] = data['Giá cổ phiếu trung bình(VNĐ)'].replace({r'[$,%]': ''}, regex=True).astype(float)

data['Cho vay (Tỷ đồng)'] = data['Cho vay (Tỷ đồng)'].replace({r'[$,%]': ''}, regex=True).astype(float)

data['Nợ xấu (%)'] = data['Nợ xấu (%)'].replace({r'[$,%]': ''}, regex=True).astype(float)

data['Lợi nhuận thuần (Tỷ đồng)'] = data['Lợi nhuận thuần (Tỷ đồng)'].replace({r'[$,%]': ''}, regex=True).astype(float)

data['Tổng tài sản (Tỷ đồng)'] = data['Tổng tài sản (Tỷ đồng)'].replace({r'[$,%]': ''}, regex=True).astype(float)

# Đổi tên các cột thành tiếng Việt

# Đổi tên các cột thành tiếng Việt

data.rename(columns={

'Giá cổ phiếu trung bình(VNĐ)': 'Giá cổ phiếu (VNĐ)',

'Cho vay (Tỷ đồng)': 'Cho vay (tỷ đồng)',

'Nợ xấu (%)': 'Nợ xấu (%)',

'Lợi nhuận thuần (Tỷ đồng)': 'Lợi nhuận thuần (tỷ đồng)',

'Tổng tài sản (Tỷ đồng)': 'Tổng tài sản (tỷ đồng)'

}, inplace=True)

#### **3.2.2.1 Phân tích biểu đồ scatter plot matrix**

Code phân tích Scatter plot :

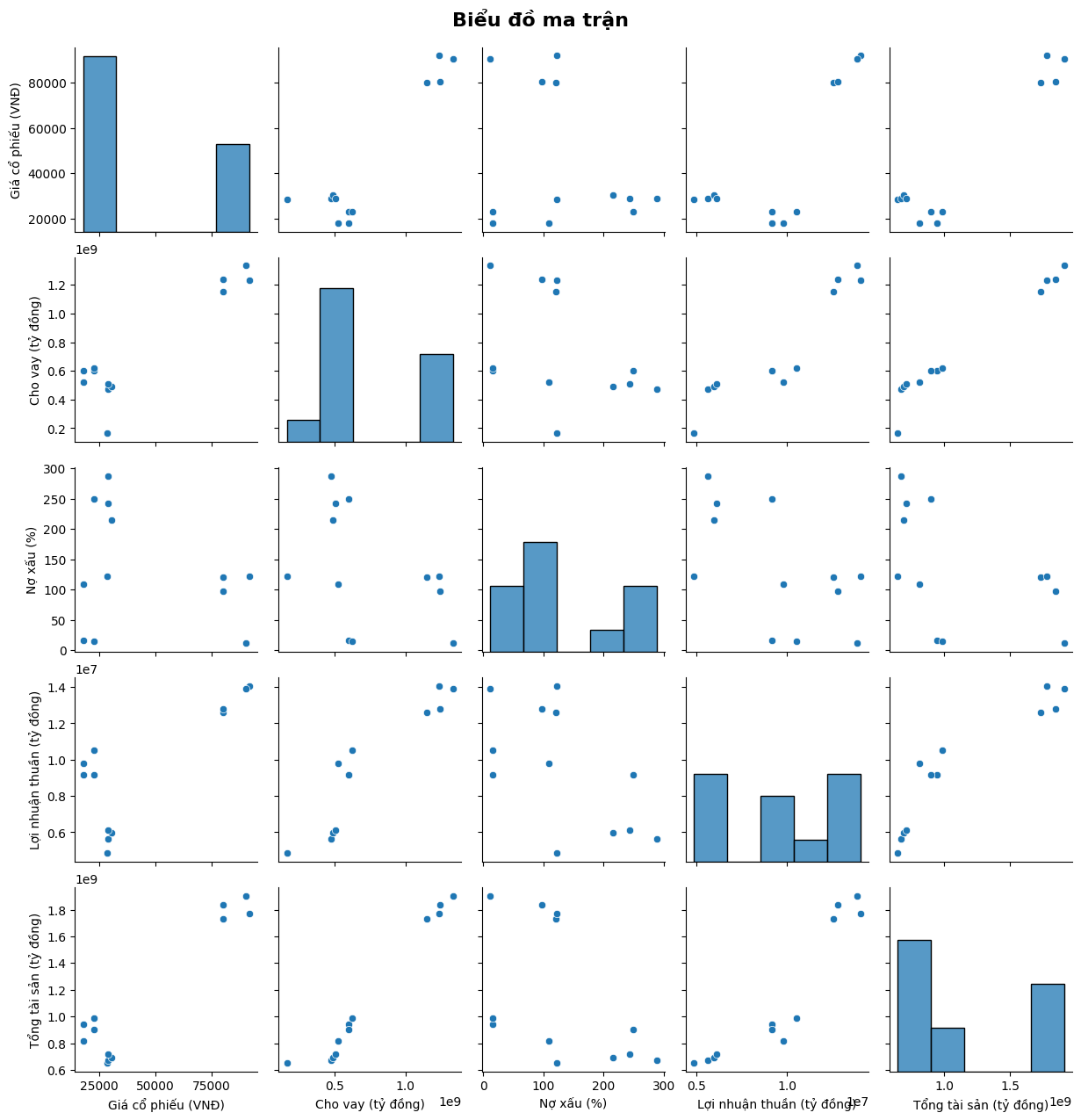
# Vẽ biểu đồ Scatter Plot

sns.pairplot(data[['Giá cổ phiếu (VNĐ)', 'Cho vay (tỷ đồng)',

                   'Nợ xấu (%)', 'Lợi nhuận thuần (tỷ đồng)', 'Tổng tài sản (tỷ đồng)']])

plt.suptitle("Biểu đồ ma trận", y=1.02, fontsize=16, fontweight='bold')

plt.show()



#### Hình 3.13 Biểu đồ ma trận

**Nhận xét:**

Phân tích biểu đồ ma trận

Biểu đồ ma trận cung cấp nhiều thông tin quan trọng về mối quan hệ giữa các biến và đặc điểm phân phối dữ liệu. Một số mối quan hệ có vẻ tuyến tính, chẳng hạn như giữa Lợi nhuận thuần (Tỷ đồng) và Tổng tài sản (Tỷ đồng), gợi ý rằng khi tổng tài sản tăng, lợi nhuận thuần cũng tăng. Tuy nhiên, các biến khác, như Giá cổ phiếu trung bình (VND), không thể hiện mối tương quan rõ ràng với các biến còn lại hoặc có mối quan hệ phi tuyến tính. Các histogram trên đường chéo của biểu đồ chỉ ra đặc điểm phân phối của từng biến, ví dụ, Giá cổ phiếu trung bình (VND) tập trung chủ yếu trong khoảng 25.000-75.000 VND, trong khi Cho vay (Tỷ đồng) và Tổng tài sản (Tỷ đồng) có xu hướng tập trung ở một số giá trị lớn. Biến Nợ xấu (%) có sự phân tán dữ liệu rộng với nhiều giá trị ngoại lệ vượt xa phần còn lại, tương tự như mối quan hệ giữa Cho vay (Tỷ đồng) và Tổng tài sản (Tỷ đồng), nơi một số điểm cao có thể ảnh hưởng mạnh đến kết quả phân tích. Một số biến, như Lợi nhuận thuần (Tỷ đồng) và Nợ xấu (%), có mức biến động lớn, cho thấy sự khác biệt đáng kể giữa các đối tượng trong tập dữ liệu. Đồng thời, nhiều cặp biến không cho thấy mối tương quan rõ ràng hoặc có tính không tuyến tính, ví dụ như giữa Giá cổ phiếu trung bình (VND) và Cho vay (Tỷ đồng).

#### **3.2.2.2. Phân tích biểu đồ ma trận tương quan**

Tổng quan về ý nghĩa biểu đồ Biểu đồ Ma trận tương quan thể hiện mối tương quan giữa các biến với nhau thông qua hệ số tương quan Pearson, dao động từ -1 đến 1:

-Giá trị dương cao (gần 1): Mối tương quan thuận mạnh, khi một biến tăng thì biến kia cũng tăng.

-Giá trị âm thấp (gần -1): Mối tương quan nghịch mạnh, khi một biến tăng thì biến kia giảm.

-Giá trị gần 0: Không có mối tương quan rõ ràng giữa hai biến.

# Tính toán ma trận tương quan

# Tính toán ma trận tương quan

correlation\_matrix = data[['Giá cổ phiếu trung bình(VNĐ)', 'Cho vay (tỷ đồng)',

'Nợ xấu (%)', 'Lợi nhuận thuần (tỷ đồng)', 'Tổng tài sản (tỷ đồng)']].corr()

# Vẽ heatmap cho ma trận tương quan

plt.figure(figsize=(10, 8))

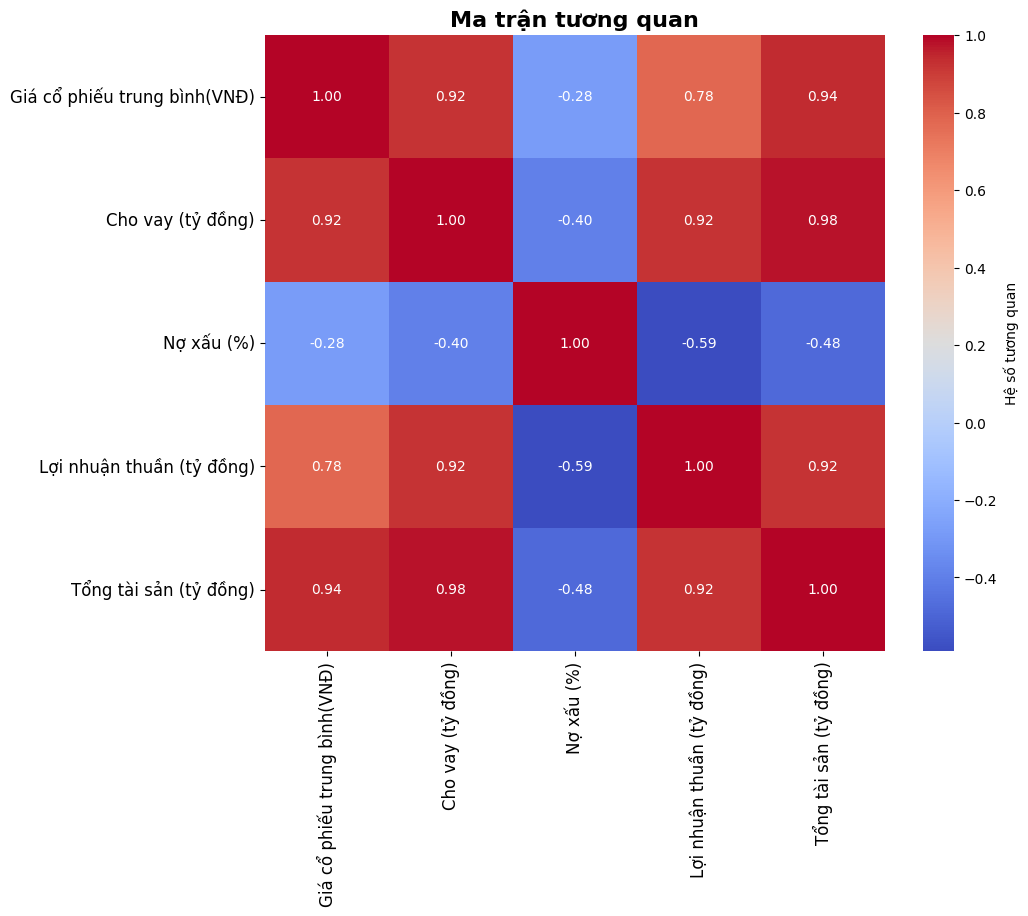
sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f", cbar\_kws={'label': 'Hệ số tương quan'})

plt.title("Ma trận tương quan", fontsize=16, fontweight='bold')

plt.xticks(fontsize=12)

plt.yticks(fontsize=12)

plt.show()



#### Hình 3.14 biểu đồ ma trận tương quan

**Nhận xét:**

Phân tích các mối tương quan giữa các biến trong dữ liệu cho thấy một số mối quan hệ quan trọng. Tổng tài sản (Tỷ đồng) và Cho vay (Tỷ đồng) có hệ số tương quan 0.98, thể hiện mối quan hệ tuyến tính rất mạnh, hợp lý vì tài sản thường bao gồm các khoản cho vay. Giá cổ phiếu trung bình (VND) và Tổng tài sản (Tỷ đồng) cũng có mối tương quan mạnh với hệ số 0.94, gợi ý rằng công ty có tổng tài sản lớn thường được thị trường định giá cao hơn. Tương tự, mối quan hệ giữa Cho vay (Tỷ đồng) và Lợi nhuận thuần (Tỷ đồng) với hệ số 0.92 cho thấy cho vay nhiều hơn thường đi kèm lợi nhuận cao hơn, dù cần xem xét thêm các yếu tố như nợ xấu. Mối quan hệ trung bình được ghi nhận giữa Giá cổ phiếu trung bình (VND) và Lợi nhuận thuần (Tỷ đồng) với hệ số tương quan 0.78, cho thấy công ty lợi nhuận cao thường có giá cổ phiếu cao. Ngoài ra, Nợ xấu (%) có mối quan hệ nghịch trung bình với Tổng tài sản (hệ số -0.48), phản ánh khả năng kiểm soát nợ tốt hơn ở các công ty lớn. Các mối quan hệ yếu hoặc không đáng kể bao gồm Nợ xấu (%) với Giá cổ phiếu trung bình (-0.28), và Nợ xấu (%) với Lợi nhuận thuần (-0.59), cho thấy tỷ lệ nợ xấu cao làm giảm lợi nhuận nhưng không ảnh hưởng rõ ràng đến giá cổ phiếu.

Từ phân tích này, có thể rút ra rằng Tổng tài sản và Lợi nhuận thuần là các yếu tố chính tác động đến giá cổ phiếu, do đó việc cải thiện các chỉ số này sẽ gia tăng giá trị thị trường. Mặt khác, nợ xấu ảnh hưởng tiêu cực đến lợi nhuận và tài sản, dù tác động lên giá cổ phiếu không rõ ràng. Vì vậy, việc kiểm soát nợ xấu là cần thiết để bảo vệ sức khỏe tài chính doanh nghiệp.

## **3.3 PHÂN TÍCH PCA**

Mục tiêu

PCA giảm số chiều của dữ liệu từ nhiều biến (cột trong dữ liệu gốc) xuống một số lượng nhỏ hơn (trong ví dụ này là 2), giúp chúng ta:

Giữ lại phần lớn thông tin quan trọng.

Loại bỏ nhiễu hoặc thông tin ít quan trọng.

Dễ dàng trực quan hóa dữ liệu trong không gian 2D hoặc 3D.

Phần trăm phương sai giải thích:

Thành phần chính 1 (PC1): Giải thích 82.80% phương sai của dữ liệu. Điều này có nghĩa là PC1 chứa hầu hết thông tin quan trọng từ dữ liệu gốc.

Thành phần chính 2 (PC2): Giải thích 12.45% phương sai của dữ liệu. PC2 thêm thông tin bổ sung không trùng lặp với PC1.

Tổng cộng hai thành phần chính giải thích 95.25% phương sai. Điều này cho thấy hai thành phần này đại diện tốt cho toàn bộ dữ liệu.

Ý nghĩa của các thành phần chính:

Principal Component 1 (PC1): Đây là trục chính mang phần lớn thông tin quan trọng. PC1 là sự kết hợp tuyến tính của tất cả các biến (các cột dữ liệu gốc). Các điểm dữ liệu có giá trị PC1 gần nhau thường có đặc điểm tương tự nhau về các biến trong dữ liệu gốc.

Principal Component 2 (PC2): Là trục thứ hai vuông góc với PC1. PC2 cung cấp thông tin bổ sung về sự khác biệt mà PC1 chưa giải thích.

\*Trung bình trượt:

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

# Đoc file csv

df1 = pd.read\_csv('/content/bang\_thong\_ke\_2.csv')

print(df1)

Ngân hàng Quý Ngày Mở cửa Đóng cửa KLDD khớp lệnh

0 MBB 2/2024 28/06/2024 22,55 22,20 9,571,200

1 MBB NaN 27/06/2024 22,45 22,50 6,168,000

2 MBB NaN 26/06/2024 22,60 22,50 11,084,000

3 MBB NaN 25/06/2024 22,70 22,60 8,882,600

4 MBB NaN 24/06/2024 23,10 22,65 25,433,000

.. ... ... ... ... ... ...

355 SCB NaN 08/01/2024 29,60 29,20 15,853,600

356 SCB NaN 05/01/2024 29,00 29,30 31,291,500

357 SCB NaN 04/01/2024 28,60 28,60 31,414,000

358 SCB NaN 03/01/2024 27,75 28,50 19,990,700

359 SCB NaN 02/01/2024 28,00 27,75 14,813,300

[360 rows x 6 columns]

# Thay thế giá trị NaN bằng giá trị phía trên nó

df1.fillna(method='ffill', inplace=True)

# Kiểm tra lại kết quả

print(df1)

Ngân hàng Quý Ngày Mở cửa Đóng cửa KLDD khớp lệnh

0 MBB 2/2024 28/06/2024 22,55 22,20 9,571,200

1 MBB 2/2024 27/06/2024 22,45 22,50 6,168,000

2 MBB 2/2024 26/06/2024 22,60 22,50 11,084,000

3 MBB 2/2024 25/06/2024 22,70 22,60 8,882,600

4 MBB 2/2024 24/06/2024 23,10 22,65 25,433,000

.. ... ... ... ... ... ...

355 SCB 1/2024 08/01/2024 29,60 29,20 15,853,600

356 SCB 1/2024 05/01/2024 29,00 29,30 31,291,500

357 SCB 1/2024 04/01/2024 28,60 28,60 31,414,000

358 SCB 1/2024 03/01/2024 27,75 28,50 19,990,700

359 SCB 1/2024 02/01/2024 28,00 27,75 14,813,300

[360 rows x 6 columns]

# Kiểm tra các dòng trùng lặp

print('Các dòng trùng lặp:')

print(df1.duplicated().sum())

Các dòng trùng lặp:

0

# Kiểm tra các giá trị bị thiếu

print(df1.isnull().sum())

Ngân hàng 0

Quý 0

Ngày 0

Mở cửa 0

Đóng cửa 0

KLDD khớp lệnh 0

dtype: int64

# Chuyển đổi cột "Đóng cửa": loại bỏ đơn vị tiền tệ và dấu phân cách, chuyển sang float

df1['Đóng cửa'] = df1['Đóng cửa'].str.replace(' đ', '').str.replace(',', '').astype(float)

# Hiển thị dữ liệu đã xử lý

print(df1)

Ngân hàng Quý Ngày Mở cửa Đóng cửa KLDD khớp lệnh

0 MBB 2/2024 28/06/2024 22,55 2220.0 9,571,200

1 MBB 2/2024 27/06/2024 22,45 2250.0 6,168,000

2 MBB 2/2024 26/06/2024 22,60 2250.0 11,084,000

3 MBB 2/2024 25/06/2024 22,70 2260.0 8,882,600

4 MBB 2/2024 24/06/2024 23,10 2265.0 25,433,000

.. ... ... ... ... ... ...

355 SCB 1/2024 08/01/2024 29,60 2920.0 15,853,600

356 SCB 1/2024 05/01/2024 29,00 2930.0 31,291,500

357 SCB 1/2024 04/01/2024 28,60 2860.0 31,414,000

358 SCB 1/2024 03/01/2024 27,75 2850.0 19,990,700

359 SCB 1/2024 02/01/2024 28,00 2775.0 14,813,300

[360 rows x 6 columns]

# Chuyển cột "Ngày" về kiểu datetime để vẽ biểu đồ chính xác

df1['Ngày'] = pd.to\_datetime(df1['Ngày'], format='%d/%m/%Y')

# Tính trung bình trượt 30 phiên cho từng ngân hàng

df1['Trung bình trượt 30 ngày'] = df1.groupby('Ngân hàng')['Đóng cửa'].transform(lambda x: x.rolling(window=30).mean())

# Tạo biểu đồ

plt.figure(figsize=(12, 6))

# Duyệt qua từng ngân hàng và vẽ biểu đồ đường

for bank, group in df1.groupby('Ngân hàng'):

plt.plot(group['Ngày'], group['Trung bình trượt 30 ngày'], label=bank)

# Thiết lập các thuộc tính biểu đồ

plt.title('Trung bình trượt 30 ngày của Đóng cửa theo từng ngân hàng năm 2024')

plt.xlabel('Thời gian')

plt.ylabel('Giá trung bình trượt (VND)')

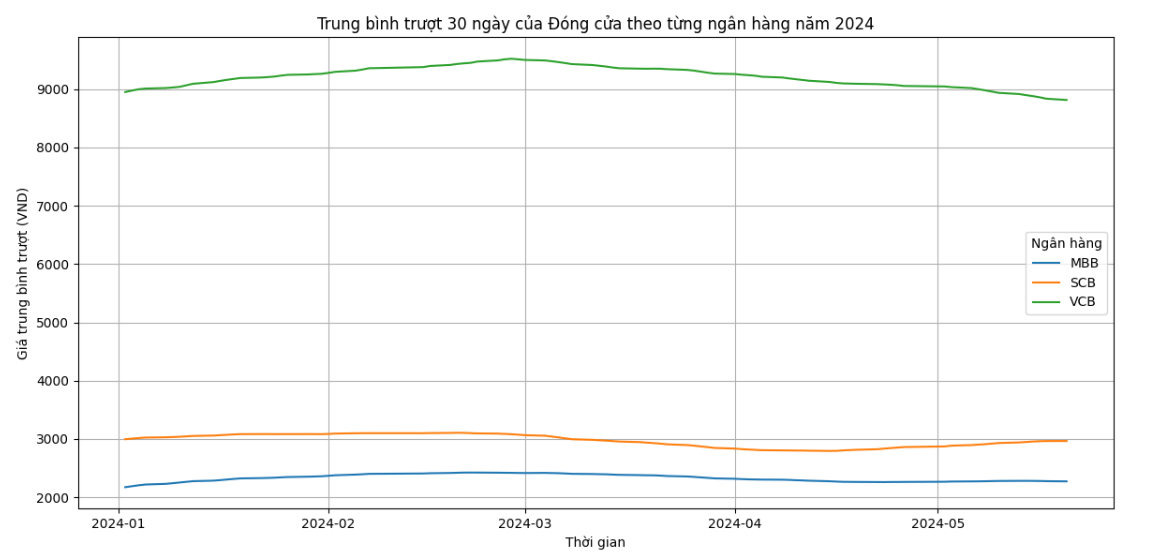
plt.legend(title='Ngân hàng')

plt.grid(True)

plt.tight\_layout()

# Hiển thị biểu đồ

plt.show()



#### Hình 3.15 Biểu đồ trung bình trượt 30 ngày của 3 ngân hàng

Phân tích thành phần chính :

Code:

# Chuẩn hóa dữ liệu (loại bỏ cột không cần thiết và chuẩn hóa các giá trị)

features = ["Giá cổ phiếu trung bình(VNĐ)", "Cho vay (Tỷ đồng)", "Nợ xấu (%)", "Lợi nhuận thuần (Tỷ đồng)", "Tổng tài sản (Tỷ đồng)"]

x = df[features]

x = StandardScaler().fit\_transform(x)

# Áp dụng PCA

pca = PCA(n\_components=2)  # Chọn 2 thành phần chính

principal\_components = pca.fit\_transform(x)

# Đưa kết quả vào DataFrame

pca\_df = pd.DataFrame(data=principal\_components, columns=['PC1', 'PC2'])

pca\_df['Ngân Hàng'] = df['Ngân Hàng']

pca\_df['Quý'] = df['Quý']

# In kết quả

print("Explained Variance Ratio (Tỷ lệ phương sai giải thích):", pca.explained\_variance\_ratio\_)

print("\nDữ liệu PCA:")

print(pca\_df)

# Vẽ biểu đồ PCA

plt.figure(figsize=(10, 6))

for bank in df["Ngân Hàng"].unique():

    subset = pca\_df[pca\_df["Ngân Hàng"] == bank]

    plt.scatter(subset["PC1"], subset["PC2"], label=bank)

plt.title('PCA ')

plt.xlabel('PC1')

plt.ylabel('PC2')

plt.legend()

plt.grid()

plt.show()

Kết quả:

Explained Variance Ratio (Tỷ lệ phương sai giải thích): [0.82848603 0.12568952]

Dữ liệu PCA:

PC1 PC2 Ngân Hàng Quý

0 -0.678526 -1.349473 Mbank 3\_2023

1 -0.823490 -0.503384 Mbank 4\_2023

2 -1.271550 0.834948 Mbank 1\_2024

3 -0.420474 -0.588290 Mbank 2\_2024

4 -1.929156 -1.345651 Sacombank 3\_2023

5 -2.291218 1.334891 Sacombank 4\_2023

6 -1.782741 0.296414 Sacombank 1\_2024

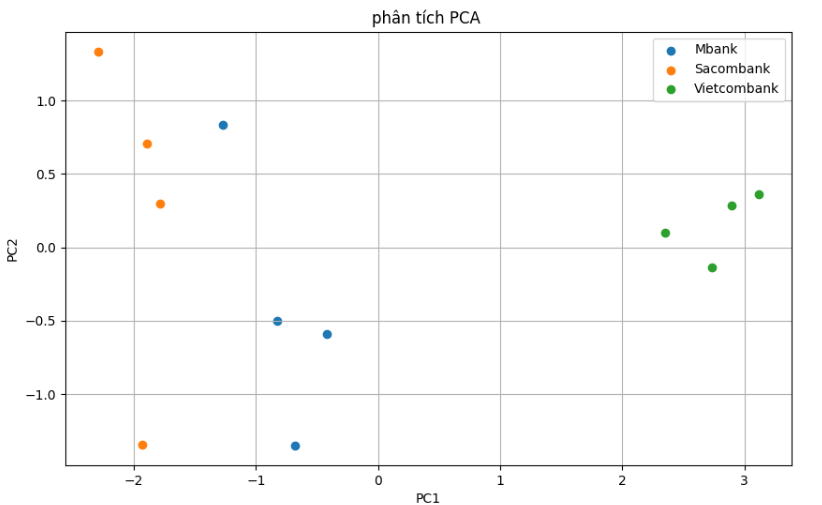
7 -1.891039 0.709164 Sacombank 2\_2024

8 2.347692 0.100073 Vietcombank 3\_2023

9 2.731433 -0.135562 Vietcombank 4\_2023

10 2.893552 0.287545 Vietcombank 1\_2024

11 3.115518 0.359326 Vietcombank 2\_2024



#### Hình 3.16. Biểu đồ phân tích PCA

**Nhận xét:**

Tỷ lệ phương sai :

Phân tích PCA cho thấy 82.85% phương sai được giữ lại ở thành phần chính đầu tiên (PC1) và 12.57% ở thành phần thứ hai (PC2). Tổng cộng, hai thành phần chính này đã giữ lại 95.42% thông tin từ dữ liệu gốc. Điều này chứng tỏ rằng việc sử dụng PCA để giảm chiều dữ liệu đã hiệu quả trong việc nắm bắt phần lớn thông tin quan trọng.

Phân bố của các ngân hàng:

MBBank: Các điểm dữ liệu của Mbank nằm ở vùng có giá trị âm trên cả PC1 và PC2. Điều này cho thấy Mbank có những đặc điểm khác biệt rõ rệt so với Sacombank và Vietcombank, đặc biệt là về các chỉ số như dư nợ cho vay, tổng tài sản, hoặc lợi nhuận. Mặc dù có sự thay đổi nhỏ giữa các quý, các điểm dữ liệu vẫn nằm gần nhau, thể hiện sự ổn định trong hoạt động của ngân hàng này.

Sacombank: Các điểm dữ liệu của Sacombank chủ yếu tập trung ở vùng có PC1 âm và PC2 gần 0 hoặc dương. Điều này cho thấy Sacombank có đặc điểm trung hòa giữa quy mô và hiệu quả hoạt động, không quá vượt trội nhưng cũng không thua kém. Các quý có sự phân tán nhẹ, điều này gợi ý rằng ngân hàng có một số thay đổi theo thời gian nhưng vẫn trong một phạm vi tương đối ổn định.

Vietcombank: Các điểm dữ liệu của Vietcombank tập trung ở vùng giá trị dương cao của PC1, tách biệt rõ ràng so với MBBank và Sacombank. Điều này cho thấy Vietcombank có quy mô lớn hơn và hiệu quả vượt trội hơn, đặc biệt ở các chỉ số như lợi nhuận thuần và tổng tài sản. Các điểm dữ liệu của Vietcombank nằm gần nhau, thể hiện sự ổn định đáng kể qua các quý.

**Ý nghĩa của các thành phần chính**

Thành phần chính đầu tiên (PC1), chiếm 82.85% phương sai, có thể đại diện cho các yếu tố liên quan đến quy mô tài sản và hiệu quả hoạt động, chẳng hạn như tổng tài sản và lợi nhuận thuần. Vietcombank có giá trị PC1 cao nhất, cho thấy sự vượt trội về quy mô và hiệu quả hoạt động. Thành phần chính thứ hai (PC2), chiếm 12.57% phương sai, có thể phản ánh các yếu tố liên quan đến chất lượng tín dụng hoặc mức độ rủi ro, như tỷ lệ nợ xấu. Sự khác biệt ở PC2 cho thấy MBBank và Sacombank có thể có những thay đổi trong chất lượng tín dụng hoặc mức độ rủi ro giữa các quý.

**Kết luận**

Từ phân tích trên, có thể rút ra rằng Vietcombank dẫn đầu rõ rệt về quy mô và hiệu quả hoạt động với sự ổn định qua các quý. MBBank có quy mô nhỏ hơn và hoạt động khác biệt đáng kể so với các ngân hàng còn lại, gợi ý rằng ngân hàng này vẫn đang trong giai đoạn cải thiện hiệu quả. Sacombank có vị trí trung gian, vừa phải về quy mô và hiệu quả, với một số thay đổi nhẹ qua các quý.

# **CHƯƠNG 4: TỔNG KẾT BÁO CÁO**

## **4.1. MỤC TIÊU VÀ PHẠM VI CỦA ĐỒ ÁN**

Mục tiêu chính của đồ án là sử dụng thành thạo các công cụ phân tích và trực quan hóa dữ liệu bằng ngôn ngữ lập trình Python, thông qua việc phân tích bộ dữ liệu “bang\_thong\_kê.csv” được thu thập trên các trang web chính thức của các ngân hàng và các trang sàn giao dịch. Phạm vi nghiên cứu bao gồm việc thu thập, làm sạch dữ liệu, thăm dò dữ liệu, trực quan hóa các phát hiện quan trọng bằng biểu đồ và đồ thị, cũng như xây dựng mô hình dự đoán. Qua đó, giúp chúng ta hiểu rõ hơn về các yếu tố ảnh hưởng đến giao dịch khoản vay của ba ngân hàng MBbank, Vietcombank, Sacombank.

## **4.2. QUY TRÌNH VÀ PHƯƠNG PHÁP SỬ DỤNG**

Quá trình thực hiện đồ án bắt đầu bằng việc thu thập dữ liệu từ trang web chính thức của các ngân hàng, nơi cung cấp một bộ dữ liệu phong phú và đa dạng về giao dịch khoản vay . Bộ dữ liệu này bao gồm nhiều biến số liên quan đến giao dịch khoản vay, như giá cổ phiếu trung bình, cho vay, nợ xấu, lợi nhuận thuần và tổng tài sản.

Sau khi thu thập dữ liệu, bước tiếp theo là làm sạch và xử lý dữ liệu. Quá trình này bao gồm việc kiểm tra và loại bỏ các giá trị thiếu, giá trị ngoại lệ và các bản ghi bị lỗi. Do dữ liệu tìm được khá sạch và chuẩn chỉ còn một số định dạng cần điều chỉnh cho phù hợp trong việc kiểm tra chuẩn hóa và trực quan dữ liệu. Đây là một bước quan trọng để đảm bảo rằng dữ liệu được chuẩn bị kỹ lưỡng trước khi tiến hành các phân tích sâu hơn.

Để có cái nhìn tổng quát hơn về dữ liệu sau khi đã làm sạch dữ liệu, chúng tôi sử dụng các hàm thống kê và một số biểu đồ đơn giản. Tiếp theo, chúng tôi sử dụng các công cụ và thư viện mạnh mẽ của ngôn ngữ lập trình Python, như pandas, matplotlib, seaborn, và nhiều thư viện khác, để tiến hành thăm dò và trực quan hóa dữ liệu. Matplotlib và seaborn là một trong những công cụ mạnh mẽ nhất để tạo ra các biểu đồ và đồ thị chất lượng cao. Chúng tôi sử dụng nó để tạo ra các biểu đồ tán xạ, biểu đồ hộp, biểu đồ cột, biểu đồ đường, và nhiều loại biểu đồ khác. Mục đích là để minh họa rõ ràng mối quan hệ giữa các biến số và cung cấp cái nhìn sâu sắc về dữ liệu. Các biểu đồ và đồ thị này giúp ta nhận diện các xu hướng, mẫu hình và mối quan hệ tiềm năng giữa các yếu tố khác nhau ảnh hưởng đến giao dịch khoản vay của ba ngân hàng Mbbank, Vietcombank, Sacombank.

Sau khi có được cái nhìn tổng quan từ các phân tích thăm dò, chúng tôi tiến hành xây dựng một mô hình cây quyết định để dự đoán giao dịch khoản vay của các ngân hàng. Quá trình này bắt đầu bằng việc chọn lựa các biến số quan trọng và xây dựng mô hình phân loại dựa trên cây quyết định. Chúng tôi tinh chỉnh các tham số của mô hình và đánh giá hiệu suất của nó bằng các chỉ số như ma trận nhầm lẫn độ chính xác, độ nhạy và độ chính xác cân bằng để đảm bảo rằng mô hình đáp ứng được yêu cầu của bài toán phân tích giao dịch khoản vay của ba ngân hàng Mbbank, Vietcombank, Sacombank.

Cuối cùng, chúng tôi so sánh và phân tích kết quả từ các mô hình khác nhau để hiểu rõ hơn về các yếu tố ảnh hưởng đến giao dịch khoản vay của ba ngân hàng Mbbank, Vietcombank, Sacombank. Kết quả phân tích cho thấy Vietcombank là ngân hàng có hiệu quả quản lý tín dụng tốt nhất với tỷ lệ nợ xấu thấp và ổn định. Sacombank có sự biến động lớn về tỷ lệ nợ xấu, cần tập trung nhiều hơn vào việc kiểm soát rủi ro tín dụng. MBbank có tỷ lệ nợ xấu ổn định ở mức trung bình cần cải thiện thêm để đạt được mức độ an toàn cao hơn.Quá trình này không chỉ giúp chúng tôi đạt được mục tiêu ban đầu của đồ án mà còn mở ra nhiều hướng nghiên cứu mới để tiếp tục khám phá các yếu tố ảnh hưởng đến giao dịch khoản vay của ba ngân hàng MBBank, Vietcombank, Sacombank.

## **4.3. ĐƯA RA KẾT QUẢ CHÍNH VÀ PHÁT HIỆN**

Kết quả phân tích cho thấy giao dịch khoản vay của ba ngân hàng MBBank, Vietcombank và Sacombank, tỷ lệ nợ xấu là một yếu tố quan trọng ảnh hưởng trực tiếp đến kết quả tài chính của từng ngân hàng. Vietcombank duy trì tỷ lệ nợ xấu thấp, dao động từ 1.09% (quý 3, 2023) đến 1.5% (quý 2, 2024), cho thấy khả năng quản lý tín dụng hiệu quả. Trong khi đó, Sacombank có tỷ lệ nợ xấu cao hơn, dao động từ 1.22% đến 2.43%, điều này phản ánh việc ngân hàng gặp phải nhiều thách thức trong việc kiểm soát chất lượng tín dụng. MBBANK có tỷ lệ nợ xấu dao động từ 1.09% đến 2.49%, với sự biến động khá lớn giữa các quý.

Về tổng số cho vay, Vietcombank có tổng dư nợ cho vay cao nhất trong ba ngân hàng, từ 1,150,000 tỷ đồng (quý 3, 2023) đến 1,334,000 tỷ đồng (quý 2, 2024), trong khi MBBank có mức tăng trưởng cho vay ổn định, từ 523,969 tỷ đồng (quý 3, 2023) lên 622,585 tỷ đồng (quý 2, 2024). Sacombank có mức cho vay từ 451,000 tỷ đồng đến 563,000 tỷ đồng trong các quý, với sự tăng trưởng chậm hơn so với hai ngân hàng còn lại. Kết quả phân tích cũng chỉ ra rằng mối quan hệ giữa cho vay và giá cổ phiếu có sự khác biệt. Vietcombank, mặc dù có mức cho vay cao và ổn định, lại có giá cổ phiếu tăng trưởng mạnh mẽ, từ 80,000 VND (quý 3, 2023) lên 91,830 VND (quý 1, 2024). Trong khi đó, MBBank, dù có mức cho vay tăng đáng kể, lại chỉ ghi nhận mức giá cổ phiếu dao động từ 17,950 VND đến 22,970 VND, không tăng trưởng mạnh như Vietcombank. Sacombank có mức giá cổ phiếu dao động từ 28,500 VND đến 29,000 VND, thấp hơn nhiều so với hai ngân hàng còn lại.

Những phát hiện này cho thấy rằng việc duy trì tỷ lệ nợ xấu thấp và chiến lược cho vay hiệu quả là những yếu tố quan trọng giúp các ngân hàng như Vietcombank duy trì sự ổn định và gia tăng giá trị cổ phiếu. Trong khi đó, các ngân hàng như Sacombank và MBBANK cần cải thiện các chiến lược tín dụng và quản lý rủi ro để nâng cao khả năng sinh lời và giảm thiểu rủi ro tài chính.

## **4.4. THẢO LUẬN VỀ Ý NGHĨA VÀ ỨNG DỤNG KẾT QUẢ**

Kết quả phân tích giao dịch khoản vay của ba ngân hàng MBBank, Vietcombank và Sacombank có nhiều ứng dụng thực tiễn quan trọng. Đối với nhà đầu tư, kết quả giúp đánh giá sức khỏe tài chính và quyết định đầu tư, trong khi đối với các ngân hàng, nó hỗ trợ điều chỉnh chiến lược tín dụng và cải thiện hiệu quả quản lý rủi ro. Các cơ quan quản lý có thể dựa trên kết quả để phát triển chính sách hỗ trợ ngân hàng, trong khi khách hàng vay được hưởng lợi từ các gói vay ưu đãi hơn. Các nhà phân tích tài chính có thể sử dụng thông tin này để dự báo xu hướng phát triển của thị trường tín dụng.

## **4.5. NÊU RA CÁC HẠN CHẾ VÀ ĐỀ XUẤT HƯỚNG PHÁT TRIỂN**

Mặc dù báo cáo đã cung cấp những kết quả quan trọng về phân tích giao dịch khoản vay của ba ngân hàng MBBank, Vietcombank và Sacombank, nhưng vẫn còn một số hạn chế như dữ liệu được sử dụng trong nghiên cứu còn chưa đầy đủ, đặc biệt là thiếu thông tin về các yếu tố kinh tế vĩ mô, chính sách tín dụng cụ thể của từng ngân hàng và tình hình tài chính của khách hàng vay, điều này có thể làm giảm độ chính xác trong phân tích và dự báo. Phạm vi nghiên cứu chỉ tập trung vào ba ngân hàng lớn, trong khi việc mở rộng nghiên cứu ra các ngân hàng khác hoặc so sánh với các ngân hàng quốc tế có thể mang lại cái nhìn tổng thể hơn về ngành ngân hàng và các chiến lược quản lý rủi ro. Sự biến động mạnh mẽ của thị trường chứng khoán và các yếu tố vĩ mô như lãi suất, chính sách nhà nước hay các sự kiện quốc tế có thể ảnh hưởng đến kết quả phân tích mà không thể dự đoán hết được.

Để khắc phục những hạn chế này, trong tương lai, nghiên cứu có thể mở rộng phạm vi dữ liệu bằng cách thu thập thông tin từ các ngân hàng nhỏ và vừa, cũng như bổ sung các yếu tố vĩ mô như tỷ lệ lạm phát và lãi suất để làm phong phú thêm phân tích. Ngoài ra, ứng dụng các công nghệ phân tích dữ liệu tiên tiến như học máy (machine learning) và trí tuệ nhân tạo (AI) sẽ giúp xây dựng các mô hình dự báo nợ xấu và đánh giá chất lượng tín dụng chính xác hơn. Các mô hình này có thể giúp các ngân hàng quản lý rủi ro hiệu quả hơn và đưa ra quyết định cho vay đúng đắn. Bên cạnh đó, ngân hàng cần tăng cường đầu tư vào các công cụ quản lý rủi ro tín dụng để theo dõi và kiểm soát tỷ lệ nợ xấu, giảm thiểu rủi ro tài chính. Các nhà đầu tư và ngân hàng cần hướng đến các chiến lược tài chính bền vững, chú trọng vào việc giảm thiểu rủi ro dài hạn và duy trì khả năng sinh lời ổn định thay vì chỉ tập trung vào các yếu tố ngắn hạn.

# **TÀI LIỆU THAM KHẢO**

[1]. Cao Diệp Thắng, Tài liệu học tập Tin học cơ sở: Trực quan hóa dữ liệu bằng Python, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.

[2]. Cao Diệp Thắng, Đỗ Tuấn Hạnh, Tài liệu học tập Python nâng cao, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.

[3]. Trần Thị Kim Thanh, Trần Chí Lê (2023), Tài liệu học tập Thống kê Toán học cho ngành Khoa học dữ liệu, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.

[4]. Link nguồn dữ liệu:

<https://finance.vietstock.vn/STB-ngan-hang-tmcp-sai-gon-thuong-tin.htm>

<https://finance.vietstock.vn/MBB-ngan-hang-tmcp-quan-doi.htm>

<https://finance.vietstock.vn/VCB-ngan-hang-tmcp-ngoai-thuong-viet-nam.htm>

[5]. Trần Thị Hoàng Yến, Bùi Văn Tân, Chu Bình Minh (2024), Tài liệu học tập Nhập môn Trí tuệ Nhân tạo, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.

[6]. Hỗ trợ của Chat GPT 4.0

Link: https://chatgpt.com/