

TRƯỜNG ĐẠI HỌC KINH TẾ ĐÀ NẴNG

KHOA THƯƠNG MẠI ĐIỆN TỬ



BÁO CÁO ĐỀ ÁN THỰC HÀNH 1

CHỦ ĐỀ

BÁO CÁO HÀNG TỒN KHO

GVHD: Trương Hồng Tuấn

Lớp: 47K29.1

Nhóm:

Họ và tên	Mã sinh viên	Phần trăm đóng góp
Nguyễn Thị Minh Phượng	211124.29129	100%
Nguyễn Đỗ Kim Thi	211124029134	100%

Đà Nẵng, ngày 1 tháng 5 năm 2024

MỤC LỤC

PHẦN 1: GIỚI THIỆU TỔNG QUAN.....	5
1. Lí do chọn đề tài:	5
1.1. Tầm quan trọng của báo cáo hàng tồn kho:	5
1.2. Nhu cầu thực tế:	5
1.3. Khả năng ứng dụng cao:	5
1.4. Khả năng nghiên cứu:	5
2. Mục tiêu của đề tài:	6
3. Phương pháp thực hiện:	6
3.1. Nghiên cứu tổng quan:	6
3.2. Thu thập và tích hợp dữ liệu:	6
3.3. Xây dựng công cụ trực quan hóa dữ liệu- dashboard:	6
3.4. Xây dựng mô hình:	7
3.5. Kiểm tra, đánh giá và hoàn thiện:	7
4. Đối tượng nghiên cứu:	7
4.1. Đối tượng báo cáo:	7
4.2. Phạm vi báo cáo:	7
PHẦN 2: CƠ SỞ LÝ THUYẾT.....	8
1. Báo cáo hàng tồn kho:	8
1.1. Hàng tồn kho là gì:	8
1.2. Báo cáo hàng tồn kho là gì:	8
1.3. Vai trò của báo cáo hàng tồn kho:	8
2. Các phương pháp quản lý hàng tồn kho:	9
3. Dashboard:	10
3.1. Khái niệm:	10
3.2. Các chức năng:	11
3.3. Các bước xây dựng Dashboard:	11
PHẦN 3: DỮ LIỆU.....	13
1. Bộ dữ liệu AdventureWorks:	13
1.1. Giới thiệu bộ dữ liệu:	13
1.1.1. Mô tả ngữ cảnh doanh nghiệp:	13

1.2. Chuẩn bị dữ liệu:	14
2. Mô tả dữ liệu:	19
2.1. Bảng DimProduct:	20
2.2. Bảng FactProductInventory:	21
2.3. Bảng DimProductCategory:	21
2.4. Bảng DimProductSubcategory:	21
2.5. Bảng DimDate:	22
3. Tiền xử lý dữ liệu:	23
3.1. Bảng DimProduct:	23
3.2. Bảng DimProductInventory:	24
3.3. Bảng DimProductCategory:	25
3.4. Bảng DimProductSubcategory:	26
3.5. Bảng DimDate:	27
PHẦN 4: TRỰC QUAN HÓA:	28
1. Tổng quan hàng tồn kho:	28
2. Biểu đồ xu hướng hàng tồn kho qua các năm:	29
3. Biểu đồ tỷ lệ hàng tồn kho theo danh mục sản phẩm:	31
4. Biểu đồ Top 10 danh mục sản phẩm con tồn kho nhiều nhất:	33
5. Biểu đồ Top 10 danh mục sản phẩm con tồn kho ít nhất:	34
6. Biểu đồ số lượng nhập/ xuất theo năm:	35
7. Biểu đồ số lượng và chi phí hàng tồn kho theo danh mục sản phẩm:	36
8. Số lượng hàng tồn kho theo màu sắc:	38
9. Tỷ lệ hàng tồn kho theo lớp sản phẩm:	39
10. Tỷ lệ hàng tồn kho theo phong cách:	40
11. Mức tồn kho an toàn vs Điểm đặt hàng theo danh mục sản phẩm con:	41
12. Dashboard báo cáo hàng tồn kho:	42
PHẦN 5: XÂY DỰNG MÔ HÌNH.	43
1. Tiền xử lý dữ liệu:	43
2. Xây dựng mô hình Random Forest:	46
3. Đánh giá mô hình:	46
4. Dự đoán hàng tồn kho:	48

PHẦN 6: KẾT LUẬN.....	50
1. Kết quả dự án đạt được:	50
1.1. Đối với doanh nghiệp:	51
1.2. Đối với người thực hiện dự án:	51
2. Hạn chế và hướng cải thiện của dự án:	51
PHỤ LỤC.....	52

DANH MỤC HÌNH ẢNH

Chart 1. Tổng quan hàng tồn kho.	28
Chart 2. Xu hướng hàng tồn kho qua các năm.	29
Chart 3. Tỷ lệ hàng tồn kho theo danh mục sản phẩm.	31
Chart 4. Top 10 danh mục sản phẩm con tồn kho nhiều nhất.	33
Chart 5. Top 10 danh mục sản phẩm con tồn kho ít nhất.	34
Chart 6. Số lượng hàng nhập/ xuất theo năm.	35
Chart 7. Biểu đồ số lượng và chi phí hàng tồn kho theo danh mục sản phẩm.	37
Chart 8. Giá trị trung bình của các danh mục sản phẩm.	38
Chart 9. Số lượng hàng tồn kho theo màu sắc của sản phẩm.	38
Chart 10. Tỷ lệ hàng tồn kho theo phân khúc sản phẩm (H = High, M = Medium, L = Low).	39
Chart 11. Tỷ lệ hàng tồn kho theo phong cách (W = Women's, M = Men's, U = Universal).	40
Chart 12. Mức tồn kho an toàn vs Điểm đặt hàng theo Subcategory.	41
Chart 13. Dashboard báo cáo hàng tồn kho.	42

PHẦN 1: GIỚI THIỆU TỔNG QUAN.

1. Lí do chọn đề tài:

1.1. Tầm quan trọng của báo cáo hàng tồn kho:

- Báo cáo hàng tồn kho đóng vai trò then chốt trong việc quản lý hiệu quả tài sản lưu động, ảnh hưởng trực tiếp đến hiệu quả sản xuất, kinh doanh và lợi nhuận của doanh nghiệp.
- Cung cấp thông tin chi tiết về số lượng, giá trị, tình trạng hàng hóa tồn kho, hỗ trợ theo dõi, ra quyết định, đánh giá hiệu quả hoạt động.

1.2. Nhu cầu thực tế:

- Việc quản lý hàng tồn kho hiệu quả là vấn đề quan tâm của mọi doanh nghiệp, đòi hỏi có hệ thống báo cáo khoa học, chính xác.
- Nắm vững kiến thức về báo cáo hàng tồn kho giúp bản thân đáp ứng nhu cầu thực tế trong công việc, hỗ trợ doanh nghiệp quản lý hiệu quả..

1.3. Khả năng ứng dụng cao:

- Kiến thức về báo cáo hàng tồn kho có thể áp dụng trong nhiều lĩnh vực như kế toán, quản lý kho, tài chính, kiểm toán,...
- Giúp đánh giá hiệu quả hoạt động kinh doanh, xác định điểm mạnh, điểm yếu, đề xuất giải pháp tối ưu hóa quản lý hàng tồn kho, nâng cao lợi nhuận.
- Tư vấn cho doanh nghiệp về cách thức lập báo cáo hiệu quả, hỗ trợ họ quản lý tốt hơn.

1.4. Khả năng nghiên cứu:

- Có nhiều nguồn tài liệu, số liệu phong phú về báo cáo hàng tồn kho, từ sách vở, báo cáo tài chính doanh nghiệp đến các bài viết khoa học.
- Cho phép nghiên cứu đa dạng các khía cạnh, từ phương pháp lập báo cáo, phân tích dữ liệu đến ứng dụng thực tế.

- Nâng cao kiến thức và kỹ năng lập báo cáo chính xác, hiệu quả, phục vụ cho công việc và học tập.
- Với những lý do trên, đề tài "Báo cáo hàng tồn kho" là lựa chọn phù hợp cho nghiên cứu, giúp hiểu rõ tầm quan trọng, vai trò và cách thức lập báo cáo hiệu quả, từ đó áp dụng kiến thức thực tế, góp phần nâng cao hiệu quả hoạt động kinh doanh của doanh nghiệp.

2. Mục tiêu của đề tài:

- Mục tiêu chính của đề tài là xây dựng một công cụ phân tích và trực quan hóa dữ liệu hàng tồn kho hiệu quả, giúp các doanh nghiệp nâng cao khả năng theo dõi, đánh giá và ra quyết định về quản lý hàng tồn kho. Cung cấp cho Ban lãnh đạo doanh nghiệp bức tranh toàn cảnh về tình hình hàng tồn kho của các sản phẩm và danh mục sản phẩm. Từ đó, Ban lãnh đạo doanh nghiệp đưa ra quyết định sáng suốt về quản lý hàng tồn kho, đảm bảo đáp ứng nhu cầu thị trường một cách hiệu quả nhất với mục đích tối ưu hoá chi phí lưu kho và nâng cao lợi nhuận cho doanh nghiệp.

3. Phương pháp thực hiện:

Để đạt được các mục tiêu đã đề ra, đề tài sẽ được triển khai theo các bước sau:

3.1. Nghiên cứu tổng quan:

- Tìm hiểu các khái niệm, lý thuyết liên quan đến quản lý hàng tồn kho và trực quan hóa dữ liệu
- Khảo sát các giải pháp hiện có trên thị trường về trực quan hóa hàng tồn kho

3.2. Thu thập và tích hợp dữ liệu:

- Xác định các nguồn dữ liệu hàng tồn kho bao gồm dữ liệu hàng xuất/ nhập vào kho, dữ liệu thời gian và dữ liệu liên quan đến sản phẩm.
- Tích hợp dữ liệu và tiền xử lý

3.3. Xây dựng công cụ trực quan hóa dữ liệu- dashboard:

- Lựa chọn công cụ trực quan hoá phù hợp và giao diện dễ sử dụng, dễ hiểu để sử dụng
- Thiết kế và phát triển giao diện trực quan biểu diễn dữ liệu bằng biểu đồ, đồ thị, bản đồ

- Tích hợp các tính năng lọc, tùy chỉnh trực quan theo nhu cầu người dùng

3.4. Xây dựng mô hình:

- Tiến hành tiền xử lý để chuẩn bị cho việc xây dựng mô hình.
- Lựa chọn mô hình phù hợp để xây dựng.
- Huấn luyện mô hình trên dữ liệu đã được chuẩn bị và đánh giá hiệu suất của mô hình bằng các chỉ số như MSE, R-squared.

3.5. Kiểm tra, đánh giá và hoàn thiện:

- Kiểm tra lại toàn bộ quá trình thực hiện đề tài
- Đánh giá khó khăn - thuận lợi khi thực hiện đề tài
- Tính khả thi của đề tài

4. Đối tượng nghiên cứu:

4.1. Đối tượng báo cáo:

Báo cáo này dành cho Ban lãnh đạo doanh nghiệp, đặc biệt là bộ phận Quản lý kho và Quản lý chuỗi cung ứng, nhằm cung cấp thông tin chi tiết về tình hình hàng tồn kho hiện tại, dự báo nhu cầu trong tương lai và đề xuất các giải pháp tối ưu hóa quản lý hàng tồn kho cho các danh mục sản phẩm có trong doanh nghiệp.

4.2. Phạm vi báo cáo:

Dự án phân tích dữ liệu hàng tồn kho theo các tiêu chí: thời gian, đặc tính sản phẩm, danh mục sản phẩm, phân khúc sản phẩm,...

Dự báo nhu cầu hàng tồn kho trong tương lai dựa trên xu hướng thị trường, kế hoạch bán hàng và các yếu tố tác động khác.

Dựa vào dữ liệu, đề xuất các giải pháp tối ưu hoá quản lý hàng tồn kho cho dự án, bao gồm:

- Chiến lược nhập hàng, xuất hàng phù hợp.
- Mức tồn kho tối ưu cho từng danh mục sản phẩm.
- Phương pháp quản lý hàng tồn kho hiệu quả.

PHẦN 2: CƠ SỞ LÝ THUYẾT.

1. Báo cáo hàng tồn kho:

1.1. Hàng tồn kho là gì:

Hàng tồn kho là những mặt hàng sản phẩm sẽ được doanh nghiệp giữ lại để bán ra sau cùng. Nói một cách dễ hiểu hơn thì hàng tồn kho chính là những mặt hàng dự trữ mà doanh nghiệp sản xuất để bán và kèm theo những thành phần khác tạo ra sản phẩm.

Dựa vào khái niệm có thể thấy được hàng tồn kho chính là sự liên kết của việc sản xuất và bán sản phẩm. Đồng thời còn là một phần của tài sản ngắn hạn, chiếm tỷ trọng lớn và có vai trò quan trọng trong hoạt động sản xuất, kinh doanh của chính doanh nghiệp. Vậy nên nếu biết cách quản trị hàng tồn kho một cách hiệu quả sẽ giúp cho doanh nghiệp giảm bớt các khoản phí không cần thiết và tăng thêm lợi nhuận hiệu quả khi sản xuất, kinh doanh.

1.2. Báo cáo hàng tồn kho là gì:

Báo cáo hàng tồn kho là một công cụ quản lý quan trọng giúp doanh nghiệp theo dõi số lượng và giá trị hàng hóa còn lại trong kho tại một thời điểm nhất định. Báo cáo này thường được thực hiện định kỳ, ví dụ như hàng tháng, hàng quý hoặc hàng năm.

1.3. Vai trò của báo cáo hàng tồn kho:

- Báo cáo hàng tồn kho đóng vai trò quan trọng trong việc quản lý kho hàng và hoạt động kinh doanh của doanh nghiệp. Cụ thể:

+ *Giúp theo dõi số lượng và giá trị hàng tồn kho:* Báo cáo hàng tồn kho cung cấp cho doanh nghiệp thông tin chính xác về số lượng và giá trị hàng hóa còn lại trong kho.

Nhờ đó, doanh nghiệp có thể dễ dàng nắm bắt tình hình hàng hóa và đưa ra quyết định phù hợp về việc nhập kho, xuất kho và sản xuất.

+ *Hỗ trợ quản lý dòng tiền:* Việc theo dõi hàng tồn kho giúp doanh nghiệp quản lý dòng tiền hiệu quả hơn. Doanh nghiệp có thể biết được số tiền đang bị ứ đọng trong hàng tồn kho và có kế hoạch thu hồi vốn nhanh hơn.

- *Giảm thiểu rủi ro:* Báo cáo hàng tồn kho giúp doanh nghiệp xác định các sản phẩm bán chậm, ế ẩm hoặc hết hạn sử dụng. Nhờ đó, doanh nghiệp có thể đưa ra các biện pháp kịp thời để giảm thiểu rủi ro tồn kho và thất thoát.

- *Tối ưu hóa hoạt động kho hàng:* Báo cáo hàng tồn kho giúp doanh nghiệp tối ưu hóa hoạt động kho hàng bằng cách sắp xếp hàng hóa hợp lý, sử dụng không gian kho hiệu quả và tiết kiệm chi phí lưu kho.

2. Các phương pháp quản lý hàng tồn kho:

Quản lý hàng tồn kho là những hoạt động lập kế hoạch kiểm soát, sắp xếp, bảo quản hàng hoá, xuất nhập hàng hoá, lưu trữ và sử dụng hàng tồn kho. Việc này giúp đảm bảo các hoạt động trong kho diễn ra trơn tru và có quy trình, kế hoạch cụ thể, đảm bảo doanh nghiệp luôn có đủ hàng để bán ra và tránh tồn đọng quá nhiều hàng hoá.

Có 9 phương pháp quản lý hàng tồn kho hiệu quả nhất:

- *Phương pháp thiết lập vị trí kho hàng:*

- + Sắp xếp hàng hóa khoa học, logic theo nhóm, chủng loại, tần suất xuất nhập kho.
- + Vị trí dễ dàng tiếp cận, thuận tiện cho việc kiểm tra, lấy hàng.
- + Áp dụng các phương pháp sắp xếp như ABC, XYZ,...

- *Ứng dụng phương pháp quản lý kho FIFO hoặc LIFO:*

- + FIFO (Nhập trước xuất trước): Sử dụng nguyên liệu, thành phẩm nhập kho trước.
- + LIFO (Nhập sau xuất trước): Sử dụng nguyên liệu, thành phẩm nhập kho sau.
- + Lựa chọn phương pháp phù hợp với ngành hàng, mục tiêu quản lý.

- *Phương pháp đặt mức tồn kho cho mỗi sản phẩm:*

- + Xác định mức tồn kho tối thiểu, tối đa cho từng sản phẩm.
- + Dựa trên nhu cầu tiêu thụ, dự báo bán hàng, thời gian nhập hàng.
- + Đảm bảo đủ hàng hóa đáp ứng nhu cầu, tránh tồn kho dư thừa hoặc thiếu hụt.

- *Mã hoá và dán nhãn tất cả hàng tồn kho:*

- + Sử dụng hệ thống mã code, tem nhãn để phân biệt sản phẩm.
- + Thông tin trên nhãn: mã sản phẩm, tên sản phẩm, nhà cung cấp, hạn sử dụng,...
- + Dễ dàng quản lý, theo dõi, kiểm kê hàng hóa.

- *Kiểm soát quy trình xuất - nhập kho:*

- + Quy trình xuất nhập kho rõ ràng, chặt chẽ.

- + Sử dụng hóa đơn, phiếu xuất kho, phiếu nhập kho.
- + Cập nhật số lượng tồn kho theo từng lần xuất nhập.
- *Kiểm kê cho định kì:*
 - + Kiểm tra số lượng, tình trạng hàng hóa thực tế so với sổ sách.
 - + Phát hiện sai sót, thất thoát, hư hỏng kịp thời.
 - + Điều chỉnh số liệu tồn kho cho chính xác.
- *Quản lý mối quan hệ với nhà cung cấp:*
 - + Lựa chọn nhà cung cấp uy tín, đảm bảo chất lượng hàng hóa.
 - + Duy trì mối quan hệ hợp tác lâu dài, có lợi cho cả hai bên.
 - + Đàm phán giá cả, điều kiện thanh toán hợp lý.
- *Quản lý kho có kế hoạch dự phòng:*
 - + Lưu trữ lượng hàng nhất định để dự phòng cho trường hợp thiếu hụt đột xuất.
 - + Đảm bảo hoạt động sản xuất, kinh doanh diễn ra suôn sẻ.
 - + Giảm thiểu thiệt hại do thiếu hàng hóa.
- *Sử dụng phần mềm quản lý kho:*
 - + Tự động hóa các quy trình quản lý kho hàng.
 - + Cung cấp báo cáo, phân tích dữ liệu tồn kho chính xác, hiệu quả.
 - + Tiết kiệm thời gian, chi phí, nâng cao năng suất làm việc.

Việc áp dụng hiệu quả 9 phương pháp quản lý hàng tồn kho được nêu trên sẽ giúp doanh nghiệp tối ưu hóa nguồn lực, giảm thiểu chi phí, nâng cao lợi nhuận và tăng khả năng cạnh tranh trong thị trường đầy biến động. Doanh nghiệp cần lựa chọn và triển khai các phương pháp phù hợp với đặc thù ngành hàng, quy mô hoạt động và mục tiêu kinh doanh của mình để đạt được hiệu quả quản lý tối ưu nhất.

3. Dashboard:

3.1. Khái niệm:

Dashboard là một bảng điều khiển kỹ thuật số mà tại đó toàn bộ thông tin được mã hóa và hiển thị dưới dạng biểu đồ, số liệu và hình ảnh. Dựa trên kết quả hiển thị trên dashboard, doanh nghiệp có thể dễ dàng theo dõi, nắm bắt và quản lý một cách tổng quan những dữ liệu quan trọng phục vụ cho các hoạt động kinh doanh.

3.2. Các chức năng:

Trong kỷ nguyên số 4.0, việc sử dụng dashboard đã trở thành một phần tất yếu của hoạt động doanh nghiệp và quản lý thông tin. Vai trò của dashboard đối với doanh nghiệp là:

- + *Theo dõi nhiều KPI và chỉ số cùng một lúc*: Dashboard giúp tổng hợp thông tin và thống kê từ nhiều nguồn theo thời gian thực tế. Từ đó, doanh nghiệp có thể nắm bắt và quản lý các hoạt động trong doanh nghiệp hiệu quả hơn mà không tốn kém nhiều thời gian và công sức.
- + *Trình bày số liệu trực quan và dễ hiểu hơn*: Thay vì phải đọc những con số rời rạc, Dashboard thể hiện dữ liệu dưới dạng biểu đồ và hình ảnh trực quan và nổi bật. Điều này giúp doanh nghiệp nhanh chóng hiểu được thông tin.
- + *Hỗ trợ tính năng chia sẻ và cộng tác*: Dashboard cho phép người dùng chia sẻ đến những đối tượng có liên quan theo mong muốn. Khi đó, nhà quản lý và nhân sự có thể nắm bắt thông tin kịp thời, nhanh chóng và đồng bộ.
- + *Giúp cho việc làm báo cáo dễ dàng hơn*: Dashboard giúp doanh nghiệp tạo tự động các báo cáo dựa trên dữ liệu đã có. Doanh nghiệp sẽ không cần phải tổng hợp hay tìm kiếm các thông tin từ đầu.
- + *Tối ưu cho các thiết bị di động*: Dashboard cho phép theo dõi và quản lý từ cả điện thoại và laptop. Khi đó, người dùng có thể theo dõi thông tin mọi lúc, mọi nơi và cập nhật dữ liệu thời gian thực.

3.3. Các bước xây dựng Dashboard:

- Bước 1. Xác định mục tiêu và đối tượng.

Đầu tiên, bạn cần xác định mục tiêu rõ ràng cho Dashboard. Để biết được mục tiêu sử dụng dashboard, bạn hãy trả lời các câu hỏi: Bạn sử dụng dashboard để là gì? Bạn muốn đạt được điều gì khi tạo Dashboard? Ví dụ: Một doanh nghiệp tạo Dashboard với mục tiêu như theo dõi doanh thu hàng tháng, quản lý tiến độ dự án hay theo dõi chỉ số hiệu suất chung.

Sau đó, bạn tiếp tục xác định đối tượng sẽ sử dụng Dashboard. Hãy làm rõ hai vấn đề chính bao gồm: họ là ai và những quyền hạn của họ khi sử dụng Dashboard. Ví dụ,

giám đốc điều hành muốn theo dõi tổng quan doanh nghiệp, còn quản lý dự án cần theo dõi tiến trình của từng dự án cụ thể.

- Bước 2. Chọn dữ liệu và nguồn dữ liệu.

Dựa vào mục tiêu, bạn hãy xác định những dữ liệu quan trọng và cần thiết để phục vụ cho mục tiêu bạn mong muốn. Ví dụ các dữ liệu về doanh thu, số lượng đơn hàng, tỷ lệ chuyển đổi, tồn kho,...

Đồng thời, trong bước thứ 2 bạn cần phải xác định rõ nguồn dữ liệu cho Dashboard. Dữ liệu có thể đến từ hệ thống CRM, cơ sở dữ liệu SQL, tệp Excel, Google Analytics,...

- Bước 3. Chọn biểu đồ và bố trí.

Bước tiếp theo là lựa chọn loại biểu đồ phù hợp với dữ liệu. Dựa vào loại dữ liệu bạn có, lựa chọn biểu đồ thích hợp nhất. Các loại biểu đồ phổ biến như: biểu đồ cột, biểu đồ đường, biểu đồ hình tròn,...

Các thông tin cần được sắp xếp một cách khoa học và hợp lý. Đặt các thành phần trực quan sao cho người đọc dễ dàng theo dõi, từ thông tin quan trọng đến thông tin chi tiết.

- Bước 4. Xây dựng và tùy chỉnh.

Khi đã có hệ thống dữ liệu hoàn chỉnh, bạn có thể sử dụng các công cụ tạo Dashboard như Microsoft Power BI, Tableau, Google Data Studio,... để xây dựng Dashboard. Bạn có thể tùy chỉnh giao diện và cài đặt các yếu tố như màu sắc, font chữ, biểu tượng để phù hợp với thương hiệu của doanh nghiệp.

- Bước 5. Kiểm tra và cải thiện.

Cuối cùng, bạn hãy kiểm tra kỹ dữ liệu sử dụng để tạo Dashboard để đảm bảo tính chính xác. Đồng thời, kiểm tra cách hiển thị trên các thiết bị khác nhau như máy tính, điện thoại di động, máy tính bảng.

Để các Dashboard mang lại hiệu quả cao nhất, bạn không thể bỏ qua hoạt động thu thập phản hồi từ người sử dụng mẫu Dashboard. Thông qua đó, bạn có thể chỉnh sửa và tối ưu các Dashboard.

4. Random Forest:

- Random Forest (Rừng Ngẫu Nhiên) là một thuật toán học máy phổ biến và mạnh mẽ được sử dụng trong phân loại và dự đoán. Nó là một biến thể của cây quyết định (decision tree) và sử dụng kỹ thuật "ensemble learning" (học tổ hợp) để cải thiện hiệu suất của mô hình dự đoán.

- Ưu điểm: Random Forest có khả năng giảm quá khớp, xử lý dữ liệu nhiễu và xác định biến quan trọng. Nó cũng không yêu cầu nhiều tiền xử lý dữ liệu và thường cho ra kết quả tốt mà không cần tinh chỉnh tham số phức tạp.

PHẦN 3: DỮ LIỆU.

1. Bộ dữ liệu AdventureWorks:

1.1. Giới thiệu bộ dữ liệu:

Cơ sở dữ liệu AdventureWorks do Microsoft SQL Server cung cấp. Đây là cơ sở dữ liệu tiêu chuẩn được sử dụng để hiển thị quá trình xử lý giao dịch hàng ngày cho một doanh nghiệp. Các kịch bản bao gồm bán hàng, quản lý khách hàng, quản lý sản phẩm và nguồn nhân lực.

1.1.1. Mô tả ngữ cảnh doanh nghiệp:

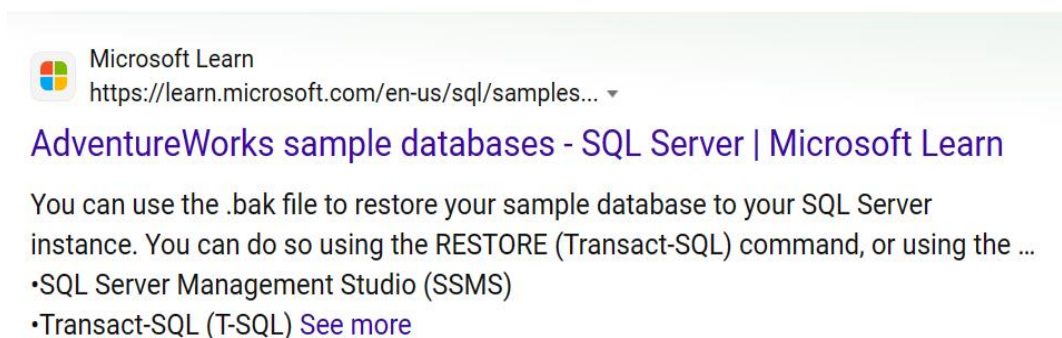
- Cơ sở dữ liệu minh họa AdventureWorks lấy bối cảnh trên dữ liệu của 1 công ty hư cấu có tên Adventure Works Cycles. Đây được cho là 1 công ty sản xuất đa quốc gia có qui mô lớn. Công ty sản xuất và kinh doanh xe đạp làm từ kim loại và các chất chiệu tổng hợp. Thị trường của công ty này bao gồm cả khu vực Bắc Mỹ, Châu Âu và Châu Á. Trong khi trụ sở chính của công ty lại được đặt ở Bothell, Washington gồm có 290 nhân viên, công ty cũng có 1 vài nhóm nhân viên kinh doanh khu vực ở các thị trường hoạt động của mình
- Vào năm 2000, công ty Adventure Works Cycles có mua lại 1 nhà máy sản xuất nhỏ là Importadores Neptuno ở Mexico. Nhà máy này tham gia sản xuất 1 vai thành phần thiết yếu trong toàn bộ dây chuyền sản xuất ra thành phẩm của công ty. Các thành phần được nhà máy này sản xuất sẽ được chuyển tới trụ sở chính ở

Bothell để lắp ráp thành thành phố. Năm 2001, Importtadores Neptuneo trở thành nhà máy duy nhất sản xuất phà phân phối dòng sản phẩm xe đạp du lịch.

- Kết thúc 1 năm tài chính thành công, công ty Adventure Works Cycles đang tìm cách mở rộng thị phần bằng cách tập trung vào các hoạt động bán hàng cho khách hàng quan trọng nhất của họ, mở rộng thông tin sản phẩm thông qua hệ thống Website đồng thời giảm chi phí bán hàng bằng cách giảm chi phí sản xuất

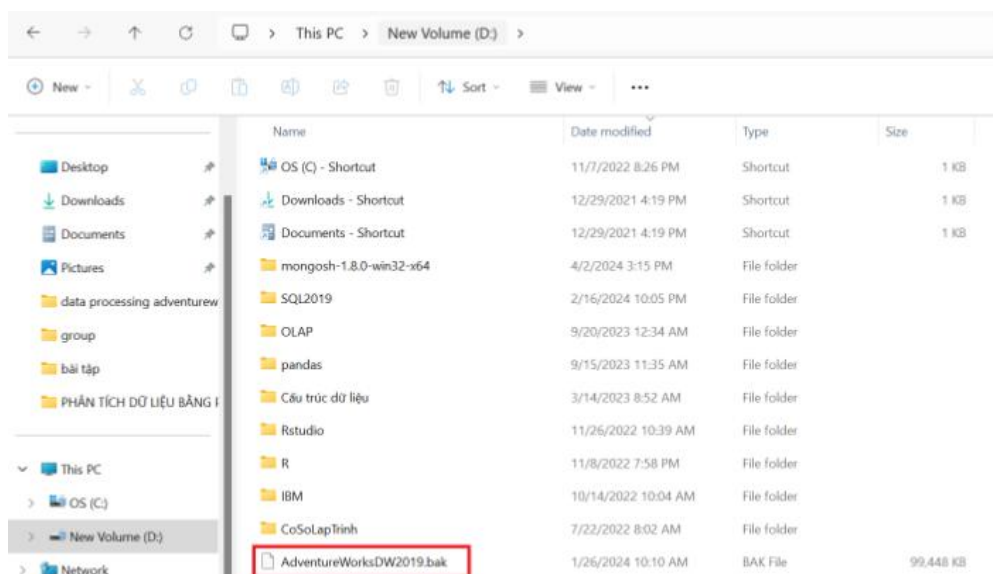
1.2. Chuẩn bị dữ liệu:

- Bước 1: Truy cập vào trang website: [AdventureWorks sample databases - SQL Server | Microsoft Learn](https://learn.microsoft.com/en-us/sql/samples/adventureworksdb/)

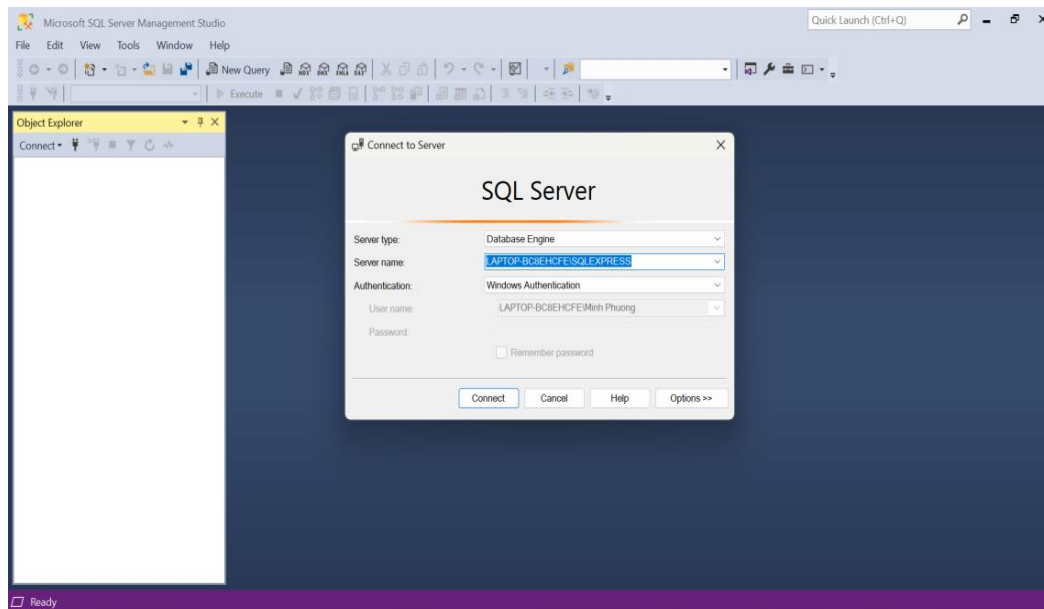


- Bước 2: Tải AdventureWorksDW2019.bak

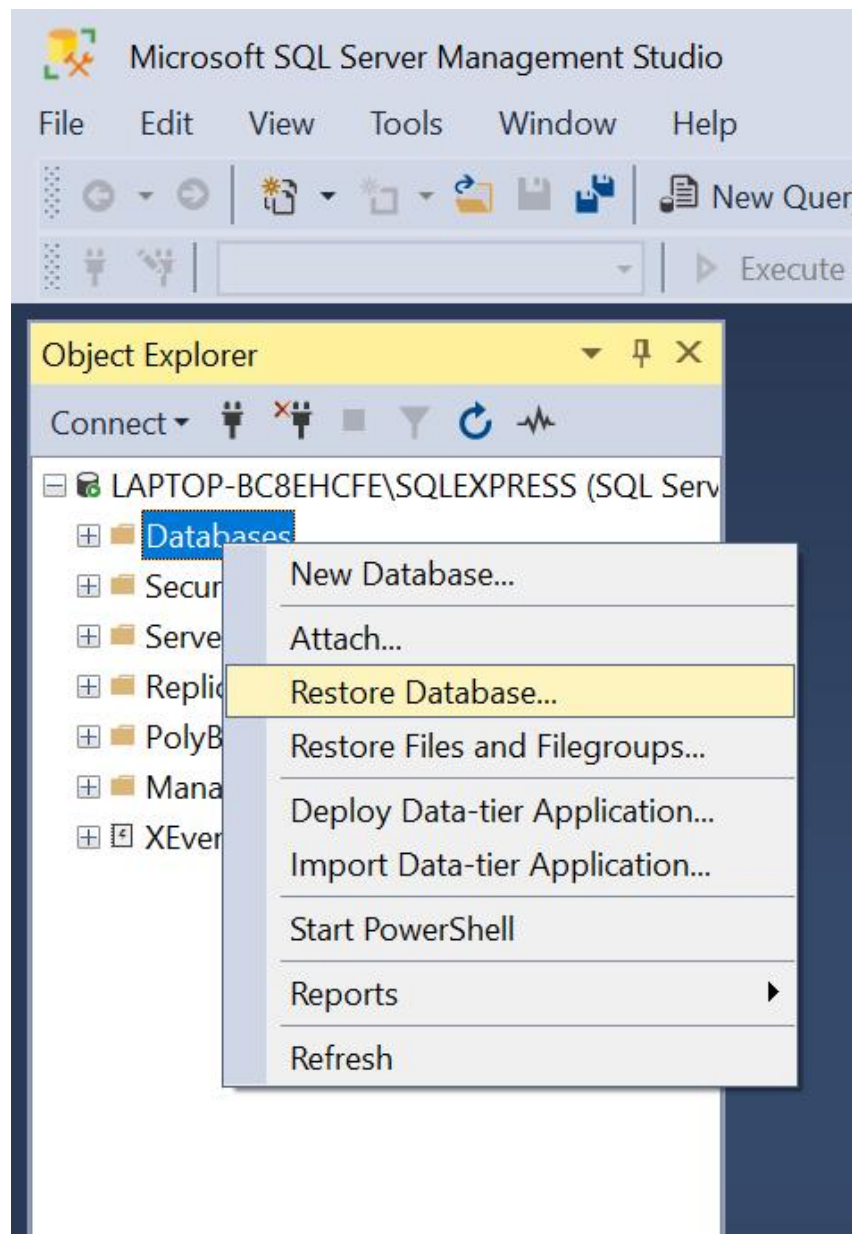
- Bước 3: Di chuyển tệp đến vị trí sao lưu SQL Server của bạn.



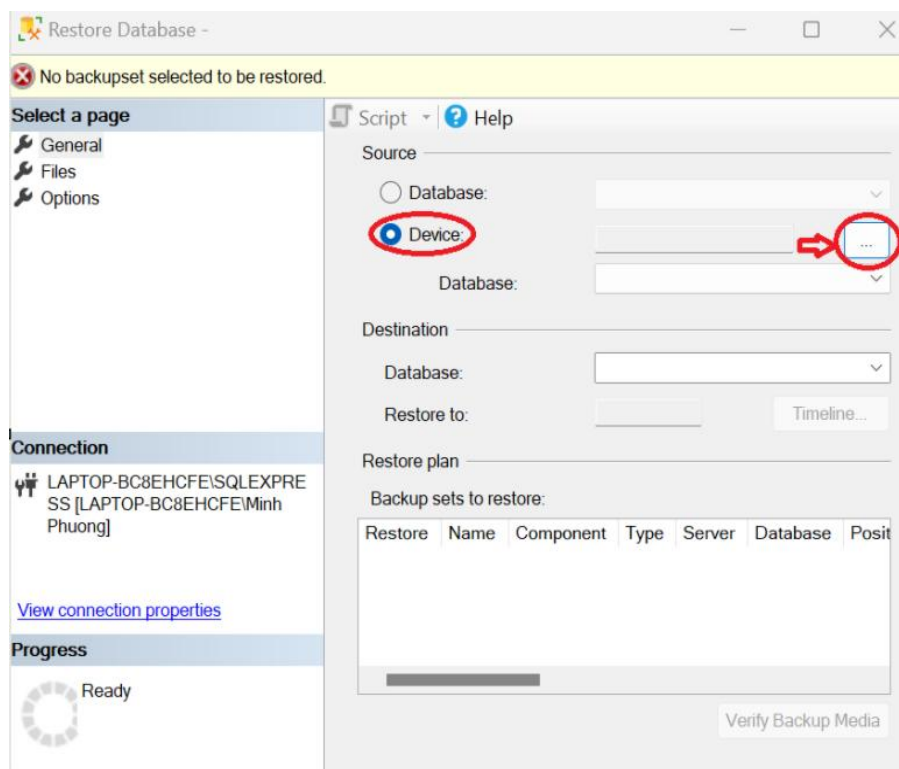
- Bước 4: Mở SSMS và kết nối đến phiên bản SQL Server của bạn.



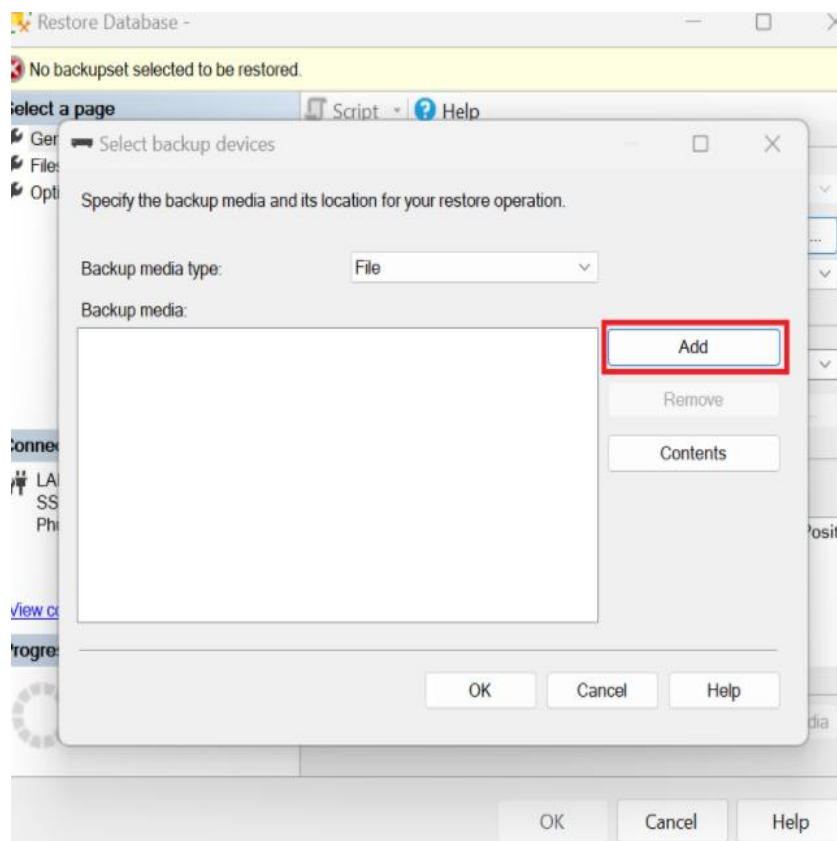
- Bước 5: Nhấp chuột phải vào Databases trong Object Explorer > Restore Database... để khởi chạy trình hướng dẫn Restore Database.

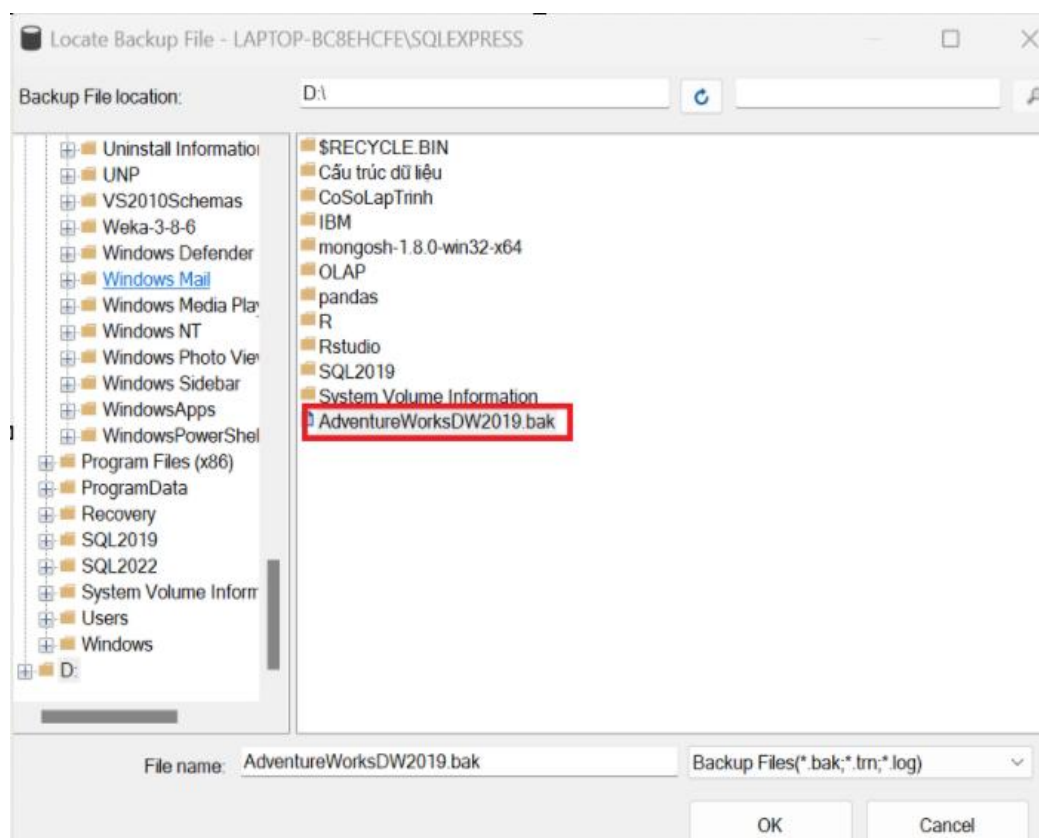


- Bước 6: Chọn Device rồi chọn dấu chấm lửng (...) để chọn thiết bị.

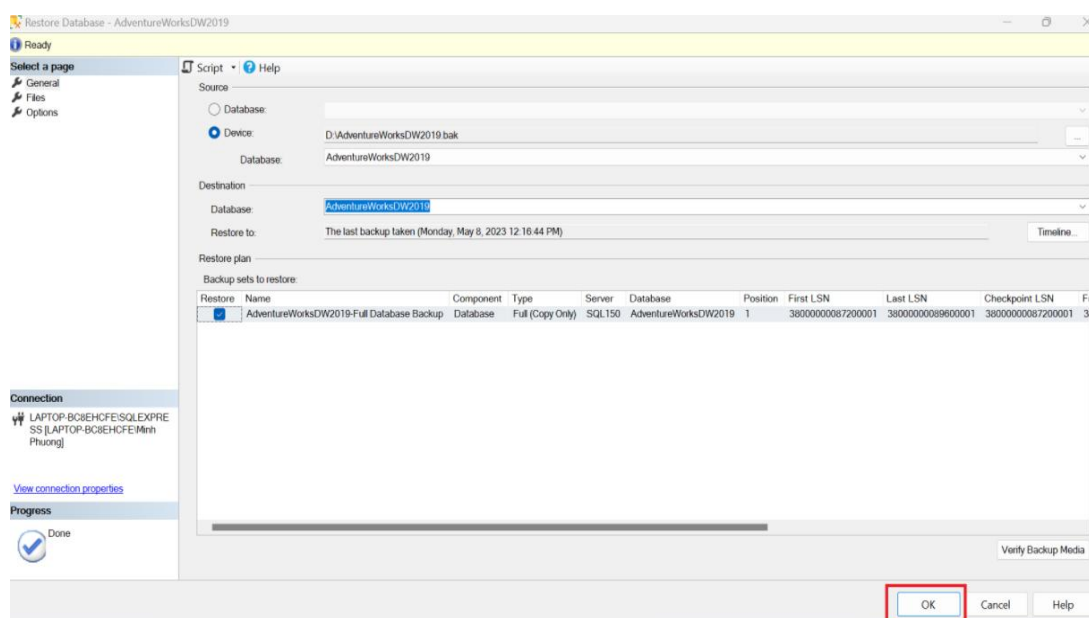


- Bước 7: Chọn Add rồi chọn tệp bạn đã di chuyển đến vị trí sao lưu

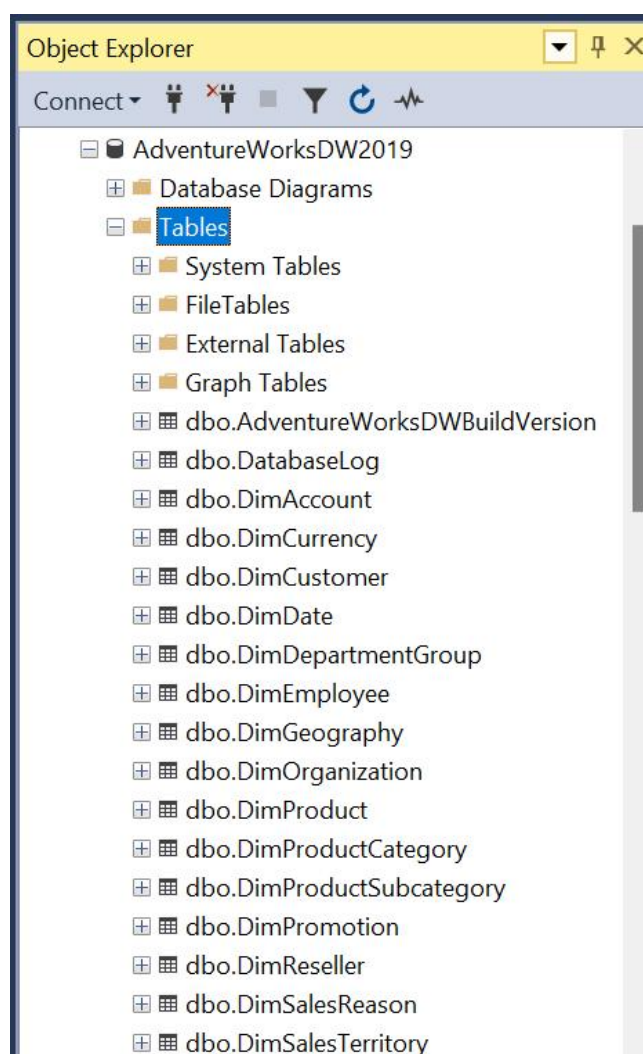




- Bước 8: Chọn OK để xác nhận lựa chọn sao lưu cơ sở dữ liệu của bạn và đóng cửa sổ Select Backup Devices
- Bước 9: Kiểm tra tab Files để xác nhận Khôi phục dưới dạng vị trí và tên tệp khớp với vị trí dự định và tên tệp của bạn trong trình hướng dẫn Khôi phục Cơ sở dữ liệu. Chọn OK để khôi phục cơ sở dữ liệu của bạn.



+ Kết quả sẽ như sau:



2. Mô tả dữ liệu:

- Đối với chủ đề báo cáo hàng tồn kho, ở đây chúng em sẽ chọn ra các bảng phù hợp với chủ đề trong bộ dữ liệu AdventureWorks2019 để thực hiện gồm các bảng sau:

- + DimProduct, gồm 606 dòng, 36 cột
- + FactProductInventory, gồm 776,286 dòng, 7 cột
- + DimDate gồm 3652 dòng, 19 cột
- + DimProductCategory gồm 4 hàng, 5 cột
- + DimProductSubcategory gồm 37 dòng, 6 cột

- Sau khi chọn ra các bảng dữ liệu, chúng em sẽ Export Data từ SQL Server sang file Excel để tiện xử lý nếu cần.

- Các bảng mô tả dữ liệu phía dưới là mô tả đối với dữ liệu đã xóa các cột không cần thiết.

2.1. Bảng DimProduct:

	Column	Description
1	ProductKey	Mã sản phẩm.
2	ProductSubcategoryKey	Mã danh mục sản phẩm con. Khóa ngoại tới DimProductSubcategory.ProductSubcategoryKey.
3	EnglishProductName	Tên sản phẩm.
4	StandardCost	Chi phí tiêu chuẩn của sản phẩm.
5	FinishedGoodsFlag	0 = Sản phẩm không phải là mặt hàng bán được. 1 = Sản phẩm là mặt hàng bán được.
6	Color	Màu sắc của sản phẩm.
7	SafetyStockLevel	Số lượng tồn kho tối thiểu.
8	ReorderPoint	Mức tồn kho gửi đơn hàng hoặc lệnh sản xuất.(Mức tồn kho mà khi đạt đến, yêu cầu phải đặt hàng mới để đảm bảo không gặp phải tình trạng thiếu hàng)
9	ListPrice	Giá bán.
10	Size	Kích thước của sản phẩm.
11	DaysToManufacture	Số ngày cần thiết để sản xuất sản phẩm.
12	DealerPrice	Giá bán cho các nhà phân phối hoặc đại lý.
13	Class	Lớp sản phẩm. (H = High, M = Medium, L = Low).
14	Style	Phong cách sản phẩm. (W = Women's, M = Men's, U = Universal).

15	ModelName	Tên mô hình của sản phẩm.
16	Status	Trạng thái hiện tại của sản phẩm

2.2. Bảng FactProductInventory:

	Column	Description
1	ProductKey	Mã sản phẩm. Khóa ngoại tới DimProduct.ProductKey.
2	DataKey	Mã thời gian. Khóa ngoại tới DimDate.DataKey
3	MovementDate	Ngày mà sản phẩm được nhập hoặc xuất
4	UnitCost	Chi phí của mỗi đơn vị sản phẩm.
5	UnitsIn	Số lượng sản phẩm được nhập vào kho.
6	UnitsOut	Số lượng sản phẩm được xuất khỏi kho
7	UnitsBalance	Số lượng sản phẩm tồn kho.

2.3. Bảng DimProductCategory:

	Column	Description
1	ProductCategoryKey	Mã danh mục sản phẩm.
2	ProductCategoryAlternateKey	Mã thay thế cho mã danh mục sản phẩm.
3	EnglishProductCategoryName	Tên của danh mục sản phẩm.

2.4. Bảng DimProductSubcategory:

	Column	Description
--	--------	-------------

1	ProductSubcategoryKey	Mã danh mục con sản phẩm.
2	ProductSubcategoryAlternateKey	Mã thay thế cho mã danh mục sản phẩm con.
3	EnglishProductSubcategoryName	Tên danh mục sản phẩm con.
4	ProductCategoryKey	Mã của danh mục sản phẩm. Khóa ngoại đến DimProductCategory.ProductCategoryKey

2.5. Bảng DimDate:

	Column	Description
1	DateKey	Mã thời gian.
2	FullDateAlternateKey	Mã thay thế cho mã thời gian.
3	DayNumberOfWeek	Số thứ tự của ngày trong tuần, 1 cho Chủ Nhật, 2 cho Thứ Hai,.....
4	EnglishDayNameOfWeek	Tên của ngày trong tuần.
5	DayNumberOfMonth	Số thứ tự của ngày trong tháng
6	DayNumberOfYear	Số thứ tự của ngày trong năm
7	WeekNumberOfYear	Số thứ tự của tuần trong năm
8	EnglishMonthName	Tên của tháng.
9	MonthNumberOfYear	Số thứ tự của tháng trong năm
10	CalendarQuarter	Số thứ tự của quý trong năm theo lịch
11	CalendarYear	Năm theo lịch
12	CalendarSemester	Học kỳ trong năm theo lịch, 1 hoặc 2

13	FiscalQuarter	Số thứ tự của quý trong năm tài chính
----	---------------	---------------------------------------

3. Tiền xử lý dữ liệu:

- Sau khi chọn ra các bảng dữ liệu phù hợp, chúng em sẽ xuất dữ liệu từ SQL Server sang file Excel.

3.1. Bảng DimProduct:

- Đọc dữ liệu từ file excel:

```

1 url='https://docs.google.com/spreadsheets/d/1IZ1N2a-bGqU69zZNYc_KazdG_wqkTYXN/edit?usp=drive_link&ouid=111630126831229986868&rtfpof=true&sd=true'
2 DimProduct = pd.read_excel('https://drive.google.com/uc?id=' + url.split('/')[-2])
3 display(DimProduct.head(5))

```

	ProductKey	ProductAlternateKey	ProductSubcategoryKey	WeightUnitMeasureCode	SizeUnitMeasureCode	EnglishProductName	SpanishProductName	French
0	1	AR-5381	NaN	NaN	NaN	Adjustable Race	NaN	
1	2	BA-8327	NaN	NaN	NaN	Bearing Ball	NaN	
2	3	BE-2349	NaN	NaN	NaN	BB Ball Bearing	NaN	
3	4	BE-2908	NaN	NaN	NaN	Headset Ball Bearings	NaN	
4	5	BL-2036	NaN	NaN	NaN	Blade	NaN	

5 rows x 36 columns

- Xóa một số cột không cần thiết:

```

1 Drop_Columns_DimProduct = ['ProductAlternateKey', 'WeightUnitMeasureCode', 'SizeUnitMeasureCode', 'SpanishProductName', 'FrenchProductName',
2                             'SizeRange', 'Weight', 'ProductLine', 'LargePhoto', 'EnglishDescription', 'FrenchDescription', 'ChineseDescription',
3                             'ArabicDescription', 'HebrewDescription', 'ThaiDescription', 'GermanDescription', 'JapaneseDescription',
4                             'TurkishDescription', 'StartDate', 'EndDate']
5 DimProduct.drop(Drop_Columns_DimProduct, axis = 1, inplace = True)

```

- Kiểm tra giá trị NULL:


```
1 DimProduct.isnull().sum()

ProductKey          0
ProductSubcategoryKey 209
EnglishProductName   0
StandardCost         211
FinishedGoodsFlag    0
Color                254
SafetyStockLevel     0
ReorderPoint         0
ListPrice            211
Size                 307
DaysToManufacture    0
DealerPrice          211
Class                276
Style                305
ModelName            209
Status               200
dtype: int64
```

- Sau khi kiểm tra, ta thấy có khá nhiều giá trị NULL trong bảng, vì thế ta sẽ thay thế những giá trị này bằng giá trị phù hợp.

- Thay thế giá trị NULL theo kiểu dữ liệu:

```
# Thay thế giá trị NULL bằng 0 đối với kiểu dữ liệu float64 trong bảng DimProduct
Replace_NULL_Values_0 = ['ProductSubcategoryKey', 'StandardCost', 'ListPrice', 'DealerPrice']
DimProduct[Replace_NULL_Values_0] = DimProduct[Replace_NULL_Values_0].fillna(0)

# Thay thế giá trị NULL bằng 'No' đối với kiểu dữ liệu object trong bảng DimProduct
Replace_NULL_Values_No = ['Color', 'Size', 'Class', 'Style', 'ModelName', 'Status']
DimProduct[Replace_NULL_Values_No] = DimProduct[Replace_NULL_Values_No].fillna('No')
```

- Sau khi tiền xử lý, bảng DimProduct còn 606 dòng, 16 cột.

3.2. Bảng DimProductInventory:

- Đọc dữ liệu từ file excel:

```

1 url='https://docs.google.com/spreadsheets/d/1NSJcuYZ-xhxXy5Mhw1MVqE6Wpy5r4sKR/edit?usp=sharing&ouid=111630126831229986868&rtpof=true&sd=true'
2 DimProductInventory = pd.read_excel('https://drive.google.com/uc?id=' + url.split('/')[2])
3 display(DimProductInventory.head(5))

```

	ProductKey	DateKey	MovementDate	UnitCost	UnitsIn	UnitsOut	UnitsBalance
0	1	20101228	40540	0.32	0	0	875
1	1	20101229	40541	0.32	0	0	875
2	1	20101230	40542	0.32	0	0	875
3	1	20101231	40543	0.32	0	0	875
4	1	20110101	40544	0.32	0	0	875

- Kiểm tra giá trị NULL:

```
1 DimProductInventory.isnull().sum()
```

```

ProductKey      0
DateKey         0
MovementDate    0
UnitCost        0
UnitsIn         0
UnitsOut        0
UnitsBalance    0
dtype: int64

```

- Không có giá trị NULL.

3.3. Bảng DimProductCategory:

- Đọc dữ liệu từ file excel:

```

1 url='https://docs.google.com/spreadsheets/d/1qH-Elc8yaCF9f_70imH2lDpIu4mAYs0h/edit?usp=drive_link&ouid=111630126831229986868&rtpof=true&sd=true'
2 DimProductCategory = pd.read_excel('https://drive.google.com/uc?id=' + url.split('/')[2])
3 display(DimProductCategory.head(5))

```

	ProductCategoryKey	ProductCategoryAlternateKey	EnglishProductCategoryName	SpanishProductCategoryName	FrenchProductCategoryName
0	1	1	Bikes	Bicicleta	Vélo
1	2	2	Components	Componente	Composant
2	3	3	Clothing	Prenda	Vêtements
3	4	4	Accessories	Accesorio	Accessoire

- Xóa một số cột dữ liệu không cần thiết:

```

1 Drop_Columns_DimProductCategory = ['SpanishProductCategoryName', 'FrenchProductCategoryName']
2 DimProductCategory.drop(Drop_Columns_DimProductCategory, axis = 1, inplace = True)

```

- Kiểm tra giá trị NULL:

```
1 DimProductCategory.isnull().sum()
```

```
ProductCategoryKey          0
ProductCategoryAlternateKey  0
EnglishProductCategoryName   0
dtype: int64
```

- Không có giá trị NULL.

- Sau khi tiền xử lý, bộ dữ liệu DimProductCategory còn 4 dòng, 3 cột

3.4. Bảng DimProductSubcategory:

- Đọc dữ liệu từ file excel:

```
1 url='https://docs.google.com/spreadsheets/d/1kDGgTtXKP-ZtXv4b-HCI1Bk_3ZUjr1QA/edit?usp=drive_link&ouid=111630126831229986868&rtf=true&sd=true'
2 DimProductSubcategory = pd.read_excel('https://drive.google.com/uc?id=' + url.split('/')[-2])
3 display(DimProductSubcategory.head(5))
```

	ProductSubcategoryKey	ProductSubcategoryAlternateKey	EnglishProductSubcategoryName	SpanishProductSubcategoryName	FrenchProductSubcategoryName
0	1	1	Mountain Bikes	Bicicleta de montaña	VTT
1	2	2	Road Bikes	Bicicleta de carretera	Vélo de route
2	3	3	Touring Bikes	Bicicleta de paseo	Vélo de randonnée
3	4	4	Handlebars	Barra	Barre d'appui
4	5	5	Bottom Brackets	Eje de pedalier	Axe de pédalier

- Xóa một số cột không cần thiết:

```
1 Drop_Columns_DimProductSubcategory = ['SpanishProductSubcategoryName', 'FrenchProductSubcategoryName']
2 DimProductSubcategory.drop(Drop_Columns_DimProductSubcategory, axis = 1, inplace = True)
```

- Kiểm tra giá trị NULL:

```
1 DimProductSubcategory.isnull().sum()
```

```
ProductSubcategoryKey          0
ProductSubcategoryAlternateKey  0
EnglishProductSubcategoryName  0
ProductCategoryKey             0
dtype: int64
```

- Không có giá trị NULL:
- Sau khi tiền xử lý bảng DimProductSubcategory còn 37 dòng, 4 cột.

3.5. Bảng DimDate:

- Đọc dữ liệu từ file excel:

```
1 url='https://docs.google.com/spreadsheets/d/1Zr4hGtp76PbliuhXottRHem1o1K_hQt7/edit?usp=drive_link&ouid=111630126831229986868&rtpof=true&sd=true'
2 DimDate = pd.read_excel('https://drive.google.com/uc?id=' + url.split('/')[2])
3 display(DimDate.head(5))
```

	DateKey	FullDateAlternateKey	DayNumberOfWeek	EnglishDayNameOfWeek	SpanishDayNameOfWeek	FrenchDayNameOfWeek	DayNumberOfMonth	DayNumberOfYear
0	20050101	2005-01-01	7	Saturday	Sábado	Samedi	1	1
1	20050102	2005-01-02	1	Sunday	Domingo	Dimanche	2	2
2	20050103	2005-01-03	2	Monday	Lunes	Lundi	3	3
3	20050104	2005-01-04	3	Tuesday	Martes	Mardi	4	4
4	20050105	2005-01-05	4	Wednesday	Miércoles	Mercredi	5	5

- Xóa một số cột không cần thiết:

```
1 Drop_Columns_DimDate = ['SpanishDayNameOfWeek', 'FrenchDayNameOfWeek', 'SpanishMonthName', 'FrenchMonthName']
2 DimDate.drop(Drop_Columns_DimDate, axis = 1, inplace = True)
```

- Kiểm tra giá trị NULL:

1	DimDate.isnull().sum()
DateKey	0
FullDateAlternateKey	0
DayNumberOfWeek	0
EnglishDayNameOfWeek	0
DayNumberOfMonth	0
DayNumberOfYear	0
WeekNumberOfYear	0
EnglishMonthName	0
MonthNumberOfYear	0
CalendarQuarter	0
CalendarYear	0
CalendarSemester	0
FiscalQuarter	0
FiscalYear	0
FiscalSemester	0
dtype:	int64

- Không có giá trị NULL
- Sau khi tiền xử lý, bảng DimDate còn 3652 dòng, 15 cột

PHẦN 4: TRỰC QUAN HÓA:

1. Tổng quan hàng tồn kho:

Số lượng hàng tồn kho 1,300,527	Chi phí hàng tồn kho 128,657,992
Số lượng hàng nhập 270,969	Số lượng hàng xuất 274,776

Chart 1. Tổng quan hàng tồn kho.

Dữ liệu hàng tồn kho được cập nhật từ ngày 28/12/2010 đến ngày 30/6/2014, trong khoảng thời gian này tổng số lượng hàng tồn kho là 1,300,527 đơn vị sản phẩm và chi phí để lưu trữ số lượng hàng tồn kho này là 128,657,992 đơn vị tiền tệ, chênh lệch

giữa số lượng hàng xuất (274,776 đơn vị sản phẩm) và số lượng hàng nhập (270,969 đơn vị sản phẩm) không lớn, cho thấy việc quản lý hàng tồn kho của doanh nghiệp tương đối ổn định và không gặp phải lỗ hổng nghiêm trọng trong quá trình quản lý kho.

Trong dashboard có thể lọc các giá trị ngày theo thời gian (tính bằng năm) và theo danh mục sản phẩm (bốn danh mục sản phẩm) để có cái nhìn chi tiết hơn về tình hình tồn kho của doanh nghiệp

2. Biểu đồ xu hướng hàng tồn kho qua các năm:



Chart 2. Xu hướng hàng tồn kho qua các năm.

Bộ dữ liệu adventureworksDW2019 cập nhật dữ liệu kho từ 1/12/2010 đến tháng 30/6/2014. Với mỗi mã sản phẩm ở bảng FactProductInventory sẽ được cập nhật chỉ số “UnitsBalance” theo từng ngày từ 1/12/2010-30/6/2014, tuy nhiên có những ngày chỉ số hàng tồn kho “UnitsBalance” không đổi vì không có sự thay đổi giữa số lượng hàng nhập “UnitsIn” và số lượng hàng xuất “UnitsOut”. Nên việc theo dõi xu hướng hàng tồn kho dựa trên bảng FactProductInventory của dữ liệu gốc là không phù hợp.

Để phù hợp cho việc trực quan hoá để theo dõi xu hướng hàng tồn kho qua các năm xử lý bảng FactProductInventory được thực hiện như sau:

- Chỉ lấy những chỉ số “UnitsBalance” ngày cuối cùng của năm là 31/12/2010, 31/12/2011, 31/12/2012, 31/12/2013, 30/6/2014.
- Sau khi lọc xong những ngày cuối cùng của năm, không hiển thị những cột không có ý nghĩa sử dụng trong chart (UnitsCost) hoặc không đúng, cột “UnitsIn” và “UnitsOut” không còn phù hợp để sử dụng cho những trường hợp khác (giải thích: không thể tổng được số lượng hàng xuất và nhập dựa trên những ngày cuối cùng của năm mà phải tổng số lượng ở tất cả khoảng thời gian trong năm) nên sẽ được loại bỏ.
- Trích xuất bảng dữ liệu bảng dữ liệu. (Trường hợp này nhóm sử dụng SQL để truy vấn dữ liệu)

```
SELECT Productkey, MovementDate, UnitsBalance, DateKey
FROM FactProductInventory
WHERE
    (DAY(MovementDate) = 31 AND MONTH(MovementDate) = 12 AND YEAR(MovementDate) IN (2010, 2011, 2012, 2013))
    or
    (DAY(MovementDate) = 30 AND MONTH(MovementDate) = 6 AND YEAR(MovementDate) = 2014)
ORDER BY MovementDate
```

- Quan sát Chart 1, ta nhận thấy qua 4 năm thì số lượng hàng tồn kho không có nhiều thay đổi, số lượng chỉ giao động trong một phạm vi nhỏ từ 258.794 - 262.601 sản phẩm. Điều này có thể phản ánh một số điểm về doanh nghiệp:
 - + Doanh nghiệp có khả năng dự đoán và quản lý hàng tồn kho một cách hiệu quả. Điều này giúp tránh tình trạng hàng tồn kho dư thừa hoặc thiếu hụt, giữ cho hoạt động kinh doanh diễn ra một cách suôn sẻ hơn.
 - + Phản ánh sự ổn định trong quá trình sản xuất và cung ứng của doanh nghiệp. Nếu doanh nghiệp có thể duy trì mức độ ổn định này qua thời gian, điều này có thể góp phần vào việc duy trì chất lượng sản phẩm, giảm thiểu rủi ro trong chuỗi cung ứng và tăng cường lòng tin của khách hàng.
 - + Phản ánh sự ổn định trong chính sách và chiến lược kinh doanh của doanh nghiệp. Nó có thể chỉ ra rằng doanh nghiệp có các chiến lược kinh doanh dài hạn và quy trình hoạt động ổn định, giúp duy trì sự ổn định trong hàng tồn kho.

- Đây là một dấu hiệu tích cực về hiệu suất và ổn định của doanh nghiệp trong quản lý và vận hành hàng tồn kho.

3. Biểu đồ tỷ lệ hàng tồn kho theo danh mục sản phẩm:

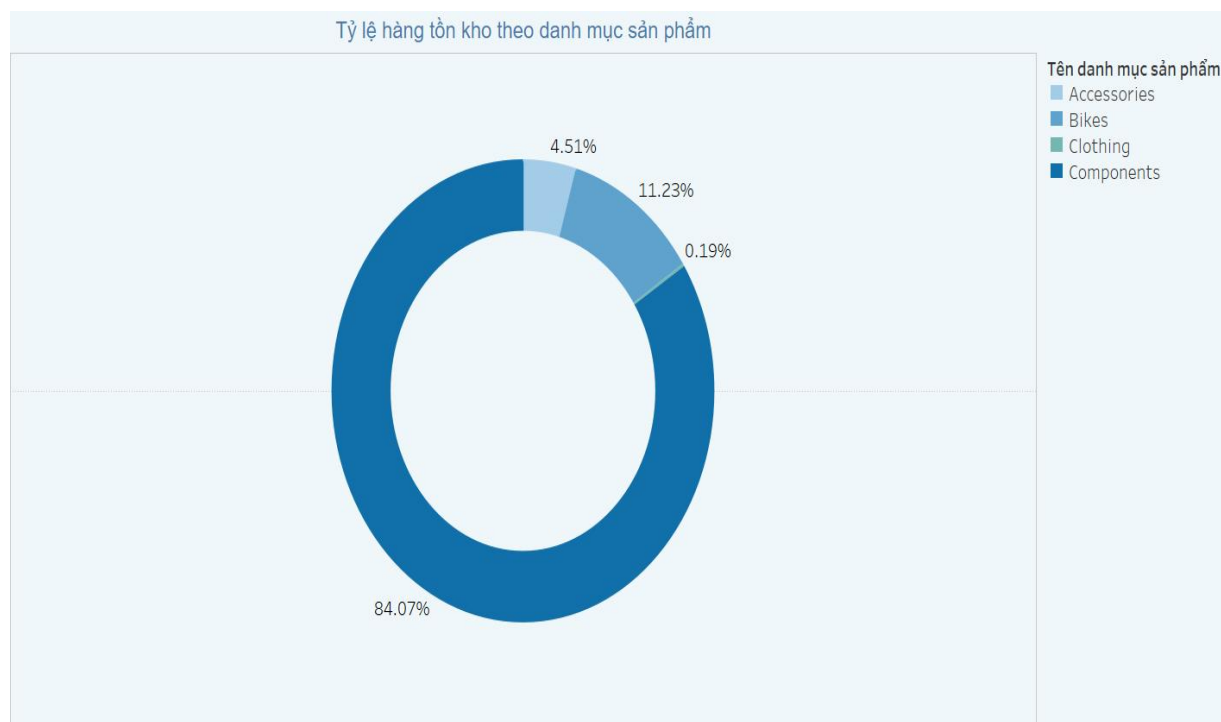


Chart 3. Tỷ lệ hàng tồn kho theo danh mục sản phẩm.

Biểu đồ tròn ở chart 3 thể hiện tỷ lệ hàng tồn kho theo 4 danh mục sản phẩm gồm Accessories (Phụ kiện), Bikes (Xe đạp), Clothing (Quần áo thể thao), Components (Phụ tùng). Trong đó 84,07% sản phẩm trong kho là sản phẩm thuộc danh mục phụ tùng và 11,23% là sản phẩm thuộc danh mục xe đạp. Để lý giải cho điều này có 1 số lý do sau mà nhóm phân tích được sau khi quan sát biểu đồ kết hợp với thực tế:

- Tỷ lệ lớn hơn của danh mục phụ tùng có thể chỉ ra rằng có một nhu cầu lớn hoặc một chiến lược kinh doanh cụ thể đằng sau việc tập trung vào loại sản phẩm này. Có thể do phụ tùng có nhu cầu sử dụng cao hơn so với các loại sản phẩm khác hoặc do lợi nhuận từ bán hàng phụ tùng cao hơn.
- Phân phối hàng tồn kho theo danh mục sản phẩm có thể phản ánh chiến lược kinh doanh và tiếp thị của doanh nghiệp. Có thể doanh nghiệp đang tập trung vào việc cung cấp và tiếp thị các sản phẩm phụ tùng và xe đạp hơn là các loại sản phẩm khác.

- Xe đạp và phụ tùng là 2 danh mục sản phẩm mà trong đó sản phẩm ít khi có sự thay đổi hoặc update hoặc thời gian update dài. Việc lưu trữ các sản phẩm trong 2 danh mục này đảm bảo rằng sản phẩm không bị lỗi thời. Ngược lại, danh mục sản phẩm quần áo và phụ kiện dễ thay đổi mẫu mã theo xu hướng thời trang nên doanh nghiệp cân nhắc việc tồn kho số lượng lớn mặt hàng thuộc danh mục này.
- Có thể doanh nghiệp đang tập trung vào việc cung cấp và tiếp thị các sản phẩm phụ tùng hơn là các loại sản phẩm khác hoặc đây là sản phẩm bán chính của doanh nghiệp.

Tuy nhiên, khi nhìn vào biểu đồ ta thấy phân phối không cân đối. Lời khuyên cho doanh nghiệp là nên đa dạng hoá nguồn lợi nhuận bằng cách nhập thêm sản phẩm của các danh mục khác vào kho và áp dụng các chính sách thúc đẩy bán hàng. Điều này đảm bảo sự an toàn cho doanh nghiệp bởi giảm thiểu rủi ro do phụ thuộc quá nhiều vào một nguồn lợi nhuận cụ thể và nếu một lĩnh vực/danh mục sản phẩm không phát triển như mong đợi, sự đa dạng hóa sản phẩm có thể giúp bù đắp mất mát từ các lĩnh vực khác.

4. Biểu đồ Top 10 danh mục sản phẩm con tồn kho nhiều nhất:

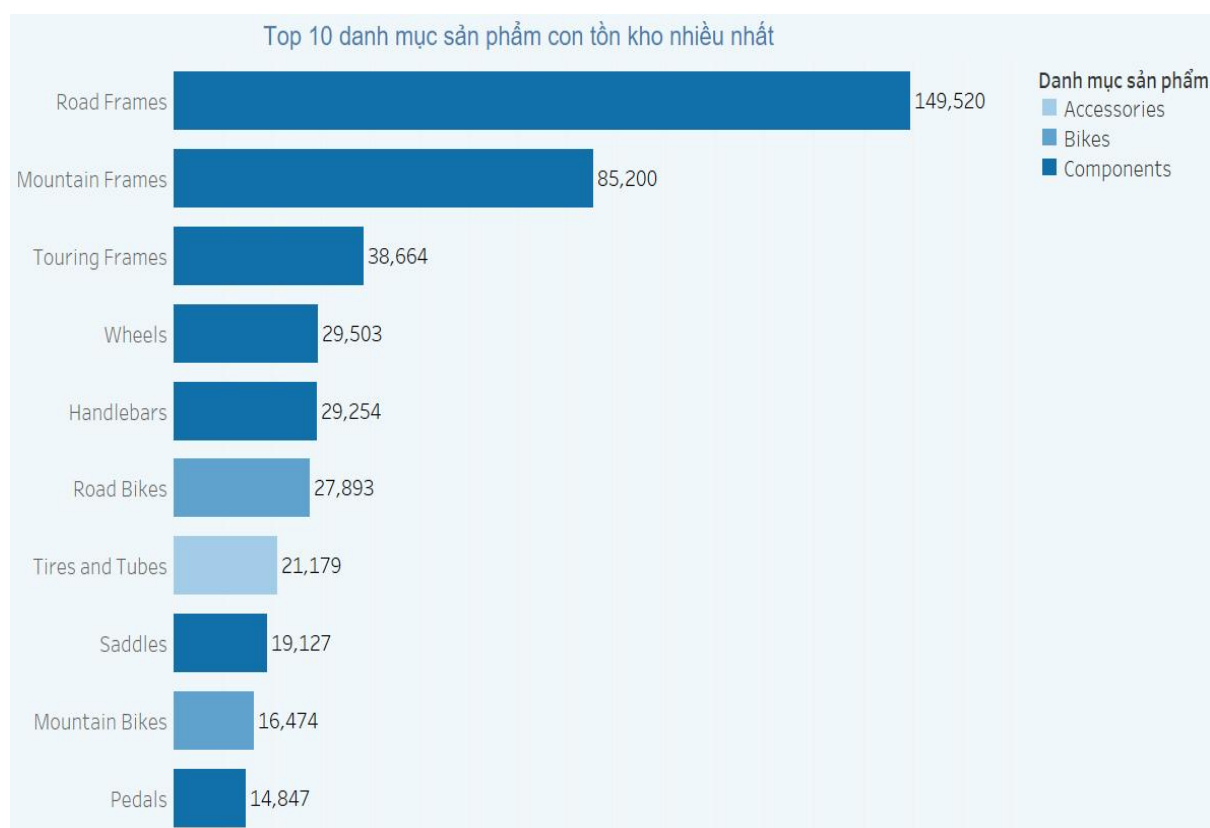


Chart 4. Top 10 danh mục sản phẩm con tồn kho nhiều nhất.

Ở chart 3, ta quan sát được sự đóng góp lớn của danh mục sản phẩm phụ tùng trong tổng số lượng sản phẩm tồn kho của doanh nghiệp đa quốc gia mà nhóm đang nghiên cứu, để biết được sản phẩm nào chiếm phần lớn sự đóng góp đó thì nhóm trực quan được trên chart 4. Danh mục sản phẩm con khung xe đạp đường phố (Road Frames) thuộc danh mục phụ tùng có số lượng tồn kho vượt trội so với những danh mục con khác, có 2 danh mục con sản phẩm của danh mục xe đạp là Road Bikes (xe đạp đường phố) và Mountain Bikes (xe đạp địa hình). Điều dễ hiểu ở đây là khách hàng ưa chuộng và có nhu cầu cao đối với sản phẩm xe đạp đường phố nên khung xe đạp đường phố bán chạy hơn vì mỗi xe đạp có thể thay thế được nhiều bộ khung phù hợp. Trong 4 danh mục sản phẩm chỉ có 3 danh mục xuất hiện trong top 10 danh mục con có số lượng tồn kho lớn nhất, ngoại trừ danh mục quần áo thể thao (Clothing).

Sự tập trung của số lượng tồn kho vào danh mục sản phẩm phụ tùng, đặc biệt là danh mục con khung xe đạp đường phố (Road Frames), có thể phản ánh một số điều sau:

- Xe đạp đường phố thường là một lựa chọn phổ biến đối với người tiêu dùng trong việc sử dụng hàng ngày hoặc trong các hoạt động thể thao. Do đó, khung xe đạp đường phố có thể là một phần quan trọng trong việc sản xuất và bán hàng cho các doanh nghiệp liên quan đến xe đạp.
- Khung xe đạp đường phố thường được thiết kế để có khả năng tùy chỉnh cao, giúp người dùng có thể lựa chọn và thay đổi các linh kiện khác nhau để phù hợp với nhu cầu cụ thể của họ. Điều này có thể tạo ra nhu cầu lớn cho khung xe đạp đường phố trong việc thay thế hoặc nâng cấp các bộ phận của xe đạp.
- Sự tập trung vào các sản phẩm có nhu cầu cao như khung xe đạp đường phố có thể là một chiến lược phân phối tồn kho hiệu quả. Bằng cách tập trung sản xuất và tồn kho các sản phẩm mà thị trường đang yêu cầu nhiều nhất, doanh nghiệp có thể tối ưu hóa việc sử dụng nguồn lực và tăng cường hiệu quả kinh doanh.

5. Biểu đồ Top 10 danh mục sản phẩm con tồn kho ít nhất:

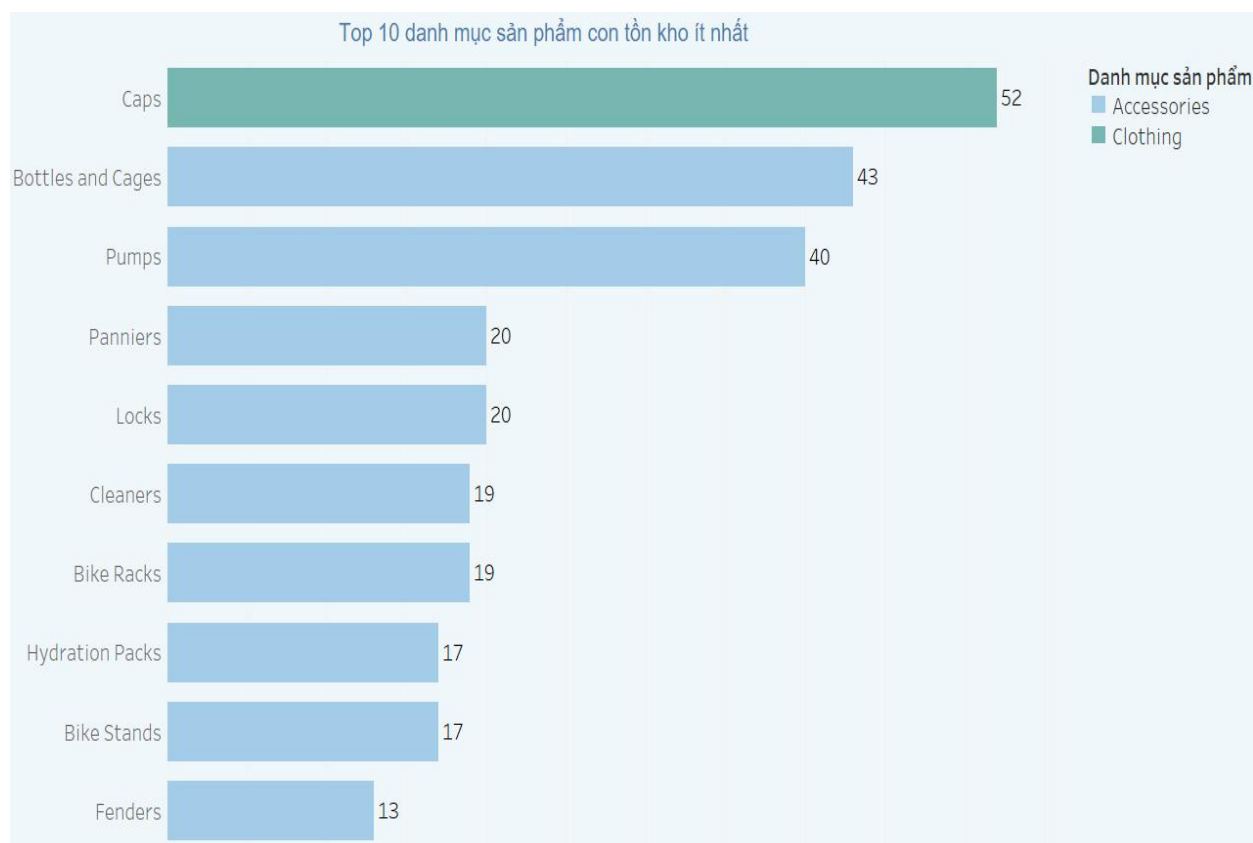


Chart 5. Top 10 danh mục sản phẩm con tồn kho ít nhất.

Ngược lại với danh mục phụ tùng, ta có thể thấy được top 10 danh mục sản phẩm con tồn kho ít nhất thì đến 9 danh mục con thuộc danh mục Clothing (quần áo thể thao).

Giữa danh mục sản phẩm con tồn kho ít nhất và tỉ lệ danh mục sản phẩm con tồn kho ít nhất có một khoảng cách quá lớn (tỉ lệ của chúng xấp xỉ 0.00009). Điều này cho thấy sản phẩm thuộc danh mục quần áo thể thao không được khách hàng ưa chuộng ở doanh nghiệp và dẫn đến số lượng tồn kho rất ít, lời khuyên đưa ra là có thể bỏ kinh doanh mặt hàng này hoặc cần đưa ra các chiến lược thu hút khách hàng. Nếu bỏ kinh doanh mặt hàng này thì tập trung nguồn lực vào việc phát triển sản phẩm ở danh mục phụ kiện (Accessories) vì trong top 10 danh mục sản phẩm con xuất hiện 1 danh mục sản phẩm con thuộc Accessories.

Đối với trường hợp quần áo là thứ dễ lỗi mode và thị trường thường xuyên thay đổi nên doanh nghiệp hạn chế tồn kho những sản phẩm thuộc danh mục này thì không phù hợp vì số lượng tồn kho thấp nhất cũng chỉ từ 13 - 20 sản phẩm

6. Biểu đồ số lượng nhập/ xuất theo năm:

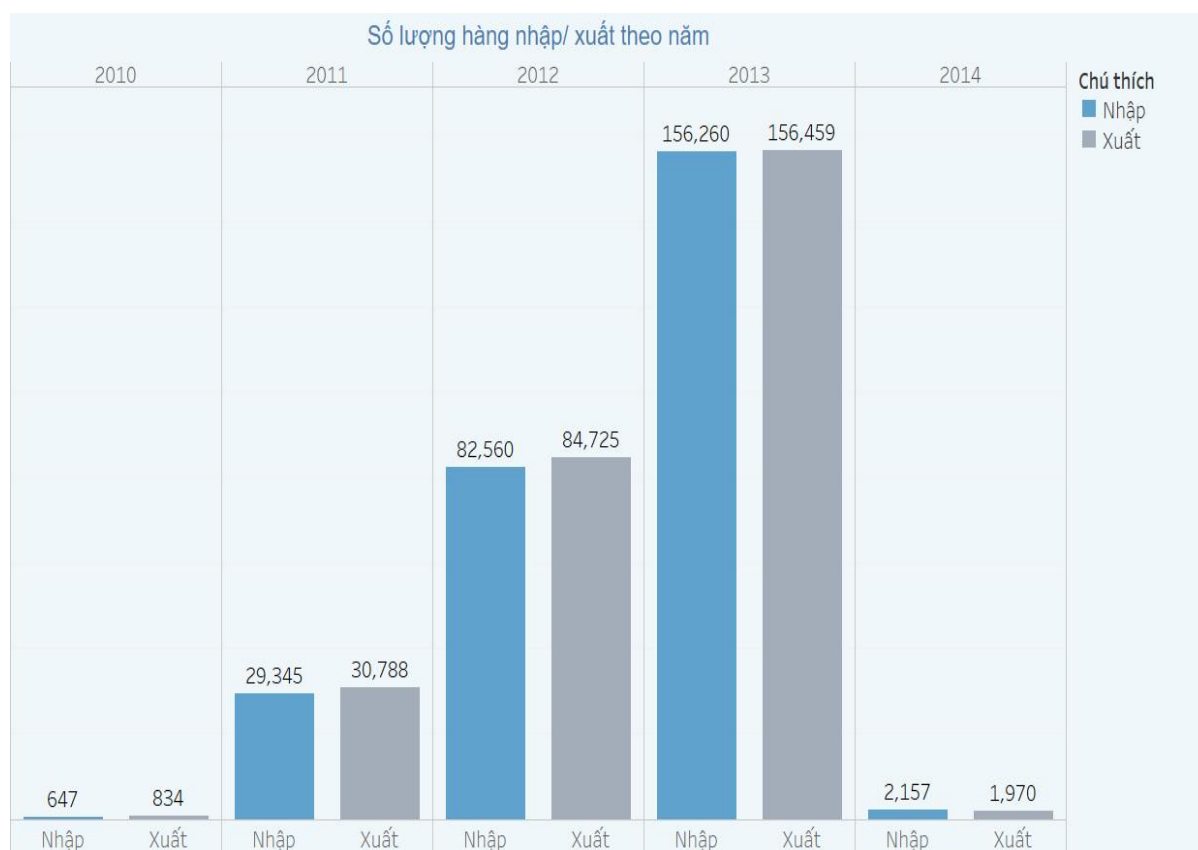


Chart 6. Số lượng hàng nhập/ xuất theo năm.

Như đã giải thích ở chart 2, bộ dữ liệu AdventureWorks2019 cập nhật giá trị hàng tồn kho ở khoảng thời gian từ ngày 1/12/2010 đến ngày 30/6/2014 nên số lượng hàng

nhập/xuất kho ở 2 năm 2010 (số liệu trong 1 tháng cuối năm) và năm 2014 (tính đến hết tháng 6 của năm) ít hơn so với những năm 2011, 2012, 2013.

Số lượng nhập và xuất hàng tăng từ năm 2011 đến năm 2013, cho thấy doanh nghiệp đang trải qua một giai đoạn tăng trưởng. Tăng trưởng này có thể phản ánh sự mở rộng hoạt động kinh doanh hoặc tăng cường nhu cầu từ thị trường.

Tuy số lượng nhập và xuất hàng đều tăng từ năm 2011 đến năm 2013, nhưng chúng có sự chênh lệch nhất định. Chẳng hạn, trong năm 2011, số lượng xuất hàng ít hơn so với số lượng nhập (30,788 xuất so với 29,345 nhập), trong khi trong năm 2013, số lượng xuất hàng gần bằng số lượng nhập (156,459 xuất so với 156,260 nhập). Điều này có thể gợi ý rằng có sự thay đổi trong quy trình quản lý tồn kho hoặc trong mô hình kinh doanh của doanh nghiệp.

Sự chênh lệch giữa số lượng nhập và xuất hàng cũng có thể ảnh hưởng đến hiệu suất tồn kho của doanh nghiệp. Nếu có sự chênh lệch lớn giữa số lượng nhập và xuất hàng, có thể doanh nghiệp đang phải đối mặt với vấn đề về tồn kho dư thừa hoặc thiếu hụt, ảnh hưởng đến lợi nhuận và hiệu suất hoạt động tổng thể.

Việc theo dõi và phân tích số liệu nhập và xuất hàng từng năm có thể giúp doanh nghiệp dự đoán và lập kế hoạch cho các hoạt động kinh doanh trong tương lai. Điều này có thể giúp đảm bảo rằng doanh nghiệp có đủ tồn kho để đáp ứng nhu cầu của thị trường mà không gặp phải tình trạng tồn kho dư thừa hoặc thiếu hụt.

7. Biểu đồ số lượng và chi phí hàng tồn kho theo danh mục sản phẩm:

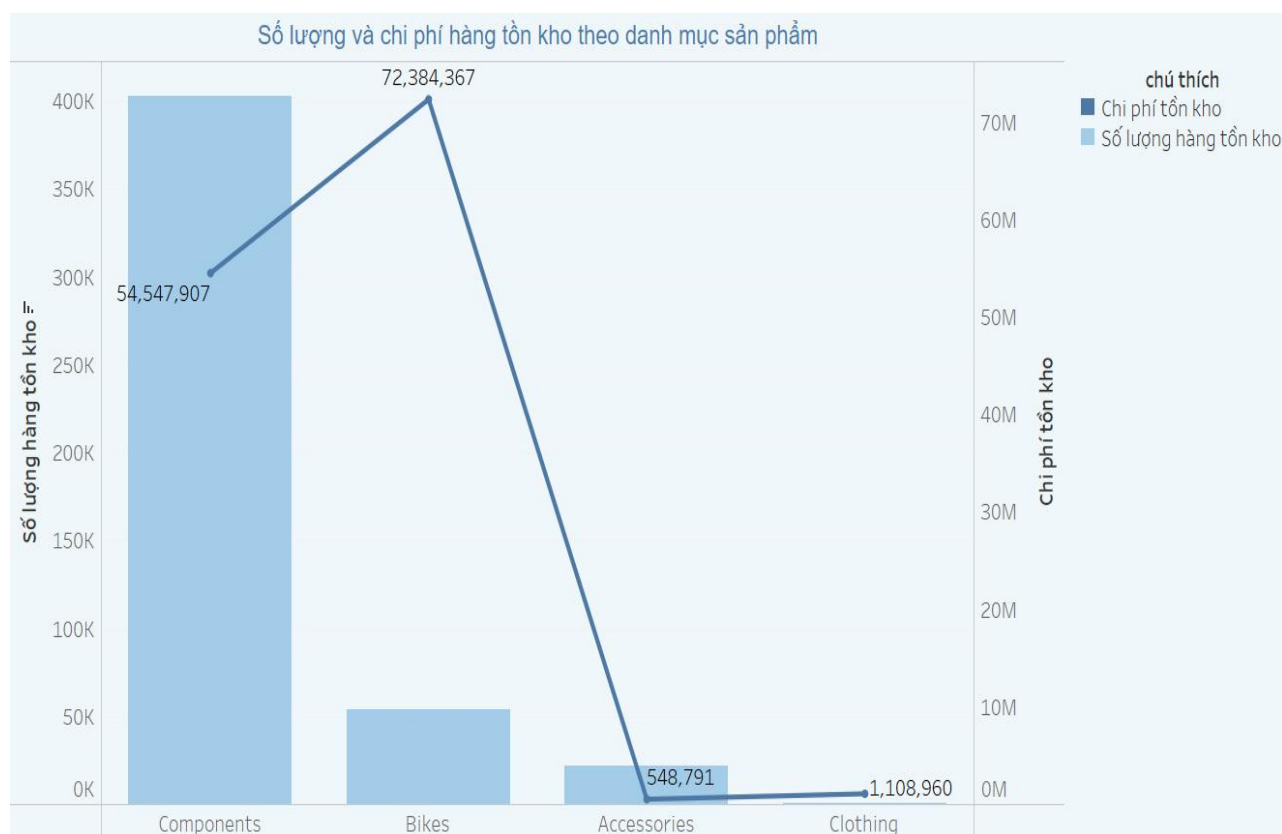


Chart 7. Biểu đồ số lượng và chi phí hàng tồn kho theo danh mục sản phẩm.

Số lượng hàng tồn kho không ảnh hưởng đến chi phí tồn kho, phần lớn sản phẩm của doanh nghiệp thuộc danh mục sản phẩm Components (84,07%) nhưng chi phí tồn kho của danh mục này lại không giữ vị trí cao nhất, trong khi đó chi phí của danh mục sản phẩm Bikes đứng đầu với 72.384.367 (đơn vị sản phẩm) với tỉ lệ số lượng hàng tồn kho là 11,23% trong tổng số lượng hàng tồn kho của cả 4 danh mục sản phẩm. Các sản phẩm thuộc danh mục Clothing ít tồn kho nhất nhưng chi phí tồn kho lại cao gần gấp hai lần so với danh mục Accessories.

Để lí giải điều này, chúng tôi thực hiện kiểm tra giá trị trung bình của các danh mục sản phẩm (chart 8), nhận thấy được rằng giá trị trung bình của danh mục sản phẩm Bikes, Components, Clothing, Accessories lần lượt là 1.525, 536, 47 và 34 (đơn vị sản phẩm). Có thể đưa ra kết luận giá trị trung bình của sản phẩm sẽ ảnh hưởng đến chi phí tồn kho, với giá trị sản phẩm càng cao thì chi phí tồn kho càng cao.

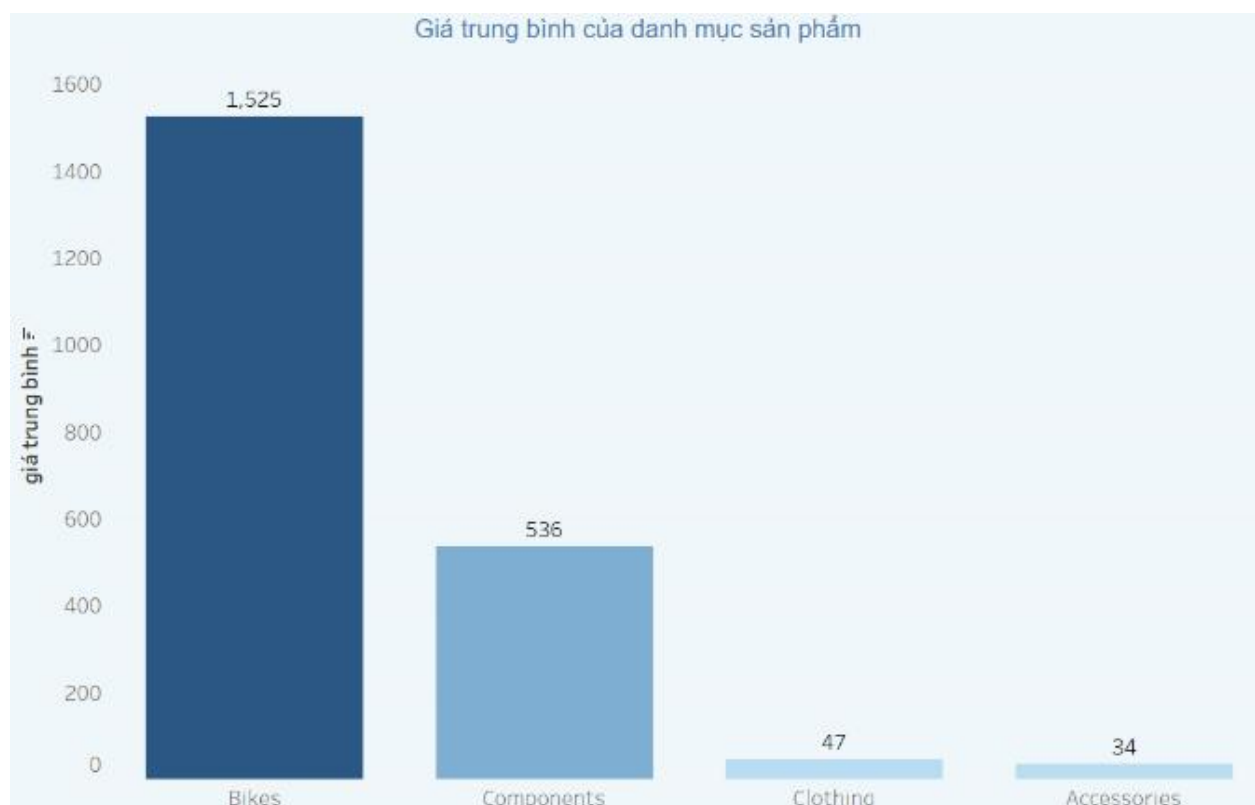


Chart 8. Giá trị trung bình của các danh mục sản phẩm.

8. Số lượng hàng tồn kho theo màu sắc:

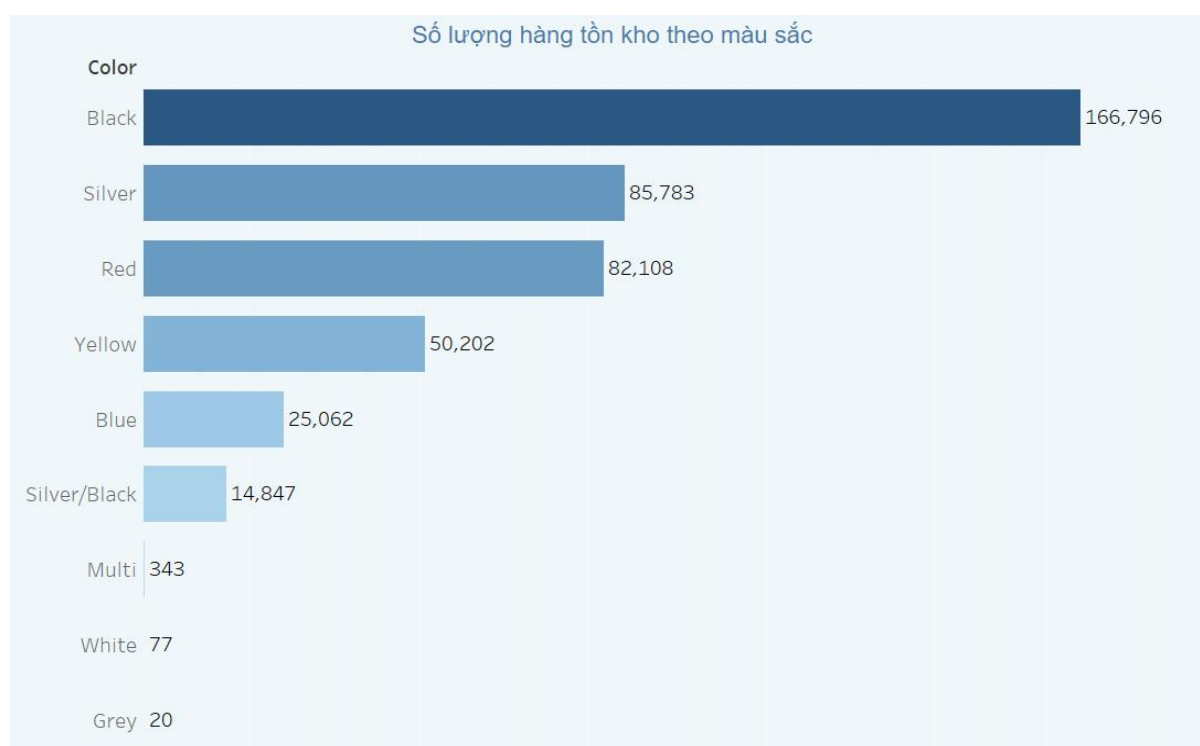


Chart 9. Số lượng hàng tồn kho theo màu sắc của sản phẩm.

Để xem xét việc hàng tồn kho và đặc tính sản phẩm có ảnh hưởng lẫn nhau như thế nào chúng tôi thực hiện trực quan hoá dữ liệu trên chart 9, chart 10 và chart 11.

Chart 9 thể hiện số lượng hàng tồn kho theo màu sắc của sản phẩm thì đưa ra nhận định rằng những sản phẩm có màu sắc là “Black”, “Sliver”, “Red” có số lượng lưu kho nhiều, chiếm tỉ lệ trên 50% tổng số lượng tồn kho. Điều này khá dễ hiểu vì đây là doanh nghiệp kinh doanh trên sàn thương mại điện tử nên những sản phẩm tồn kho nhiều là những sản phẩm bán chạy, những mặt hàng có màu sắc là đen, bạc và đỏ là những màu thông dụng và dễ sử dụng, màu mà hầu hết người mua hàng đều ưu tiên lựa chọn khi mua một sản phẩm nào đó nên số lượng tồn kho đối với đặc tính 3 loại màu sắc này chiếm tỉ lệ cao. Theo thống kê thế giới màu đỏ thuộc top màu được yêu thích nhiều nhất, vậy nên thuộc tính này còn phụ thuộc vào gu thẩm mỹ, sở thích và thói quen sử dụng của người mua để quyết định số lượng tồn kho.

9. Tỷ lệ hàng tồn kho theo lớp sản phẩm:

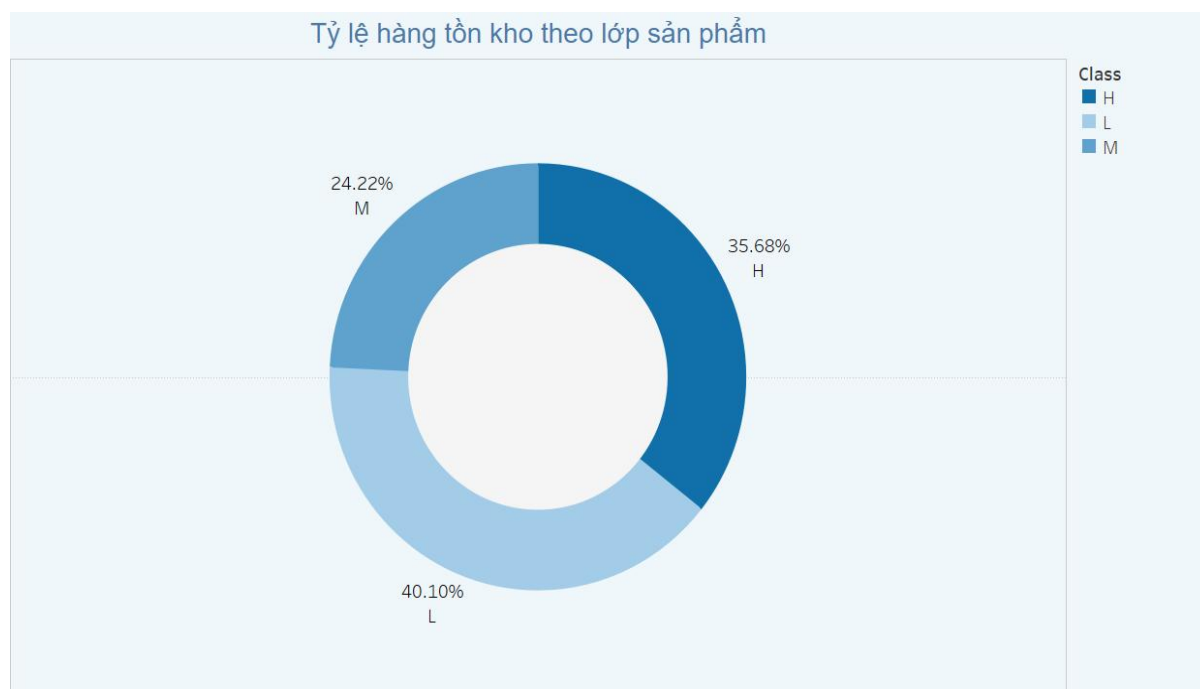


Chart 10. Tỷ lệ hàng tồn kho theo phân khúc sản phẩm (H = High, M = Medium, L = Low).

Cấu trúc tỷ lệ hàng tồn kho hiện tại của doanh nghiệp đang nghiêng về phân khúc sản phẩm Thấp, với tỷ lệ tồn kho cao hơn so với phân khúc sản phẩm Cao cấp và Trung

bình. Điều này cho thấy doanh nghiệp cần điều chỉnh chiến lược quản lý hàng tồn kho để tập trung vào phân khúc sản phẩm Cao cấp và Trung bình, nơi có lợi nhuận cao hơn và rủi ro thấp hơn.

Có thể một số nguyên nhân có thể dự đoán để dẫn đến cấu trúc tỷ lệ hàng tồn kho như trên, bao gồm:

- Nhu cầu thị trường đối với phân khúc sản phẩm Thấp có thể cao hơn so với phân khúc sản phẩm Cao cấp và Trung bình.
- Doanh nghiệp có thể áp dụng chính sách bán hàng khuyến khích tiêu thụ với phân khúc sản phẩm Thấp, dẫn đến lượng hàng tồn kho cao hơn.
- Doanh nghiệp chưa tập trung quản lý hàng tồn kho đối với phân khúc sản phẩm Cao cấp và Trung bình.

10. Tỷ lệ hàng tồn kho theo phong cách:

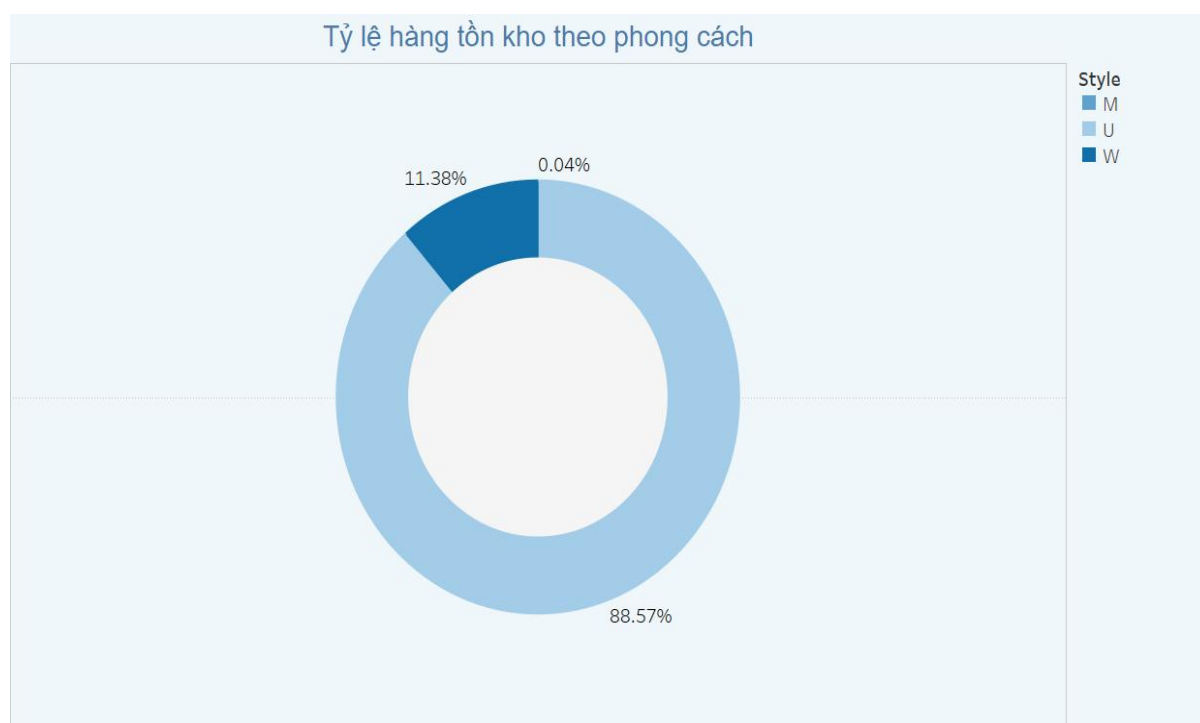


Chart 11. Tỷ lệ hàng tồn kho theo phong cách (W = Women's, M = Men's, U = Universal).

Doanh nghiệp chủ yếu kinh doanh về các sản phẩm liên quan đến thể thao như xe đạp, linh kiện, phụ kiện thể thao nên khách hàng thuộc giới tính nào cũng có thể tiếp cận

được, dẫn đến phong cách ‘Universal’ chiếm tỷ lệ cao nhất với 88,57% trên tổng lượng hàng tồn kho. Điều này có thể cho thấy sản phẩm phong cách unisex hoặc không phân biệt giới tính là phân khúc chính của công ty.

Phong cách ‘Women’ chiếm 11,38% hàng tồn kho, cao hơn đáng kể so với phong cách ‘Men’ chỉ chiếm 0,04%. Điều này gợi ý rằng công ty có thể tập trung nhiều hơn vào thị trường nam giới hoặc cần xem xét chiến lược để cân bằng lại tồn kho giữa hai phân khúc nam và nữ.

Một số đề xuất:

Doanh nghiệp cần xem xét điều chỉnh chiến lược sản phẩm và quảng bá để cân bằng lượng tồn kho giữa các phân khúc sản phẩm cho nam, nữ và unisex.

Nghiên cứu nhu cầu thị trường để điều chỉnh hoạt động nhập hàng cho phù hợp với từng phân khúc sản phẩm.

Đánh giá hiệu quả của quản lý hàng tồn kho của từng phân khúc để tối ưu hoá chi phí và giảm tồn kho không cần thiết.

11. Mức tồn kho an toàn vs Điểm đặt hàng theo danh mục sản phẩm con:



Chart 12. Mức tồn kho an toàn vs Điểm đặt hàng theo Subcategory.

- Chart 12 so sánh Mức tồn kho an toàn với Điểm đặt hàng theo danh mục sản phẩm con (Subcategory), có thể nhận thấy một số điểm sau:

Mức tồn kho an toàn cao nhất cho 3 danh mục sản phẩm con “Road Frames”, “Mountain Frames”, “Touring Frames” là từ 10000 đến 35000 đơn vị sản phẩm, cho thấy nhu cầu cao hoặc rủi ro thiếu hàng lớn cho các danh mục này, nên cần dự trữ nhiều hơn những sản phẩm thuộc ba danh mục sản phẩm con này.

Điểm đặt hàng cao nhất cũng tương ứng với các danh mục trên, lần lượt là 26250, 15000, 6750 đơn vị sản phẩm. Điều này cho thấy điểm đặt hàng phù hợp với mức tồn kho an toàn nhằm đáp ứng nhu cầu và tránh thiếu hàng.

Các danh mục khác có mức tồn kho an toàn và điểm đặt hàng thấp hơn đáng kể, chỉ khoảng 375 - 2625 đơn vị, phản ánh nhu cầu thấp hơn và không có rủi ro thiếu hàng cao nên doanh nghiệp không đề phòng về số lượng lưu kho.

Một số danh mục có điểm đặt hàng chỉ là 3 đơn vị, rất thấp so với mức tồn kho an toàn, có thể gây ra rủi ro thiếu hàng nếu không được quản lý tốt.

12. Dashboard báo cáo hàng tồn kho:

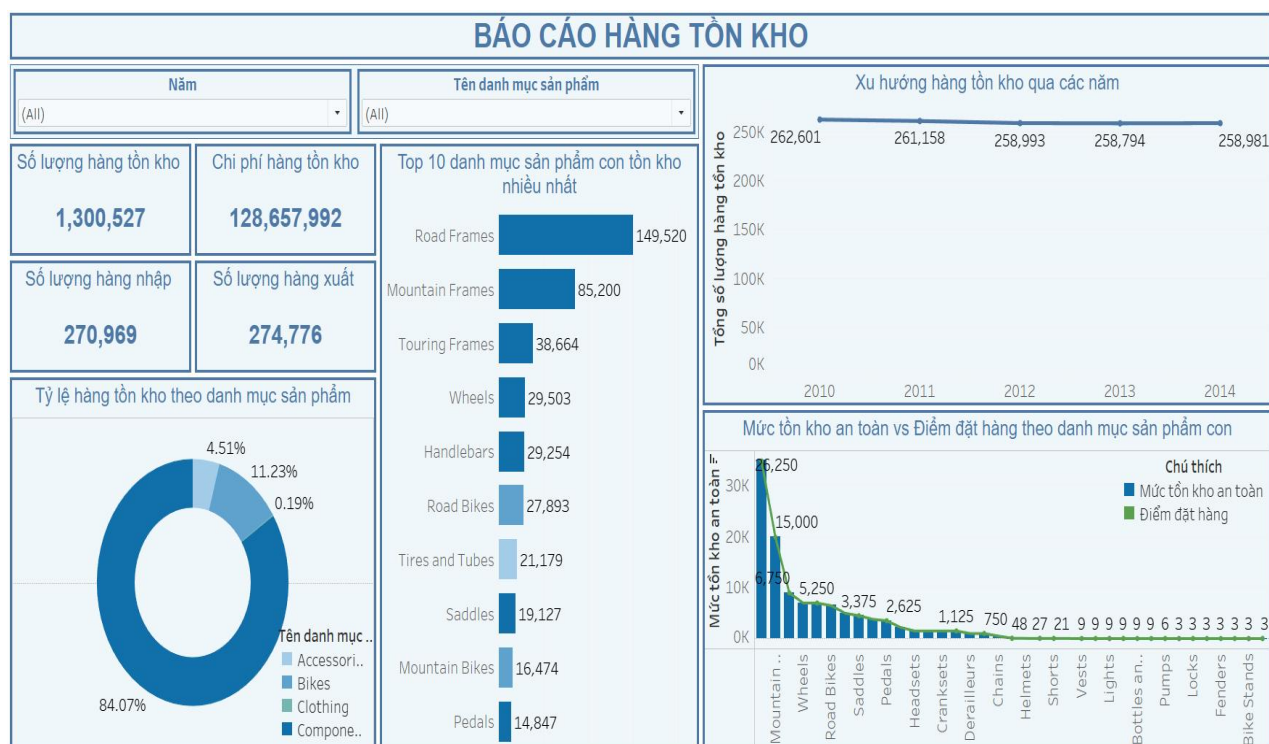


Chart 13. Dashboard báo cáo hàng tồn kho.

PHẦN 5: XÂY DỰNG MÔ HÌNH.

1. Tiền xử lý dữ liệu:

- Merge dữ liệu: Nhóm sử dụng 4 bảng ở đây là FactInternetSales, DimProduct, FactProductInventory và DimDate. Nối các bảng lại dựa trên khóa chung là ProductKey, DateKey.

```
# Merge
Merged = Sales.merge(Product, on = 'ProductKey')
Merged = Merged.merge(Inventory, on = ['ProductKey', 'DateKey'])
Merged = Merged.merge(Date, on = 'DateKey')
```

- Tổng hợp dữ liệu theo danh mục con và tháng, sau đó tính tổng số lượng hàng tồn kho và số lượng hàng đã bán cho mỗi loại sản phẩm và tháng trong năm.

```
# Tổng hợp data theo Subcategory và Month
Grouped = Merged.groupby(['ProductSubcategoryKey', 'MonthNumberOfYear']).agg({'UnitsBalance': 'sum',
                                                                              'UnitsOut': 'sum'}).reset_index()
```

- Dữ liệu sau khi tổng hợp: Gồm 204 dòng và 4 cột.

1 Grouped					
	ProductSubcategoryKey	MonthNumberOfYear	UnitsBalance	UnitsOut	
0	1	1	22101	1009	
1	1	2	19866	827	
2	1	3	23516	802	
3	1	4	24854	586	
4	1	5	27592	1763	
...	
199	37	8	488058	10104	
200	37	9	461605	10021	
201	37	10	544525	12301	
202	37	11	510968	10804	
203	37	12	541164	10991	

204 rows × 4 columns

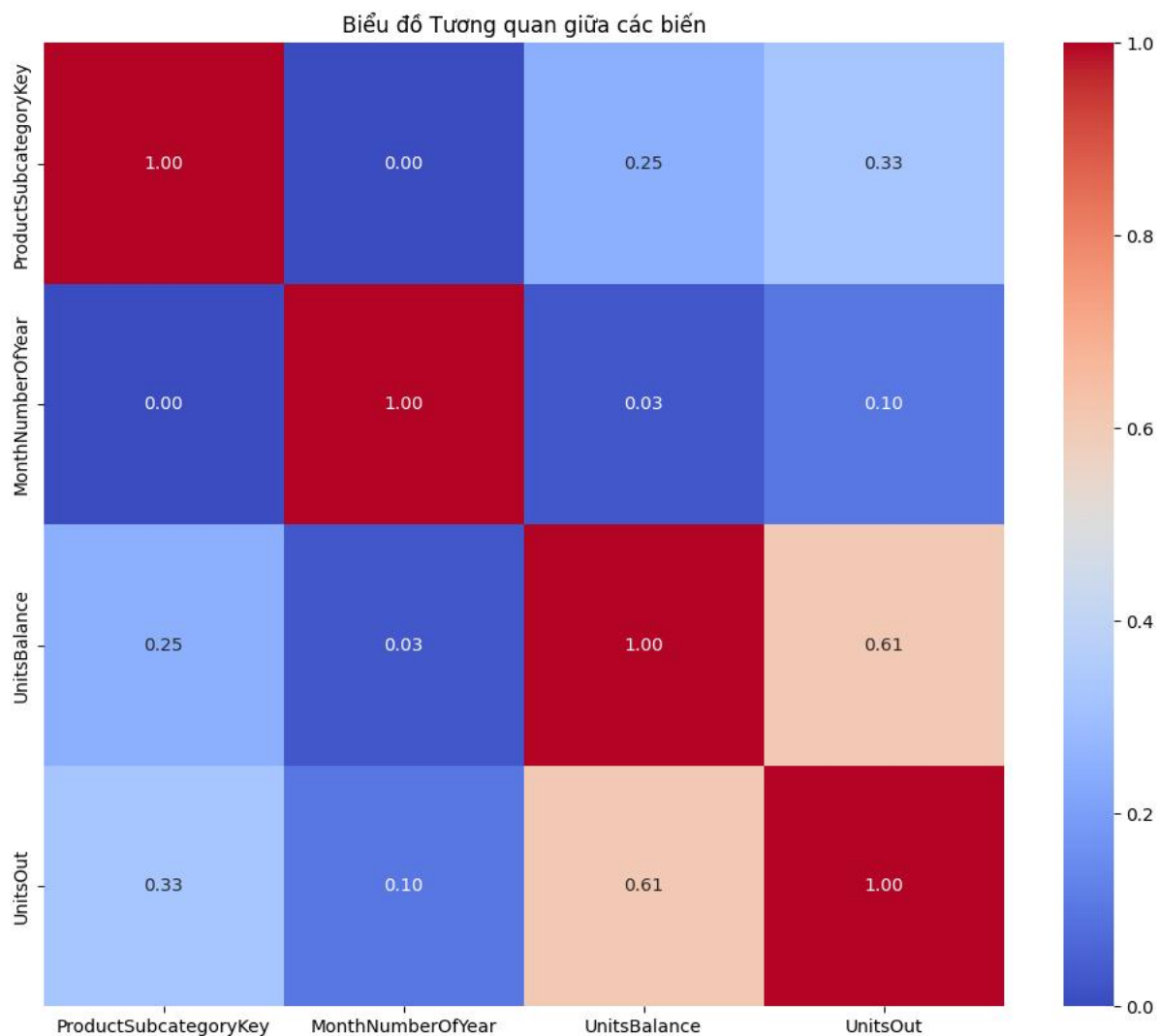
- Tính ma trận tương quan:

```
1 # Tính ma trận tương quan
2 correlation_matrix = Grouped.corr()
3 display(correlation_matrix )
```

	ProductSubcategoryKey	MonthNumberOfYear	UnitsBalance	UnitsOut
ProductSubcategoryKey	1.000000e+00	1.364549e-16	0.249747	0.327068
MonthNumberOfYear	1.364549e-16	1.000000e+00	0.027280	0.100190
UnitsBalance	2.497474e-01	2.728040e-02	1.000000	0.607990
UnitsOut	3.270677e-01	1.001905e-01	0.607990	1.000000

- Vẽ biểu đồ heatmap:

```
# Vẽ biểu đồ heatmap
plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot =True, cmap = 'coolwarm', fmt = ".2f")
plt.title('Biểu đồ Tương quan giữa các biến')
display()
```



- Lấy độ tương quan của tất cả các biến với biến "UnitsBalance".

```
1 # Lấy độ tương quan của tất cả các biến với biến "UnitsBalance"
2 correlation_with_UnitsBalance = correlation_matrix("UnitsBalance")
3 display(correlation_with_UnitsBalance)
```

```
ProductSubcategoryKey    0.249747
MonthNumberOfYear        0.027280
UnitsBalance             1.000000
UnitsOut                 0.607990
Name: UnitsBalance, dtype: float64
```

- Tính số lượng tồn kho và chuẩn bị biến mục tiêu:


```

Grouped['Inventory'] = Grouped['UnitsBalance'] - Grouped['UnitsOut']
# chuẩn bị biến mục tiêu
X = Grouped[['ProductSubcategoryKey', 'MonthNumberOfYear']]
y = Grouped['Inventory']

```

2. Xây dựng mô hình Random Forest:

- Import các thư viện cần thiết.

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor

```

- Xây dựng mô hình Random Forest.

```

1 from sklearn.preprocessing import StandardScaler
2 from sklearn.ensemble import RandomForestRegressor

```

```

1 # Chia tập dữ liệu thành tập huấn luyện và tập kiểm tra
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 42)

```

```

1 scaler = StandardScaler()
2 X_train_scaled = scaler.fit_transform(X_train)
3 X_test_scaled = scaler.transform(X_test)

```

```

1 # Tạo và huấn luyện mô hình Random Forest
2 Model = RandomForestRegressor(n_estimators = 100, random_state = 42)
3 Model .fit(X_train_scaled, y_train)

```

```

▼ RandomForestRegressor
RandomForestRegressor(random_state=42)

```

```

1 # Đưa ra dự đoán trên tập test đã chia chỉ số
2 y_pred = Model.predict(X_test_scaled)

```

3. Đánh giá mô hình:

- Các giá trị đánh giá mô hình:

```
1 # Đánh giá mô hình
2 mse = mean_squared_error(y_test, y_pred)
3 r2 = r2_score(y_test, y_pred)
4 print(f"Mean Squared Error: {mse:.2f}")
5 print(f"R-squared: {r2:.2f}")
```

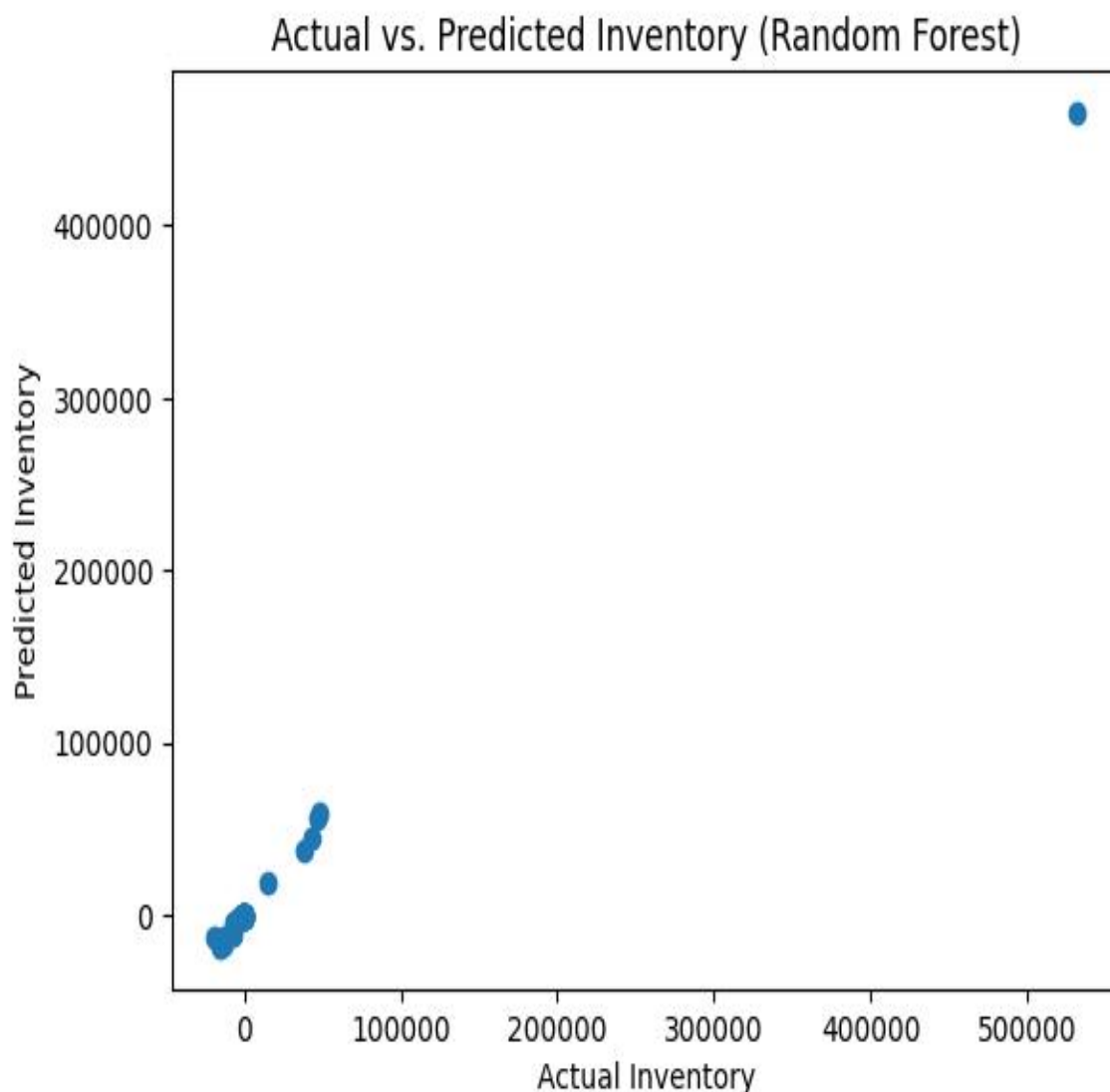
```
Mean Squared Error: 115986535.56
R-squared: 0.98
```

+ MSE (Mean Squared Error): là một phép đo lường khoảng cách trung bình giữa giá trị thực tế và giá trị dự đoán. Trong trường hợp này, giá trị MSE cao cho thấy sự chênh lệch giữa giá trị dự đoán và giá trị thực tế là khá lớn.

+ R-squared: là một phép đo mức độ giải thích của mô hình đối với biến phụ thuộc. Giá trị của R-squared nằm trong khoảng từ 0 đến 1. Trong trường hợp này, Giá trị R-squared là 0.98, giá trị R-squared càng gần 1, mô hình càng tốt.

- Mối quan hệ giữa giá trị thực tế và giá trị dự đoán:

```
# Actual vs predicted values
plt.scatter(y_test, y_pred)
plt.xlabel('Actual Inventory')
plt.ylabel('Predicted Inventory')
plt.title('Actual vs. Predicted Inventory (Random Forest)')
plt.show()
```

- Trục hoành hiển thị giá trị thực tế của biến phụ thuộc ('Actual Inventory'), trục tung hiển thị giá trị dự đoán của biến phụ thuộc ('Predicted Inventory'). Mỗi điểm trên đồ thị đại diện cho một điểm dữ liệu trong tập kiểm tra.

Dựa vào biểu đồ trên ta có thể đánh giá mức độ khớp giữa dự đoán và thực tế. Các điểm phân tán khá gần với đường chéo vì vậy mô hình này dự đoán khá tốt.

4. Dự đoán hàng tồn kho:

```
def predict_inventory(product_subcategory, month):
    input_data = pd.DataFrame({
        'ProductSubcategoryKey': [product_subcategory],
        'MonthNumberOfYear': [month]
    })

    # Dự đoán
    inventory = model.predict(input_data)[0]
    return inventory

# Dự đoán tồn kho cho product subcategory và month
product_subcategory = 1
month = 1
inventory = predict_inventory(product_subcategory, month)
print(f"Predicted inventory for product subcategory {product_subcategory} in month {month}: {inventory:.2f}")

month_list = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
inventory_list = []
for month in month_list:
    inventory = predict_inventory(product_subcategory, month)
    inventory_list.append(int(inventory))

# Tạo df và lưu kết quả
actual_inventory_list = []
for month in month_list:
    actual_inventory = Grouped[(Grouped['ProductSubcategoryKey'] == product_subcategory) & (Grouped['MonthNumberOfYear'] == month)][['Inventory']].values[0]
    actual_inventory_list.append(actual_inventory)
results = pd.DataFrame({
    'MonthNumberOfYear': month_list,
    'PredictedInventory': inventory_list,
    'ActualInventory': actual_inventory_list
})
print("Predict Inventory for Product Subcategory:", product_subcategory)
results
```

Predicted inventory for product subcategory 1 in month 1: 20909.73

Predict Inventory for Product Subcategory: 1

	MonthNumberOfYear	PredictedInventory	ActualInventory
0	1	20909	21092
1	2	20026	19039
2	3	22087	22714
3	4	24027	24268
4	5	26914	25829
5	6	35718	39529
6	7	33964	32024
7	8	35089	35478
8	9	35921	35786
9	10	37963	38475
10	11	43879	44617
11	12	45643	49657



- Việc phân tích và đưa ra đề xuất dựa trên dashboard hiện tại không cung cấp cho doanh nghiệp số liệu cụ thể để dễ dàng tham khảo và đưa ra quyết định, nó chỉ dựa vào mối quan hệ để dự đoán được nguyên nhân và đề xuất dựa trên kinh nghiệm dẫn đến độ chính xác không cao. Đây là lí do Dự án bổ sung việc thực hiện xây dựng mô hình dự đoán để cung cấp được cho doanh nghiệp số liệu cụ thể về số lượng tồn kho của các sản phẩm trong danh mục sản phẩm theo thời gian.

Số liệu được dự đoán theo từng danh mục con, với số lượng và thời gian. Lấy một ví dụ về output của mô hình ở ảnh trên, với những sản phẩm thuộc mã danh mục sản phẩm con là '1' thì số lượng hàng tồn kho dự đoán cho tháng 1 là 20909.73 đơn vị sản phẩm. Mỗi danh mục sản phẩm con đều có một con số dự đoán về hàng tồn kho theo từng tháng trong năm phù hợp với thực tế.

Độ chính xác của mô hình cao dẫn đến số lượng dự báo và số lượng thực tế không chênh lệch nhiều, mô hình phù hợp để đưa ra đề xuất về số lượng hàng tồn kho theo mỗi tháng trong năm, có thể áp dụng mô hình để doanh nghiệp đưa ra quyết định về việc xuất/ nhập hàng cho phù hợp nhằm tăng hiệu suất và chất lượng quản lý hàng tồn kho.

Dữ liệu train mô hình khá phù hợp với thực tế khi số lượng hàng tồn kho những tháng cuối của năm có nhỉnh hơn so với những tháng khác, thực tế người tiêu dùng có khả năng mua sắm cao hơn vào những tháng cuối năm nên việc tồn kho nhiều giúp đáp ứng được nhu cầu kịp thời nhằm tăng lợi ích kinh tế cho doanh nghiệp. Trong mô hình, số lượng hàng tồn kho dự đoán của các tháng 11 là 43879 đơn vị sản phẩm, tháng 12 là 45643 đơn vị sản phẩm số lượng hàng hóa gấp đôi so với những tháng 1, tháng 2 và tháng 3.

Kết luận: Mô hình dự đoán số lượng hàng tồn kho trong tương lai của Dự án phù hợp để ra quyết định xác định số lượng hàng hóa lưu kho theo từng tháng trong năm giúp doanh nghiệp quản lý hàng tồn kho tốt hơn.

PHẦN 6: KẾT LUẬN.

1. Kết quả dự án đạt được:

1.1. Đối với doanh nghiệp:

Đã xây dựng dashboard trực quan, dễ hiểu, giúp Ban lãnh đạo doanh nghiệp nắm bắt nhanh chóng tình hình hàng tồn kho theo các tiêu chí: sản phẩm, thời gian, các đặc tính sản phẩm, nhu cầu thị trường,...

Nhận diện được các vấn đề tồn tại trong việc quản lý hàng tồn kho như: tỷ lệ tồn kho cao đối với một số sản phẩm, tình trạng thiếu hụt hàng hóa cho một số sản phẩm khác.

Đã xây dựng thành công mô hình dự đoán số lượng tồn kho trong tương lai với độ chính xác cao.

Mô hình có thể dự đoán chính xác nhu cầu hàng tồn kho cho từng sản phẩm, danh mục sản phẩm theo từng khoảng thời gian cụ thể.

1.2. Đối với người thực hiện dự án:

Tham gia vào dự án giúp người thực hiện dự án nâng cao kiến thức và kỹ năng về:

- Phân tích dữ liệu: Kỹ năng thu thập, xử lý, phân tích dữ liệu hàng tồn kho bằng các công cụ và phương pháp thống kê.
- Lập trình: Kỹ năng lập trình để xây dựng dashboard và mô hình học máy.
- Nâng cao năng lực giải quyết vấn đề: Tham gia vào dự án giúp người thực hiện dự án rèn luyện khả năng tư duy logic, phân tích vấn đề và đưa ra giải pháp hiệu quả.

2. Hạn chế và hướng cải thiện của dự án:

Dự án đã đạt được những kết quả nhất định, nhưng cũng gặp phải một số hạn chế và thách thức cần được giải quyết để nâng cao hiệu quả và chất lượng báo cáo.

- Hạn chế về tính đầy đủ dữ liệu: dữ liệu không có các biến liên quan đến địa lý, thời gian cập nhật dữ liệu hàng tồn kho ngắn, việc đầy đủ dữ liệu sẽ làm bài báo cáo được phân tích một cách toàn diện hơn, chặt chẽ hơn.
- Hạn chế về quy trình xử lý dữ liệu và thực hiện dự án: dự án dường như được thực hiện theo từng bước riêng lẻ, có thể dẫn đến sự gián đoạn và thiếu tính liên tục trong quy trình.

PHỤ LỤC.

[Link Tableau Dashboard Báo cáo hàng tồn kho](#)

[Link code mô hình dự báo hàng tồn kho](#)

-----*The End*-----