ĐẠI HỌC UEH

TRƯỜNG CÔNG NGHỆ VÀ THIẾT KẾ

KHOA CÔNG NGHỆ THÔNG TIN KINH DOANH

ĐỒ ÁN MÔN HỌC

HỆ HỖ TRỢ QUẢN TRỊ THÔNG MINH

ĐỀ TÀI: Phân tích tình hình kinh doanh dựa trên bộ dữ liệu Global Store

Giảng viên: ThS. Phạm Thị Thanh Tâm

Mã lớp học phần: 24D1INF50908501

Sinh viên - MSSV: Ngô Gia Bảo - 31211027630

Lý Gia Thuận - 31211020753

Trần Hoàng Trung Đức - 31211027635

Đặng Nhật Huy - 31211027641

Hoàng Đức Dân - 31211027635

Nguyễn Tân Niên - 31211027635

MỤC LỤC

LỜI CẢM ƠN

Trong quá trình thực hiện đồ án này, chúng em đã nhận được sự hướng dẫn tận tình, chi tiết từ cô Phạm Thị Thanh Tâm. Với kiến thức chuyên môn sâu rộng cùng phương pháp giảng dạy khoa học, cô đã giúp chúng em không chỉ hoàn thành đồ án một cách tốt nhất mà còn trau dồi được nhiều kiến thức quý báu, nhất là trong lĩnh vực Hệ hỗ trợ quản trị thông minh, một ngành có nhiều ứng dụng thiết thực trong thời đại số.

Chúng em xin bày tỏ lòng biết ơn sâu sắc và lòng kính trọng đối với cô Phạm Thị Thanh Tâm. Kinh nghiệm và kiến thức mà cô đã truyền đạt không chỉ là hành trang vững chắc cho chúng em trên con đường học vấn mà còn là kim chỉ nam cho sự nghiệp tương lai của mỗi thành viên trong nhóm. Chúng em xin hứa sẽ tiếp tục nỗ lực học tập, nghiên cứu để không phụ lòng cô.

Một lần nữa, chân thành cảm ơn cô Phạm Thị Thanh Tâm về tất cả.

CHƯƠNG 1: TỔNG QUAN

1.1. Lý do chọn đề tài

Trong thời đại số hóa, dữ liệu không chỉ là nguồn thông tin mà còn là tài sản quý giá giúp doanh nghiệp hiểu rõ thị trường và khách hàng hơn. "Global Store", một bộ dữ liệu toàn diện về hoạt động kinh doanh quốc tế, cung cấp cái nhìn sâu sắc về xu hướng mua sắm, hành vi khách hàng và hiệu suất kinh doanh. Sự phức tạp và tính ứng dụng cao của bộ dữ liệu này đã thúc đẩy chúng em chọn nó làm chủ đề cho đồ án cuối kỳ, nhằm khám phá và phân tích tình hình kinh doanh qua lăng kính dữ liệu.

1.2. Mục đích, mục tiêu của đề tài

Chúng em áp dụng kiến thức từ khóa học Hệ hỗ trợ quản trị thông minh để mang lại cái nhìn mới mẻ và đa chiều về bối cảnh kinh doanh qua bộ dữ liệu "Global Store" từ 2011 đến 2015. Cụ thể, đề tài này hướng đến việc:

- Sử dụng Power BI và Excel để trích xuất thông tin hữu ích, giúp quản trị viên nắm bắt được bức tranh kinh doanh toàn diện. Quá trình này bao gồm việc áp dụng các kỹ thuật phân tích dữ liệu chuyên sâu, từ đó rút ra những hiểu biết sâu sắc về xu hướng và mẫu hành vi.

- Tạo dựng các báo cáo và dashboard trực quan từ Power BI để đánh giá chi tiết về hiệu suất bán hàng, độ phổ biến của sản phẩm, và hành vi tiêu dùng. Việc này giúp làm nổi bật các thông tin cốt lõi, qua đó phát hiện cơ hội và thách thức.
- Biến dữ liệu thành thông tin có giá trị, phục vụ cho việc lập kế hoạch và thực thi chiến lược kinh doanh. Áp dụng kiến thức và công cụ học được, chúng em nhấn mạnh vào việc cung cấp các insight đúng đắn và thời sự để hỗ trợ quyết định chiến lược.

Qua đó, mục tiêu của chúng em không chỉ là áp dụng lý thuyết vào thực tiễn thông qua việc phân tích bộ dữ liệu kinh doanh. Chúng em còn được nâng cao kỹ năng sử dụng công cụ phân tích dữ liệu, chuẩn bị cho sự nghiệp trong lĩnh vực quản trị thông tin và dữ liệu.

1.3. Đối tượng và phạm vị thực hiện

Bộ dữ liệu "Global Store" lưu trữ thông tin đa dạng từ doanh số bán hàng, thông tin sản phẩm, đến dữ liệu khách hàng và xu hướng thị trường trên quy mô toàn cầu trong giai đoạn 2011-2015. Chúng em không chỉ đánh giá hiệu suất bán hàng tổng thể mà còn nhận diện cơ hội tăng trưởng, hiểu biết sâu sắc về hành vi khách hàng, và tối ưu hóa dòng sản phẩm. Phạm vi cụ thể bao gồm:

- Phân tích doanh số bán hàng: Đi sâu vào việc phân tích doanh số để khám phá mẫu về mùa vụ, sự biến động theo năm, và ảnh hưởng của các sự kiện cụ thể lên doanh thu.
- Đánh giá sản phẩm: Phân loại và đánh giá hiệu suất của từng loại sản phẩm, nhóm sản phẩm, và dòng sản phẩm để xác định các yếu tố thành công và những khu vực cần cải thiện.
- Khám phá hành vi khách hàng: Sử dụng dữ liệu để phân tích xu hướng mua sắm, độ trung thành của khách hàng, và đặc điểm demografic, cung cấp insights quan trọng cho việc phát triển sản phẩm và chiến lược tiếp thị.

- Nghiên cứu thị trường: Tìm hiểu và so sánh hiệu suất kinh doanh trên các thị trường khác nhau, xác định các thị trường có tiềm năng cao và các yếu tố ảnh hưởng đến sự thành công tại mỗi thị trường.
- Phát triển chiến lược: Dựa trên phân tích, mục tiêu cuối cùng là đề xuất chiến lược để cải thiện hiệu suất kinh doanh, em ưu hóa dòng sản phẩm, và tăng cường mối quan hệ với khách hàng.

Mục tiêu của việc phân tích là không chỉ cung cấp cái nhìn tổng quan về hiệu suất kinh doanh của "Global Store" trong khoảng thời gian đã nêu mà còn nhận diện được các yếu tố chính yếu ảnh hưởng đến sự thành công và thất bại, từ đó phát triển các chiến lược cụ thể cho việc mở rộng và em ưu hóa hoạt động kinh doanh trong tương lai.

1.4. Phương pháp thực hiện

Chúng em sẽ áp dụng các bước cụ thể sau đây:

- Thu thập và tiền xử lý dữ liệu: Làm sạch dữ liệu và chuẩn bị chúng cho quá trình phân tích, bao gồm việc loại bỏ dữ liệu ngoại lai, điền vào các giá trị thiếu, và định dạng dữ liệu cho phù hợp.
- Phân tích dữ liệu: Sử dụng Excel cho các phân tích định lượng cơ bản và Power BI để tạo ra các báo cáo và dashboard trực quan. Qua đó, hiểu rõ hơn về tình hình kinh doanh qua các chỉ số như doanh số bán hàng, hiệu xuất các sản phẩm, và mô hình hành vi khách hàng.
- Trích xuất và trình bày Insights: Nhấn mạnh vào việc trình bày insights dễ hiểu, có giá trị thực tiễn cho các nhà quản trị. Sử dụng biểu đồ, bảng, và các công cụ trực quan khác trong Power BI để làm nổi bật thông tin quan trọng.
- Phát triển báo cáo: Cuối cùng, tổng hợp tất cả phân tích và insights vào một báo cáo cuối cùng hoặc một dashboard trực quan, cung cấp cái nhìn tổng quan và sâu sắc về tình hình kinh doanh dựa trên bộ dữ liệu Global Store.

1.5. Bố cục của đề tài

Đồ án được cấu trúc thành các chương nhằm đảm bảo tính rõ ràng, mạch lạc và hệ thống trong việc trình bày, phân tích dữ liệu từ bộ dữ liệu "Global Store". Cụ thể, cấu trúc của đồ án gồm các phần sau:

Chương 1: Tổng Quan

Chương này cung cấp cái nhìn khái quát về đề tài, bao gồm lý do chọn đề tài, mục tiêu của đồ án, đối tượng và phạm vi nghiên cứu, phương pháp thực hiện và phân công công việc cho từng thành viên trong nhóm.

Chương 2: Mô Tả Doanh Nghiệp

Phần này giới thiệu về doanh nghiệp được phân tích trong đề tài, bao gồm lịch sử hình thành, cấu trúc tổ chức, và thực trạng hiện tại của doanh nghiệp. Đồng thời, chương này cũng đề cập đến các vấn đề mà doanh nghiệp đang phải đối mặt cùng các mục tiêu cần giải quyết.

Chương 3: Quá Trình ETL

Trình bày chi tiết các bước Trích xuất (Extract), Chuyển đổi (Transform) và Nạp dữ liệu (Load):

- Giai đoạn 1: Trích xuất dữ liệu - Mô tả nguồn dữ liệu và cách thức trích xuất dữ liệu.
- Giai đoạn 2: Chuyển đổi dữ liệu - Diễn giải các kỹ thuật và phương pháp được áp dụng để chuyển đổi dữ liệu thô thành dữ liệu sạch và có cấu trúc.
- Giai đoạn 3: Nạp dữ liệu - Giới thiệu cách thức dữ liệu được nạp vào kho dữ liệu để phục vụ quá trình phân tích.

Chương 4: Phân Tích Dữ Liệu

Phần này phân tích chi tiết dữ liệu đã được chuyển đổi, bao gồm:

- Dashboard tổng quan - Cung cấp cái nhìn toàn cảnh về tình hình kinh doanh của doanh nghiệp qua các chỉ số chính.
- Phân tích chuyên sâu - Đánh giá hiệu quả hoạt động kinh doanh qua các phân khúc, sản phẩm, và thị trường.

Chương 5: Đề Xuất Giải Pháp

Dựa trên kết quả phân tích, chương này sẽ trình bày các giải pháp và chiếnlược cải thiện hiệu quả hoạt động, phát triển sản phẩm, và mở rộng thị trường cho doanh nghiệp.

Chương 6: Kết Luận và Hướng Phát Triển

Tóm tắt các phát hiện chính và đề xuất hướng phát triển tương lai cho doanh nghiệp dựa trên các kết quả và giải pháp đã nêu trong đồ án.

Mỗi chương đều được hỗ trợ bằng đầy đủ dữ liệu, bảng biểu, và minh họa trực quan để tăng cường tính thuyết phục và dễ hiểu cho người đọc. Cấu trúc này không chỉ giúp người đọc theo dõi dễ dàng mà còn thúc đẩy việc khai thác hiệu quả dữ liệu trong việc đưa

1.6. Phân công công việc

BẢNG PHÂN CÔNG

STT	Họ Tên	Công việc phụ trách	MĐHT
1	Lý Gia Thuận	● Chương 3: ETL ● Vẽ dashboard theo độ ưu tiên đơn đặt hàng	100%
2	Trần Hoàng Trung Đức	● Chương 1: Tổng quan ● Vẽ dashboard theo thị trường	100%
3	Ngô Gia Bảo	● Chương 4: Phân tích dữ liệu ● Vẽ dashboard overview	100%
4	Hoàng Đức Dân	● Chương 2: Mô tả doanh nghiệp ● Vẽ dashboard phân khúc khách hàng	100%
5	Nguyễn Tân Niên	● Chương 6: Kết luận ● Vẽ dashboard danh mục con	100%
6	Đặng Nhật Huy	● Chương 5: Đề xuất giải pháp ● Vẽ dashboard danh mục sản phẩm	100%

CHƯƠNG 2: MÔ TẢ DOANH NGHIỆP

2.1. Giới thiệu doanh nghiệp

Global Superstore, có trụ sở chính tại thành phố New York sôi động, nổi bật lên bởi cách thức hoạt động như một công ty bán lẻ toàn cầu bao gồm rất nhiều sản phẩm đa dạng. Với mục tiêu phục vụ như điểm đến cuối cùng cho khách hàng, siêu thị phục vụ một đối tượng khách hàng đa dạng từ 147 quốc gia khác nhau bao gồm cả người tiêu dùng và các doanh nghiệp lớn nhỏ.

Công ty Global Superstore kinh doanh 3 mặt hàng chính: Công nghệ; Nội thất; Văn phòng phẩm được thành lập vào năm 2011 và hoạt động chủ yếu ở thị trường trực tuyến. Để đáp ứng được phân khúc và quy mô thị trường đang nhắm tới, Global Superstore có mạng lưới phân phối rộng, ghi nhận các đơn hàng sản phẩm từ khắp nơi trên thế giới với nhiều hình thức vận chuyển khác nhau mặc dù quy mô của doanh nghiệp vẫn chỉ ghi nhận ở mức vừa và nhỏ với tổng đơn hàng ở mức hơn 50 nghìn đơn

Tuy nhiên, điều này mang lại một thách thức khá lớn khi doanh thu của công ty chỉ đạt khoảng 10 triệu USD trong khoảng thời gian này và đây cũng là 1 con số khá khiêm tốn. Đối với thị trường cạnh tranh trong lĩnh vực bán lẻ, Global Superstore hiện đang phải dốc sức để có thể bắt kịp tốc độ cũng như quy mô của các công ty tên tuổi khác như Walmart, Amazon,... Điều này có thể sẽ trở nên khó khăn khi đây cũng là những công ty nắm được thị phần rất lớn trong thị trường nước Mỹ và cả thế giới.

Vì vậy, việc củng cố những dữ liệu đầu vào và đưa ra các chiến lược kinh doanh cũng như tiếp thị sẽ rất quan trọng để có thể nâng cao hiệu suất và lợi nhuận tổng thể của cả công ty cũng như chiến lược phát triển lâu dài và bền vững

2.2. Thực trạng của doanh nghiệp

Trong khoảng thời gian từ 2011 đến 2015 của bộ dữ liệu, Global Superstore hoạt động chính trên khắp lãnh thổ nước Mỹ khi chiếm khoảng ⅕ tổng số đơn hàng ghi nhận thực hiện và phần còn lại phân bố ở khắp toàn cầu. Điều này có thể khiến cho chi phí vận chuyển của công ty chưa thực sự được tối ưu bởi sự phân bố là không đều trên toàn thế giới khiến cho mạng lưới logistics của công ty sẽ gặp khá nhiều khó khăn để có được một chuỗi cung ứng hiệu quả giữa cung và cầu

Tổng lợi nhuận: Thống kê từ bộ dữ liệu trên, Global Superstore sau khi thành lập đã đạt tổng cộng 1,467 triệu USD

Tổng doanh thu ghi nhận được của công ty là khoảng 12,6 triệu USD. Điều này có thể cho chúng ta thấy rằng tỷ lệ lợi nhuận chỉ đạt khoảng 1/10 tổng doanh thu - một con số khá thấp trong công cuộc kinh doanh của cả doanh nghiệp. Nguyên nhân có thể kể đến sự chưa tối ưu hóa trong quy trình hoạt động cũng như chi phí phải bỏ ra. Nhìn chung, đây là một con số chưa thực sự nổi bật khi hai doanh nghiệp bán lẻ hàng đầu nước Mỹ là Walmart và Amazon đều đạt được những con số cực kì nổi trội

- Amazon: Từ Q1 2015 đến hiện tại, Amazon đã thể hiện sự tăng trưởng về doanh thu, với hơn 25 tỷ đô la doanh thu trong Q3 2015.

  Tuy nhiên, lợi nhuận của Amazon trong giai đoạn 2011-2015 vẫn chưa thực sự ấn tượng và bỏ xa các đối thủ khác. Tổng lợi nhuận kết hợp của Amazon từ 1995 đến 2015 chỉ là 2.56 tỷ đô la

- Walmart: Trong năm 2015, Walmart đã tạo ra tổng cộng 482 tỷ đô la doanh thu.

  Tuy nhiên, lợi nhuận của Walmart trong giai đoạn này không thể so sánh với Amazon, Walmart chỉ tăng trưởng lợi nhuận với tỷ lệ trung bình 1% mỗi năm.

  Chi phí vận chuyển của công ty ghi nhận ở khoảng 1,4 triệu USD. Đây cũng là một con số không nhỏ khi chỉ mỗi dịch vụ vận

chuyển hàng hóa đến tay khách hàng đã chiếm tổng cộng 1/10 doanh thu. Việc chỉ tập trung vào thị trường trực tuyến nhưng lại không chú tâm đến việc xuất khẩu và quảng bá sản phẩm khi là một đơn vị phân phối sản phẩm toàn cầu có thể là nguyên do cho việc số lượng đơn hàng ở các nước ngoài Hoa Kỳ vẫn còn trong tình trạng rất hạn chế. Song, việc này còn dễ mang lại sự giảm sút của tỷ lệ lợi nhuận khi không thể tối ưu hóa chi phí từ các đơn hàng nhỏ lẻ cũng như không mang lại hiệu quả trong kinh doanh.

2.3. Bài toán của doanh nghiệp và mục tiêu cần giải quyết

Vì vậy, để làm mục tiêu cho đề tài nghiên cứu cũng như phương hướng giải quyết các vấn đề gặp phải trong việc phát triển kinh doanh nói chung và tỷ lệ doanh thu, lợi nhuận với các chi phí nói riêng. Nhóm nghiên cứu đã phân tích và nhận thấy rằng doanh nghiệp cần đạt được những mục tiêu sau đây:

- Xem xét việc thực hiện kinh doanh mua bán trên cả hai thị trường: trực tuyến và phân phối sản phẩm tại các cửa hàng vật lý. Việc giảm trừ chi phí vận chuyển tại Hoa Kỳ cũng như tăng cường việc xuất khẩu hàng hóa đến các quốc gia khác để đảm bảo 1 mạng lưới kinh doanh bền vững góp phần hạn chế những đơn hàng nhỏ lẻ ở các quốc gia ghi nhận mức đơn hàng thấp cũng là rất cần thiết.
- Tăng cường phân tích thêm về hành vi khách hàng, áp dụng các công nghệ đổi mới về gom cụm cũng như phân loại các phân khúc khách hàng theo các mức độ ưu tiên cũng như mức tiêu dùng. Việc này sẽ giúp ích cho doanh nghiệp có một chiến lược tiếp cận khách hàng và đưa ra những quyết định để làm tăng số lượng đơn đặt hàng

- Ngoài ra, việc theo dõi xu hướng mua sắm theo mùa, theo các dịp lễ cũng là điểm mà doanh nghiệp nên tận dụng để nâng cao mức nhận thức cũng như độ nhận diện thương hiệu đối với khách hàng. Các chiến lược quảng bá và marketing cũng được nhắm đến khi có thể thấy độ bao phủ trên toàn cầu là chưa cao, có thể áp dụng các chiến lược digital marketing hoặc nhắm vào các thị trường trọng điểm nằm ở các khu vực lận cận.

CHƯƠNG 3: QUÁ TRÌNH ETL

3.1. Giai đoạn 1: Trích xuất dữ liệu (Extract)

3.1.1. Trích xuất dữ liệu

Bộ dữ liệu "Global Superstore" được thu thập từ Kaggle, một nền tảng web dành cho khoa học dữ liệu và được đăng tải bởi tác giả có tên là APOORVA MAHALINGAPPA.

Tác giả: Profile

Link bộ dữ liệu: Global Superstore

3.1.2. Mô tả bộ dữ liệu

Bộ dữ liệu có tên là Global Superstore, chứa các thông tin về doanh số bán hàng, thông tin sản phẩm, dữ liệu khách hàng và xu hướng thị trường trên quy mô toàn cầu chủ yếu từ năm 2011-2015. Bộ dữ liệu chứa khoảng 51 nghìn bản ghi nhận số liệu.

Data card:

Row ID	ID duy nhất cho mỗi hàng
Category	Danh mục của sản phẩm được đặt hàng
Sub-Category	Danh mục con của sản phẩm được đặt hàng

Product ID	ID duy nhất của Sản phẩm
Product Name	Tên của Sản phẩm
Market	Thị trường hoạt động
Region	Khu vực
Country	Quốc gia cư trú
State	Bang
City	Thành phố cư trú
Postal Code	Mã bưu chính của mỗi Khách hàng
Customer ID	ID duy nhất để xác định mỗi Khách hàng
Customer Name	Tên của Khách hàng
Segment	Phân khúc Khách hàng
Order ID	ID đặt hàng duy nhất cho đơn hàng
Order Date	Ngày đặt hàng của sản phẩm
Order Priority	Thứ tự ưu tiên đặt hàng

Ship Date	Ngày giao hàng của Sản phẩm.
Ship mode	Phương thức giao hàng
Sales	Doanh số của Sản phẩm
Discount	Chiết khấu được cung cấp
Profit	Lợi nhuận
Quantity	Số lượng của Sản phẩm
Shipping Cost	Chi phí vận chuyển

3.1.3. Đánh giá chất lượng của dữ liệu

Bộ dữ liệu được tổ chức rất tốt gồm 24 đặc trưng về các thông tin của xu hướng thị trường và các dòng dữ liệu được qua xử lý và nhập vào gọn gàng, nhưng vẫn chưa thực sự tối ưu cho việc phân tích dữ liệu vì chỉ nằm ở trong một bảng duy nhất.

3.1.4. Mô tả cơ bản

“Global Superstore” gồm 51291 dòng và 24 cột và chỉ có một cột Postal Code chứa các giá trị rỗng, với 41296 giá trị, có thể bởi vì bộ dữ liệu chỉ ghi nhận những mã bưu chính trong Hoa Kỳ nói riêng.

```
for col in df:

    if df[col].isnull().sum() > 0:
```

```
        print(f"Missing of {col}: {df[col].isnull().sum()}")



Missing of Postal Code: 41296
```

3.2. Giai đoạn 2: Chuyển đổi dữ liệu (Transform)

Ngôn ngữ được sử dụng chính cho quá trình ETL là Python.

3.2.1. Chuyển đổi Category_Dim

- Xử lý: Trích xuất thuộc tính Category, xóa các giá trị trùng lặp và sau đó tạo khóa CategoryID bằng dạng chuỗi "C-" với ba ký tự đầu của các giá trị trong biến Category

```
category_dim = df[['Category']]

category_dim.drop_duplicates(subset=['Category'],
   inplace=True)

category_dim = category_dim.sort_values('Category')

category_dim.reset_index(drop=True, inplace=True)

category_dim['CategoryID'] = 'C-' +
   category_dim['Category'].apply(lambda x:
   x[:3].upper()


   if len(x.split(' ')) == 1 else
   ''.join(word[0].upper() for word in x.split(' ')))

category_dim = category_dim[['CategoryID', 'Category']]
```

	CategoryID	Category
0	C-FUR	Furniture
1	C-OS	Office Supplies
2	C-TEC	Technology

- Mô tả: Bảng Category_Dim Gồm 2 cột
  - CategoryID: Khóa cho hạng mục sản phẩm, chứa ba giá trị duy nhất và có ý nghĩa cho tên của hạng mục sản phẩm
  - Category: Tên hạng mục sản phẩm

3.2.2. Chuyển đổi Sub_Category_Dim

- Xử lý: Trích xuất các thuộc tính liên quan đến hạng mục con, xóa các giá trị trùng lặp và sau đó tạo khóa Sub-CategoryID bằng dạng chuỗi "SC-" với ba ký tự đầu của các giá trị trong thuộc tính Category, sau đó sử dụng hàm map để ghép các giá trị của Category tương ứng với Sub-Category đó.

```
sub_category_dim = df[['Sub-Category','Category']]

sub_category_dim = df[['Sub-Category',
   'Category']].drop_duplicates(subset=['Sub-Category'])

sub_category_dim = sub_category_dim.sort_values('Sub-
   Category')

sub_category_dim.reset_index(drop=True, inplace=True)

sub_category_dim['Sub-CategoryID'] =  'SC-' +
   sub_category_dim['Sub-Category'].apply(lambda x:
   x[:3].upper()


   if len(x.split(' ')) == 1 else
   ''.join(word[0].upper() for word in x.split(' ')))

sub_category_dim['CategoryID'] =
   sub_category_dim['Category'].map(dict(zip(category_di
   m['Category'], category_dim['CategoryID'])))
```

```
sub_category_dim = sub_category_dim[['Sub-CategoryID',
   'Sub-Category', 'CategoryID']]
```

	Sub-CategoryID	Sub-Category	CategoryID
0	SC-ACC	Accessories	C-TEC
1	SC-APP	Appliances	C-OS
2	SC-ART	Art	C-OS
3	SC-BIN	Binders	C-OS
4	SC-BOO	Bookcases	C-FUR
5	SC-CHA	Chairs	C-FUR
6	SC-COP	Copiers	C-TEC
7	SC-ENV	Envelopes	C-OS
8	SC-FAS	Fasteners	C-OS
9	SC-FUR	Furnishings	C-FUR
10	SC-LAB	Labels	C-OS
11	SC-MAC	Machines	C-TEC
12	SC-PAP	Paper	C-OS
13	SC-PHO	Phones	C-TEC
14	SC-STO	Storage	C-OS
15	SC-SUP	Supplies	C-OS
16	SC-TAB	Tables	C-FUR

- Mô tả: Gồm 3 cột:
    - Sub-CategoryID: Khóa của hạng mục con của sản phẩm, gồm 16 giá trị duy nhất và có ý nghĩa cho tên hạng mục con đó
    - Sub-Category: Tên hạng mục con
    - CategoryID: Khóa ngoại, nơi mà các hạng mục con thuộc một hạng mục cha, có ý nghĩa liên kết về bảng cha của bảng là Category_Dim.

3.2.3. Chuyển đổi Product_Dim:

- Xử lý: Trích xuất các thuộc tính về sản phẩm, xóa các giá trị trùng lặp của bảng, do đã có cột Product ID trong bảng ban đầu nên không cần tạo khóa cho bảng, sử dụng hàm map để gán hạng mục con Sub-Category mà sản phẩm thuộc về.

```
product_dim = df[['Product ID', 'Product Name', 'Sub-
    Category']]

product_dim['Sub-CategoryID'] = product_dim['Sub-
    Category'].map(dict(zip(sub_category_dim['Sub-
    Category'],
    sub_category_dim['Sub-CategoryID'])))



product_dim = product_dim[['Product ID', 'Product Name',
    'Sub-CategoryID']]

product_dim = product_dim.drop_duplicates('Product ID')

product_dim.reset_index(drop=True, inplace=True)
```

	Product ID	Product Name	Sub-CategoryID
0	TEC-AC-10003033	Plantronics CS510 - Over-the-Head monaural Wir...	SC-ACC
1	FUR-CH-10003950	Novimex Executive Leather Armchair, Black	SC-CHA
2	TEC-PH-10004664	Nokia Smart Phone, with Caller ID	SC-PHO
3	TEC-PH-10004583	Motorola Smart Phone, Cordless	SC-PHO
4	TEC-SHA-10000501	Sharp Wireless Fax, High-Speed	SC-COP
...	...	...	...
10287	OFF-FA-10004112	Stockwell Staples, 12 Pack	SC-FAS
10288	OFF-BI-10003253	Ibico Index Tab, Economy	SC-BIN
10289	OFF-BI-10002510	Acco Index Tab, Clear	SC-BIN
10290	FUR-ADV-10002329	Advantus Light Bulb, Erganomic	SC-FUR
10291	OFF-AP-10002203	Eureka Disposable Bags for Sanitaire Vibra Gro...	SC-APP

- Mô tả: Gồm 3 cột:
    - ProductID: Khóa chính của sản phẩm, gồm 10292 giá trị duy nhất và có ý nghĩa cho tên của sản phẩm
    - Product Name: Tên sản phẩm
    - Sub-CategoryID: Khóa ngoại, nơi mà các sản phẩm thuộc một hạng mục con, có ý nghĩa liên kết về bảng cha của bảng Product_Dim là Sub_Category_Dim.

3.2.4. Chuyển đổi Location_Dim:

- Xử lý:

- Trích xuất các thuộc tính có ý nghĩa về địa điểm, xóa các giá trị trùng lặp theo subset là City và Postal, việc làm này sẽ tạo ra các giá trị duy nhất theo khả năng kết hợp của thành phố và mã bưu chính bởi vì có thể một thành phố có thể có nhiều mã bưu chính, sau đó thực hiện mã hóa cho bộ kết hợp của City_Postal bằng thư viện LabelEncoder thành biến City_Postal_Encoded. Việc duy nhất hóa theo hai cột này là bởi vì đây là hai giá trị chứa có mức ý nghĩa phân bổ địa điểm nhỏ nhất (mô tả đúng chi tiết nơi chốn)
- Tạo một hàm để trích xuất ID của thành phố với điều kiện: (*Lấy 3 ký tự đầu nếu thành phố có một chữ, Lấy các ký tự đầu nếu thành phố có 2 chữ trở lên*)
- Tạo một biến Abbre như là viết tắt của State
- Sau đó tạo khóa chính LocationID cho Location_Dim bằng tất chuỗi “City + Abbre + City_Postal_Encoded”
- Chọn lại các thuộc tính cần thiết

```
from sklearn.preprocessing import LabelEncoder



location_dim = df[['City', 'State', 'Country',
   'Postal Code', 'Market', 'Region']]

location_dim['City'] =
   location_dim['City'].apply(lambda x: x.strip())

location_dim =
   location_dim.drop_duplicates(subset=['City',
   'Postal Code']).reset_index(drop=True)

location_dim['Abbre'] =
   location_dim['Market'].replace({'Africa':
   'AFR', 'Canada': 'CAN'})



location_dim['Postal Code'] = location_dim['Postal
```

```
   Code'].astype(str)

location_dim['City_Postal'] = location_dim['City']
   + '_' + location_dim['State'] + '_' +
   location_dim['Country'] + '_' +
   location_dim['Postal Code']

label_encoder = LabelEncoder()

location_dim['City_Postal_Encoded'] =
   label_encoder.fit_transform(location_dim['City_
   Postal'])




def extract_id_city(city):

    if len(city.split(' ')) == 1:

        return city[:3].upper()

    elif len(city.split(' ')) > 1:

        return ''.join(word[0].upper() for word in
   city.split(' '))



location_dim['City ID'] =
   location_dim['City'].apply(extract_id_city)

location_dim['Location ID'] = location_dim['City
   ID'] + '-' + location_dim['Abbre'] + '-' +
   location_dim['City_Postal_Encoded'].astype(str)

location_dim = location_dim[['Location ID',
   'City', 'State', 'Country', 'Postal Code',
   'Market', 'Region', 'City_Postal_Encoded']]

location_dim['Postal Code'] = location_dim['Postal
```

```
    Code'].replace({'nan': np.NaN})
```

	Location ID	City	State	Country	Postal Code	Market ID	Region	City_Postal_Encoded
0	NYC-US-2377	New York City	New York	United States	10024.0	US	East	2377
1	WOL-APAC-3651	Wollongong	New South Wales	Australia	NaN	APAC	Oceania	3651
2	BRI-APAC-508	Brisbane	Queensland	Australia	NaN	APAC	Oceania	508
3	BER-EU-384	Berlin	Berlin	Germany	NaN	EU	Central	384
4	DAK-AFR-884	Dakar	Dakar	Senegal	NaN	AFR	Africa	884
...	...	...	...	...	...	...	...	...
3766	SLO-US-2973	San Luis Obispo	California	United States	93405.0	US	West	2973
3767	ABI-US-13	Abilene	Texas	United States	79605.0	US	Central	13
3768	FEL-EMEA-1110	Felahiye	Kayseri	Turkey	NaN	EMEA	EMEA	1110
3769	LEW-US-1906	Lewiston	Idaho	United States	83501.0	US	West	1906
3770	VF-AFR-3518	Victoria Falls	Matabeleland North	Zimbabwe	NaN	AFR	Africa	3518

- Mô tả: Gồm 8 Cột:
    - LocationID: Là khóa chính cho bảng Location_Dim, gồm 3771 giá trị duy nhất và có ý nghĩa cho mỗi địa điểm duy nhất.
    - City: Thành Phố
    - State: Bang
    - Country: Đất nước
    - Market ID: Khóa ngoại và liên kết về bảng cha Market_Dim
    - Region: Vùng

3.2.5. Chuyển đổi Market_Dim:

- Xử lý: Tạo khóa chính Market ID bằng biến Market từ Location_Dim, sau đó tạo cột Market với các tên đầy đủ của Market đó. Cuối cùng xóa các giá trị trùng lặp của Market ID

```
market_dim = location_dim[['Market']]

market_dim.rename(columns={'Market': 'Market ID'},
   inplace=True)
```

```
market_dim['Market'] = market_dim['Market
   ID'].replace({'US': 'United States', 'APAC': 'Asia-
   Pacific', 'EU': 'Europe', 'EMEA': 'Europe, Middle
   East and Africa', 'LATAM': 'Latin America'})

market_dim['Market ID'] = market_dim['Market
   ID'].replace({'Africa': 'AFR', 'Canada': 'CAN'})

location_dim.rename(columns={'Market': 'Market ID'},
   inplace=True)

location_dim['Market ID'] = location_dim['Market
   ID'].replace({'Africa': 'AFR', 'Canada': 'CAN'})

market_dim = market_dim.drop_duplicates(subset=['Market
   ID']).reset_index(drop=True)
```

	Market ID	Market
0	US	United States
1	APAC	Asia-Pacific
2	EU	Europe
3	AFR	Africa
4	EMEA	Europe, Middle East and Africa
5	LATAM	Latin America
6	CAN	Canada

- Mô tả:
  - Market ID: Khóa chính của các thị trường, gồm 6 giá trị duy nhất
  - Market: Tên của các thị trường hoạt động

3.2.6. Chuyển đổi Customer_Dim:

- Xử lý: Chọn các thuộc tính thuộc về khách hàng, xóa các giá trị trùng lặp của CustomerID vì đã có khóa có trong bảng ban đầu.

```
customer_dim = df[['Customer ID', 'Customer Name',
   'Segment']]
```

```
customer_dim = customer_dim.drop_duplicates('Customer
    ID')

customer_dim.reset_index(drop=True,inplace=True)
```

3.2.7 Chuyển đổi Segment_Dim:

- Xử lý: Chọn thuộc tính Segment và xóa các giá trị trùng lặp, sau đó tạo khóa bằng viết tắt của Segment.

```
segment_dim = customer_dim[['Segment']]

segment_dim['Segment ID'] =
    segment_dim['Segment'].replace({"Consumer": "CONSR",
    "Corporate": "CORP", "Home Office": "H-OFFICE"})

segment_dim.drop_duplicates('Segment ID', inplace=True)

segment_dim
```

	Segment	Segment ID
0	Consumer	CONSR
1	Corporate	CORP
3	Home Office	H-OFFICE

- Mô tả:
  - Segment: Tên các loại phân khúc khách hàng
  - Segment ID: Khóa chính của các loại phân khúc khách hàng

** Tái Chuyển đổi cột Customer_Dim:*

- Xử lý: Chèn Khóa ngoại Segment ID vào bảng Customer ID:

```
customer_dim['Segment'] =
    customer_dim['Segment'].replace({"Consumer": "CONSR",
    "Corporate": "CORP", "Home Office": "H-OFFICE"})

customer_dim.rename(columns={"Segment": "Segment ID"},
```

```
    inplace = True)
```

	Customer ID	Customer Name	Segment ID
0	RH-19495	Rick Hansen	CONSR
1	JR-16210	Justin Ritter	CORP
2	CR-12730	Craig Reiter	CONSR
3	KM-16375	Katherine Murray	H-OFFICE
4	RH-9495	Rick Hansen	CONSR
...	...	...	...
1585	SC-10800	Stuart Calhoun	CONSR
1586	BD-1500	Bradley Drucker	CONSR
1587	RC-9825	Roy Collins	CONSR
1588	MG-7890	Michael Granlund	H-OFFICE
1589	ZC-11910	Zuschuss Carroll	CONSR

- Mô tả:
    - Customer ID: ID của từng khách hàng và là Khóa chính của bảng Customer_Dim có ý nghĩa cho từng tên khách hàng, gồm 1589 giá trị duy nhất.
    - Customer Name: Tên khách hàng
    - Segment ID: Khóa ngoại và liên kết bảng Segment_Dim

3.2.8 Chuyển đổi Order_Dim:

- Xử lý: Chọn các thuộc tính thuộc về order/ship, xóa các giá trị trùng lặp theo Order ID - đã có sẵn trong bộ dữ liệu ban đầu, sau đó chuyển đổi các giá trị thời gian về dạng datetime để đồng nhất dữ liệu.

```
order_dim = df[['Order ID', 'Order Date', 'Ship Date',
    'Ship Mode']]

order_dim = order_dim.drop_duplicates('Order
    ID').reset_index(drop=True)

order_dim['Order Date'] = order_dim['Order
    Date'].str.replace('-', '/')

order_dim['Order Date'] =
    pd.to_datetime(order_dim['Order Date'],
```

```
    format='%d/%m/%Y')

order_dim['Ship Date'] = order_dim['Ship
    Date'].str.replace('-', '/')

order_dim['Ship Date'] = pd.to_datetime(order_dim['Ship
    Date'], format='%d/%m/%Y')
```

	Order ID	Order Date	Ship Date	Ship Mode
0	CA-2012-124891	2012-07-31	2012-07-31	Same Day
1	IN-2013-77878	2013-02-05	2013-02-07	Second Class
2	IN-2013-71249	2013-10-17	2013-10-18	First Class
3	ES-2013-1579342	2013-01-28	2013-01-30	First Class
4	SG-2013-4320	2013-11-05	2013-11-06	Same Day
...	...	...	...	...
25030	ZI-2011-4350	2011-03-21	2011-03-26	Standard Class
25031	MX-2014-169530	2014-06-09	2014-06-11	First Class
25032	IN-2014-72327	2014-05-30	2014-05-30	Same Day
25033	IN-2014-57662	2014-08-05	2014-08-10	Standard Class
25034	MX-2012-134460	2012-05-22	2012-05-26	Second Class

3.2.9 Chuyển đổi Ordtime_Dim và Shiptime_Dim:

- Xử lý: chọn các thuộc tính về order và ship, xóa các giá trị trùng lặp ở hai bảng theo "Order Date" và "Ship Date" và sắp xếp theo giá trị tăng dần sau đó tạo khóa cho Ordtime_Dim theo dạng chuỗi: "ORD-" và Shiptime_Dim: "SHIP-" với ngày và hai số cuối của năm

```
ordtime_dim = order_dim[['Order Date']]

shiptime_dim = order_dim[['Ship Date']]

ordtime_dim = ordtime_dim.drop_duplicates('Order Date')

ordtime_dim = ordtime_dim.sort_values('Order Date',
    ascending = True)

ordtime_dim.reset_index(drop='True', inplace=True)
```

```
shiptime_dim = shiptime_dim.drop_duplicates('Ship Date')

shiptime_dim = shiptime_dim.sort_values('Ship Date',
   ascending = True)

shiptime_dim.reset_index(drop='True', inplace=True)



ordtime_dim['Order Date ID'] = 'ORD-' +
   ordtime_dim['Order
   Date'].dt.strftime('%m').str.zfill(2) +
   ordtime_dim['Order
   Date'].dt.strftime('%d').str.zfill(2) +
   ordtime_dim['Order Date'].dt.strftime('%y')

shiptime_dim['Ship Date ID'] = 'SHIP' + '-' +
   shiptime_dim['Ship
   Date'].dt.strftime('%m').str.zfill(2) +
   shiptime_dim['Ship
   Date'].dt.strftime('%d').str.zfill(2) +
   shiptime_dim['Ship Date'].dt.strftime('%y')
```

	Order Date	Order Date ID
0	2011-01-01	ORD-010111
1	2011-01-02	ORD-010211
2	2011-01-03	ORD-010311
3	2011-01-04	ORD-010411
4	2011-01-05	ORD-010511
...	...	...
1423	2014-12-27	ORD-122714
1424	2014-12-28	ORD-122814
1425	2014-12-29	ORD-122914
1426	2014-12-30	ORD-123014
1427	2014-12-31	ORD-123114

	Ship Date	Ship Date ID
0	2011-01-03	SHIP-010311
1	2011-01-05	SHIP-010511
2	2011-01-06	SHIP-010611
3	2011-01-07	SHIP-010711
4	2011-01-08	SHIP-010811
...	...	...
1459	2015-01-03	SHIP-010315
1460	2015-01-04	SHIP-010415
1461	2015-01-05	SHIP-010515
1462	2015-01-06	SHIP-010615
1463	2015-01-07	SHIP-010715

3.2.10 Xử lí Month_Dim:

- Xử lý: Xóa các giá trị trùng lặp của tháng và tạo khóa chính cho là chữ viết tắt của tháng

```
month_val = ordtime_dim['Order
   Date'].dt.strftime('%b').drop_duplicates().reset_inde
   x(drop=True)

month_dim = pd.DataFrame(ordtime_dim['Order
   Date'].dt.strftime("%m").drop_duplicates().reset_inde
   x(drop=True))

month_dim['MonthID'] = month_val

month_dim.rename(columns={'Order Date': 'Month'},
   inplace = True)

month_dim = month_dim[['MonthID', 'Month']]

month_dim['Month'] = month_dim['Month'].astype(int)

month_dim
```

	MonthID	Month
0	Jan	1
1	Feb	2
2	Mar	3
3	Apr	4
4	May	5
5	Jun	6
6	Jul	7
7	Aug	8
8	Sep	9
9	Oct	10
10	Nov	11
11	Dec	12

- Mô tả:
    - MonthID: Khóa chính của tháng và có ý nghĩa cho các số tháng, gồm 12 tháng.

- Month: Tháng theo số

3.2.11 Xử lí Year_Dim:

- Xử lý: Chọn các giá trị năm, xóa các giá trị trùng lặp của năm, và xem đây là khóa chính vì sẽ không có ý nghĩa khi tạo thêm một khóa chính cho năm.

```
year_dim = pd.DataFrame("20" + shiptime_dim['Ship
   Date'].dt.strftime("%y").drop_duplicates().reset_inde
   x(drop=True))

year_dim.rename(columns={'Ship Date': 'Year'}, inplace =
   True)
```

- Mô tả:
  - Year: các năm hiện hữu từ năm 2011 tới 2015

* *Tái chuyển đổi bảng Ordtime_Dim và Shiptime_Dim:*

- Xử lý: Thêm các giá trị tháng và năm vào trong hai bảng

```
ordtime_dim['Year'] = ordtime_dim['Order Date'].dt.year

shiptime_dim['Year'] = shiptime_dim['Ship Date'].dt.year



ordtime_dim['Month'] = ordtime_dim['Order
   Date'].dt.month

shiptime_dim['Month'] = shiptime_dim['Ship
   Date'].dt.month
```

	Order Date	Order Date ID	Year	Month
0	2011-01-01	ORD-1111	2011	1
1	2011-01-02	ORD-1211	2011	1
2	2011-01-03	ORD-1311	2011	1
3	2011-01-04	ORD-1411	2011	1
4	2011-01-05	ORD-1511	2011	1
...	...	...	...	...
1423	2014-12-27	ORD-122714	2014	12
1424	2014-12-28	ORD-122814	2014	12
1425	2014-12-29	ORD-122914	2014	12
1426	2014-12-30	ORD-12314	2014	12
1427	2014-12-31	ORD-123114	2014	12

	Ship Date	Ship Date ID	Year	Month
0	2011-01-03	SHIP-1311	2011	1
1	2011-01-05	SHIP-1511	2011	1
2	2011-01-06	SHIP-1611	2011	1
3	2011-01-07	SHIP-1711	2011	1
4	2011-01-08	SHIP-1811	2011	1
...	...	...	...	...
1459	2015-01-03	SHIP-1315	2015	1
1460	2015-01-04	SHIP-1415	2015	1
1461	2015-01-05	SHIP-1515	2015	1
1462	2015-01-06	SHIP-1615	2015	1
1463	2015-01-07	SHIP-1715	2015	1

- Mô tả:
  - Order Date: Ngày đặt hàng
  - Order Date ID: Khóa chính cho bảng Ordtime_Dim, gồm 1427 giá trị duy nhất cho bảng Ordtime_Dim
  - Ship Date: Ngày giao hàng
  - Ship Date ID: Khóa chính cho bảng Shiptime_Dim, gồm 1463 giá duy nhất cho bảng Shiptime_Dim
  - Year: Năm hiện hữu
  - Month: Tháng của năm hiện hữu

** Tái chuyển đổi bảng Order_Dim:*

- Xử lý: sử dụng các hàm map để thêm các khóa chính vào bảng và xóa cột order date và ship date

```
order_dim['Ship Date ID']  = order_dim['Ship
   Date'].map(dict(zip(shiptime_dim['Ship Date'],
   shiptime_dim['Ship Date ID'])))

order_dim['Order Date ID']  = order_dim['Order
   Date'].map(dict(zip(ordtime_dim['Order Date'],
   ordtime_dim['Order Date ID'])))



order_dim = order_dim.drop(columns=['Order Date', 'Ship
```

```
    Date'])
```

	Order ID	Ship Mode	Ship Date ID	Order Date ID
0	CA-2012-124891	Same Day	SHIP-73112	ORD-73112
1	IN-2013-77878	Second Class	SHIP-2713	ORD-2513
2	IN-2013-71249	First Class	SHIP-11813	ORD-11713
3	ES-2013-1579342	First Class	SHIP-1313	ORD-12813
4	SG-2013-4320	Same Day	SHIP-11613	ORD-11513
...	...	...	...	...
25030	ZI-2011-4350	Standard Class	SHIP-32611	ORD-32111
25031	MX-2014-169530	First Class	SHIP-61114	ORD-6914
25032	IN-2014-72327	Same Day	SHIP-5314	ORD-5314
25033	IN-2014-57662	Standard Class	SHIP-8114	ORD-8514
25034	MX-2012-134460	Second Class	SHIP-52612	ORD-52212

- Mô tả:
  - Order ID: Khóa chính và là có ý nghĩa cho mỗi đơn đặt hàng duy nhất, gồm 25034 dòng
  - Ship Mode: Hình thức giao
  - Order Date ID: Khóa ngoại và liên kết về bảng Ordtime_Dim
  - Ship Date ID: Khóa ngoại và liên kết về bảng Shiptime_Dim

3.2.12 Chuyển đổi bảng Order_Priority_Dim:

- Xử lý: Chọn thuộc tính Order Priority, xóa các giá trị trùng lặp và tạo khóa chính cho loại ưu tiên đặt hàng bằng chữ cái đầu tiên

```
order_priority_dim = df[['Order Priority']]

order_priority_dim =
    order_priority_dim.drop_duplicates().reset_index(drop
    =True)

order_priority_dim['Order Priority ID'] =
    order_priority_dim['Order Priority'].apply(lambda x:
```

```
    x[0].upper())
```

	Order Priority	Order Priority ID
0	Critical	C
1	Medium	M
2	High	H
3	Low	L

- Mô tả:
    - Order Priority: Thứ tự ưu tiên của đơn hàng
    - Order ID: Khóa chinh cho thứ tự ưu tiên đơn hàng, gồm 4 giá trị duy nhất

3.3. Giai đoạn 3: Nạp dữ liệu vào kho dữ liệu (Load)

3.3.1. Giới thiệu các bảng Dim

Các bảng được chia theo sơ đồ Snowflake, gồm 5 nhánh chính và 13 bảng (Dim) và một bảng Fact

- Nhánh 1 về các thông tin của sản phẩm, hạng mục và hạng mục con: Product_Dim, Category_Dim, Sub_Category_Dim
- Nhánh 2 về các thông tin về đỉa điểm và thị trường: Location_Dim và Market_Dim
- Nhánh 3 về các thông tin về đặt và giao hàng: Order_Dim, Ordtime_Dim, Shiptime_Dim, Month_Dim, Year_Dim
- Nhánh 4 chỉ có thông tin về thứ tự ưu tiên đơn hàng: Order_Priority_Dim
- Nhánh 5 chứa các thông tin về khách hàng và phân khúc khách hàng: Customer_Dim, Segment_Dim

3.3.2. Giới thiệu bảng Fact

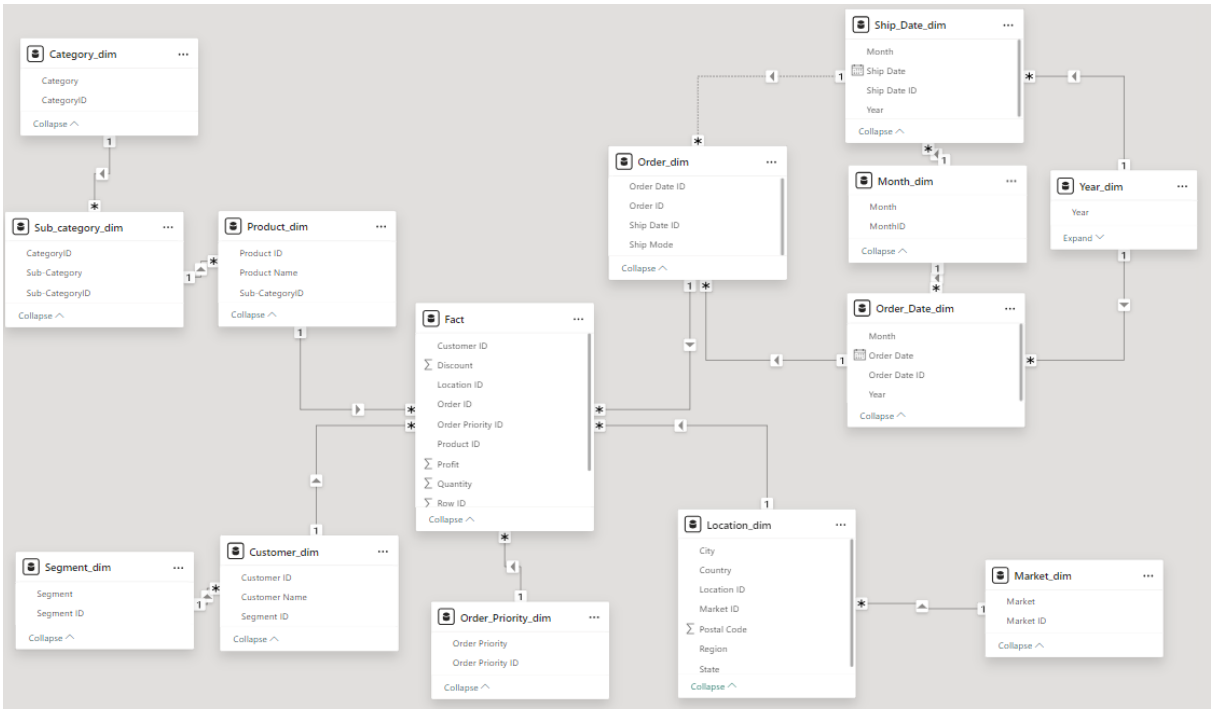
Bảng Fact bao gồm các thông số kinh tế: Profit, Sales, Discount, Shipping Cost và chứa các tất cả thông tin của các quan hệ bảng được liên kết bởi các khóa phụ.

	Row ID	Order ID	Customer ID	Product ID	Location ID	Order Priority ID	Sales	Quantity	Discount	Profit	Shipping Cost
0	32298	CA-2012-124891	RH-19495	TEC-AC-10003033	NYC-US-2377	C	2309.650	7	0.0	762.1845	933.57
1	26341	IN-2013-77878	JR-16210	FUR-CH-10003950	WOL-APAC-3651	C	3709.395	9	0.1	-288.7650	923.63
2	25330	IN-2013-71249	CR-12730	TEC-PH-10004664	BRI-APAC-508	M	5175.171	9	0.1	919.9710	915.49
3	13524	ES-2013-1579342	KM-16375	TEC-PH-10004583	BER-EU-384	M	2892.510	5	0.1	-96.5400	910.16
4	47221	SG-2013-4320	RH-9495	TEC-SHA-10000501	DAK-AFR-884	C	2832.960	8	0.0	311.5200	903.04
...	...	...	...	...	...	...	...	...	...	...	...
51285	29002	IN-2014-62366	KE-16420	OFF-FA-10000746	KUR-APAC-1784	M	65.100	5	0.0	4.5000	0.01
51286	35398	US-2014-102288	ZC-21910	OFF-AP-10002906	HOU-US-1453	M	0.444	1	0.8	-1.1100	0.01
51287	40470	US-2013-155768	LB-16795	OFF-EN-10001219	OXN-US-2501	H	22.920	3	0.0	11.2308	0.01
51288	9596	MX-2012-140767	RB-19795	OFF-BI-10000806	VAL-LATAM-3474	M	13.440	2	0.0	2.4000	0.00
51289	6147	MX-2012-134460	MC-18100	OFF-PA-10004155	TIP-LATAM-3350	H	61.380	3	0.0	1.8000	0.00

Mô tả:

- Row ID: ID của từng hàng được định nghĩa trong dataset ban đầu, để đảm bảo tính vẹn nguyên của dữ liệu thì việc loại hay không bỏ sẽ được thi hành trong việc phân tích.
- Order ID: ID của từng đơn hàng (khóa chính của Order_dim)
- Customer ID: ID của từng khách hàng (khóa chính của Customer_dim)
- Product ID: ID của từng loại sản phẩm (khóa chính của Product_dim)
- Location ID: ID của từng địa điểm (khóa chính của Location_dim)
- Order Priority ID: ID cho từng thứ tự ưu tiên của đơn hàng (khóa chính của Order_Priority_dim)
- Sales: Doanh thu
- Quantity: Số lượng sản phẩm bán được
- Discount: Phần trăm chiết khấu
- Profit: Lợi Nhuận
- Shipping Cost: Chi phí vận chuyển

3.3.3. Mô hình dữ liệu (Data Model)



CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU

4.1. Phân tích tổng quan tình hình doanh nghiệp (Dashboard Overview)

4.1.1. Tổng quan:

Dashboard overview của Global Superstore cung cấp một cái nhìn toàn diện về các khía cạnh khác nhau của hiệu suất kinh doanh. Nó hiển thị tổng doanh thu, tổng lợi nhuận, số lượng đơn hàng, doanh thu và lợi nhuận cho từng mặt hàng, doanh thu và lợi nhuận cho từng mặt hàng, doanh thu theo khu vực, phân phối doanh thu theo phương thức vận chuyển và doanh thu cho các phân khúc khách hàng khác nhau.

4.1.2. Mục đích của Dashboard:

Mục đích của dashboard này là để theo dõi và phân tích hiệu quả hoạt động của doanh nghiệp qua các năm. Nó cho phép nhìn nhận nhanh các yếu tố quan trọng như doanh thu, lợi nhuận, và xu hướng thị trường, giúp ban lãnh đạo có thể đưa ra quyết định dựa trên dữ liệu.

4.1.3. Phân tích các chỉ số:

Doanh số và lợi nhuận qua các năm: Đồ thị cho thấy sự tăng trưởng liên tục trong doanh số từ năm 2011 đến 2014, điều này cho thấy sự phát triển tích cực và bền vững của doanh nghiệp.

Phân tích lợi nhuận theo phân khúc: Phân khúc 'Consumer' chiếm tỷ lệ lớn nhất trong tổng lợi nhuận, cho thấy sự ưu tiên và hiệu quả trong việc phục vụ khách hàng cá nhân.

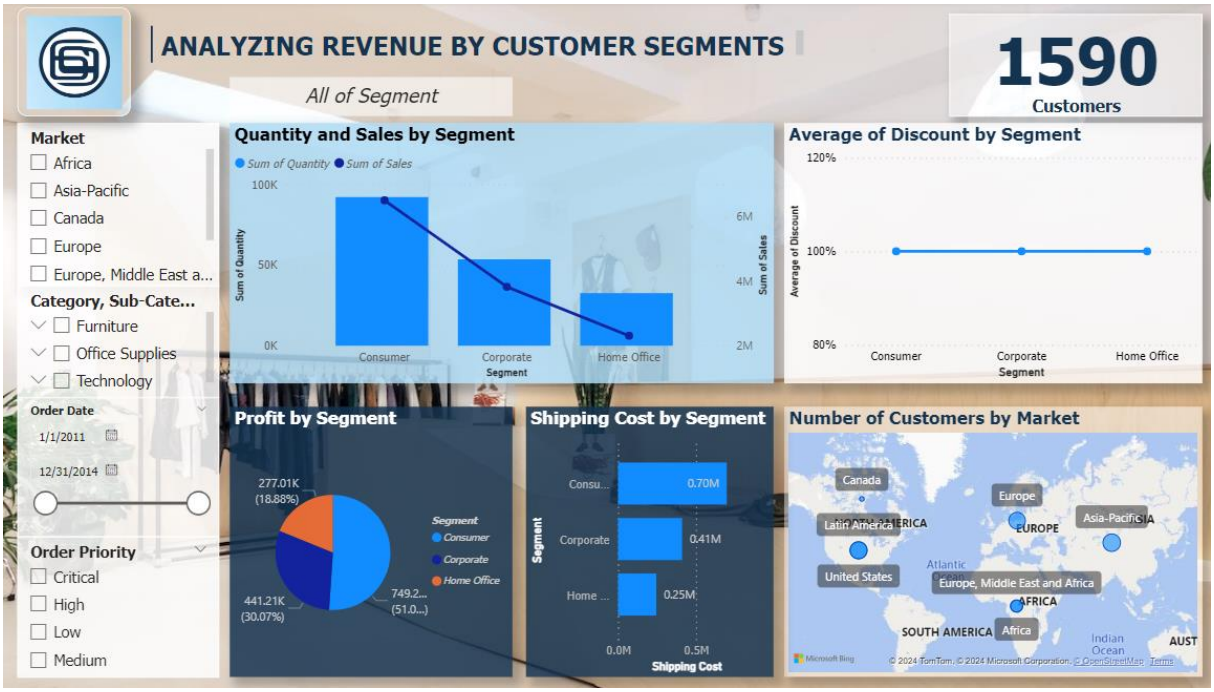
Doanh số theo độ ưu tiên của đơn hàng: Đơn hàng được xếp hạng ưu tiên 'Medium' và 'High' có tổng doanh số cao, phản ánh nhu cầu cao và sự quan trọng của việc xử lý nhanh chóng các đơn hàng này.

Doanh số và lợi nhuận theo danh mục sản phẩm: Các sản phẩm công nghệ mang lại doanh thu cao nhất nhưng mục 'Office Supplies' lại cho thấy sự gia tăng lợi nhuận đột biến, có lẽ do chiến lược giá hoặc quản lý chi phí tốt hơn.

Doanh số theo khu vực địa lý: Doanh số đặc biệt mạnh ở Bắc Mỹ và châu Âu, cho thấy hai thị trường này là trọng điểm của doanh nghiệp.

Kết luận: Công ty đã có một năm tăng trưởng mạnh mẽ, với doanh thu và lợi nhuận tăng lên. Mặc dù vậy, có sự khác biệt đáng kể giữa các quý và các phân khúc sản phẩm cũng như khu vực, điều này cần được nghiên cứu kỹ lưỡng để tối ưu hóa chiến lược kinh doanh. Việc ưu tiên đơn hàng và phương thức vận chuyển cho thấy tiềm năng quản lý chi phí. Cuối cùng, việc tập trung vào phân khúc khách hàng "Consumer" đã mang lại lợi nhuận đáng kể, nhưng cũng cần xem xét các cơ hội để đa dạng hóa và duy trì lợi nhuận trong các phân khúc khác.

4.2. Phân tích doanh thu theo phân khúc khách hàng



4.2.1.Mục đích:

Phân tích doanh thu, lợi nhuận và các chi phí khác liên quan đến các phân khúc khách hàng khác nhau, để hiểu rõ nhu cầu và hành vi mua sắm của từng nhóm.

4.2.2. Tổng quan:

Dashboard này cung cấp thông tin chi tiết về số lượng bán hàng, lợi nhuận và chi phí vận chuyển cho mỗi phân khúc khách hàng, giúp xác

định phân khúc nào là lợi nhuận nhất và chiếm tỷ trọng lớn trong doanh thu.Phân tích các chỉ số chính:

4.2.3. Phân tích về doanh thu theo phân khúc khách hàng

Số lượng và doanh số:

Biểu đồ: Cho thấy số lượng và doanh số bán hàng chia theo các phân khúc khách hàng (Consumer, Corporate, Home Office).

Nhận xét: Số lượng bán hàng cao nhất thuộc về phân khúc 'Consumer', nhưng có sự giảm dần về doanh số khi chuyển từ phân khúc này sang 'Corporate' và 'Home Office'. Điều này có thể cho thấy mức độ ưu tiên và hiệu quả kinh doanh khác nhau giữa các phân khúc.

Doanh thu:

Biểu đồ hình tròn: Hiển thị tỷ lệ phần trăm lợi nhuận từ mỗi phân khúc.

Nhận xét:Phân khúc 'Home Office' tạo ra lợi nhuận cao nhất (51%), mặc dù số lượng bán hàng không phải là cao nhất. Điều này có thể cho thấy tính hiệu quả của phân khúc này về mặt chuyển đổi doanh số thành lợi nhuận.

Chi phí vận chuyển:

Biểu đồ cột: Thể hiện chi phí vận chuyển cho mỗi phân khúc.

Nhận xét: Chi phí vận chuyển thấp nhất cho phân khúc 'Home Office', tuy nhiên đây là phân khúc tạo ra lợi nhuận cao nhất, cho thấy sự quản lý chi phí hiệu quả.

Trung bình khuyến mãi:

Biểu đồ đường: Phản ánh mức trung bình của chiết khấu được cung cấp cho mỗi phân khúc khách hàng.

Nhận xét: Các phân khúc khách hàng nhận được mức chiết khấu tương đương nhau. Điều này cho thấy sự nhất quán trong chính sách giá của doanh nghiệp đối với các phân khúc khác nhau.

Số lượng khách hàng:

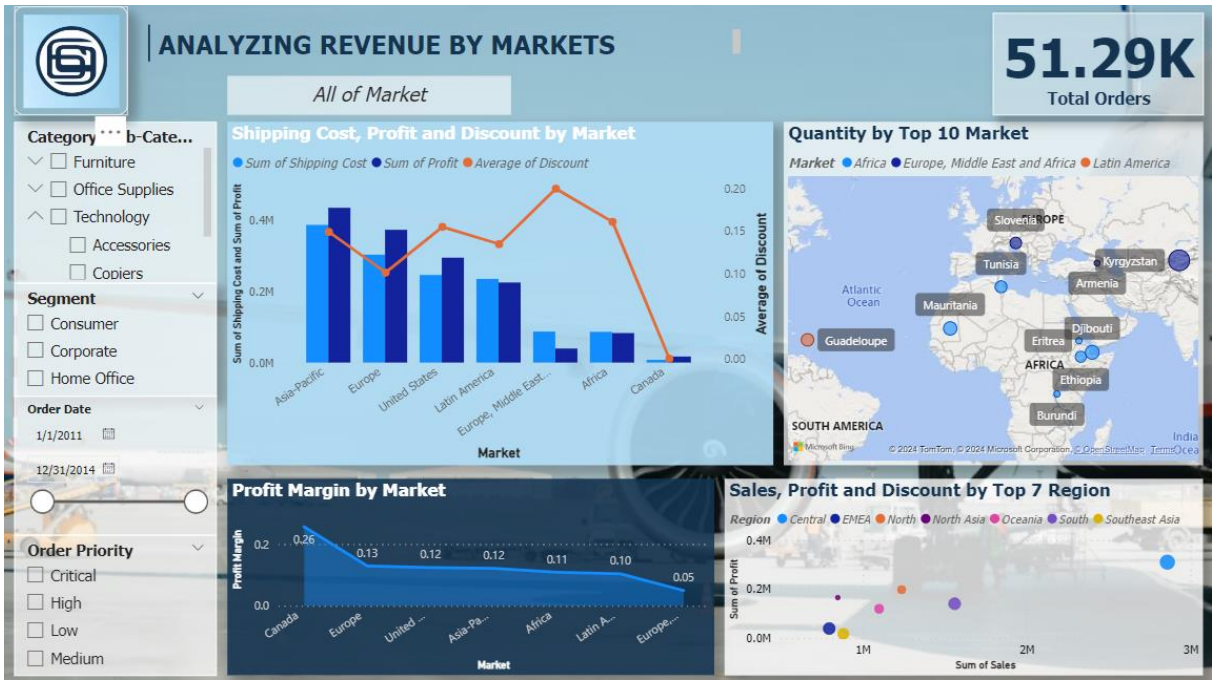
Biểu đồ bản đồ: Hiển thị số lượng khách hàng tại từng khu vực địa lý.

Nhận xét: Phần lớn khách hàng tập trung ở Bắc Mỹ và châu Âu, chỉ ra hai thị trường chính của Global Superstore. Khu vực Latinh Mỹ và châu Á - Thái Bình Dương cũng có một lượng khách hàng đáng kể.

Kết luận:

Dashboard này cung cấp cái nhìn sâu sắc về cách thức doanh nghiệp phân bổ nguồn lực và tạo ra doanh thu từ các phân khúc khách hàng khác nhau. Nổi bật là hiệu quả lợi nhuận tốt ở phân khúc 'Home Office', đòi hỏi cần sự chú trọng đặc biệt để mở rộng hoặc cải thiện các phân khúc khác. Đồng thời, sự quản lý chi phí tốt ở phân khúc này có thể là điểm mấu chốt để tăng cường hiệu quả kinh doanh chung.

4.3. Phân tích doanh thu theo thị trường



4.3.1. Mục đích:

Đánh giá hiệu quả kinh doanh ở các thị trường khác nhau, từ đó phát triển chiến lược tiếp cận phù hợp cho mỗi khu vực.

4.3.2. Tổng quan:

Dashboard này biểu diễn tổng doanh thu, lợi nhuận và chi phí vận chuyển tại các thị trường quốc tế, bao gồm cả mức độ chiết khấu áp dụng, để định hướng cho các quyết định mở rộng hoặc điều chỉnh kinh doanh.

4.3.3. Phân tích về doanh thu theo thị trường

Chi phí vận chuyển, lợi nhuận và khuyến mãi:

Biểu đồ cột và đường: Thể hiện chi phí vận chuyển, lợi nhuận và mức chiết khấu trung bình tại các thị trường khác nhau.

Nhận xét:

- Chi phí vận chuyển và lợi nhuận: Cao nhất ở châu Á-Thái Bình Dương và châu Âu. Điều này có thể do khoảng cách vận chuyển lớn và khối lượng đơn hàng cao.

- Mức chiết khấu: Giảm dần từ châu Á-Thái Bình Dương sang Canada. Mức chiết khấu cao có thể liên quan đến chiến lược thâm nhập thị trường hoặc cạnh tranh cao.

Sản lượng top 10:

Bản đồ: Hiển thị lượng hàng hóa được bán ra tại 10 thị trường hàng đầu.

Nhận xét: Các thị trường lớn như Ethiopia và Tunisia hiển thị số lượng giao dịch nhiều, cho thấy sự hiện diện và thành công của Global Superstore tại các thị trường này.

Tỷ suất lợi nhuận:

Biểu đồ cột: Cho thấy tỷ suất lợi nhuận tại các thị trường khác nhau.

Nhận xét: Canada có tỷ suất lợi nhuận cao nhất (0.26), trong khi các thị trường như châu Á-Thái Bình Dương và châu Âu có tỷ suất lợi nhuận ở mức trung bình. Điều này có thể phản ánh hiệu quả quản lý chi phí hoặc giá cả cạnh tranh tại từng thị trường.

Doanh thu, lợi nhuận, và chiết khấu tại 7 khu vực chính**:**

Biểu đồ scatter: Biểu diễn doanh thu, lợi nhuận, và chiết khấu tại 7 khu vực chính.

Nhận xét:

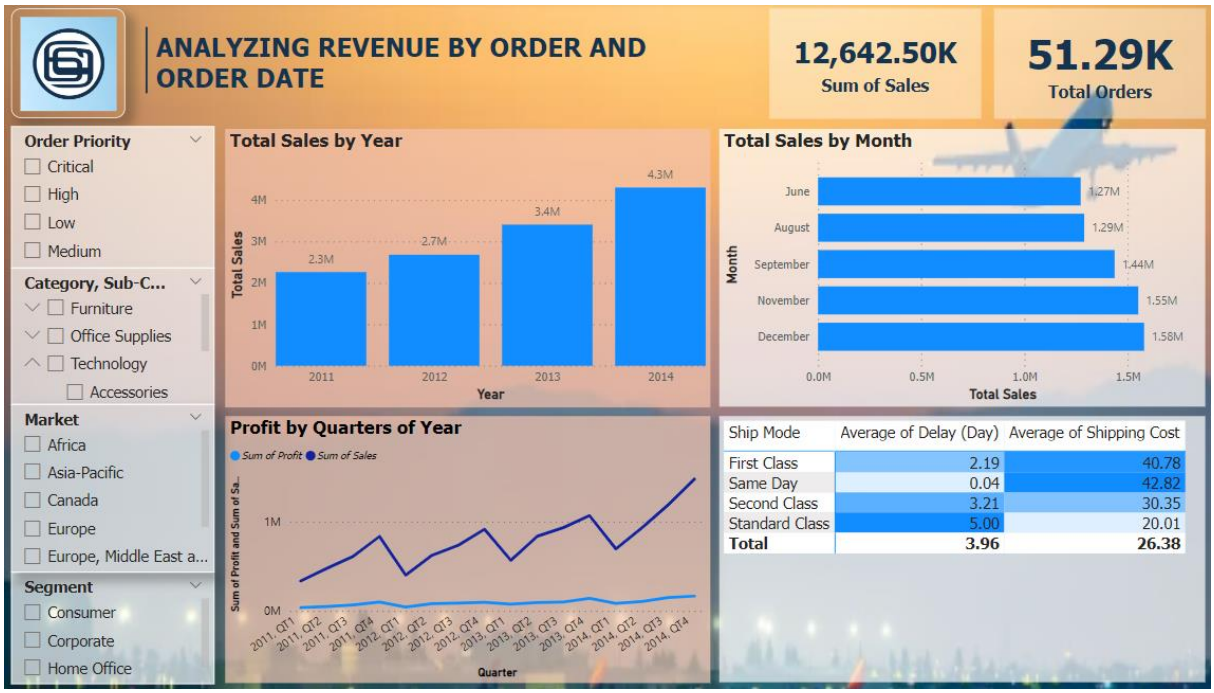
- Doanh thu: Cao nhất tại khu vực Đông Nam Á, tương ứng với lượng hàng bán ra lớn.

- Lợi nhuận và chiết khấu: Có sự khác biệt đáng kể giữa các khu vực, phản ánh chiến lược giá và khuyến mãi khác nhau. Chẳng hạn, mặc dù doanh thu cao nhưng lợi nhuận tại Đông Nam Á không phải là cao nhất, có thể do mức chiết khấu cao.

Kết luận:

Dashboard này cung cấp cái nhìn toàn diện về hiệu quả kinh doanh theo địa lý, từ đó cho phép Global Superstore hiểu rõ về các điểm mạnh và điểm yếu tại từng thị trường cụ thể. Phân tích này giúp định hướng chiến lược phát triển thị trường, quản lý chi phí và xác định các cơ hội tăng trưởng. Chiến lược chiết khấu và vận chuyển cần được điều chỉnh cho phù hợp với từng thị trường để tối ưu hóa lợi nhuận.

4.4. Phân tích doanh thu theo đơn đặt hàng



4.4.1. Mục đích:

Phân tích tổng quan về doanh thu qua các năm dựa trên các đơn đặt hàng để hiểu xu hướng phát triển chung của công ty.

4.4.2. Tổng quan:

Dashboard này hiển thị sự thay đổi trong tổng doanh số, lợi nhuận và số lượng đơn đặt hàng theo từng năm, cung cấp cái nhìn sâu sắc về mức độ tăng trưởng và các yếu tố ảnh hưởng đến sự thay đổi này.

4.4.3. Phân tích về doanh thu theo đơn đặt hàng:

Doanh số:

Biểu đồ cột: Cho thấy doanh số bán hàng theo từng năm từ 2011 đến 2014.

Nhận xét: Có sự tăng trưởng đều đặn qua các năm, từ 2.3 triệu đô la năm 2011 lên đến 4.3 triệu đô la năm 2014, cho thấy sự phát triển mạnh mẽ và tăng trưởng ổn định của doanh nghiệp.

Lợi nhuận và doanh số các quý:

Biểu đồ đường: Thể hiện lợi nhuận và doanh số qua các quý.

Nhận xét: Lợi nhuận có xu hướng tăng theo từng quý, với những đỉnh điểm và thấp điểm rõ rệt qua các năm. Doanh thu cao nhất ở Quý 4 hằng năm, cụ thể hơn, doanh số của Global Superstore đạt đỉnh điểm ở Quý 4 năm 2014, thấp nhất ở Quý 1 năm 2011, các năm khác cùng thời điểm cũng ở thấp điểm. Điều này có thể phản ánh các chiến dịch khuyến mãi thành công, đặc biệt là mùa lễ cuối năm nhu cầu của người phương tây tăng cao.

Doanh thu hàng tháng:

Biểu đồ cột: Hiển thị doanh số bán hàng theo từng tháng.

Nhận xét: Doanh số tăng đáng kể vào các tháng cuối năm, đặc biệt là tháng 11 và 12, phù hợp với mùa mua sắm nghỉ lễ. Điều này giúp xác định rõ mùa cao điểm để tối ưu hóa chiến lược bán hàng và quản lý hàng tồn kho

Phương thức vận chuyển:

Bảng dữ liệu: Cho thấy tổng số ngày giao hàng và trung bình chi phí giao theo từng phương thức vận chuyển (First Class, Same Day, Second Class, Standard Class).

Nhận xét:

- Standard Class có số lượng đơn hàng và tổng số ngày giao hàng cao nhất cũng như chi phí giao hàng thấp nhất. Và cũng dễ hiểu điều này dẫn tới Tiêu chuẩn được ưa chuộng nhất do chi phí hợp lý và lượng khách hàng sử dụng nhiều.

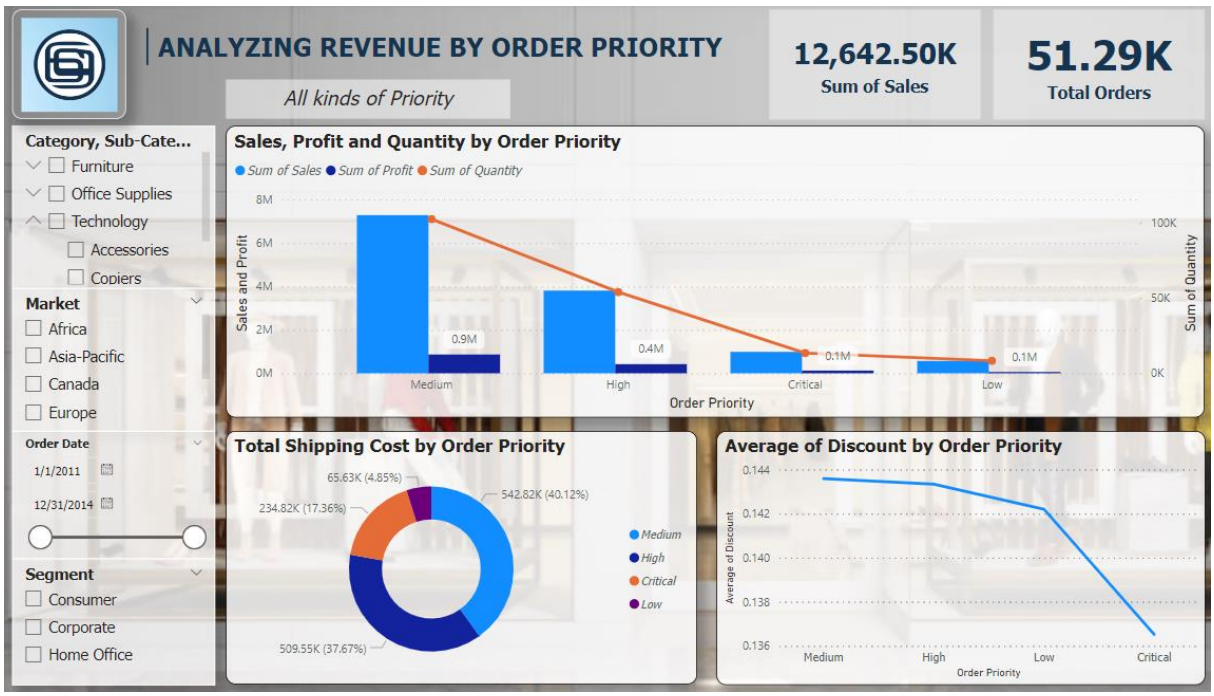
- Same Day (Hỏa tốc) và First Class có số ngày giao ít hơn, chi phí giao cũng cao nhất, rõ ràng đây là 2 giá trị nghịch với nhau. Và số đơn hàng cũng là thấp hơn đáng kể. Vì những phương thức vận chuyển này

có chi phí cao và khách hàng cũng quan tâm tới ví tiền của mình nhiều hơn là thời gian giao hàng.

Kết luận:

Dashboard này cung cấp cái nhìn toàn diện về mô hình doanh thu của Global Superstore theo năm và theo các quý, cũng như ảnh hưởng của các phương thức vận chuyển đến lợi nhuận và số lượng đơn hàng. Sự phát triển đều qua các năm và lợi nhuận tăng theo từng quý cho thấy sự hiệu quả của chiến lược kinh doanh hiện tại. Tuy nhiên, việc phân tích sâu hơn vào chi phí vận chuyển và lợi nhuận của từng phương thức vận chuyển có thể giúp tối ưu hóa chi phí và cải thiện lợi nhuận tổng thể.

4.5. Phân tích doanh thu theo độ ưu tiên đơn hàng



4.5.1. Mục đích:

- Đánh giá ảnh hưởng của mức độ ưu tiên đơn hàng đến doanh thu, lợi nhuận và chi phí vận chuyển, nhằm cải thiện hiệu quả xử lý đơn hàng.

4.5.2. Tổng quan:

Dashboard này cung cấp thông tin về cách mức độ ưu tiên của đơn hàng ảnh hưởng đến kết quả kinh doanh, từ doanh thu đến chi phí vận chuyển và mức chiết khấu, giúp tối ưu hoá các quyết định về xếp độ ưu tiên và xử lý đơn hàng.

4.5.3. Phân tích về doanh thu theo độ ưu tiên đơn hàng

Doanh thu, lợi nhuận và số lượng:

Biểu đồ cột và đường: Hiển thị tổng doanh số, lợi nhuận và số lượng sản phẩm theo mức độ ưu tiên của đơn hàng (Medium, High, Critical, Low).

Nhận xét:

- Doanh số và lợi nhuận cao nhất ở mức độ ưu tiên Medium, điều này cho thấy đa số đơn hàng được thực hiện với mức độ ưu tiên này, có lẽ do sự cân bằng giữa chi phí và tốc độ giao hàng.

- Đơn hàng Critical và Low có doanh số và lợi nhuận thấp, phản ánh số lượng đơn hàng thấp hoặc ít được ưu tiên.

Tổng chi phí của từng mức độ:

Biểu đồ tròn: Thể hiện tổng chi phí vận chuyển cho từng mức độ ưu tiên đơn hàng.

Nhận xét:

- Medium priority chiếm phần lớn chi phí vận chuyển, tương ứng với số lượng đơn hàng và doanh số cao nhất.

- High và Critical priorities có chi phí thấp hơn nhiều, do số lượng đơn hàng ít hơn điều này cũng dễ hiểu vì những đơn hàng này khá đặc biệt nên sẽ không chiếm ưu thế so với những đơn hàng bình thường.

Mức chiết khấu của mỗi mức độ:

Biểu đồ đường: Biểu thị mức chiết khấu trung bình cho mỗi mức độ ưu tiên

Nhận xét:

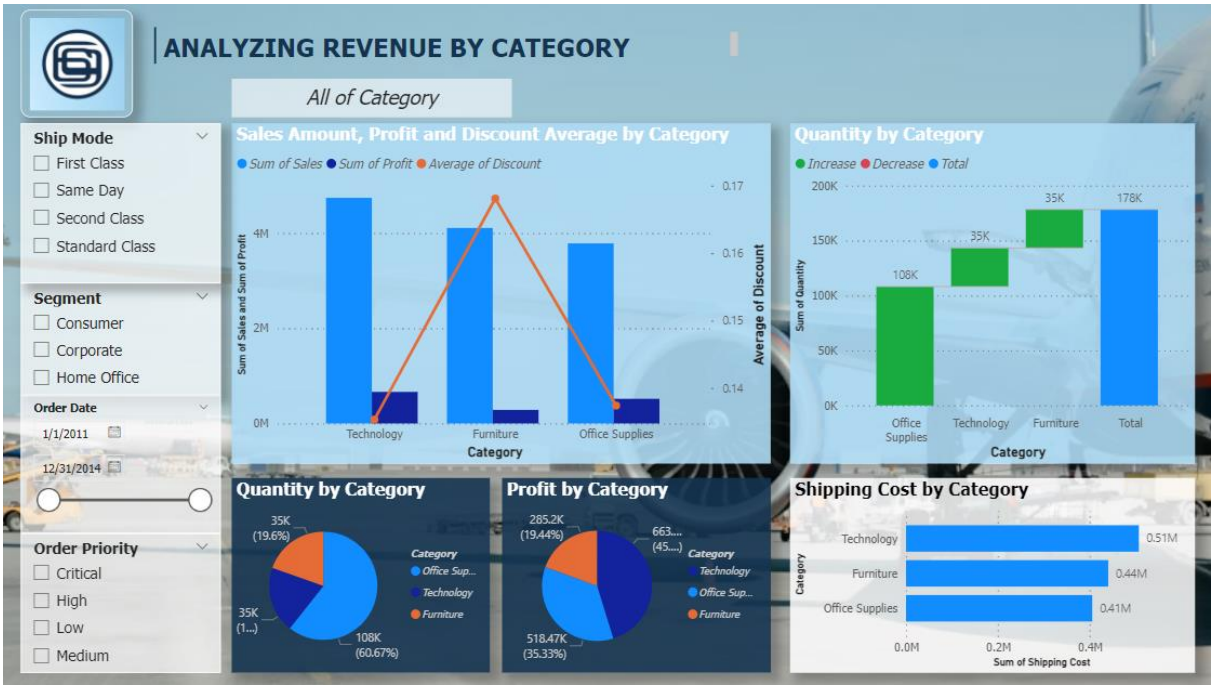
- Giảm dần từ Medium đến High cho thấy mức chiết khấu cao hơn cho các đơn hàng không quá khẩn cấp, có lẽ để khuyến khích đặt hàng trước hoặc lựa chọn giao hàng không gấp.

- Mức độ Low và Critical có chiết khấu tương tự nhau, có thể là để xử lý các đơn hàng dư thừa hoặc không đủ khẩn cấp.

Kết luận

Dashboard này cung cấp cái nhìn chi tiết về cách thức phân bổ doanh thu, chi phí vận chuyển, và mức chiết khấu dựa trên độ ưu tiên của đơn hàng tại Global Superstore. Dữ liệu này hỗ trợ cho việc đưa ra quyết định chiến lược về cách thức xử lý và ưu tiên các đơn hàng để tối ưu hóa lợi nhuận và hiệu quả vận hành. Sự hiểu biết về mối liên hệ giữa các mức độ ưu tiên với chi phí và lợi nhuận có thể giúp cải thiện chất lượng dịch vụ và sự hài lòng của khách hàng.

4.6. Phân tích doanh thu theo danh mục sản phẩm



4.6.1. Mục đích:

Phân tích hiệu quả kinh doanh dựa trên các danh mục sản phẩm khác nhau, từ đó xác định các cơ hội và thách thức trong việc quản lý hàng hóa, chiết khấu, và chi phí vận chuyển.

4.6.2. Tổng quan:

Dashboard này cung cấp cái nhìn toàn diện về cách thức các danh mục sản phẩm khác nhau ảnh hưởng đến doanh số, lợi nhuận và chi phí vận chuyển. Thông tin này hỗ trợ việc đưa ra quyết định về việc điều chỉnh giá, chương trình khuyến mãi, và chiến lược phân phối, nhằm tối ưu hóa lợi nhuận và hiệu quả kinh doanh.

4.6.3. Phân tích về doanh thu theo danh mục sản phẩm

Doanh thu, lợi nhuận và mức khuyến mãi trung bình:

Biểu đồ cột và đường: Hiển thị doanh số, lợi nhuận và mức chiết khấu trung bình cho mỗi danh mục sản phẩm (Technology, Furniture, Office Supplies).

Nhận xét:

- Technology: Có doanh số cao nhất và lợi nhuận cao, nhưng cũng có mức chiết khấu cao nhất.
- Furniture: Lợi nhuận cao nhất nhưng doanh số thấp hơn so với Technology.
- Office Supplies: Doanh số và lợi nhuận thấp hơn các danh mục khác, chiết khấu tương đối thấp, phản ánh sự khác biệt về giá trị và chiến lược giá.

Số lượng bán ra:

Biểu đồ cột: Cho thấy số lượng sản phẩm bán ra cho mỗi danh mục.

Nhận xét:

- Furniture: Mặc dù lợi nhuận cao nhưng số lượng bán ra ít nhất, cho thấy giá trị trung bình cao cho mỗi sản phẩm.
- Office Supplies: Số lượng bán ra lớn nhất, nhưng tổng lợi nhuận thấp, chỉ ra rằng đây là các mặt hàng giá rẻ, lợi nhuận thấp.

Lợi nhuận:

Biểu đồ tròn: Thể hiện tỷ lệ phần trăm lợi nhuận mỗi danh mục đóng góp vào tổng lợi nhuận.

Nhận xét:

- Furniture: Chiếm tỷ lệ lợi nhuận cao (hơn 50%), cho thấy mặc dù số lượng bán ra không cao nhưng mỗi sản phẩm đóng góp lợi nhuận đáng kể.

Chi phí vận chuyển:

Biểu đồ cột: Hiển thị tổng chi phí vận chuyển cho mỗi danh mục.

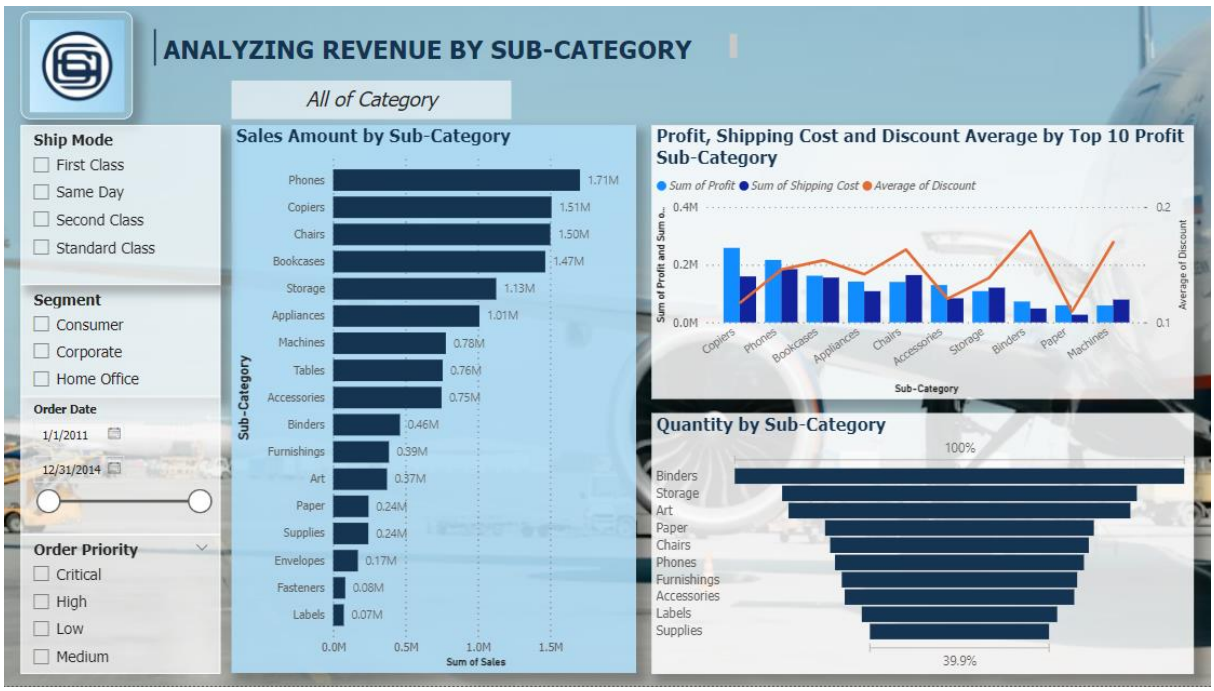
Nhận xét:

- Technology: Chi phí vận chuyển cao nhất, có thể do yêu cầu đóng gói đặc biệt hoặc vận chuyển cẩn thận hơn.
- Furniture: Chi phí vận chuyển tiếp theo cao, phù hợp với kích thước lớn và khối lượng của các sản phẩm này.
- Office Supplies: Chi phí vận chuyển thấp nhất, phản ánh trọng lượng và kích thước nhỏ hơn của các mặt hàng này.

Kết luận:

Dashboard này cung cấp cái nhìn toàn diện về ảnh hưởng của các danh mục sản phẩm đến doanh thu, lợi nhuận, và chi phí vận chuyển của Global Superstore. Sự phân tích này giúp nhận diện được các danh mục có hiệu quả kinh doanh cao và các danh mục cần được cải thiện về mặt chiến lược và quản lý. Cụ thể, danh mục Technology, mặc dù đem lại doanh số lớn nhưng chi phí chiết khấu và vận chuyển cũng cao, đòi hỏi sự điều chỉnh về giá và chi phí để tăng lợi nhuận. Trong khi đó, Furniture, mặc dù bán ít nhưng lại có tỷ lệ lợi nhuận cao, cho thấy tiềm năng tốt cho việc mở rộng. Office Supplies, mặc dù có số lượng bán ra cao, nhưng lợi nhuận thấp hơn nhiều, cần chiến lược giá và khuyến mãi hiệu quả hơn.

4.7. Phân tích doanh thu theo danh mục phụ.



4.7.1. Mục đích:

Dashboard này được thiết kế để phân tích cụ thể hiệu quả kinh doanh của từng phân mục sản phẩm trong danh mục Công nghệ của Global Superstore. Mục đích chính là xác định doanh số, lợi nhuận, chi phí vận chuyển và mức độ ưu đãi cho từng loại sản phẩm như Điện thoại, Máy in, Phụ kiện và Máy móc. Phân tích này giúp nhận diện các sản phẩm chủ chốt góp phần vào doanh thu cao cũng như các sản phẩm có chi phí cao hoặc mức chiết khấu đặc biệt, từ đó hỗ trợ việc ra quyết định chiến lược về sản phẩm và marketing.

4.7.2. Tổng quan:

Dashboard này cung cấp một cái nhìn toàn diện về hiệu quả từng phân khúc sản phẩm trong danh mục Công nghệ, từ đó cho thấy sự phân bổ nguồn lực và ảnh hưởng đến kết quả kinh doanh chung của công ty. Ngoài ra, dashboard này có thực hiện chức năng drill through với *Category* của bảng *Category_dim*, nhằm liên hệ với dashboard ngay trước nó có tên *"category"* Thông qua việc phân tích doanh số, lợi nhuận, chi phí vận chuyển và mức chiết khấu của từng loại sản phẩm, dashboard giúp:

- Đánh giá hiệu quả của chiến lược giá và khuyến mãi.

- Nhận diện các cơ hội tối ưu hóa chi phí và tăng cường lợi nhuận.
- Điều chỉnh chiến lược phân phối dựa trên chi phí vận chuyển và mức độ ưa chuộng của sản phẩm.

Kết hợp những hiểu biết này, Global Superstore có thể cải thiện quy trình hoạch định và thực thi chiến lược để đảm bảo tăng trưởng doanh thu bền vững và tối ưu hóa lợi nhuận trong môi trường cạnh tranh.

4.7.3. Phân tích về doanh thu theo danh mục con trong Công nghệ

Doanh số bán hàng theo danh mục phụ:

- Biểu đồ cột: Thể hiện doanh số bán hàng cho các phân mục Điện thoại, Máy in, Phụ kiện và Máy móc.
- Nhận xét:
- Điện thoại: Đạt doanh số cao nhất, lên tới 1.71 triệu đô la, cho thấy sức mua mạnh mẽ trong phân khúc này.
- Máy in: Cũng có doanh số khá cao, 1.51 triệu đô la, phản ánh nhu cầu cao đối với các sản phẩm này trong môi trường văn phòng.
- Máy móc và Phụ kiện: Có doanh số thấp hơn nhiều so với hai nhóm trên, cho thấy đây có thể là các sản phẩm bổ trợ hoặc có giá trị cao nhưng số lượng bán ra ít.

Lợi nhuận, chi phí vận chuyển và khuyến mãi:

- Biểu đồ cột và đường: Hiển thị lợi nhuận, chi phí vận chuyển và mức chiết khấu trung bình cho mỗi phân mục.
- Nhận xét:
- Copiers (Máy in): Có lợi nhuận cao và chi phí vận chuyển thấp, dù mức chiết khấu cao, cho thấy hiệu quả cao từ việc bán các sản phẩm này.
- Phones (Điện thoại) và Accessories (Phụ kiện): Mặc dù điện thoại có doanh số cao nhưng mức chiết khấu cao có thể ảnh hưởng đến tổng lợi nhuận thu được.

Số lượng bán ra:

- Biểu đồ cột: Thể hiện số lượng bán ra cho mỗi phân mục.
- Nhận xét:
- Máy in: Chiếm tỷ lệ lớn về số lượng bán ra, điều này cùng với lợi nhuận cao cho thấy đây là sản phẩm kinh doanh chính lực.

- Điện thoại và Phụ kiện: Số lượng bán ra ít hơn nhưng doanh số cao cho thấy giá bán trên mỗi đơn vị cao.

Kết luận

Dashboard này cung cấp cái nhìn chi tiết về hiệu quả kinh doanh của từng phân mục sản phẩm trong danh mục Công nghệ. Mặc dù Điện thoại đạt doanh số cao nhất, nhưng Máy in lại mang lại lợi nhuận cao với chi phí vận chuyển thấp, cho thấy sự cân bằng giữa doanh số và lợi nhuận là cần thiết. Các kết quả này có thể giúp Global Superstore điều chỉnh chiến lược sản phẩm và tiếp thị để tối ưu hóa lợi nhuận và đáp ứng nhu cầu thị trường một cách hiệu quả hơn.

CHƯƠNG 5: ĐỀ XUẤT GIẢI PHÁP

5.1. Hạn chế

Bộ dữ liệu gặp phải một số hạn chế nhất định. Thứ nhất, nó chỉ bao gồm các đơn hàng trực tuyến trong khoảng thời gian từ 2011 đến 2015, dẫn đến việc không thể đưa ra các kết luận về hiệu suất kinh doanh của doanh nghiệp trong các giai đoạn gần đây. Thứ hai, bộ dữ liệu chỉ chứa thông tin về các đơn hàng đã hoàn thành, thiếu thông tin về các đơn hàng đang chờ xử lý hoặc đã bị hủy, điều này có thể làm hạn chế trong việc đánh giá toàn diện về quy trình kinh doanh của doanh nghiệp.

5.2. Đề xuất cải tiến

Thực hiện cập nhật bộ dữ liệu, mở rộng phạm vi thời gian bộ dữ liệu đến thời điểm gần nhất, giúp phân tích được thực trạng phản ánh chính xác nhất về tình hình kinh doanh hiện tại của doanh nghiệp.

Thực hiện các chiến dịch tiếp thị, ghi nhận phản hồi của khách hàng, luôn cập nhật xu hướng thị trường mới nhất. Việc này giúp tạo ra các phân tích có ý nghĩa hơn, cải thiện hiệu suất hoạt động kinh doanh.

Xây dựng mô hình dự báo, mô hình dữ báo doanh thu để dự đoán xu hướng và doanh số bán hàng trong tương lai. Mô hình giúp doanh nghiệp trong việc lập kế hoạch kinh doanh và dự đoán nguồn lực cần thiết.

Phân tích chi tiết về hành vi khách hàng để hiểu rõ hơn về nhu cầu và mong muốn của họ. Thông tin này có thể hữu ích phát triển các chiến lược tiếp thị và tùy chỉnh sản phẩm, từ đó tăng cơ hội bán hàng và tăng doanh thu.

CHƯƠNG 6: KẾT LUẬN

6.1. Kết luận

Mặc dù bộ dữ liệu vẫn còn thiếu xót, nhất là chưa thể thu thập được số liệu từ năm 2015 trở đi, nhưng bài phân tích vẫn đủ cơ sở để đưa ra cái nhìn rõ ràng hơn về tình hình của doanh nghiệp. Qua việc phân tích các dashboard khác nhau, từ danh mục sản phẩm, đơn hàng theo ngày, phân khúc khách hàng, đến khu vực địa lý, chúng ta đã thu được những hiểu biết sâu sắc về hoạt động kinh doanh của Global Superstore. Các sản phẩm công nghệ, đặc biệt là điện thoại và máy in, liên tục dẫn đầu về doanh thu và lợi nhuận, nhưng cũng đi kèm với chi phí vận chuyển cao và chiết khấu đáng kể. Điều này chỉ ra rằng mặc dù đây là nguồn doanh thu chính nhưng cũng cần quản lý chi phí một một cách phù hợp. Ngoài ra doanh thu tăng mạnh vào các tháng cuối năm, điều này phù hợp với mùa mua sắm lễ hội. Thị trường Bắc Mỹ và Châu Âu là những khu vực có doanh thu cao nhất, cho thấy tầm quan trọng của việc tập trung vào những khu vực này để đẩy mạnh doanh số. Phương thức vận chuyển có ảnh hưởng đáng kể đến chi phí và sự hài lòng của khách hàng, với các dịch vụ giao hàng Tiêu chuẩn dĩ nhiên chiếm số nhiều bởi vì khách hàng họ vẫn ưu tiên tiết kiệm chi phí giao hàng.

6.2. Hướng phát triển

- Tối ưu hóa chi phí và hiệu suất cho sản phẩm công nghệ: Đối với các sản phẩm công nghệ như điện thoại và máy in, cần xem xét lại chi phí vận chuyển và chiết khấu để đảm bảo rằng lợi nhuận vẫn được duy trì, ngoài ra cũng kéo theo việc giữ chân khách hàng cũ.
- Việc mở các cửa hàng bán lẻ hoặc phân phối cho các đại lý lớn nhỏ cũng là cách để giảm thiểu chi phí vận chuyển và làm tăng tỷ lệ tiếp cận của khách hàng đến với sản phẩm. Có thể tham khảo các chiến lược về cửa hàng vật lý của Amazon và Walmart để có thể đem lại sự tiện ích cho phân khúc khách hàng khó tiếp cận với thị trường trực tuyến
- Tận dụng cơ hội trong thời gian mua sắm lễ hội: Tận dụng các chiến dịch tiếp thị và khuyến mãi vào các tháng cuối năm để tăng cường doanh số và lợi nhuận. Ngoài ra việc tham gia vào hoạt động mua sắm lễ hội cũng là

cơ hội để doanh nghiệp tăng cường nhận thức thương hiệu. Từ đó gián tiếp tăng doanh thu ở các thời điểm khác trong năm.

- Tăng cường tiếp thị và tọa nhiều ưu đãi cho các khách hàng có mức độ ưu tiên cao. Bởi đây là nguồn khách có lợi nhuận trung bình ở mức cao nhưng vẫn chưa thể tối ưu hóa được số lượng đơn hàng từ họ.
- Tập trung làm Marketing thương hiệu vào thị trường Bắc Mỹ và Châu Âu: Tập trung vào phát triển kinh doanh và tiếp thị tại thị trường Bắc Mỹ và Châu Âu, nơi có doanh thu cao nhất. Bắc Mỹ và Châu Âu cũng là hai trong những trung tâm công nghệ hàng đầu thế giới, với sự phát triển nhanh chóng của ngành công nghiệp công nghệ thông tin và viễn thông. Ngoài ra vấn đề cạnh tranh cũng khốc liệt nhưng vẫn lành mạnh và tạo động lực cho doanh nghiệp nâng cao chất lượng sản phẩm và dịch vụ. Ngoài ra cũng cần đảm bảo việc phân phối sản phẩm đến các thị trường trọng điểm một cách hợp lý để đảm bảo mạng lưới logistics bền vững.

--- HẾT ---

TRÍCH DẪN

Accuracy vs. precision vs. recall in machine learning: what's the difference? (n.d.). Evidently AI. Retrieved December 10, 2023, from https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall

Accuracy vs. precision vs. recall in machine learning: what's the difference? (n.d.). Evidently AI. Retrieved December 10, 2023, from https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall

Loukas, S. (2020, October 12). *Text Classification Using Naive Bayes: Theory & A Working Example*. Towards Data Science. Retrieved December 10, 2023, from https://towardsdatascience.com/text-classification-using-naive-bayes-theory-a-working-example-2ef4b7eb7d5a

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019, October 2). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv*. https://arxiv.org/abs/1910.01108

Sharma, N. (2023, June 10). *Understanding and Applying F1 Score: A Deep Dive with Hands-On Coding*. Arize AI. Retrieved December 10, 2023, from https://arize.com/blog-course/f1-score/

Smolic, H. (2022, September 16). *Precision Versus Recall - Essential Metrics in Machine Learning*. Graphite Note. Retrieved December 10, 2023, from https://graphite-note.com/precision-versus-recall-machine-learning

Stanford NLP. (2009, 7 4). *Properties of Naive Bayes*. Stanford NLP Group. Retrieved December 10, 2023, from https://nlp.stanford.edu/IR-book/html/htmledition/properties-of-naive-bayes-1.html

Vaswani. (2017, June 12). [1706.03762] Attention Is All You Need. *arXiv*. https://arxiv.org/abs/1706.03762