TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**CAO ĐĂNG TÌNH - 52100936**

**TRƯƠNG THÁI ĐAN HUY – 52100222**

**CAO HOÀNG OANH - 52100917**

**PHÂN TÍCH HÀNH VI KHÁCH HÀNG**

**BÁO CÁO CUỐI KÌ**

**HỆ THỐNG THƯƠNG MẠI THÔNG MINH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**CAO ĐĂNG TÌNH - 52100936**

**TRƯƠNG THÁI ĐAN HUY – 52100222**

**CAO HOÀNG OANH - 52100917**

**PHÂN TÍCH HÀNH VI KHÁCH HÀNG**

**BÁO CÁO CUỐI KÌ**

**HỆ THỐNG THƯƠNG MẠI THÔNG MINH**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Trước tiên, em xin gửi lời cảm ơn chân thành và lòng biết ơn sâu sắc đến ***ThS Dương Hữu Phúc***. Thầy là người đã luôn hỗ trợ và hướng dẫn tận tình cho em trong suốt quá trình nghiên cứu và hoàn thành đồ án.

Tiếp theo, em xin gửi lời cảm ơn đến khoa Công Nghệ Thông Tin trường Đại học Tôn Đức Thắng vì đã tạo điều kiện cho em được học tập và nghiên cứu môn học này. Khoa đã luôn sẵn sàng chia sẻ các kiến thức bổ ích cũng như chia sẻ các kinh nghiệm tham khảo tài liệu, giúp ích không chỉ cho việc thực hiện và hoàn thành đề tài nghiên cứu mà còn giúp ích cho việc học tập và rèn luyện trong quá trình thực hành tại trường ***Đại học Tôn Đức Thắng*** nói chung.

Cuối cùng, sau khoảng thời gian học tập trên lớp em đã hoàn tất đề tài nghiên cứu nhờ vào sự hướng dẫn, giúp đỡ và những kiến thức học hỏi được từ thầy. Do giới hạn về mặt kiến thức và khả năng lý luận nên nhóm em vẫn còn nhiều thiếu sót và hạn chế, kính mong sự chỉ dẫn và đóng góp của thầy để bài nghiên cứu của em được hoàn thiện hơn. Hơn nữa, nhờ những góp ý từ thầy cô và các bạn hữu, em sẽ hoàn thành tốt hơn ở những bài nghiên cứu trong tương lai. Em mong thầy và các bạn bè – những người luôn quan tâm và hỗ trợ em – luôn tràn đầy sức khỏe và sự bình an.

**NHÓM NGHIÊN CỨU XIN CHÂN THÀNH CẢM ƠN !**

*TP. Hồ Chí Minh, ngày 26 tháng 11 năm 2024.*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Cao Hoàng Oanh*

*Trương Thái Đan Huy*

*Cao Đăng Tình*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ***ThS Dương Hữu Phúc***. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 26 tháng 11 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Cao Hoàng Oanh*

*Trương Thái Đan Huy*

*Cao Đăng Tình*

**PHÂN TÍCH HÀNH VI KHÁCH HÀNG**

# TÓM TẮT ĐỀ TÀI

Phân tích hành vi khách hàng là một trong những lĩnh vực quan trọng của *Business Intelligence (BI),* tập trung vào việc thu thập, xử lý và khai thác dữ liệu để hiểu rõ hơn về sở thích, hành vi và nhu cầu của khách hàng. Với đề tài này cốt yếu ta cần phải thu thập dữ liệu từ nhiều nguồn như giao dịch mua bán, tương tác trực tuyến, khảo sát, và mạng xã hội. Tiếp theo, phát hiện các mẫu hành vi phổ biến (behavioral patterns) như thói quen mua sắm, sự trung thành và phản hồi đối với sản phẩm. Việc thực hành đề tài này sẽ giúp ta hiểu rõ hơn về khách hàng mục tiêu, giúp doanh nghiệp xây dựng các chiến lược cá nhân hóa hiệu quả hơn. Ngoài ra ta có thể ứng dụng phân tích dữ liệu để tối ưu hóa các chiến lược tiếp thị, cá nhân hóa trải nghiệm khách hàng, và xây dựng chương trình khách hàng thân thiết để tăng hiệu quả tiếp cận và doanh thu

**MỤC LỤC**

[TÓM TẮT ĐỀ TÀI iii](#_Toc183797796)

[DANH MỤC HÌNH ẢNH ix](#_Toc183797797)

[DANH MỤC BẢNG BIỂU xi](#_Toc183797798)

[CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1](#_Toc183797799)

[1.1 Lý do chọn đề tài 1](#_Toc183797800)

[1.2 Mục tiêu thực hiện đề tài 1](#_Toc183797801)

[CHƯƠNG 2. XỬ LÝ DỮ LIỆU 3](#_Toc183797802)

[2.1 Cài đặt thư viện 4](#_Toc183797803)

[2.2 Xử lý dữ liệu 4](#_Toc183797804)

[CHƯƠNG 3. TRỰC QUAN HÓA DỮ LIỆU 10](#_Toc183797805)

[3.1 Phân tích xu hướng sử dụng ưu đãi giảm giá. 10](#_Toc183797806)

[3.1.1 Giới thiệu biểu đồ 12](#_Toc183797807)

[3.1.2 Phân tích biểu đồ 12](#_Toc183797808)

[3.1.3 Kết luận 13](#_Toc183797809)

[3.2 Phân tích sở thích mua sắm theo từng nhóm tuổi dựa trên sản phẩm. 13](#_Toc183797810)

[3.2.1 Giới thiệu biểu đồ 15](#_Toc183797811)

[3.2.2 Phân tích biểu đồ 15](#_Toc183797812)

[3.2.3 Kết luận 16](#_Toc183797813)

[3.3 Phân tích đánh giá trung bình về sản phẩm cho từng thành phố. 16](#_Toc183797814)

[3.3.1 Giới thiệu biểu đồ 17](#_Toc183797815)

[3.3.2 Phân tích biểu đồ 18](#_Toc183797816)

[3.3.3 Kết luận 19](#_Toc183797817)

[3.4 Phân tích tần suất sử dụng ưu đãi theo giới tính. 19](#_Toc183797818)

[3.4.1 Giới thiệu biểu đồ 21](#_Toc183797819)

[3.4.2 Phân tích biểu đồ 21](#_Toc183797820)

[3.4.3 Kết luận 22](#_Toc183797821)

[3.5 Phân tích tổng doanh thu của sản phẩm trong các khu vực. 22](#_Toc183797822)

[3.5.1 Giới thiệu biểu đồ 24](#_Toc183797823)

[3.5.2 Phân tích biểu đồ 24](#_Toc183797824)

[3.5.3 Kết luận 24](#_Toc183797825)

[3.6 Phân tích số lượng sản phẩm được mua theo từng mùa. 25](#_Toc183797826)

[3.6.1 Giới thiệu biểu đồ 27](#_Toc183797827)

[3.6.2 Phân tích biểu đồ 27](#_Toc183797828)

[3.6.3 Kết luận 28](#_Toc183797829)

[3.7 Phân tích tổng doanh thu theo từng bang trong khu vực Midwest. 29](#_Toc183797830)

[3.7.1 Giới thiệu bản đồ 29](#_Toc183797831)

[3.7.2 Phân tích biểu đồ 29](#_Toc183797832)

[3.7.3 Kết luận 30](#_Toc183797833)

[3.8 Phân tích tổng doanh thu theo từng bang trong khu vực Northeast. 30](#_Toc183797834)

[3.8.1 Giới thiệu biểu đồ 31](#_Toc183797835)

[3.8.2 Phân tích biểu đồ 31](#_Toc183797836)

[3.8.3 Kết luận 32](#_Toc183797837)

[3.9 Phân tích tổng doanh thu theo từng bang trong khu vực South. 32](#_Toc183797838)

[3.9.1 Tổng quan biểu đồ 33](#_Toc183797839)

[3.9.2 Phân tích biểu đồ 33](#_Toc183797840)

[3.9.3 Kết luận 34](#_Toc183797841)

[3.10 Phân tích tổng doanh thu theo từng bang trong khu vực West. 34](#_Toc183797842)

[3.10.1 Tổng quan biểu đồ 35](#_Toc183797843)

[3.10.2 Phân tích biểu đồ 35](#_Toc183797844)

[3.10.3 Kết luận 36](#_Toc183797845)

[3.11 So sánh giá trị mua hàng tối thiểu và giá trị mua hàng tối đa cho từng danh mục sản phẩm. 36](#_Toc183797846)

[3.11.1 Giới thiệu biểu đồ 37](#_Toc183797847)

[3.11.2 Phân tích biểu đồ 37](#_Toc183797848)

[3.11.3 Kết luận 38](#_Toc183797849)

[3.12 Phân tích mức độ sử dụng mã khuyến mãi dựa trên giới tính. 38](#_Toc183797850)

[3.12.1 Giới thiệu biểu đồ 40](#_Toc183797851)

[3.12.2 Phân tích biểu đồ 40](#_Toc183797852)

[3.12.3 Kết luận 40](#_Toc183797853)

[3.13 Phân tích tổng số tiền chi tiêu theo từng bang. 41](#_Toc183797854)

[3.13.1 Giới thiệu biểu đồ 41](#_Toc183797855)

[3.13.2 Phân tích biểu đồ 42](#_Toc183797856)

[3.13.3 Phân tích sâu hơn 42](#_Toc183797857)

[3.13.4 Kết luận 42](#_Toc183797858)

[3.14 Mối quan hệ giữa mùa thu và các sản phẩm quần áo được mua 42](#_Toc183797859)

[3.14.1 Giới thiệu biểu đồ 43](#_Toc183797860)

[3.14.2 Phân tích biểu đồ 43](#_Toc183797861)

[3.14.3 Xu hướng dự đoán 44](#_Toc183797862)

[3.14.4 Kết luận 44](#_Toc183797863)

[3.15 Mối quan hệ giữa mùa xuân và các sản phẩm được mua. 44](#_Toc183797864)

[3.15.1 Giới thiệu biểu đồ 45](#_Toc183797865)

[3.15.2 Phân tích biểu đồ 45](#_Toc183797866)

[3.15.3 Xu hướng dự đoán 46](#_Toc183797867)

[3.15.4 Kết luận 46](#_Toc183797868)

[3.16 Mối quan hệ giữa mùa hè và các sản phẩm được mua. 46](#_Toc183797869)

[3.16.1 Giới thiệu biểu đồ 47](#_Toc183797870)

[3.16.2 Phân tích biểu đồ 47](#_Toc183797871)

[3.16.3 Xu hướng dự đoán 48](#_Toc183797872)

[3.16.4 Kết luận 48](#_Toc183797873)

[3.17 Mối quan hệ giữa mùa đông và các sản phẩm được mua. 49](#_Toc183797874)

[3.17.1 Tổng quan biểu đồ 49](#_Toc183797875)

[3.17.2 Phân tích biểu đồ 49](#_Toc183797876)

[3.17.3 Xu hướng dự đoán 50](#_Toc183797877)

[3.17.4 Kết luận 50](#_Toc183797878)

[3.18 Phân tích số lượng kích cỡ theo mùa và giới tính. 51](#_Toc183797879)

[3.18.1 Tổng quan biểu đồ 53](#_Toc183797880)

[3.18.2 Phân tích biểu đồ 53](#_Toc183797881)

[3.18.3 Dự đoán xu hướng mua sắm 54](#_Toc183797882)

[3.18.4 Kết luận 54](#_Toc183797883)

[3.19 Mối quan hệ giữa các khu vực địa lý và sản phẩm. 54](#_Toc183797884)

[3.19.1 Tổng quan biểu đồ 56](#_Toc183797885)

[3.19.2 Phân tích biểu đồ 56](#_Toc183797886)

[3.19.3 Kết luận 56](#_Toc183797887)

[3.20 Phân tích mùa mua sắm tốt nhất. 57](#_Toc183797888)

[3.20.1 Tổng quan biểu đồ 57](#_Toc183797889)

[3.20.2 Phân tích biểu đồ 57](#_Toc183797890)

[3.20.3 Kết luận 58](#_Toc183797891)

[3.21 Dự đoán số lượng quần áo được dự trữ theo kích cỡ 58](#_Toc183797892)

[3.21.1 Tổng quan biểu đồ 60](#_Toc183797893)

[3.21.2 Phân tích biểu đồ 60](#_Toc183797894)

[3.21.3 Kết luận 61](#_Toc183797895)

[CHƯƠNG 4. TỔNG KẾT 61](#_Toc183797896)

[4.1 Tổng kết đề tài 61](#_Toc183797897)

[4.1.1 Mùa ảnh hưởng đáng kể đến xu hướng tiêu dùng 61](#_Toc183797898)

[4.1.2 Khác biệt trong hành vi mua sắm theo giới tính 61](#_Toc183797899)

[4.1.3 Sự đa dạng và nhu cầu về kích thước sản phẩm 62](#_Toc183797900)

[4.1.4 Hiệu quả của mã khuyến mãi theo mùa 62](#_Toc183797901)

[4.2 Bài học thông qua đề tài 62](#_Toc183797902)

[4.2.1 Hiểu sâu hơn về nhu cầu của khách hàng 62](#_Toc183797903)

[4.2.2 Tăng cường chiến lược khuyến mãi 62](#_Toc183797904)

[TÀI LIỆU THAM KHẢO 63](#_Toc183797905)

# DANH MỤC HÌNH ẢNH

[Ảnh 1 : Cài đặt thư viện 4](#_Toc183750549)

[Ảnh 2 : Hiển thị 3 dòng dữ liệu đầu tiên 4](#_Toc183750550)

[Ảnh 3 : Mở rộng dữ liệu 5](#_Toc183750551)

[Ảnh 4 : Thêm một cột id 5](#_Toc183750552)

[Ảnh 5 : Hiển thị các giá trị trong cột Location 6](#_Toc183750553)

[Ảnh 6 : Tạo từ điển ánh xạ các bang đến danh sách thành phố và khu vực 7](#_Toc183750554)

[Ảnh 7 : Bổ sung thông tin thành phố và vùng địa lý vào tập dữ liệu 8](#_Toc183750555)

[Ảnh 8 : Xuất tập dữ liệu vừa xử lý 9](#_Toc183750556)

[Ảnh 9 : Xu hướng sử dụng ưu đãi giảm giá 12](#_Toc183750557)

[Ảnh 10 : Sở thích mua sắm theo từng nhóm tuổi 14](#_Toc183750558)

[Ảnh 11 : Đánh giá trung bình về sản phẩm cho từng thành phố 16](#_Toc183750559)

[Ảnh 12 : Tần suất sử dụng ưu đãi dựa theo giới tính 20](#_Toc183750560)

[Ảnh 13 : Tổng giá trị mua hàng của từng loại sản phẩm trong các khu vực địa lý 23](#_Toc183750561)

[Ảnh 14 : Phân tích số lượng sản phẩm được mua theo từng mua 26](#_Toc183750562)

[Ảnh 15 : Tổng giá trị mua hàng trong khu vực Midwest 28](#_Toc183750563)

[Ảnh 16 : So sánh giá trị mua hàng tối thiểu và giá trị mua hàng tối đa 30](#_Toc183750564)

[Ảnh 17 : Tổng giá trị mua hàng theo từng bang trong khu vực Northeast. 32](#_Toc183750565)

[Ảnh 18 : Mức độ sử dụng mã khuyến mãi dựa trên giới tính. 35](#_Toc183750566)

[Ảnh 19 : Tổng số tiền chi tiêu theo từng bang 36](#_Toc183750567)

[Ảnh 20 : Mối quan hệ giữa mùa thu và các sản phẩm. 38](#_Toc183750568)

[Ảnh 21 : Mối quan hệ giữa mùa xuân và các sản phẩm 40](#_Toc183750569)

[Ảnh 22 : Mối quan hệ giữa mùa hè và các sản phẩm 42](#_Toc183750570)

[Ảnh 23 : Mối quan hệ giữa mùa đông và các sản phẩm 44](#_Toc183750571)

[Ảnh 24 : Số lượng kích cỡ theo mùa và giới tính 48](#_Toc183750572)

[Ảnh 25 : Tổng giá trị mua hàng theo từng bang trong khu vực phía Nam 50](#_Toc183750573)

[Ảnh 26 : Mối quan hệ giữa khu vực địa lý và sản phẩm 53](#_Toc183750574)

[Ảnh 27 : Mùa mua sắm tốt nhất 53](#_Toc183750575)

[Ảnh 28 : Dự đoán số lượng quần áo được dự trữ theo kích cỡ 57](#_Toc183750576)

[Ảnh 29 : Tổng giá trị mua hàng theo từng bang trong khu vực phía Tây 58](#_Toc183750577)

# DANH MỤC BẢNG BIỂU

[Bảng 1 : Mô tả thuộc tính 4](#_Toc183741591)

# MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

## Lý do chọn đề tài

Trong bối cảnh kinh doanh hiện đại, việc hiểu và đáp ứng nhu cầu của khách hàng là yếu tố then chốt để doanh nghiệp tồn tại và phát triển. Đề tài ***"Phân tích hành vi khách hàng, nâng cao chiến lược và tìm xu hướng"*** được chọn vì nó không chỉ mang lại lợi ích thiết thực cho doanh nghiệp mà còn góp phần cải thiện chất lượng cuộc sống của người tiêu dùng.

Phân tích hành vi khách hàng giúp doanh nghiệp tối ưu hóa chiến lược kinh doanh, từ đó cung cấp các sản phẩm và dịch vụ phù hợp hơn, cá nhân hóa trải nghiệm và tăng sự hài lòng của khách hàng. Đồng thời, việc dự đoán xu hướng tiêu dùng trong tương lai còn giúp giảm lãng phí tài nguyên, cải thiện hiệu quả sản xuất và phân phối.

Với người tiêu dùng, đề tài này tạo cơ hội để họ nhận được những sản phẩm, dịch vụ đáp ứng đúng nhu cầu, nâng cao trải nghiệm mua sắm và tiếp cận nhiều lựa chọn phù hợp hơn. Vì thế, nghiên cứu và áp dụng các giải pháp từ đề tài này không chỉ có ý nghĩa kinh tế mà còn mang lại giá trị xã hội đáng kể.

## Mục tiêu thực hiện đề tài

* *Hiểu rõ hành vi khách hàng:*

Thu thập và phân tích dữ liệu từ các nguồn như giao dịch mua bán, hành vi trực tuyến, và phản hồi của khách hàng để xác định thói quen, sở thích và nhu cầu.

Phân loại khách hàng thành các nhóm mục tiêu cụ thể dựa trên hành vi tiêu dùng.

* *Nâng cao hiệu quả chiến lược kinh doanh:*

Xây dựng các chiến lược tiếp thị cá nhân hóa để tiếp cận đúng đối tượng khách hàng với thông điệp phù hợp.

* *Dự đoán xu hướng và nhu cầu trong tương lai:*

Ứng dụng các mô hình phân tích dự đoán để phát hiện xu hướng mới trong hành vi tiêu dùng.

* *Cải thiện trải nghiệm khách hàng:*

Nâng cao mức độ hài lòng và trung thành thông qua việc cá nhân hóa sản phẩm, dịch vụ và chăm sóc khách hàng.

* *Ứng dụng công nghệ hiện đại:*

Sử dụng các công cụ Business Intelligence (BI), AI, và Machine Learning để phân tích và trực quan hóa dữ liệu một cách hiệu quả.

Tích hợp các giải pháp BI vào quá trình ra quyết định kinh doanh.

# XỬ LÝ DỮ LIỆU

Để đảm bảo tính đồng nhất của dữ liệu, tránh làm sai lệch kết quả phân tích và cải thiện hiệu suất của tập dữ liệu và mô hình. Trước hết ta phải thực hiện chuẩn hóa dữ liệu. Với bộ dữ liệu này sẽ gồm các cột và ý nghĩa của từng cột như sau :

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| *Customer ID* | Mã định danh của từng khách hàng. |
| *Age* | Tuổi của khách hàng. |
| *Gender* | Giới tính của khách hàng (Male/Female) |
| *Item Purchased* | Tên sản phẩm được mua. |
| *Category* | Danh mục của sản phẩm (ví dụ: Clothing, Footwear, v.v.). |
| *Purchase Amount (USD)* | Số tiền mua hàng (tính bằng USD). |
| *Location* | Địa chỉ của khách hàng. |
| *Size* | Kích cỡ của sản phẩm (S, M, L, v.v.). |
| *Color* | Màu sắc của sản phẩm. |
| *Season* | Mùa trong năm mà sản phẩm được mua (Winter, Spring, v.v.). |
| *Review Rating* | Điểm đánh giá của khách hàng (trung bình, dạng số thập phân). |
| *Subscription Status* | Trạng thái đăng ký (Yes/No). |
| *Shipping Type* | Loại hình vận chuyển (Free Shipping, Express, v.v.). |
| *Discount Applied* | Áp dụng giảm giá (Yes/No). |
| *Promo Code Used* | Sử dụng mã khuyến mãi (Yes/No). |
| *Previous Purchases* | Số lần mua hàng trước đó. |
| *Payment Method* | Phương thức thanh toán (Credit Card, Cash, v.v.). |
| *Frequency of Purchases* | Tần suất mua hàng (Weekly, Monthly, Annually, v.v.). |

Bảng 1 : Mô tả thuộc tính

## Cài đặt thư viện

Để thực hiện chuẩn hóa dữ liệu ta cần phải cài đặt các thư viện cần như thiết như :

A screenshot of a computer

Description automatically generated

Ảnh 1 : Cài đặt thư viện

## Xử lý dữ liệu

A screenshot of a computer

Description automatically generated

Ảnh 2 : Hiển thị 3 dòng dữ liệu đầu tiên

Sau khi đã cài đặt thư viện thành công, ta tiến hành đọc từ máy tính tập dữ liệu “shopping\_behavior\_update.csv” và tiến hành hiển thị ra màn hình 3 dòng dữ liệu đầu tiên nhằm kiểm tra chắc chắn máy đã đọc được dữ liệu thành công. Với bộ dữ liệu này ta có tổng cổng 18 cột và 3900 dòng.

A screen shot of a computer

Description automatically generated

Ảnh 3 : Mở rộng dữ liệu

Với đoạn mã này, ta sẽ tạo một DataFrame mới có tên data\_expand và với DataFrame mới này, mỗi hàng trong tập dữ liệu *data* trước đó sẽ được lặp lại 30 lần nhằm tạo ra một tập dữ liệu mở rộng với mỗi dòng ban đầu được sao chép 30 lần. Sau khi thực hiện việc lặp như trên, chỉ số (index) sẽ được thiết lập lại về mặc định (0, 1, 2, ...). Ngoài ra (drop=True) sẽ đảm bảo rằng chỉ số cũ không được giữ lại trong DataFrame mới.

Sau khi thực hiện lặp lại như trên ta tiến hành dùng shape() để kiểm tra kích thước dữ liệu và với dữ liệu đã lặp ta sẽ có 117000 dòng và 18 cột.

A screenshot of a computer

Description automatically generated

Ảnh 4 : Thêm một cột id

Ở đoạn mã này việc của ta sẽ tạo một mảng số nguyên bắt đầu từ 1 đến *len(data\_expand),* tương ứng với số dòng của data\_expand. Sau đó gán mảng vừa tạo vào cột *id* trong DataFrame *data\_expand.* Cuối cùng ta tiến hành dùng head(4) để hiển thị 4 dòng đầu tiên của dữ liệu sau khi đã thêm cột *id.*

A screenshot of a computer

Description automatically generated

Ảnh 5 : Hiển thị các giá trị trong cột Location

Với đoạn code này ta sẽ dùng unique() để hiển thị ra màn hình tất cả các giá trị trong cột Location của DataFrame data\_expand.

A white paper with blue text

Description automatically generated

Ảnh 6 : Tạo từ điển ánh xạ các bang đến danh sách thành phố và khu vực

Với đoạn code này, ta tiến hành tạo một từ điển ánh xạ các bang ở Hoa Kỳ đến danh sách các thành phố chính và mỗi bang có một danh sách các thành phố lớn trong bang đó. Ví dụ: Bang *Kentucky* có các thành phố *Louisville, Lexington, và Bowling Green.* Ngoài ra,mỗi bang được gắn nhãn vùng region tương ứng gồm : Northeast, Midwest, South, hoặc West.

*A screenshot of a computer code

Description automatically generated*

Ảnh 7 : Bổ sung thông tin thành phố và vùng địa lý vào tập dữ liệu

Để hoàn thành bộ dữ liệu hoàn chỉnh, ta tiến hành tạo một hàm có tên *get\_random\_city\_and\_region* và hàm này lấy một bang (state) làm tham số đầu vào. Hàm này hoạt động bằng cách kiểm tra xem *state* có trong từ điển *state\_info\_mapping* hay không. Nếu có sẽ lấy ngẫu nhiên một thành phố từ danh sách *cities* của bang, lấy vùng địa lý từ danh sách *region* của bang và trả về *tuple (city, region).*

Ngược lại, nếu bang không tồn tại trong từ điển *state\_info\_mapping*, hàm trả về *(None, None).*

Tiếp theo ta tiến hành thêm 2 cột *City* và *Region* vào DataFrame *data\_expand* bằng cách áp dụng hàm *get\_random\_city\_and\_region(state)* lên từng giá trị trong cột *Location*. Vì hàm *get\_random\_city\_and\_region(state)* trả về kết quả là một tuple *(city, region)* vì vậy ta sử dụng *pd.Series* để kết quả *(city, region)* có thể tách thành hai cột. Cuối cùng ta tiến hành in ra màn hình tập dữ liệu để kiểm tra.

A white rectangular object with black text

Description automatically generated

Ảnh 8 : Xuất tập dữ liệu vừa xử lý

Cuối cùng, để thực hiện trực quan hóa tập dữ liệu đã xử lý ta tiến hành tạo một tệp csv có tên “data\_behavior\_expand\_x30.csv” trong đó tệp này sẽ chứa các thông tin trong DataFrame data\_expand. Việc sử dụng csv.QUOTE\_ALL giúp bảo vệ dữ liệu chứa các kí tự đặc biệt không bị sai định dạng.

# TRỰC QUAN HÓA DỮ LIỆU

## Phân tích xu hướng sử dụng ưu đãi giảm giá.

A graph of a bar chart

Description automatically generated with medium confidence

Ảnh 9 : Xu hướng sử dụng ưu đãi giảm giá

### Giới thiệu biểu đồ

Mục tiêu: Phân tích xu hướng sử dụng ưu đãi giảm giá (discount) dựa trên các yếu tố:

* Tuổi (Age)
* Giới tính (Gender)
* Quyết định sử dụng ưu đãi hay không (Yes/No)

Dữ liệu hiển thị:

* Số lượng người (Count) được biểu diễn trên trục hoành.
* Từng nhóm tuổi (Age) được chia nhỏ theo giới tính (Male/Female) và trạng thái sử dụng ưu đãi (Yes/No).

### Phân tích biểu đồ

#### **Tương quan giới tính và việc sử dụng ưu đãi:**

Nữ giới (Female):

* Nữ giới có xu hướng **ít sử dụng ưu đãi** hơn so với nam giới trong hầu hết các nhóm tuổi.
* Đây có thể phản ánh rằng nữ giới không bị ảnh hưởng nhiều bởi các chương trình giảm giá hoặc có mức độ quan tâm thấp hơn đến các ưu đãi.

Nam giới (Male):

* Nam giới có tỷ lệ sử dụng ưu đãi (Yes) cao hơn nữ giới, đặc biệt rõ rệt ở các nhóm tuổi trẻ (18–25) và trung niên (40–50).

#### **Tương quan theo nhóm tuổi:**

Nhóm tuổi trẻ (18–25):

* Tỷ lệ sử dụng ưu đãi cao hơn, đặc biệt với nam giới, cho thấy giới trẻ dễ bị thu hút bởi giảm giá.
* Có thể giải thích bởi giới trẻ thường có thu nhập thấp hơn và tìm kiếm các ưu đãi để tiết kiệm chi tiêu.

Nhóm tuổi trung niên (30–50):

* Đây là nhóm có xu hướng sử dụng ưu đãi giảm dần, đặc biệt với nữ giới.
* Lý do có thể thói quen tiêu dùng ít bị chi phối bởi giảm giá.

Nhóm người cao tuổi (55+):

* Xu hướng sử dụng ưu đãi tăng lên với cả nam và nữ. Điều này có thể do những người lớn tuổi thường muốn tối ưu hóa chi tiêu.

#### **Tương quan tổng thể giữa "Yes" và "No":**

Trong tổng thể:

* Số lượng người không sử dụng ưu đãi (No) cao hơn đáng kể so với những người sử dụng ưu đãi (Yes).
* Điều này có thể cho thấy rằng không phải tất cả các chiến dịch giảm giá đều hiệu quả trong việc thu hút người tiêu dùng.

### Kết luận

Từ biểu đồ trên cho thấy nam giới có xu hướng sử dụng ưu đãi cao hơn, đặc biệt ở nhóm tuổi trẻ và trung niên. Nữ giới ít bị ảnh hưởng bởi ưu đãi, điều này cần được xem xét khi thiết kế các chương trình khuyến mãi.

Ngoài ra, tỉ lệ người trẻ (18–25) và người già (55+) dễ tiếp cận ưu đãi hơn nhóm người trung niên (30–50).

## Phân tích sở thích mua sắm theo từng nhóm tuổi dựa trên sản phẩm.

A graph of a number of colored bars

Description automatically generated with medium confidence

Ảnh 10 : Sở thích mua sắm theo từng nhóm tuổi

### Giới thiệu biểu đồ

Mục tiêu: Hiểu được sở thích mua sắm theo từng nhóm tuổi (Age) dựa trên các loại sản phẩm (Item Categories).

Dữ liệu hiển thị:

Trục ngang: Số lượng người mua (Count).

Trục dọc: Nhóm tuổi (Age).

Mỗi màu sắc: Đại diện cho loại sản phẩm khác nhau.

### Phân tích biểu đồ

#### **Xu hướng tiêu dùng theo độ tuổi**

Nhóm tuổi trẻ (18–25) là nhóm tuổi có sự đa dạng sản phẩm nhất, trong đó bao gồm:

* Công nghệ (Electronics): Nhóm này thường mua các sản phẩm như điện thoại, laptop, và phụ kiện công nghệ.
* Quần áo (Clothing): Nhu cầu cao với các mặt hàng thời trang, phản ánh xu hướng tiêu dùng của giới trẻ.

Nhóm trung niên (30–50) : Đây là nhóm tuổi mua sắm ít đa dạng hơn và tập trung vào các mặt hàng thiết yếu hoặc liên quan đến gia đình:

* Gia dụng (Home Appliances): Các mặt hàng như thiết bị nhà bếp hoặc sản phẩm tiện ích gia đình chiếm ưu thế.
* Đồ nội thất (Furniture): Nhóm này có nhu cầu cao hơn cho các sản phẩm dài hạn, thể hiện xu hướng ổn định trong cuộc sống.

Nhóm tuổi lớn (55+): Là nhóm tuổi mua sắm ít hơn và tập trung vào các mặt hàng phục vụ sức khỏe hoặc tiêu dùng hàng ngày:

* Thực phẩm chức năng (Health Products): Nổi bật ở nhóm tuổi lớn, phản ánh sự quan tâm đến sức khỏe.
* Hàng tiêu dùng (Groceries): Các sản phẩm thiết yếu như thực phẩm, đồ uống được ưa chuộng.

#### **So sánh các loại sản phẩm giữa các nhóm tuổi**

Một số sản phẩm phổ biến ở mọi nhóm tuổi như quần áo (Clothing) hoặc điện tử (Electronics) xuất hiện đồng đều trong tất cả các nhóm. Ngoài ra, một số sản phẩm chỉ phổ biến ở nhóm tuổi cụ thể đồ nội thất (Furniture) chủ yếu xuất hiện ở nhóm tuổi trung niên. Và cuối cùng sản phẩm công nghệ cao (High-tech) chỉ phổ biến ở nhóm trẻ tuổi.

#### **Tương quan giữa tuổi và sự đa dạng sản phẩm**

Nhóm tuổi trẻ (18–30) thể hiện sự đa dạng lớn nhất về sản phẩm, cho thấy sở thích tiêu dùng phong phú. Ngược lại nhóm tuổi lớn (55+) có xu hướng tập trung vào các mặt hàng thiết yếu hơn và ít mua sắm hơn các độ tuổi khác.

### Kết luận

Từ biểu đồ trên ta có thể nhận định rằng người trẻ tuổi có xu hướng yêu thích và mua sắm đồ công nghệ và thời trang. Nhóm người trung niên thường tập trung vào những sản phẩm về gia đình và tiện ích trong gia đình. Cuối cùng nhóm người lớn tuổi thường chọn những sản phẩm liên quan đến sức khỏe và nhu cầu cơ bản.

## Phân tích đánh giá trung bình về sản phẩm cho từng thành phố.

A chart of different colored lines

Description automatically generated

Ảnh 11 : Đánh giá trung bình về sản phẩm cho từng thành phố

### Giới thiệu biểu đồ

Mục tiêu: Phân tích đánh giá trung bình (Avg. Review Rating) cho từng thành phố, từ đó hiểu được mức độ hài lòng của khách hàng tại các địa điểm khác nhau.

Dữ liệu hiển thị:

* Trục ngang: Điểm đánh giá trung bình (Avg. Review Rating).
* Trục dọc: Danh sách các thành phố (City).
* Màu sắc: Được sử dụng để phân biệt các thành phố, giúp dễ dàng nhận diện.

### Phân tích biểu đồ

#### **So sánh đánh giá giữa các thành phố**

Từ biểu đồ trên ta thấy một số thành phố nổi bật với đánh giá cao bao gồm :

* Dallas (3.172): Điểm đánh giá cao nhất, cho thấy khách hàng tại đây có mức độ hài lòng cao hơn.
* Green Bay (2.982): Một thành phố khác có đánh giá rất tích cực, phản ánh trải nghiệm tốt tại khu vực này.

Ngược lại ta thấy một số thành phố có điểm đánh giá thấp gồm :

* Huntington (2.847): Điểm đánh giá thấp nhất, cho thấy cần xem xét lại dịch vụ hoặc chất lượng tại thành phố này.
* Atlanta (2.882): Cũng có điểm thấp, phản ánh mức độ hài lòng hạn chế.

#### **Xu hướng tổng quan về đánh giá**

Ta có thể nhận thấy rằng đánh giá trung bình dao động từ khoảng 2.2 đến 6.2, cho thấy sự khác biệt đáng kể giữa các thành phố. Hầu hết các thành phố có điểm trung bình khoảng 3.0–4.0, cho thấy mức độ hài lòng ổn định nhưng không quá nổi bật.

#### **So sánh giữa khu vực**

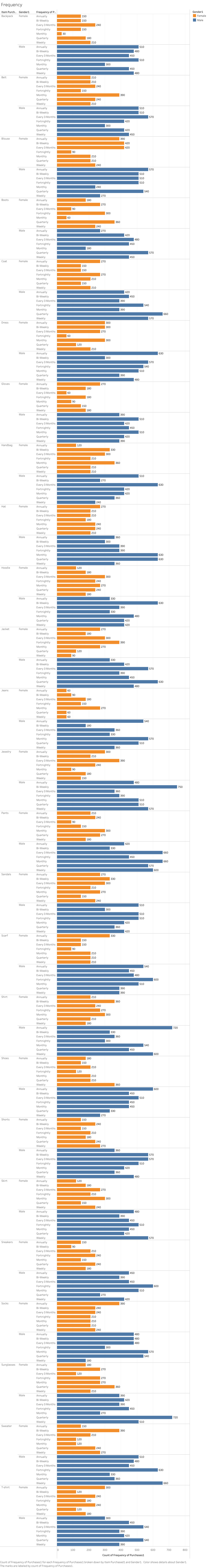
Những thành phố lớn như Chicago, Los Angeles, và New York City có điểm ở mức điểm đánh giá trung bình lần lượt là 3.264, 3.589, 3.316. Điều này có thể phản ánh sự cạnh tranh gay gắt và khó duy trì chất lượng đồng đều.

Ngược lại, những thành phố như Charleston và Aurora có đánh giá cao hơn những thành phố lớn, cho thấy khả năng phục vụ hoặc sự tập trung vào trải nghiệm khách hàng tốt hơn.

### Kết luận

Từ biểu đồ ta có thể nhận định rằng Charleston và Aurora là 2 thành phố dẫn đầu về mức độ hài lòng với đánh giá rất cao. Ngược lại, 2 thành phố Mesa và Tucson có điểm rất thấp, có thể nguyên nhân là: Dịch vụ kém, chất lượng sản phẩm không đáp ứng kỳ vọng.

## Phân tích tần suất sử dụng ưu đãi theo giới tính.



Ảnh 12 : Tần suất sử dụng mã giảm giá dựa theo giới tính và sản phẩm

### Giới thiệu biểu đồ

Mục tiêu: Tìm hiểu xu hướng mua sắm và sử dụng ưu đãi theo từng giới tính với các loại sản phẩm cụ thể.

Phân tích tần suất sử dụng ưu đãi (Discount Applied) dựa trên:

* Giới tính (Gender): Nam (Male) và nữ (Female).
* Danh mục sản phẩm (Item Category).

Biểu đồ trực quan hóa bao gồm:

* + Danh mục sản phẩm: Bao gồm các loại sản phẩm khác nhau như quần áo, phụ kiện, giày dép, v.v.
  + Giới tính: Nam (Male) và Nữ (Female)..

### Phân tích biểu đồ

#### **Sản phẩm được mua phổ biến nhất**

Nữ giới: Các sản phẩm nổi bật gồm váy (dress), túi xách (handbag), và trang sức (jewelry). Từ đó có thể phản ánh đặc điểm sở thích và nhu cầu về thời trang, phụ kiện.

Nam giới: Áo khoác (jacket), giày thể thao (sneakers), và balo (backpack) là những sản phẩm phổ biến. Điều này gắn liền với phong cách năng động và thực dụng của nam giới.

#### **Sự chênh lệch giữa nam và nữ**

Nữ giới có xu hướng thường xuyên mua những mặt hàng về phụ kiện như trang sức (Jewelry) và túi xách.

Nam giới có sở thích về những mặt hàng thời trang thường ngày như balo, áo khoác vì đây là những sản phẩm có tần suất mua hàng nhiều nhất ở nam giới.

#### **Sản phẩm có cân bằng giữa hai giới tính**

Các mặt hàng như quần jean (Jeans), áo phông (T-shirts) có sự cân bằng đáng kể giữa nam và nữ vì đây là những sản phẩm mang tính trung tính và phù hợp với cả hai giới..

### Kết luận

Nữ giới có xu hướng chi tiêu nhiều hơn vào các sản phẩm thời trang và phụ kiện, trong khi nam giới tập trung vào các sản phẩm tiện dụng và thể thao. Ngoài ra các sản phẩm cơ bản như quần áo hàng ngày được cả hai giới ưa chuộng...

## Phân tích tổng doanh thu của sản phẩm trong các khu vực.

A graph of different colored lines

Description automatically generated with medium confidence

Ảnh 13 : Tổng giá trị mua hàng của từng loại sản phẩm trong các khu vực địa lý

### Giới thiệu biểu đồ

Mục tiêu: Phân tích tổng giá trị mua hàng (Purchase Amount (USD)) của từng loại sản phẩm (Item Purchased) trong các khu vực địa lý khác nhau (Region).

Biểu đồ hiển thị:

* Trục ngang: Tổng giá trị mua hàng (USD).
* Trục dọc: Loại sản phẩm.
* Màu sắc: Đại diện cho từng loại sản phẩm.
* Vùng địa lý (Region): Midwest, Northeast, South, West.

### Phân tích biểu đồ

#### **Xu hướng mua sắm theo khu vực (Region)**

Ở khu vực Midwest, các sản phẩm như *Blouse và Boots* nổi bật với doanh thu cao lần lượt là 73.110 USD và 76.140 USD. Ngoài ra, một số sản phẩm khác như *Handbag và Hoodie* có giá trị mua hàng thấp với 47.880 USD và 32.520 USD, cho thấy ít phổ biến trong khu vực này.

Ở khu vực Northeast, hầu hết các sản phẩm đều có doanh thu giao động từ 50.000 USD đến 70.000 USD. Như vậy có thể kết luận rằng khu vực này có sự cân bằng tốt giữa các loại sản phẩm.

Ở khu vực South các sản phẩm của danh mục *Shoes và Sweater* chiếm ưu thế, phản ánh sự ưa chuộng các mặt hàng liên quan đến thời trang vì có doanh thu cao hơn 80.000USD.

Ở khu vực West, 2 danh mục Hoodie và Blouse là 2 danh mục dẫn đầu về doanh thu. Danh mục Jewelry cũng là một sản phẩm nổi bật với doanh thu 88.530 USD, cho thấy mọi người đều sở thích đối với phụ kiện cao cấp.

### Kết luận

Ta có thể kết luận rằng, West và South là hai khu vực có doanh thu cao với các sản phẩm thời trang và phụ kiện cao cấp như Shoes, Jewelry, Blouse, Sweater. Midwest và Northeast là những khu vực có xu hướng mua sắm sản phẩm giá tốt.

## Phân tích số lượng sản phẩm được mua theo từng mùa.

A graph of a bar chart

Description automatically generated with medium confidence

Ảnh 14 : Phân tích số lượng sản phẩm được mua theo từng mua

### Giới thiệu biểu đồ

Mục tiêu: Phân tích số lượng sản phẩm được mua (Count of Item Purchased) theo từng mùa (Season), danh mục chính (Category), và loại sản phẩm cụ thể (Item).

Biểu đồ hiển thị:

* Trục ngang: Số lượng sản phẩm được mua.
* Trục dọc: Danh mục sản phẩm cụ thể, các mùa trong năm.
* Màu sắc: Đại diện cho danh mục chính:
* Accessories (Phụ kiện): Màu xanh dương.
* Clothing (Trang phục): Màu cam.
* Footwear (Giày dép): Màu đỏ.
* Outerwear (Áo khoác ngoài): Màu xanh lá.

### Phân tích biểu đồ

#### **Xu hướng mua sắm theo mùa (Season)**

Với mùa thu (Fall):

* Danh mục Outerwear có tổng cộng1.620 sản phẩm Jacket được bán ra.
* Danh mục Footwear cũng được nhiều người chọn mua (930 lượt mua)
* Danh mục Accessories có sản phẩm Scarf đạt 1.200 lượt mua

Với mùa xuân (Spring):

* Danh mục Clothing có sản phẩm Skirt dẫn đầu với 1.410 lượt mua thể hiện người mua thường chọn váy là sản phẩm để mang vào mùa xuân.
* Danh mục Accessories bán chạy với sản phẩm Jewelry đạt 1.260 lượt mua thể hiện nhu cầu đeo trang sức vào các dịp quan trọng của mùa xuân như Tết dương lịch hoặc Tết nguyên đán.

Với mùa hè (Summer):

* Danh mục Accessories có sản phẩm Sunglasses dẫn đầu với 1.290 lượt mua, thể hiện rõ nhu cầu tránh nắng trong mùa hè.
* Danh mục Clothing có sản phẩm Pants đạt 1.500 lượt mua phản ánh nhu cầu thời trang nhẹ nhàng, thoải mái và có thể che nắng cho cơ thể.
* Danh mục Footwear nổi trội với sản phẩm Sandals đạt 1.380 lượt mua từ đó có thể thấy Sandals là loại giày phổ biến nhất trong các loại giày ở mùa hè.

Với mùa đông (Winter):

* Danh mục Outerwear nổi trội với sản phẩm Coat đạt 1.410 lượt mua, có thể sản phẩm này rất phù hợp với thời tiết lạnh của mùa đông.
* Danh mục Accessories có 2 sản phẩm Scarf và Gloves đạt gần 1.000 lượt mua, rất phù hợp với thời tiết lạnh.
* Danh mục Footwear có sản phẩm Boots đạt 1.140 lượt mua, là loại giày phổ biến nhất trong mùa đông.

### Kết luận

Với biểu đồ và phân tích biểu đồ trên ta có thể thấy mùa đông và mùa thu thì người mua sẽ tập trung vào các sản phẩm giữ ấm như *Jacket, Coat, Boots, Scarf*. Mùa xuân và mùa hè thì người mua sẽ ưa chuộng sản phẩm nhẹ nhàng, thoáng mát như *Skirt, Sunglasses, Sandals.*

## Phân tích tổng doanh thu theo từng bang trong khu vực Midwest.

A map of different colored states

Description automatically generated

Ảnh 15 : Tổng giá trị mua hàng trong khu vực Midwest

### Giới thiệu bản đồ

Mục tiêu: Phân tích tổng giá trị mua hàng (Purchase Amount (USD)) theo từng bang trong khu vực Midwest của Hoa Kỳ.

Biểu đồ gồm:

* Màu sắc: Phản ánh tổng giá trị mua hàng:
* Màu xanh nhạt: Giá trị thấp.
* Màu cam đậm: Giá trị cao.
* Số liệu: Tổng giá trị mua hàng (USD) của từng bang được hiển thị trực tiếp trên bản đồ.
* Khu vực: Tập trung vào các bang thuộc khu vực Midwest.

### Phân tích biểu đồ

#### **Bang có tổng giá trị mua hàng cao nhất**

Từ biểu đồ trên ta thấy Bang Illinois dẫn đầu trong khu vực Midwest về tổng giá trị mua hàng với 168.510 USD. Có thể thấy rằng Illinois là trung tâm kinh tế lớn, nhu cầu mua sắm và tiêu dùng cao.

#### **Bang có giá trị mua hàng thấp**

Từ biểu đồ trên ta thấy Kansas là bang có tổng giá trị mua hàng thấp nhất trong khu vực Midwest với doanh thu 103.110 USD. Doanh thu thấp có thể thấy người dân ở đây nhu cầu tiêu dùng và mua sắm thấp.

#### **So sánh giữa các bang khác**

Minnesota, Missouri, Michigan là những bang này có giá trị mua hàng trung bình, cho thấy mức tiêu dùng ổn định với doanh thu lần lượt là 149.310 USD, 140.730 USD, 135.990 USD

### Kết luận

Bang Illinois nổi bật với doanh thu 168.510 USD, khẳng định vị trí trung tâm kinh tế quan trọng trong khu vực. Ngoài ra các bang như Kansas, Wisconsin và South Dakota có doanh thu thấp nhất trong khu vực cho thấy sự chênh lệch lớn, phản ánh sự không đồng đều về mức độ tiêu dùng và mua sắm giữa các bang trong khu vực Midwest.

## Phân tích tổng doanh thu theo từng bang trong khu vực Northeast.

A map of the united states

Description automatically generated

Ảnh : Tổng giá trị mua hàng theo từng bang trong khu vực Northeast.

### Giới thiệu biểu đồ

Mục tiêu: Phân tích tổng giá trị mua hàng (Purchase Amount (USD)) theo từng bang trong khu vực Northeast của Hoa Kỳ.

Biểu đồ gồm:

* Màu sắc: Thể hiện tổng giá trị mua hàng:
* Màu xanh nhạt: Giá trị thấp hơn.
* Màu cam đậm: Giá trị cao hơn.
* Số liệu: Tổng giá trị mua hàng (USD) được hiển thị trên từng bang.

### Phân tích biểu đồ

#### **Bang có tổng giá trị mua hàng cao nhất**

New York là bang dẫn đầu khu vực Northeast với doanh thu 157.710 USD) có thể thấy đây là một trung tâm kinh tế quan trọng của Northeast.

#### **Các bang khác**

Pennsylvania là bang đứng sau New York với doanh thu 147.780 USD cho thấy sự ổn định về doanh thu ở khu vực Northeast. Sau Pennsylvania là các bang Maryland và Delaware có doanh thu lần lượt là 143.850 USD và 142.740 USD vượt qua các bang khác trong khu vực.

#### **Bang có giá trị thấp**

New Jersey là bang có doanh thu thấp nhất trong khu vực Northeast và chỉ có doanh thu 114.060 USD. Từ số doanh thu trên cho thấy đây là bang có ít người mua sắm nhất.

### Kết luận

Bang New York dẫn đầu về tổng giá trị mua hàng trong khu vực Northeast cho thấy đây là trung tâm kinh tế lớn của khu vực. Ngược lại bang New Jersey có tổng doanh thu thấp nhất vì vậy cần được thúc đẩy về chiến lượt marketing cho người dân tại khu vực này.

Ngoài ra do sự chênh lệch rõ ràng giữa bang New York và bang New Jersey cho thấy phản ánh sự phân hóa về mức tiêu thụ trong khu vực.

## Phân tích tổng doanh thu theo từng bang trong khu vực South.

A map of the united states

Description automatically generated

Ảnh : Tổng giá trị mua hàng theo từng bang trong khu vực phía Nam

### Tổng quan biểu đồ

Mục tiêu: Hiển thị mức chi tiêu (Purchase Amount - USD) của từng bang thuộc vùng phía Nam.

Biều đồ gồm

* Dữ liệu: Tổng chi tiêu được biểu diễn bằng mã màu từ xanh nhạt (chi tiêu thấp) đến đỏ cam (chi tiêu cao).

### Phân tích biểu đồ

#### **Tiểu bang có chi tiêu cao**

Alabama là bang dẫn đầu với mức chi tiêu cao nhất 157.830 USD cho thấy đây là trung tâm kinh tế quan trọng của khu vực phía Nam. Bang West Virginia xếp thứ hai với mức chi tiêu 155.220 USD cho thấy đây là một tiểu bang quan trọng khác trong khu vực. Các bang Mississippi (146.490 USD), Louisiana (145.440 USD), và Arkansas (144.840 USD) lần lượt là các bang có mức chi tiêu cao chỉ sau bang West Virginia

#### **Tiểu bang có chi tiêu thấp**

Oklahoma và Florida là 2 bang có có mức chi tiêu thấp hơn so với các tiểu bang khác lần lượt có mức chi tiêu là 131.280 USD và 113.930 USD cho thấy nhu cầu tiêu dùng ít nhất trong khu vực phía Nam.

### Kết luận

Alabama và West Virginia là các tiểu bang cần được tập trung phát triển thêm các chiến lược bán hàng hoặc sản phẩm mới vì đây là hai trung tâm kinh tế trong khu vực phía Nam.

Mississippi, Louisiana, và Arkansas cũng là những tiểu bang tiềm năng với mức chi tiêu cao, nên tiếp tục đầu tư và mở rộng.

Florida và Oklahoma cần được đánh giá lại để xác định nguyên nhân mức chi tiêu thấp, từ đó cải thiện chiến lược hoặc sản phẩm phù hợp hơn.

## Phân tích tổng doanh thu theo từng bang trong khu vực West.

A map of the united states

Description automatically generated

Ảnh : Tổng giá trị mua hàng theo từng bang trong khu vực phía Tây

### Tổng quan biểu đồ

Biểu đồ thể hiện tổng giá trị mua sắm (Purchase Amount) của từng bang thuộc khu vực miền Tây.

Các tiểu bang được mã hóa màu sắc tương ứng với giá trị mua sắm từ 112,560 USD đến 173,520 USD. Màu xanh đậm tương ứng với tổng giá trị mua hàng thấp nhất và màu cam đậm tương ứng với tổng giá trị mua hàng cao nhất.

### Phân tích biểu đồ

#### **Các bang có doanh thu cao nhất.**

Montana là bang đứng đầu khu vực với 173,520 USD đây là khu vực cần được ưu tiên phát triển vì là trọng tâm kinh tế của khu vực phía Tây.

California là bang xếp thứ hai với doanhh thu 168,150 USD chứng tỏ đây là một thị trường lớn với doanh thu ổn định.

#### **Các bang có doanh thu thấp nhất.**

Colorado là bang có giá trị mua sắm thấp nhất và doanh thu chỉ 126,660 USD với mức doanh thu này ta cần xem xét các chiến lược marketing để kích thích nhu cầu tiêu dùng tại đây.

#### **Xu hướng**

Các tiểu bang ven biển như California và Oregon thường có giá trị mua sắm cao hơn do dân số đông và kinh tế phát triển hơn so với các tiểu bang nội địa.

### Kết luận

Montana và California là hai thị trường quan trọng nhất của khu vực, cần được tiếp tục đầu tư.

Colorado và Oregon có giá trị mua sắm thấp nhất trong khu vực phía Tây vì thế ta cần chiến lược kích mua sắm phù hợp.

## So sánh giá trị mua hàng tối thiểu và giá trị mua hàng tối đa cho từng danh mục sản phẩm.

A close-up of a graph

Description automatically generated

Ảnh 16 : So sánh giá trị mua hàng tối thiểu và giá trị mua hàng tối đa

### Giới thiệu biểu đồ

Mục tiêu: So sánh giá trị mua hàng tối thiểu (Min Purchase Amount) và giá trị mua hàng tối đa (Max Purchase Amount) cho từng danh mục sản phẩm (Item Purchased).

Biểu đồ gồm :

* Trục ngang: Giá trị mua hàng (USD).
* Biểu đồ bên trái: Giá trị tối thiểu.
* Biểu đồ bên phải: Giá trị tối đa.
* Trục dọc: Các danh mục sản phẩm.
* Màu sắc: Đại diện cho từng danh mục sản phẩm, giúp dễ dàng phân biệt.

### Phân tích biểu đồ

#### **Giá trị mua hàng tối thiểu (Min Purchase Amount)**

Hầu hết giá trị mua hàng tối thiểu của hầu hết các sản phẩm là 20 USD, ngoại trừ một vài sản phẩm bao gồm:

* Boots, Glove và Handbag: Có giá trị mua hàng tối thiểu cao hơn (21 USD), có thể do đây là các sản phẩm cao cấp hơn.

#### **Giá trị mua hàng tối đa (Max Purchase Amount)**

Hầu hết giá trị mua hàng tối đa của các sản phẩm đạt gần mức tối đa tuyệt đối (100 USD) ngoại trừ :

* Hat, Gloves,…: Có giá trị mua hàng tối đa thấp hơn một chút, lần lượt là 98 USD và 99 USD

#### **Phân tích theo loại sản phẩm**

Danh mục phụ kiện (Accessories): Các sản phẩm Belt, Sunglasses, Scarf đều đạt giá trị tối đa 100 USD, cho thấy phụ kiện có thể có mức giá rất cao trong danh mục này.

Danh mục trang phục (Clothing): Các sản phẩm như Blouse, Skirt, T-shirt đều có sự chênh lệch lớn giữa giá trị tối thiểu và tối đa (20-100 USD), cho thấy mức giá đa dạng, phù hợp với nhiều đối tượng khách hàng.

Danh mục giày dép (Footwear): Các sản phẩm Boots và Sneakers đều đạt giá trị tối đa 100 USD nhưng Boots có giá trị tối thiểu cao hơn (21 USD) thể hiện Boots là sản phẩm cao cấp hơn.

Danh mục áo khoác ngoài (Outerwear): Các sản phẩm Coat và Jacket đều đạt giá trị tối đa 100 USD, cho thấy đây là các sản phẩm chủ lực trong danh mục Outerwear với sự đa dạng về mức giá.

### Kết luận

Phần lớn các sản phẩm có giá trị tối thiểu là 20 USD trừ các sản phẩm cao cấp khác. Ngoài ra, nhiều sản phẩm đạt giá trị tối đa 100 USD đặc biệt là các sản phẩm trong danh mục phổ biến như Clothing và Accessories.

## Phân tích mức độ sử dụng mã khuyến mãi dựa trên giới tính.

A graph of a number of people

Description automatically generated with medium confidence

Ảnh 18 : Mức độ sử dụng mã khuyến mãi dựa trên giới tính.

### Giới thiệu biểu đồ

Mục tiêu: Phân tích mức độ sử dụng mã khuyến mãi (Promo Code) dựa trên giới tính (Gender) và danh mục sản phẩm (Item Purchased).

Biểu đồ gồm :

* Trục ngang: Số lượng người sử dụng mã khuyến mãi (Count of Subscription Status).
* Màu sắc: Phân biệt giới tính:
* Màu cam: Female.
* Màu xanh: Male.
* Phân loại: Người sử dụng mã khuyến mãi (Yes) và không sử dụng mã khuyến mãi (No).

### Phân tích biểu đồ

#### **Xu hướng giới tính không sử dụng mã khuyến mãi (Promo Code: No)**

Ở nữ giới (Female) thường có xu hướng không sử dụng mã khuyến mãi đặc biệt là các sản phẩm gồm : Sandals và Shirt (1,770 người không sử mã khuyến mãi), Handbag (1,740 người không sử dụng mã khuyến mãi).

#### **Xu hướng giới tính sử dụng mã khuyến mãi (Promo Code: Yes)**

Ở nam giới (Male) có tỉ lệ sử dụng mã khuyến mãi nhiều và nổi trội ở các sản phẩm là T-shirt và Coat (2,370 người không sử mã khuyến mãi) và Socks (1,470 người không sử mã khuyến mãi).

### Kết luận

Nữ giới ít sử dụng mã khuyến mãi hơn nam giới thể hiện sự chiếm ưu thế rõ rệt trong nhóm không sử dụng mã. Ngược lại nam giới lại tận dụng mã khuyến mãi nhiều hơn để phục vụ nhu cầu mua sắm đặc biệt trong các sản phẩm thời trang giá trị cao.

Như vậy ta có thể đề nghị rằng nên tập trung vào nữ giới bằng cách tặng thêm quà hữu ích khi sử dụng mã khuyến. m, thay vì chỉ dựa vào mã khuyến mãi.

Tăng cường ưu đãi cho nam giới bằng cách đẩy mạnh chương trình mã khuyến mãi để tiếp tục khai thác xu hướng hiện tại.

## Phân tích tổng số tiền chi tiêu theo từng bang.

A map of the united states

Description automatically generated

Ảnh 19 : Tổng số tiền chi tiêu theo từng bang

### Giới thiệu biểu đồ

Mục tiêu: Phân tích tổng số tiền chi tiêu (Purchase Amount USD) theo từng bang tại Hoa Kỳ.

Biểu đồ gồm :

* Màu sắc: Thể hiện mức độ chi tiêu, từ thấp đến cao:
* Màu xanh nhạt: Chi tiêu thấp (103,110 USD).
* Màu cam đậm: Chi tiêu cao (173,520 USD).
* Các bang được đánh dấu: Hiển thị số liệu cụ thể về chi tiêu trong từng bang.

### Phân tích biểu đồ

#### **Các bang có mức chi tiêu cao**

Có thể thấy các bang California, Texas, và Florida là những bang dẫn đầu về tổng chi tiêu với giá trị gần mức cao nhất.

#### **Các bang có mức chi tiêu thấp**

Hawaii và Alaska là 2 bang có tổng chi tiêu thấp hơn các bang khác, điều này có thể do đặc điểm vị trí địa lý hoặc dân số thấp hơn.

### Phân tích sâu hơn

#### **Ảnh hưởng từ khu vực**

Các bang ở Miền Tây như California, Washington có mức chi tiêu cao, nhờ kinh tế phát triển mạnh.Ngược lại, Alaska và Hawaii có mức chi tiêu thấp hơn rõ rệt.

Ở Miền Nam các bang như Texas và Florida là những bang dẫn đầu, nhờ dân số lớn và hoạt động kinh tế mạnh.

### Kết luận

Mức chi tiêu tập trung cao nhất ở các bang đông dân và kinh tế phát triển như California, Texas, và Florida. Do đó ta cần tập trung khai thác ở các bang này nhiều hơn. Ngoài ra ta cần nghiên cứu để phát triển và mở rộng kinh tế tại các bang có chi tiêu thấp để khai thác tiềm năng còn ẩn dấu.

## Mối quan hệ giữa mùa thu và các sản phẩm quần áo được mua

A graph of a bar chart

Description automatically generated with medium confidence

Ảnh 20 : Mối quan hệ giữa mùa thu và các sản phẩm.

### Giới thiệu biểu đồ

Mục tiêu: Xác định xu hướng mua sắm các mặt hàng quần áo trong mùa thu (Season: Fall).

Biểu đồ gồm:

* Danh mục sản phẩm quần áo: Bao gồm các mặt hàng như Jacket, T-shirt, Pants, Shoes, Sweater, Dress, Scarf, v.v.
* Cột hiển thị: Tổng số lượt mua (Count of Season1) cho từng sản phẩm.

### Phân tích biểu đồ

#### **Sản phẩm được mua nhiều nhất trong mùa thu**

Áo khoác (Jacket) là mặt hàng phổ biến nhất trong mùa thu với 1.620 lượt mua vì đây là sản phẩm phù hợp với thời tiết mát mẻ và nhu cầu giữ ấm. Tiếp theo là Sweater với 1,260 lượt mua, có thể thấy Sweater có sự phổ biến cao nhờ khả năng giữ ấm mà vẫn đảm bảo phong cách. Từ đó ta kết luận rằng đây là 2 sản phẩm mang tính thiết yếu, đặc biệt tại các khu vực khí hậu lạnh.

#### **Các sản phẩm ít phổ biến hơn**

Giày và váy là 2 sản phẩm có số lượt mua tương đối thấp (780 và 1,020 lượt mua) vì có thể do chúng ít liên quan đến nhu cầu giữ ấm. Handbag và Scarf (1,440 và 1,200 lượt mua) là những phụ kiện có mức tiêu thụ khá cân đối vì đây là sản phẩm mang tính trang trí cho trang phục thêm phong cách hơn.

### Xu hướng dự đoán

Các sản phẩm được kỳ vọng tăng trưởng:

* Áo khoác và áo len (Jacket, Sweater) là sản phẩm được dự đoán sẽ tiếp tục chiếm tỷ lệ cao trong mùa thu tới, đặc biệt ở khu vực lạnh. Ngoài ra khăn quàng cổ (Scarf) cũng có thể trở thành xu hướng mạnh mẽ hơn khi là phụ kiện của thời trang mùa thu.

Sản phẩm cần cải thiện:

* Váy (Skirt) là sản phẩm được tiêu thụ kém do không có đặc điểm giữ ấm trong thời tiết lạnh trừ khi sản phẩm này được cải tiến về chất liệu cũng như độ dài che chân sẽ được người dùng chọn mua.

### Kết luận

Tăng cường quảng cáo các mặt hàng Jacket, T-shirt, và Sweater cũng như nhấn mạnh các tính năng giữ ấm, chất liệu bền bỉ của các sản phẩm này

Cải thiện chất liệu cũng như kiểu dáng của sản phẩm váy để có thể tiếp cận với người mua hơn.

## Mối quan hệ giữa mùa xuân và các sản phẩm được mua.

A graph of different colored lines

Description automatically generated with medium confidence

Ảnh 21 : Mối quan hệ giữa mùa xuân và các sản phẩm

### Giới thiệu biểu đồ

Mục tiêu: Xác định xu hướng mua sắm các mặt hàng trong mùa xuân (Season: Spring).

Biểu đồ gồm :

* Danh mục sản phẩm: Bao gồm các mặt hàng như Jacket, Shirt, Pants, Sneakers, Scarf, v.v.
* Cột hiển thị: Tổng số lượt mua (Count of Season1) cho từng sản phẩm.

### Phân tích biểu đồ

#### **Sản phẩm được mua nhiều nhất trong mùa xuân**

Áo khoác tiếp tục là sản phẩm dẫn đầu trong mùa xuân với 1.410 lượt mua cho thấy đây là sản phẩm có nhu cầu sử dụng cao cho thời tiết se lạnh vào đầu mùa. Ngoài ra đây là một sản phẩm quan trọng cho cả thời trang và chức năng giữ ấm.

Sweater, Dress, và Skirt (1,380 lượt mua mỗi loại) là 3 sản phẩm có nhiều lượt mua sau áo khoác bởi vì Sweater là lựa chọn phổ biến nhờ khả năng giữ ấm nhẹ, phù hợp thời trang. Dress và Skirt cũng được ưa chuộng hơn so với mùa đông nhờ có thiết kế nhẹ nhàng phù hợp với dịp tết.

#### **Các sản phẩm ít phổ biến hơn**

Mũ có số lượng tiêu thụ thấp nhất chỉ với 810 lượt mua do có thiết kế ít liên quan đến thời trang hoặc không có nhu cầu thực tế trong mùa xuân.

So với mùa đông ủng giảm mức tiêu thụ nhiều bởi vì thời tiết ở mùa xuân đã trở nên ấm áp hơn và nhường chỗ cho các loại giày nhẹ như Sneakers.

### Xu hướng dự đoán

Các sản phẩm có khả năng tăng trưởng:

Sweater và Jacket là sản phẩm được dự đoán tiếp tục duy trì lượng mua cao trong mùa xuân đặc biệt là khi đầu mùa

Dress và Skirt là sản phẩm được dự đoán sẽ có lượt mua tăng dần khi thời tiết ấm dần trong mùa xuân, đặc biệt là thời gian gần tết thì lượt mua của sản phẩm này sẽ tăng nhanh.

### Kết luận

Đầu tư vào Jacket, Sweater ở đầu mùa xuân vì khi đó thời tiết vẫn còn se lạnh. Tăng cường thúc đẩy các mẫu thiết kế váy và chân váy để khai thác tối đa nhu cầu của khách hàng khi kịp tết đến.

Tăng cường chiến lược marketing với các sản phẩm phụ kiện đến với người dùng vào các dịp gần tết vì đây là dịp người dùng sẽ mua sắm nhiều nhất.

## Mối quan hệ giữa mùa hè và các sản phẩm được mua.

A graph of different colored lines

Description automatically generated with medium confidence

Ảnh 22 : Mối quan hệ giữa mùa hè và các sản phẩm

### Giới thiệu biểu đồ

Mục tiêu: Đánh giá xu hướng mua sắm các mặt hàng trong mùa hè (Season: Summer).

Biểu đồ gồm :

* Danh mục sản phẩm: Bao gồm các mặt hàng như Pants, Shirt, Blouse, Dress, v.v.
* Cột hiển thị: Tổng số lượt mua (Count of Season1) của từng sản phẩm trong mùa hè.

### Phân tích biểu đồ

#### **Sản phẩm được mua nhiều nhất trong mùa hè**

Quần là sản phẩm phổ biến nhất trong mùa hè với 1.500 lượt mua thể hiện tính tiện dụng và phù hợp cho cả thời trang thường ngày lẫn công việc.

Blouse với 1.410 lượt mua là sản phẩm được ưa chuộng sau pants bởi vì Blouse phù hợp với thời tiết nóng và thường được ưa chuộng bởi nữ giới. Ngoài ra Dress là sản phẩm lý tưởng cho các hoạt động ngoài trời hoặc sự kiện trong mùa hè.

Ba lô cũng là sản phẩm được tiêu thụ cao với 1.350 lượt mua có thể do nhu cầu đi du lịch, hoạt động ngoài trời hoặc các gia đình mua cho con để chuẩn bị đi học trong mùa hè cũng như năm học mới.

#### **Các sản phẩm ít phổ biến hơn**

Giày dép đặc biệt là Boots là sản phẩm ít được ưa chuộng nhất trong mùa hè và chỉ có 930 lượt mua vì không phù hợp với thời tiết nóng. Mặc dù mũ là sản phẩm hữu ích để bảo vệ khỏi ánh nắng, nhưng số lượng mua không cao so với các mặt hàng khác (chỉ với 900 lượt mua)

### Xu hướng dự đoán

Sản phẩm có khả năng tăng trưởng

* Blouse là sản phẩm được dự đoán sẽ phổ biến bởi phong cách thoải mái và nhẹ nhàng trong mùa hè.
* Backpack cũng là sản phẩm được dự đoán sẽ tăng nhanh về lượt mua nhờ nhu cầu cho các hoạt động du lịch, dã ngoại, hoặc học hè có thể làm tăng lượng tiêu thụ của sản phẩm này.

Sản phẩm cần cải thiện:

* Cần phải có kế hoạch cải tiến thiết kế hoặc mở rộng chiến dịch quảng bá của mũ để tăng sức hút trong mùa hè.Ngoài ra ta cũng cần tập trung vào các loại giày nhẹ, thoáng khí hoặc giày thời trang để đáp ứng nhu cầu người tiêu dùng.

### Kết luận

Tập trung vào Pants, Blouse để tận dụng nhu cầu cao trong mùa hè và tăng cường quảng bá các sản phẩm như Hat, Scarf, và Sunglasses để thúc đẩy doanh số. Tăng cường chiến dịch marketing về tính năng che nắng của các sản phẩm Hat, Scarf, và Sunglasses.

## Mối quan hệ giữa mùa đông và các sản phẩm được mua.

A graph of different colored lines

Description automatically generated

Ảnh 23 : Mối quan hệ giữa mùa đông và các sản phẩm

### Tổng quan biểu đồ

Mục tiêu: Xác định xu hướng mua sắm của khách hàng trong mùa đông (Season: Winter).

Biểu đồ gồm :

* Các loại sản phẩm: Sunglasses, Pants, Shirt, Boots, Dress, T-shirt, v.v.
* Tổng số lượt mua (Count of Season1) cho từng sản phẩm trong mùa đông.

### Phân tích biểu đồ

#### Các sản phẩm được mua nhiều nhất trong mùa đông

Kính râm đứng đầu danh sách với 1.560 lượt mua cho thấy nhu cầu sử dụng kính râm không chỉ giới hạn vào mùa hè mà còn phổ biến trong mùa đông có thể để tránh tuyến rơi trúng và tránh thời tiết quá lạnh có thể ảnh hưởng đến mắt.

Quần là một trong những sản phẩm cơ bản và được mua nhiều với 1.530 lượt mua do tính tiện lợi và phù hợp với thời tiết lạnh.

Hoodie là sản phẩm lý tưởng cho thời tiết lạnh và đạt 1440 lượt mua vì phong cách thời trang thoải mái cũng như có thể giữ ấm trong suốt mùa đông

#### **Các sản phẩm ít phổ biến**

Ba lô là sản phẩm ít được mua hơn trong mùa đông chỉ có 950 lượt mua có thể vì các hoạt động ngoài trời đã bị tuyết che phủ ở mùa đông thì học sinh hầu hết sẽ được nghỉ đông.

Gloves và Scarf là sản phẩm phụ kiện thiết yếu và phù hợp với mùa đông cho mùa đông nhưng số lượng mua găng tay và khăn choàng vẫn thấp hơn kỳ vọng. Cũng như Gloves và Scarf thì số lượng mua boots tương đối thấp, có thể do thị hiếu thay đổi hoặc xu hướng chọn các loại giày nhẹ hơn.

### Xu hướng dự đoán

Sản phẩm có tiềm năng tăng trưởng:

* Hoodie là sản phẩm được dự đoán trở thành sản phẩm thiết yếu trong mùa đông bởi sự thoải mái và ấm áp.
* Kính râm cũng là sản phẩm được dự đoán sẽ có nhiều lượt mua hơn vì đây là sản phẩm phù hợp với mùa đông cũng như sản phẩm này có thể bảo vệ mặt khỏi thời tiết lạnh và tuyết rơi.

Sản phẩm cần chiến lược tập trung:

* Gloves và Scarf là sản phẩm cần phải tập trung vào cải thiện thiết kế hoặc tạo các chiến dịch quảng bá để tăng sức mua trong mùa đông

### Kết luận

Nhấn mạnh các sản phẩm như Sunglasses, Pants, và Hoodie, vì chúng có nhu cầu cao trong mùa đông.

Quảng bá Scarf và Gloves nhiều hơn để tăng doanh số bán hàng.

## Phân tích số lượng kích cỡ theo mùa và giới tính.

A close-up of a graph

Description automatically generated

Ảnh 24 : Số lượng kích cỡ theo mùa và giới tính

### Tổng quan biểu đồ

Mục tiêu: Hiển thị số lượng kích cỡ (size) của các sản phẩm phân chia theo mùa (Season) và giới tính (Male/Female).

Biểu đồ gồm :

* Chia cột theo từng Season: Fall, Spring, Summer, Winter.
* Phân loại các sản phẩm theo kích cỡ (Small, Medium, Large).
* Màu sắc biểu diễn giới tính: Màu cam đại diện cho nữ (Female), màu xanh lam đại diện cho nam (Male).

### Phân tích biểu đồ

#### **Tổng quát theo mùa**

Với mùa đông (Winter):

* Kích cỡ Large (L) được lưu trữ nhiều hơn ở cả hai giới đặc biệt cho các sản phẩm áo khoác (Coats), quần dài (Pants), và giày (Boots).
* Nữ giới có xu hướng mua size Medium (M) nhiều hơn trong mùa đông đặc biệt là áo Hoodie và khăn (Scarfs).

Với mùa hè (Summer):

* Kích cỡ Small (S) phổ biến hơn cho nữ giới với các sản phẩm nhẹ nhàng như váy (Dresses), áo phông (T-shirts), và kính râm (Sunglasses).
* Nam giới tập trung vào kích cỡ Medium (M) cho áo sơ mi (Shirts) và giày thể thao (Sneakers).

Với mùa thu (Fall):

* Kích cỡ Medium (M) và Large (L) được lưu trữ nhiều ở nam giới với các sản phẩm ấm áp như áo khoác (Coats) và quần dài (Pants).
* Ở nữ giới size Medium (M) được lưu trữ nhiều với sản phẩm Jacket và Shocks nhằm giữ ấm cơ thể tránh các bệnh liên quan đến cảm cúm.

Mùa xuân (Spring):

* Nữ giới thường lựa chọn kích cỡ Medium (M) cho váy (Skirts) và nam giới ưu tiên Medium và Large cho các sản phẩm như áo Hoodie và quần dài (Pants).

#### **So sánh theo giới tính**

Nam giới (Male) thường có xu hướng mua kích cỡ Medium (M) nhiều nhất, ngoài ra kích cỡ Large (L) cũng được lưu trữ phổ biến. Có thể thấy 2 kích thước này phù hợp với nhu cầu chung của nhóm khách hàng nam.

Nữ giới (Female) thường sẽ mua những sản phẩm kích cỡ Small (S) và Medium (M) cho các sản phẩm thời trang vì cơ thể nữ giới thường có nhỏ và mảnh mai hơn nam giới nên họ thường sẽ chọn những quần áo vừa vặn với cơ thể. Ngoài ra kích cỡ Large (L) phổ biến hơn trong các mùa lạnh (Fall, Winter) cho các sản phẩm như áo Hoodie hoặc quần dài.

### Dự đoán xu hướng mua sắm

Phát triển các sản phẩm theo mùa:

* Cung cấp thêm các sản phẩm kích cỡ Small và Medium trong các mùa ấm áp như mùa xuân và mùa hạ.
* Đa dạng hóa lựa chọn kích cỡ Large trong mùa lạnh như mùa thu và mùa đông

### Kết luận

Biểu đồ này giúp ta xác định rõ nhu cầu kích cỡ của khách hàng theo mùa và giới tính, từ đó tập trung vào xu hướng kích cỡ theo từng mùa sẽ giúp tăng cường hiệu quả kinh doanh, đồng thời giảm lãng phí trong quản lý kho hàng.

## Mối quan hệ giữa các khu vực địa lý và sản phẩm.

A graph of different colored lines

Description automatically generated with medium confidence

Ảnh 26 : Mối quan hệ giữa khu vực địa lý và sản phẩm

### Tổng quan biểu đồ

Mục tiêu: So sánh tổng mức chi tiêu (Purchase Amount - USD) theo từng loại sản phẩm ở các vùng địa lý (Midwest, Northeast, South, West).

Biểu đồ gồm :

* Trục ngang : Tổng mức chi tiêu
* Trục dọc : Từng sản phẩm ở các vùng khu vực
* Màu sắc : Từng màu sắc tượng trưng cho từng loại sản phẩm

### Phân tích biểu đồ

#### **Sản phẩm có chi tiêu cao nhất**

Hoodie tại khu vực West dẫn đầu với tổng chi tiêu 104.700 USD cho thấy đây là mặt hàng phổ biến nhất.

Jacket cũng nổi bật tại West với mức chi tiêu 100.470 USD và South với 90.570 USD nhấn mạnh tầm quan trọng của sản phẩm này ở các khu vực.

Blouse và Sweater cũng là những sản phẩm tiêu thụ mạnh tại South và West, với mức chi tiêu vượt 80.000 USD

#### **Sản phẩm có chi tiêu trung bình**

T-shirt, Pants, và Sunglasses có mức chi tiêu đều ở cả 4 khu vực, cho thấy đây là các mặt hàng phổ thông và được ưa chuộng trên toàn quốc.

Gloves có mức chi tiêu cao tại West với 79.860 USD và có mực chi tiêu thấp tại Midwest với 53.130 USD.

#### **Sản phẩm có chi tiêu thấp**

Handbag và Sandals có mức chi tiêu tương đối thấp ở các khu vực Midwest và Northeast với mức chi tiêu chỉ dao động từ 40.000 USD đến 60.000 USD

Sneakers có mức chi tiêu thấp nhất tại Midwest với 48.420 USD

### Kết luận

Ở khu vực West và South là các khu vực tiêu thụ mạnh các sản phẩm thuộc phân khúc thời trang mùa đông như Hoodie, Jacket, và Sweater.

Khu vực Midwest và Northeast có nhu cầu tập trung hơn vào các sản phẩm phổ thông chẳng hạn như Sunglasses, Pants, và Gloves.

## Phân tích mùa mua sắm tốt nhất.

A graph of sales on a white background

Description automatically generated with medium confidence

Ảnh 27 : Mùa mua sắm tốt nhất

### Tổng quan biểu đồ

Mục tiêu: So sánh mức chi tiêu và số lượng mã khuyến mãi được sử dụng (Promo Code Used) theo từng mùa trong năm để xác định mùa mua sắm sôi động nhất.

Biểu đồ gồm :

* Trục dọc : Bao gồm 4 mùa - Fall (Thu), Spring (Xuân), Winter (Đông), và Summer (Hè).
* Trục ngang : Số lượng mã khuyến mãi và tổng mức chi tiêu được hiển thị.
* Màu sắc : Mỗi mùa tương ứng với một màu sắc khác nhau

### Phân tích biểu đồ

#### **Mức chi tiêu theo mùa**

Với mùa Fall (Thu):

* Dẫn đầu với mức chi tiêu 1,800,540 USD cho thấy đây là mùa mua sắm nhộn nhịp nhất.

Với mùa Spring (Xuân):

* Đứng thứ hai với mức chi tiêu 1,760,370 USD nhấn mạnh tầm quan trọng của việc mua sắm đầu năm và xu hướng mua sắm cho dịp tết.

Với mùa Winter (Đông):

* Chi tiêu đạt 1,758,210 USD, rất sát sao với mùa Xuân vì các sự kiện như lễ Giáng sinh. Mùa Summer (Hè):
* Có mức chi tiêu thấp nhất, 1,673,310 USD, có thể do ít sự kiện mua sắm lớn trong mùa này.

#### **Số lượng mã khuyến mãi được sử dụng**

Mùa xuân có số mã khuyến mãi được sử dụng cao nhất (29,970 mã) cho thấy sự phổ biến của các chương trình khuyến mãi đầu năm trong mùa này.

Mùa thu và đông cũng có số lượng mã khuyến mãi tương đối cao (29,250 và 29,130 mã).

Mùa hạ ghi nhận số lượng mã khuyến mãi thấp nhất (28,650 mã) đồng nhất với mức chi tiêu thấp nhất.

### Kết luận

Mùa xuân là thời điểm lý tưởng cho các chiến dịch mua sắm lớn, với mức chi tiêu cao thứ 2 và số lượng mã khuyến mãi được sử dụng nhiều.

Mùa thu và đông cũng quan trọng và cần được ưu tiên trong các chiến lược khuyến mãi.

Mùa Summer có tiềm năng tăng trưởng nếu được thúc đẩy bởi các chương trình giảm giá hoặc sự kiện mua sắm hấp dẫn.

## Dự đoán số lượng quần áo được dự trữ theo kích cỡ

A graph with different colored squares

Description automatically generated

Ảnh 28 : Dự đoán số lượng quần áo được dự trữ theo kích cỡ

### Tổng quan biểu đồ

Mục tiêu: Biểu đồ này nhằm dự đoán nhu cầu lưu trữ quần áo theo từng kích cỡ (L, M, S, XL) cho từng danh mục sản phẩm cụ thể.

Biểu đồ gồm :

* Trục dọc : Tổng hợp số lượng quần áo như Backpack, Belt, Blouse, …
* Trục ngang : Các kích cỡ tương ứng ( L, M, S, XL)

### Phân tích biểu đồ

#### **Tổng số lượng theo kích cỡ**

Kích cỡ M (Medium) là kích cỡ được lưu trữ dẫn đầu trong hầu hết các danh mục sản phẩm, trong đó :

* Sweater: 2,310 sản phẩm.
* Shirt: 2,580 sản phẩm.

Kích cỡ L (Large) là kích cỡ được lưu trữ xếp thứ hai nổi bật trong một số danh mục như:

* Backpack (2,280 sản phẩm).
* Blouse (2,250 sản phẩm).

Kích cỡ S (Small) là kích cỡ được lưu trữ thấp hơn M và L nhưng vẫn có nhu cầu cao, đặc biệt với các sản phẩm như :

* Scarf (840 sản phẩm).
* Handbag (870 sản phẩm).

Kích cỡ XL (Extra Large) là kích cỡ có số lượng được lưu trữ thấp nhất với :

* Belt (570 sản phẩm).
* Shoes (450 sản phẩm).

#### **Sản phẩm nổi bật**

Shirt (Áo sơ mi) là sản phẩm có nhu cầu lưu trữ cao nhất đặc biệt là kích thước M và L.

Sweater là sản phẩm được lưu trữ nhiều nhất về kích cỡ M (2,310 sản phẩm) cho thấy đây là sản phẩm phổ biến và cần được dự trữ nhiều.

### Kết luận

Kích cỡ M và L là 2 kích cỡ chiếm phần lớn nhu cầu trong hầu hết các danh mục sản phẩm và là kích cỡ trọng tâm khi dự trữ.

Kích cỡ S và XL dù có nhu cầu lưu trữ thấp hơn những vẫn rất cần cung cấp đủ để phục vụ các đối tượng khách hàng đặc thù.

# TỔNG KẾT

## Tổng kết đề tài

Dựa trên các biểu đồ trực quan hóa dữ liệu trên, có thể rút ra một số kết luận tổng quan về hành vi mua sắm và xu hướng mua sắm như sau :

### Mùa ảnh hưởng đáng kể đến xu hướng tiêu dùng

Mùa xuân dường như là mùa mua sắm nổi bật với lượng mua sắm và tổng doanh thu so với các mùa khác. Điều này có thể liên quan đến nhu cầu mua sắm đầu năm cũng như nhu cầu mua sắm quần áo thời trang cho dịp tết

Mùa hè có giá trị mua sắm thấp hơn có thể do nhu cầu tiêu dùng tập trung vào các sản phẩm phù hợp với thời tiết nóng bức.

### Khác biệt trong hành vi mua sắm theo giới tính

Nữ giới thường có xu hướng mua sắm nhiều hơn và không sử dụng mã khuyến mãi so với nam giới. Điều này cho thấy nữ giới có thể ưu tiên giá trị sản phẩm hơn là giá cả khi quyết định mua hàng.

Nam giới sử dụng mã khuyến mãi thường xuyên hơn tập trung vào việc tận dụng giá trị tiết kiệm từ các ưu đãi.

### Sự đa dạng và nhu cầu về kích thước sản phẩm

Sản phẩm thời trang như áo phông, quần jeans, và giày dép chiếm tỉ lệ lớn trong danh mục sản phẩm được mua sắm, với sự phân bố đều ở các kích thước.

Các kích thước trung bình (M) thường được tiêu thụ nhiều nhất, cho thấy nhu cầu chủ đạo đến từ người tiêu dùng phổ thông.

### Hiệu quả của mã khuyến mãi theo mùa

Số lượng mã khuyến mãi được sử dụng không đồng đều giữa các mùa, trong đó mùa xuân và mùa thu ghi nhận sự gia tăng đáng kể. Điều này cho thấy các chiến dịch khuyến mãi trong thời gian này có hiệu quả tốt trong việc kích thích nhu cầu mua sắm của người dùng.

## Bài học thông qua đề tài

### Hiểu sâu hơn về nhu cầu của khách hàng

Hành vi tiêu dùng biến đổi theo mùa: Nhu cầu của khách hàng thay đổi theo thời gian, đặc biệt là giữa các mùa. Do đó, tacần nắm bắt các giai đoạn cao điểm để tối ưu hóa kho hàng, điều chỉnh sản phẩm phù hợp và tăng cường hoạt động tiếp thị.

Khác biệt theo giới tính: Nam giới và nữ giới có hành vi mua sắm khác nhau. Hiểu được sự khác biệt này giúp ta thiết kế chương trình khuyến mãi và sản phẩm phù hợp hơn với từng nhóm khách hàng.

### Tăng cường chiến lược khuyến mãi

Hiệu quả của mã khuyến mãi: Mã khuyến mãi không phải lúc nào cũng hấp dẫn với mọi nhóm khách hàng, đặc biệt là nữ giới vì họ thường quan tâm nhiều hơn đến sản phẩm. Do đó ta cần điều chỉnh các chương trình khuyến mãi sao cho phù hợp với thói quen của từng nhóm khách hàng.

Tận dụng các mùa cao điểm: Mùa thu và mùa xuân là thời điểm lý tưởng để tăng cường các chiến dịch tiếp thị và khuyến mãi vì đây là thời gian mua sắm sôi động nhất.

# TÀI LIỆU THAM KHẢO