BAN CƠ YẾU CHÍNH PHỦ

HỌC VIỆN KỸ THUẬT MẬT MÃ

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

ĐỀ TÀI

NGHIÊN CỨU, XÂY DỰNG ỨNG DỤNG PHÂN TÍCH DỮ LIỆU SỬ DỤNG POWER BI

Sinh viên thực hiện: Nông Thị Sâm

Khóa: CT4

Chuyên ngành: Kỹ thuật phần mềm nhúng và di động

Mã ngành: 748.02.01

Người hướng dẫn: TS. Phạm Văn Hưởng

Hà Nội - 2024

BAN CƠ YẾU CHÍNH PHỦ

HỌC VIỆN KỸ THUẬT MẬT MÃ

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC

ĐỀ TÀI

NGHIÊN CỨU, XÂY DỰNG ỨNG DỤNG PHÂN TÍCH DỮ LIỆU SỬ DỤNG POWER BI

Sinh viên thực hiện: Nông Thị Sâm

Khóa: CT4

Chuyên ngành: Kỹ thuật phần mềm nhúng và di động

Mã ngành: 748.02.01

Người hướng dẫn: TS. Phạm Văn Hưởng

Hà Nội - 2024

LỜI CẢM ƠN

Trong quá trình thực hiện đồ án tốt nghiệp này, tôi đã nhận được nhiều sự giúp đỡ, đóng góp ý kiến và sự dẫn dắt chỉ bảo nhiệt tình của thầy cô, gia đình và bạn bè.

Lời đầu tiên, tôi xin gửi lời cảm ơn chân thành nhất đến cán bộ hướng dẫn là TS. Phạm Văn Hưởng – Phó chủ nhiệm khoa Công nghệ thông tin Học viện Kỹ thuật Mật mã, người đã tận tình hướng dẫn, chỉ bảo tôi trong suốt thời gian qua.

Tôi cũng xin gửi cảm ơn tới các thầy, cô giáo trong trường Học viện Kỹ thuật Mật mã nói chung, các thầy cô trong Bộ môn Công nghệ thông tin nói riêng đã truyền đạt kiến thức về các môn đại cương cũng như các môn chuyên ngành, giúp tôi có được cơ sở lý thuyết vững vàng và tạo điều kiện giúp trong suốt quá trình tham gia học tập.

Cuối cùng, tôi xin chân thành cảm ơn gia đình và bạn bè, những người thân yêu đã luôn ở bên cạnh, động viên và hỗ trợ tôi trong quá trình hoàn thành đồ án tốt nghiệp này. Với điều kiện thời gian cũng như kinh nghiệm còn hạn chế, bài báo cáo đồ án này không thể tránh khỏi những thiếu sót. Tôi rất mong nhận được đóng góp ý kiến của các thầy cô để tôi có điều kiện bổ sung, nâng cao ý thức của mình, phục vụ tốt hơn công tác thực tế sau này.

Tôi xin chân thành cảm ơn!

Hà Nội, ngày 07 tháng 06 năm 2024

Sinh viên thực hiện

(Ký tên và ghi rõ họ tên)

Sâm

Nông Thị Sâm

LỜI CAM ĐOAN

Tôi xin cam đoan bản đồ án này do tôi tự nghiên cứu dưới sự hướng dẫn của cán bộ hướng dẫn TS. Phạm Văn Hưởng.

Để hoàn thành đồ án này, tôi chỉ sử dụng những tài liệu đã ghi trong mục tài liệu tham khảo, ngoài ra không sử dụng bất cứ tài liệu nào khác mà không được ghi.

Nếu sai, tôi xin chịu mọi hình thức kỷ luật theo quy định của Học viện.

Hà Nội, ngày 07 tháng 06 năm 2024

Sinh viên thực hiện

(Ký tên và ghi rõ họ tên)

Sâm

Nông Thị Sâm

MỤC LỤC

LỜI CẢM ƠN i
LỜI CAM ĐOAN ii
MỤC LỤC iii
DANH MỤC BẢNG BIỂU vi
DANH MỤC HÌNH VẼ vii
LỜI MỞ ĐẦU 1
Chương 1 4
TỔNG QUAN PHÂN TÍCH DỮ LIỆU VÀ POWER BI 4
1.1. Phân tích dữ liệu trong ngành bán lẻ 4
1.1.1. Ứng dụng phân tích dữ liệu trong bán lẻ ***8***
1.1.2. Quy trình phân tích dữ liệu 10
1.1.3. Xu hướng và thách thức 15
1.2. Phát biểu bài toán 16
1.3. Giải pháp và công nghệ sử dụng 17
1.3.1. Giải pháp phân tích nhúng 18
1.3.2 Giải pháp phân tích bài toán bán lẻ 22
1.3.3 Giải pháp bảo mật và nâng cao hiệu xuất xử lý dữ liệu 32
1.3.4 Công nghệ sử dụng 34
1.4. Power BI 34
1.4.1. Giới thiệu công cụ Power BI 34
1.4.2. Thành phần kiến trúc Power BI 35
1.4.3. Mô hình hóa dữ liệu trong Power Bi 36
1.5. Tổng kết chương 37
Chương 2 38
PHÂN TÍCH THIẾT KẾ HỆ THỐNG 38
2.1. Đặc tả yêu cầu 38
2.1.1. Yêu cầu chức năng 38
2.1.2. Yêu cầu phi chức năng 38
2.1.3. Tổng quan hệ thống 39
2.2. Use case 41

2.2.1. Đặc tả use case đăng nhập 41
2.2.2. Đặc tả use case xem báo cáo 42
2.2.3. Đặc tả use case tạo báo cáo 42
2.2.4. Đặc tả use case chỉnh sửa và lưu báo cáo 42
2.3. Sequence/Activity 43
2.3.1 Sequence chức năng đăng nhập 43
2.3.2 Sequence chức năng xem báo cáo 44
2.3.3 Sequence chức năng tạo báo cáo 44
2.3.4 Sequence chức năng chỉnh sửa và lưu báo cáo 45
2.4. Tổng kết chương 45
Chương 3 46
XÂY DỰNG VÀ THỬ NGHIỆM HỆ THỐNG 46
3.1. Xây dựng ứng dụng web MVC 46
3.1.1 Tạo ứng dụng web MVC với xác thực Azure AD 46
3.1.2 Tạo dự án .NET5 cho ứng dụng web MVC 47
3.1.3 Triển khai giải pháp bảo mật và hiệu xuất xử lý dữ liệu 48
3.2. Tạo báo cáo Power BI 51
3.2.1. Chuẩn bị dữ liệu 51
3.2.2. Tạo mô hình dữ liệu 52
3.2.3 Tính toán các chỉ số 56
3.2.4 Trực quan hóa dữ liệu 66
3.3. Nhúng báo cáo Power BI vào ứng dụng web MVC 67
3.4. Thực nghiệm và đánh giá kết quả 69
3.4.1. Thực nghiệm 69
3.4.2 Đánh giá kết quả 74
3.5. Tổng kết chương 76
KẾT LUẬN 77
TÀI LIỆU THAM KHẢO 78
PHỤ LỤC 1

DANH KÝ HIỆU VÀ TỪ VIẾT TẮT

Từ viết tắt	Nghĩa tiếng anh	Nghĩa tiếng việt
Azure	Microsoft Azure	Nền tảng cloud của Microsoft
BI	Business Intelligence	Thông Tin Kinh Doanh
CSS	Cascading Style Sheets	Ngôn ngữ thiết kế
CSV	Comma-Separated Values	Giá trị Phân tách bằng Dấu phẩy
DAX	Data Analysis Expressions	Biểu thức Phân tích Dữ liệu
ERP	Enterprise Resource Planning	Quy hoạch Nguồn lực tổ chức
ETL	Extract, Transform, Load	Trích Xuất, Chuyển Đổi, Tải
FTP	File Transfer Protocol	Giao thức Truyền tập tin
JSON	JavaScript Object Notation	Định dạng Đối tượng JavaScript
KPI	Key Performance Indicator	Chỉ số Hiệu suất Chính
LAN	Local Area Network	Mạng cục bộ
OLAP	Online Analytical Processing	Xử lý Phân tích Trực tuyến
OLTP	Online Transaction Processing	Xử lý Giao dịch Trực tuyến
ROI	Return on Investment	Lợi nhuận Đầu tư
SaaS	Software as a Service	Phần mềm dưới dạng dịch vụ
SQL	Structured Query Language	Ngôn ngữ Truy vấn Cấu trúc
UI	User Interface	Giao diện Người dùng
UX	User Experience	Trải nghiệm Người dùng
VPN	Virtual Private Network	Mạng riêng ảo
WAN	Wide Area Network	Mạng diện rộng
XML	Extensible Markup Language	Ngôn ngữ đánh dấu mở rộng

DANH MỤC BẢNG BIỂU

Bảng 1. 1 So sánh 2 phương pháp phân tích nhúng 20
Bảng 1. 2 Ví dụ chia phân vị 27
Bảng 2. 1 Đặc tả use case đăng nhập 41
Bảng 2. 2 Bảng đặc tả use case xem báo cáo 42
Bảng 2. 3 Đặc tả use case tạo báo cáo 42
Bảng 2. 4 Đặc tả use case chỉnh sửa và lưu báo cáo 42
Bảng 3. 1 Ý nghĩa thành phần trong bảng Fact và Dimension 53
Bảng 3. 2 Mối quan hệ trong mô hình dữ liệu Power BI 55
Bảng 3. 3 Tính toán các chỉ số của trang tổng quan 57
Bảng 3. 4 Thành phần trong bảng RFM 59
Bảng 3. 5 Phân khúc hành vi khách hàng 63
Bảng 3. 6 Các chỉ số trang phân tích bán chéo 64

DANH MỤC HÌNH VẼ

Hình 1. 1 Quy trình CRISP DM 11
Hình 2. 1 Quy trình hoạt động phương pháp phân tích nhúng 39
Hình 2. 2 Biểu đồ ca sử dụng 41
Hình 2. 3 Sequence chức năng đăng nhập 43
Hình 2. 4 Sequence chức năng xem báo cáo 44
Hình 2. 5 Sequence chức năng tạo báo cáo 44
Hình 2. 6 Sequence chức năng chỉnh sửa và lưu báo cáo 45
Hình 3. 1 Tạo nhóm bảo mật trong Azure AD 47
Hình 3. 2 Tạo dự án .NET5 cho ứng dụng MVC với xác thực Azure AD 48
Hình 3. 3 Thư viện Microsoft.Identity.Web được thêm vào ứng dụng 48
Hình 3. 4 Cấu hình token và Power BI Service 49
Hình 3. 5 Đặt bảo mật áp dụng cho nhóm cụ thể 49
Hình 3. 6 Xử lý dữ liệu xử dụng Python trong jupyter notebook 50
Hình 3. 7 Kết nối SQL Server database với Power BI Desktop 51
Hình 3. 8 Ví dụ các thành phần trong mô hình dữ liệu 53
Hình 3. 9 Mô hình dữ liệu 54
Hình 3. 10 Tính chỉ số Recency, Frequency, Monetary 59
Hình 3. 11 Tính điểm cột Recency 60
Hình 3. 12 Tính điểm cột Frequency 61
Hình 3. 13 Tính điểm cột Monetory 62
Hình 3. 14 Nhóm điểm khách hàng 62
Hình 3. 15 Xếp hạng điểm phân khúc 64
Hình 3. 16 Trực quan dữ liệu trang báo cáo tổng quan 66
Hình 3. 17 Trực quan hóa dữ liệu trang phân khúc khách hàng 67
Hình 3. 18 Trực quan hóa trang phân tích bán chéo 67
Hình 3. 19 Trang chủ sau khi đăng nhập thành công 68
Hình 3. 20 Thực nghiệm chức năng chỉnh sửa và lưu báo cáo 70
Hình 3. 21 Thực nghiệm chức năng tạo báo cáo mới 71
Hình 3. 22 Xem 20% KH tốt nhất đem lại bao nhiêu % tổng doanh thu 72
Hình 3. 23 Xem 80% doanh thu đang tới từ nhóm phân khúc nào 73
Hình 3. 24 Xem số lượng khách hàng phân khúc Loyal 73
Hình 3. 25 Thao tác bộ lọc 74
Hình 3. 26 Chức năng xem báo cáo 75

LỜI MỞ ĐẦU

Dưới sự phát triển bùng nổ của cuộc cách mạng công nghệ 4.0, dữ liệu lớn (Big Data) là một trong bốn nền tảng quan trọng nhất cùng với IoT (Internet of Things), Trí tuệ nhân tạo (Artificial Intelligence), Blockchain (Chuỗi khối). Big Data được hiểu là những dữ liệu khổng lồ, là nguồn thông tin có dung lượng lớn và đa dạng, có vận tốc cao, đòi hỏi các hình thức xử lý thông tin có hiệu quả về chi phí, để nâng cao việc đưa ra quyết định và tối ưu hóa quy trình. Nói cách khác, Big Data là một tệp dữ liệu khổng lồ không thể phân tích được bằng các công cụ và phần mềm thông thường.

"*Sự bùng nổ của dữ liệu đang khiến nhiều doanh nghiệp trong nước phải đối mặt với hàng loạt hạn chế trong việc xử lý thông tin, dẫn đến khả năng đưa ra các quyết định bị chậm trễ hoặc trì hoãn. Điều này về lâu dài chắc chắn sẽ ảnh hưởng không nhỏ đến hiệu quả của các doanh nghiệp*." Ông Nguyễn Kim Hòa, Giám đốc Công nghệ và Hệ thống IBM Việt Nam cho biết.

Hầu hết các doanh nghiệp, tổ chức gần như đều sử dụng nguồn dữ liệu lớn phân tích để tìm ra câu trả lời cho các câu hỏi liên quan đến việc giảm chi phí, giảm thời gian, phát triển sản phẩm mới, dịch vụ tối ưu và ra quyết định thông minh. Khi việc phân tích nguồn dữ liệu lớn được hỗ trợ tối đa, doanh nghiệp có thể xác định nguyên nhân gốc rễ của những quyết định hoặc chiến lược thất bại, từ đó tạo các chương trình khuyến mại hợp lý dựa trên thói quen của khách hàng trong lĩnh vực kinh doanh.

Là lĩnh vực đi đầu trong việc ứng dụng phân tích dữ liệu lớn. Phân tích dữ liệu kinh doanh (Business Data Analytics) là ngành có tính liên ngành giữa công nghệ thông tin và kinh tế. Công việc tập trung vào việc thu thập, khai thác, quản lý và xử lý bộ dữ liệu lớn, từ đó đưa ra các nhận định, dự đoán xu hướng hoạt động của tương lai. Phân tích dữ liệu kinh doanh có thể bao gồm phân tích dữ liệu thăm dò, phân tích dữ liệu xác nhận, phân tích dữ liệu định lượng và phân tích dữ liệu định tính... Đây là công việc có ý nghĩa và có tầm quan trọng lớn đối với bất

cứ tổ chức hoặc doanh nghiệp nào, đặc biệt trong các lĩnh vực như ngân hàng, tài chính, đầu tư, bảo hiểm, du lịch, quốc phòng, hàng không vũ trụ và y học.

Đối với các doanh nghiệp, việc quản lý và phân tích lượng dữ liệu lớn này trở thành một thách thức đối với các chuyên gia kinh doanh và nhóm phân tích dữ liệu. Bằng khả năng cho phép kết nối và chuyển đổi dữ liệu từ nhiều nguồn khác nhau thành một tổ hợp thông tin trực quan và có ý nghĩa, Power BI đã trở thành công cụ phân tích dữ liệu thông minh hàng đầu hiện nay. Đóng góp một phần không nhỏ trong việc định hình và triển khai chiến lược cho mỗi doanh nghiệp.

Tuy nhiên, thực tế trong quá trình vẫn hành vẫn còn nhiều hạn chế như giới hạn trong việc chia sẻ báo cáo giữa những người dùng. Chỉ những người dùng có tài khoản cùng tên miền email mới có thể chia sẻ và truy cập các báo cáo này. Điều này tạo ra rào cản đối với việc chia sẻ thông tin và hợp tác giữa các thành viên tổ chức hoặc giữa các doanh nghiệp khác nhau. Ngoài ra Power BI có giới hạn kích thước tệp và không chấp nhận các file dung lượng lớn hơn 250MB, tuy Power BI tuy được kết nối thời gian thực nhưng nguồn dữ liệu có sẵn vẫn còn hạn chế. Có thể nhận thấy việc tạo báo cáo còn gặp nhiều hạn chế và không đáp ứng đầy đủ yêu cầu của người dùng.

Nắm được nhu cầu ứng dụng công cụ Power BI để phân tích dữ liệu trong doanh nghiệp hiện nay. Việc tích hợp báo cáo phân tích dữ liệu bằng công cụ Power BI vào ứng dụng web tùy chỉnh là nghiên cứu cần thiết. Do đó nội dung của bản đồ án này tập trung vào việc phát triển và tích hợp nội dung báo cáo phân tích dữ liệu vào ứng dụng web tùy chỉnh với mục tiêu như sau:

- Xây dựng ứng dụng web tùy chỉnh cho phép nhúng nội dung báo cáo Power BI: Tạo ra một hệ thống ứng dụng web cho cho phép nhúng các nội dung Power BI (như báo cáo, bảng thông tin và hỏi đáp) vào ứng dụng với khả năng tương tác phân tích trực tiếp với dữ liệu.
- Giải quyết thách thức của bài toán mà doanh nghiệp đang đối mặt trong lĩnh vực kinh doanh: Đề xuất và triển khai phương pháp phân

tích hoàn chỉnh cho bài toán phân khúc khách hàng và bán chéo nhằm tối ưu các phương pháp cũ của doanh nghiệp hiện nay.

- Nghiên cứu và tích hợp báo cáo phân tích Power BI vào ứng dụng web: Tích hợp nội dung báo cáo vào ứng dụng web, xây dựng chức năng tương tác người dùng cuối như: Đăng nhập, xem báo cáo phân tích, tạo báo cáo phân tích mới với tùy chọn bộ dữ liệu hiện có, cho phép chỉnh sửa và lưu báo cáo trực tiếp tại ứng dụng web.

Đề tài "**Nghiên cứu, xây dựng ứng dụng phân tích dữ liệu sử dụng Power BI**" là một nghiên cứu cần thiết trong bối cảnh hiện nay, khi mà doanh nghiệp ngày càng nhận ra việc hiểu và tận dụng dữ liệu là chìa khóa để thành công.

Nội dung báo cáo được chia làm ba chương như sau:

❖ **Chương 1: Tổng quan phân tích dữ liệu và Power BI**

Chương này trình bày một cái nhìn tổng quan và tầm quan trọng của việc phân tích dữ liệu kinh doanh chủ yếu là trong ngành bán lẻ. Sau đó phát biểu bài toán, đưa ra các giải pháp và công nghệ sử dụng để thực hiện tích hợp nội dung báo cáo vào ứng dụng web tùy chỉnh.

❖ **Chương 2: Phân tích thiết kế hệ thống**

Chương này tập trung vào việc phân tích chức năng và thiết kế mô hình hệ thống

❖ **Chương 3: Xây dựng ứng dụng và thực nghiệm**

Chương này tiến hành triển khai và đánh giá hệ thống ứng dụng với mục tiêu bài toán đặt ra.

Sau thời gian thực hiện bài báo cáo này, các mục tiêu về cơ bản đã ổn định. Tuy nhiên, công nghệ thông tin là lĩnh vực khoa học rộng mở và phức tạp và với kiến thức vẫn còn hạn hẹp nên chắc chắn bài báo cáo sẽ không tránh khỏi thiếu sót. Rất mong được sự góp ý của các thầy cô, cũng như các bạn học viên để bài báo cáo này được hoàn thiện hơn.

Chương 1
TỔNG QUAN PHÂN TÍCH DỮ LIỆU VÀ POWER BI

1.1. Phân tích dữ liệu trong ngành bán lẻ

Cùng với sự phát triển của khoa học và công nghệ, dữ liệu lớn (Big Data) đã trở thành một chủ đề nghiên cứu quan trọng nhờ vào sự phổ biến, giá trị to lớn và khả năng ứng dụng trong nhiều lĩnh vực khác nhau. Từ doanh nghiệp nhỏ đến các tập đoàn lớn ngày càng có nhu cầu rất cao đối với các công cụ phân tích kinh doanh dựa trên dữ liệu lớn. Nhiều doanh nghiệp công nghệ thông tin hàng đầu đã xem dữ liệu lớn như là nền tảng cho sự phát triển của họ. Dữ liệu lớn là nền tảng để phát triển Internet vạn vật (IoT) và trí tuệ nhân tạo (AI). Doanh nghiệp nào nắm bắt và sử dụng tốt dữ liệu lớn sẽ có lợi thế trong môi trường kinh doanh cạnh tranh khốc liệt hiện nay

Big Data (dữ liệu lớn) là sự kết hợp của dữ liệu có cấu trúc và phi cấu trúc được thu thập bởi các tổ chức có thể được khai thác để lấy thông tin và được sử dụng trong các dự án máy học, mô hình dự đoán và các ứng dụng phân tích nâng cao khác. Phân tích dữ liệu lớn liên quan đến việc trích xuất, phân tích và xử lý dữ liệu lớn đến từ nhiều nguồn khác nhau. Mục đích chính là giúp doanh nghiệp, tổ chức và các cơ quan quản lý là tìm ra các chỉ số và nhân tố có ảnh hưởng đến xu hướng, hành vi và nhu cầu của khách hàng, từ đó đưara quyết định tốt hơn về việc phát triển sản phẩm, dịch vụ và chiến lược kinh doanh.

Hiện nay, thị trường bán lẻ đang có bước phát triển lớn. Cụ thể theo Tổng cục Thống kê, riêng 5 tháng đầu năm 2023 vừa rồi, tổng mức bán lẻ hàng hóa và doanh thu dịch vụ tiêu dùng cả nước ước đạt 2.527,1 nghìn tỷ đồng, tăng 12,6% so với cùng kỳ năm 2022. Ngoài ra, Tập đoàn Aeon của Nhật Bản cũng đang lên kế hoạch từ nay đến năm 2025 triển khai thêm 16 dự án tại Việt Nam, trong đó có 3 đến 4 dự án tại Hà Nội. Đồng thời sẽ ra mắt các mô hình bán lẻ mới để đáp ứng nhu cầu của khách hàng, tạo cơ hội phát triển trong tương lai.

Đây là cơ hội để ứng dụng phân tích dữ liệu lớn vào lĩnh vực này. Hệ thống phân tích dữ liệu lớn có thể mang lại nhiều lợi ích cho cả doanh nghiệp và khách hàng trong hoạt động kinh doanh bán lẻ. Cụ thể, trong cuộc khảo sát của McKinsey, các doanh nghiệp dựa vào dữ liệu để đưa ra quyết định có khả năng sinh lời cao gấp 19 lần so với những doanh nghiệp cùng ngành khác [5]. Không có gì ngạc nhiên khi doanh số bán hàng, lĩnh vực kinh doanh dựa trên số lượng được cải thiện từ những phân tích phù hợp. Nếu mục tiêu là tăng trưởng nhanh và bền vững thì đội ngũ bán hàng phải đưa ra quyết định nhanh chóng và chính xác để cải thiện hiệu suất.

Bán lẻ là một trong những ngành kinh doanh quan trọng trong nền kinh tế của mỗi quốc gia, đóng vai trò là cầu nối giữa nhà sản xuất và người tiêu dùng. Ngành bán lẻ bao gồm các hoạt động mua và bán hàng hóa và dịch vụ trực tiếp cho người tiêu dùng cuối cùng. Đây là nơi mà hàng hóa và dịch vụ được trưng bày, quảng cáo, và bán cho khách hàng trực tiếp thông qua các cửa hàng, siêu thị, trang web mua sắm trực tuyến và các kênh bán hàng khác.

Các loại hình bán lẻ phổ biến hiện nay:

- Mô hình bán lẻ qua cửa hàng: Mô hình này bao gồm nhiều cửa hàng độc lập, cửa hàng tạp hóa, cửa hàng chuyên dụng, siêu thị, trung tâm thương mại,... Là một trong các loại hình bán lẻ phổ biến nhất từ trước cho đến nay. Khách hàng có thể có cơ hội trải nghiệm thử, xem xét trực tiếp sản phẩm trước khi quyết định mua hàng. Tuy nhiên các chi phí thuê mặt bằng, thuê nhân viên, tiền điện nước là một thách thức khá lớn đối với những doanh nghiệp muốn bán lẻ qua cửa hàng truyền thống.
- Mô hình không qua cửa hàng: Mô hình sẽ không có cơ sở kinh doanh cố định, mọi giao dịch chủ yếu thông qua tivi, internet, máy bán hàng tự động,... Trừ máy bán hàng tự động, mọi loại hình còn lại đều không có địa điểm cụ thể. Trong loại hình này, không có địa điểm cụ

thể nào dành riêng cho việc giao dịch nên doanh nghiệp khó kiểm soát lượng hàng tồn kho, có thể dẫn đến tình trạng không linh hoạt khi cần hàng hoặc khi cung ứng bị gián đoạn.

- Mô hình máy bán hàng tự động: Mô hình máy bán hàng tự động thường được sử dụng ở những quốc gia phát triển. Tại Mỹ việc sử dụng máy bán hàng tự động đã tồn tại hơn một thế kỷ này và vẫn vô cùng hiệu quả. Giống như hầu hết các mô hình kinh doanh khác, doanh nghiệp cũng cần lựa chọn vị trí, thời điểm và chủng loại sản phẩm. Ưu điểm của loại hình này đó là không tốn nhiều tiền đầu tư và vận hành. Lợi nhuận đầu tư được thu về nhanh chóng. Người tiêu dùng cũng khá thích hình thức này bởi chúng tiện lợi.
- Mô hình thông qua bưu chính: Khách hàng đặt hàng thông qua website hoặc số đường dây nóng (hotline) của doanh nghiệp sau đó đơn hàng sẽ được bàn giao cho bên thứ ba là đơn vị vận chuyển để giao đến cho khách hàng. Bán lẻ qua bưu chính thường được áp dụng cho các loại hàng hóa thông thường, hàng hóa chuyên biệt, sản phẩm mới lạ và hàng đặt mua dài hạn như CD, DVD, sách báo. Phương pháp này không yêu cầu doanh nghiệp phải có văn phòng, cửa hàng, nhà kho. Nhưng quan trọng là phải xác định được địa chỉ của khách hàng để thiết lập hệ thống nhận đặt hàng và giao hàng.
- Bán lẻ chuyên biệt: Hình thức này nghe có vẻ mới lạ nhưng thực tế các cửa hàng bán lẻ chuyên biệt đã có mặt ở Việt Nam hơn 10 năm nay. Nổi bật phải kể đến hệ thống các siêu thị điện máy như Điện Máy Xanh, Thế Giới Di Động,… Ở các điểm bán lẻ chuyên nghiệp khách hàng sẽ được trải nghiệm tiện ích sản phẩm một cách tối ưu nhất.
- Bán lẻ trên sàn thương mại điện tử: Khi các nền tảng thương mại điện tử bán lẻ tại Việt Nam như Shopee, Lazada, Tiki và đặc biệt nổi

trội gần đây nhất là Tiktok Shop. Thì đây cũng là thời điểm mà các doanh nghiệp và nhà bán lẻ mở rộng hình thức kinh doanh của họ trong lĩnh vực bán lẻ. Việc bán lẻ trên sàn thương mại điện tử mở ra cơ hội tiếp cận thêm nhiều khách hàng cho doanh nghiệp, tuy nhiên sự cạnh tranh trên các sàn thuơng mại điện tử rất gay gắt.

Đối với các doanh nghiệp kinh doanh bán lẻ, việc tối ưu hóa hoạt động kinh tế luôn là bài toán rất phức tạp. Việc áp dụng phân tích dữ liệu vào doanh nghiệp có thể giúp dự đoán được tỉ lệ hàng hóa bán ra, xu hướng nhu cầu mua sắm để tăng cường đầu tư hoặc cắt giảm chi phí kịp thời. Qua đó đưa ra chiến lược quản lý hiệu quả kinh doanh tạo điều kiện cho doanh nghiệp bán lẻ ứng phó với các rủi ro để tối ưu hóa việc sử dụng tài sản, ngân sách, hiệu suất và chất lượng dịch vụ.

Một trong những loại dữ liệu quan trọng nhất đối với các doanh nghiệp bán lẻ là dữ liệu bán hàng. Khi doanh nghiệp làm chủ được nguồn dữ liệu và áp dụng phân tích bài bản thì việc giảm thiểu rủi ro trong kinh doanh, kiểm soát và phát hiện các hoạt động gian lận, tối ưu hóa giá cả, tăng doanh thu, quản lý tốt hàng tồn kho là rất hiệu quả. Ngoài ra, còn hỗ trợ kiểm tra và quản lý tất cả dữ liệu khách hàng để nâng cao trải nghiệm của họ và đưa ra phương hướng giữ khách hàng, phân tích các hoạt động của doanh nghiệp giúp cải thiện hiệu suất làm việc và vận hành có tổ chức hơn, hiệu quả hơn.

Lợi ích của phân tích dữ liệu bán hàng:

- Tăng năng suất: Quá trình bán hàng thường phức tạp và gặp nhiều trở ngại từ việc tìm kiếm khách hàng tiềm năng cho đến việc kết thúc việc bán hàng. Phân tích dữ liệu trong bán hàng giúp xác định các lĩnh vực có thể cải thiện, giúp giảm chu kỳ hay thời gian bán hàng và triển khai các chiến lược có thể dẫn đến sự tăng trưởng trung bình về năng suất.

- Xác định cơ hội mới: Phân tích dữ liệu giúp doanh nghiệp nhận biết và tận dụng các cơ hội bán hàng mới, từ đó mở rộng thị trường và tăng doanh số bán hàng.
- Lập kế hoạch mục tiêu: Phân tích và lưu trữ dữ liệu về giao dịch, cơ hội tiềm năng và độ dài chu kỳ bán hàng giúp doanh nghiệp xác định mục tiêu bán hàng cụ thể và hiệu quả.
- Cải thiện việc thu hút khách hàng: Phân tích dữ liệu về hành vi của khách hàng giúp doanh nghiệp xây dựng các chiến lược tiếp thị các nhân hóa, từ đó thu hút và giữ chân khách hàng mới.
- Khuyến khích đội ngũ bán hàng: Phân tích dữ liệu kinh doanh thiết lập các chính sách thưởng phù hợp và công bằng cho đội ngũ bán hàng, từ đó tăng động lực làm việc và hiệu suất làm việc của họ.
- Tăng khả năng giữ chân khách hàng: Phân tích dữ liệu giúp doanh nghiệp nhận biết và giải quyết các vấn đề có thể dẫn đến việc khách hàng chuyển đổi sang các đối thủ cạnh tranh hay gọi là tỷ lệ rời bỏ của khách hàng, từ đó tăng cơ hội giữ chân khách hàng và tăng giá trị cuộc sống của khách hàng.

1.1.1. Ứng dụng phân tích dữ liệu trong bán lẻ

Bán lẻ có thể hiểu là cách nhà sản xuất hàng hóa, dịch vụ đưa sản phẩm đến người tiêu dùng. Hiện nay dưới tác động của sự phát triển của công nghệ, thói quen mua sắm tiêu dùng của khách hàng liên tục thay đổi. Đòi hỏi nhà bán lẻ luôn phải nắm được nhu cầu khách hàng từ đó tối ưu nguồn lực của mình và đặt khách hàng làm trọng tâm để có thể tồn tại và phát triển bền vững. Đây chính là một thách thức lớn cho doanh nghiệp trong thời đại số hiện nay.

Giải pháp tối ưu nhất chính là áp dụng phân tích nguồn dữ liệu đến từ các thói quen tìm kiếm sản phẩm, dịch vụ, thông tin. Việc xem xét lựa chọn những sản phẩm (hoặc dịch vụ) mà khách hàng đang tìm kiếm sẽ giúp doanh nghiệp khám phá chính xác nhu cầu của họ. Sau đó, sử dụng các phân tích tối ưu hóa việc

bán hàng, mở rộng phạm vi kinh doanh sản phẩm một cách nhanh chóng và hiệu quả. Dưới đây là một số ứng dụng nổi bật của phân tích dữ liệu trong ngành bán lẻ:

1. **Dự đoán hành vi mua sắm của khách hàng**

Dữ liệu được dùng để xử lý các hành vi mua sắm, mối quan tâm đến các loại sản phẩm nào, từ đó dự đoán xu hướng mua sắm để đưa ra các khuyến nghị, quảng cáo phù hợp với nhu cầu của khách hàng. Các kết quả hiển thị gợi ý sản phẩm khách hàng quan tâm, tối ưu hóa kết quả hiển thị tìm kiếm thay cho việc khách hàng phải đi dạo các sản phẩm một cách ngẫu nhiên. Chẳng hạn như khi khách hàng chọn xem hoặc lựa chọn vào giỏ hàng sản phẩm điện thoại, website sẽ gợi ý các sản phẩm như ốp lưng, pin dự phòng, sạc điện thoại, tai nghe,…

Kết quả phân tích hành vi khách hàng sẽ giúp doanh nghiệp xác định được những yếu tố tác động đến việc khách hàng chọn mua sản phẩm hay đăng ký dịch vụ. Do đó, khi tập hợp được dữ liệu về sở thích, thói quen của khách hàng đồng nghĩa với việc gián tiếp giúp doanh nghiệp bán được nhiều hàng hóa hơn.

2. **Phân khúc khách hàng**

Phân khúc khách hàng là nhóm các đối tượng khách hàng được phân chia theo từng đặc điểm cụ thể. Từng phân khúc khách hàng khác nhau sẽ có những đặc điểm, hành vi mua hàng khác nhau, ảnh hưởng đến chiến lược kinh doanh của doanh nghiệp. Do đó, tùy vào phân khúc khách hàng khác nhau, doanh nghiệp sẽ chuẩn bị chương tình tiếp thị hoặc quảng cáo phù hợp hơn trong việc tiếp cận khách hàng. Một chiến lược tiếp thị, bán hàng phù hợp với phân khúc, đối tượng khách hàng mục tiêu sẽ mang đến hiệu quả vượt trội trong hoạt động kinh doanh của doanh nghiệp.

3. **Phân tích hành trình của khách hàng**

Sử dụng dữ liệu để diễn tả quy trình khách hàng tương tác với doanh nghiệp, khách hàng sẽ mua hàng hoặc khách hàng sẽ quảng cáo thương hiệu đến khách hàng khác. Việc phân tích quá trình tương tác giữa khách hàng và sản phẩm, khách

hàng và cửa hàng giúp doanh nghiệp nhận biết được các vấn đề của khách hàng. Từ đó tối ưu hóa từng hoạt động cụ thể để giải quyết tận gốc vấn đề khách hàng đang gặp phải, xây dựng mối quan hệ lâu bền với doanh nghiệp.

4. Xây dựng mô hình chi tiêu cho từng khách hàng

Đặc thù của dịch vụ bán lẻ là có số lượng lớn đối tượng khách hàng có cùng đặc điểm. Nếu không ứng dụng dữ liệu thì việc xây dựng mô hình chi tiêu cho từng khách hàng là việc bất khả thi. Chăm sóc từng khách hàng thông qua việc xây dựng kế hoạch chi tiêu cho từng khách hàng cụ thể đem đến hiệu quả cao cho các doanh nghiệp phát triển mảng bán lẻ. Thông thường khách hàng mua lẻ sẽ có tần suất chi tiêu cao theo từng tuần, từng tháng hoặc từng quý. Nắm bắt được nhu cầu và tần suất chi tiêu của khách hàng, doanh nghiệp sẽ có hiệu quả kinh doanh cao với lượng hàng hóa mua đều đặn từ khách hàng lẻ

5. Phân tích bán chéo

Bán chéo là một kĩ thuật bán hàng được sử dụng để khiến khách hàng chi tiêu nhiều hơn bằng cách mua thêm sản phẩm có liên quan được gợi ý từ sản phẩm đã mua trước đó. Doanh nghiệp sẽ sử dụng dữ liệu về sản phẩm của thương hiệu mình, tiến hành phân tích những cặp sản phẩm có mối liên quan nhất định từ đó dự đoán và gợi ý các sản phẩm có khả năng được mua cùng nhau bởi các khách hàng.

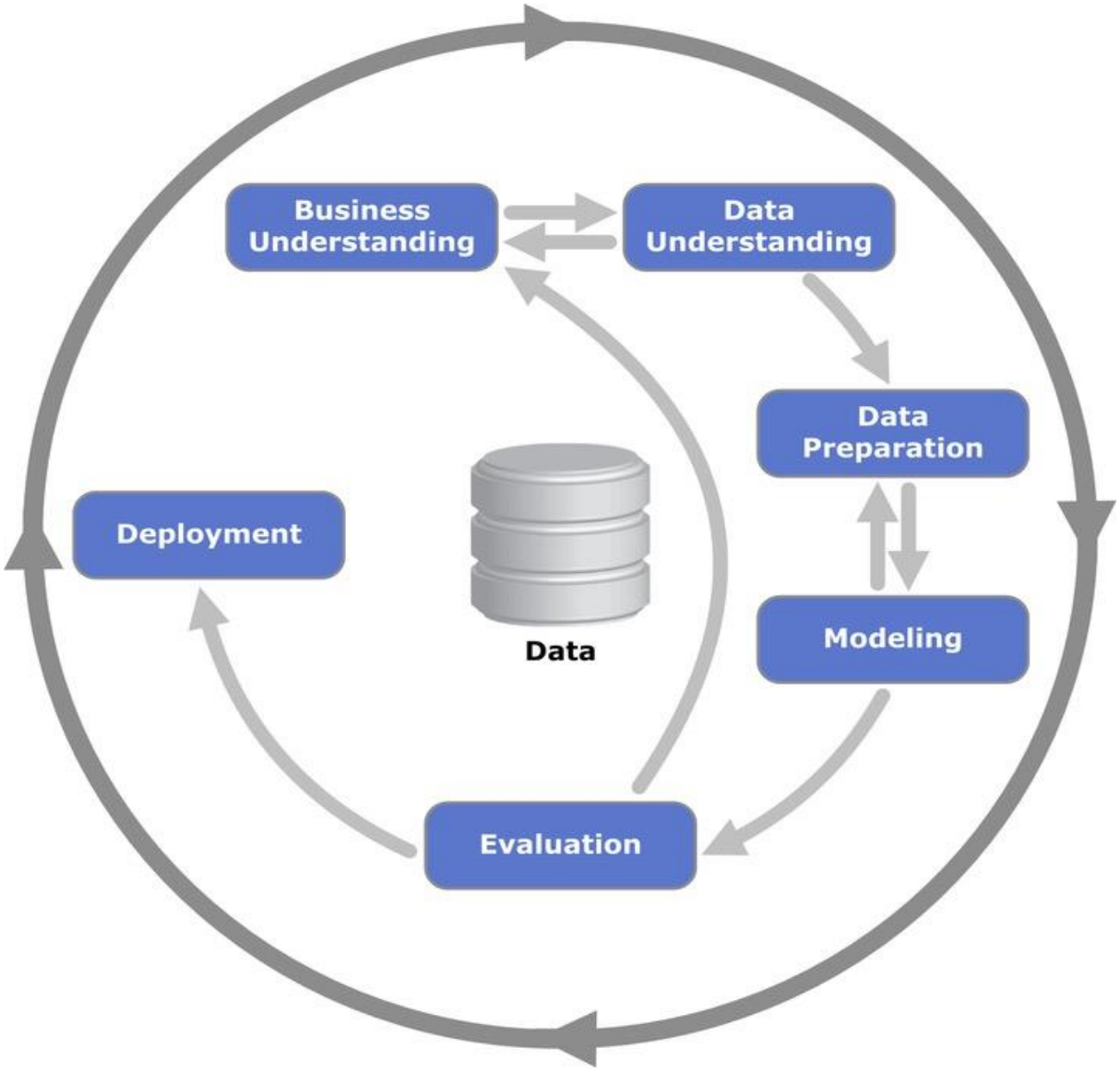
Ngoài ra còn rất nhiều phân tích khác, như phân tích thị trường, phân tích chiến dịch và quảng cáo, phân tích tương tác khách hàng…tùy thuộc vào nhu cầu và mục đích của mỗi tổ chức, doanh nghiệp từ đó xác định bài toán cần phân tích.

1.1.2. Quy trình phân tích dữ liệu

Phân tích dữ liệu bán hàng là quá trình thu thập, phân tích và diễn giải dữ liệu để hiểu biết sâu sắc về các khía cạnh khác nhau của hoạt động bán hàng. Phân tích bán hàng liên quan đến việc tận dụng các công nghệ tiên tiến và mô hình thống kê để chuyển đổi dữ liệu thô thành thông tin hữu ích nhằm giúp các tổ chức

đưa ra quyết định dựa trên dữ liệu hiệu quả hơn, xác định các cơ hội tăng trưởng doanh số, nâng cao hiệu suất tổng thể và sự hài lòng của khách hàng.

CRISP – DM (Cross Industry Standard Process for Data Mining), tạm dịch sang nghĩa tiếng việt là "quy trình tiêu chuẩn cho khai thác dữ liệu trong nhiều lĩnh vực". Quy trình này được hình thành từ năm 1996, nhưng cho đến nay có lẽ vẫn là một trong những tiêu chuẩn được sử dụng rộng rãi nhất.



Hình 1. 1 Quy trình CRISP DM

CRISP-DM là một quy trình lặp có nhiều vòng lặp bên trong quy trình và chỉ ưu tiên chứ không bắt buộc phải theo một trình tự nào. Tuy nhiên cần lưu lại

tối đa lượng thông tin đã có ở mỗi bước. Quy trình CRISP DM gồm 6 như mô tả ở (hình 1.1) ở trên bước:

Bước 1: Hiểu nhu cầu của doanh nghiệp (Business Understanding) :

Mục đích bước này là xác định mục tiêu bài toán phân tích. Thực tế thì mỗi loại hình doanh nghiệp sẽ xoay quanh một số bài toán rất đặc trưng. Những bài toán này thường phát sinh từ nhu cầu thực tế của lĩnh vực đó hoặc là đã được nghiên cứu chứng minh và áp dụng thành công.

Sau khi xác định được biến mục tiêu (phụ thuộc) thì sẽ xác định tất cả các yếu tố có khả năng ảnh hưởng tới biến mục tiêu này (biến độc lập) gọi danh sách các tính năng(feature). Việc xác định mục tiêu phân tích sẽ phụ thuộc vào mục đích, mong muốn, hoàn cảnh của doanh nghiệp. Để xác định được các feature này thì phải có sự tham gia của rất nhiều ý kiến chuyên gia họ sẽ cho ý kiến các nhân tố có thể ảnh hưởng dựa trên kinh nghiệm, hiểu biết về lĩnh vực đó. Với các bài toán kinh điển được phát triển lâu đời thì việc liệt kê danh sách những tính năng này rất chi tiết và quy chuẩn, còn các bài toán mới thì việc này khá khó khăn, mất nhiều thời gian thử nghiệm và nghiên cứu.

Ví dụ: Để có một cái nhìn tổng quan về quy trình CRISP DM thì ví dụ này sẽ được tiếp tục áp dụng cho các bước còn lại.

- Mục tiêu của dự án: Tăng năng suất bán hàng, cụ thể là tăng hiệu quả của việc tiếp thị khách hàng qua điện thoại.
- Tiêu chí cụ thể để đánh giá sự thành công của dự án: Tỷ lệ chuyển đổi khách hàng tiềm năng trở thành khách hàng thực sự (conversion rate) tăng từ 5% lên 10% (trung bình cứ mỗi 100 khách hàng được gọi, 10 khách hàng sẽ mua sản phẩm thay vì 5 khách hàng như hiện nay).
- Giải pháp đề xuất: Xây dựng mô hình dự đoán khả năng mua sản phẩm của tất cả các khách hàng, từ đó có thể tham khảo để tập trung vào nhóm khách tiềm năng nhất (nhóm khách hàng được dự đoán có khả năng mua cao nhất).

Bước 2: Hiểu về dữ liệu (Data Understading)

Sau khi đã xác định được mục tiêu của bài toán, bước tiếp theo là tìm kiếm và đánh giá xem dữ liệu có phù hợp cho giải pháp được đề ra hay không. Trường hợp dữ liệu đã có sẵn, thì sẽ cần chuẩn bị một số dữ liệu cần thiết, tiếp tục cho ví dụ ở bước 1:

- Dữ liệu về những khách hàng được tiếp thị và những khách hàng đã mua sản phẩm.
- Dữ liệu về nhân khẩu học: độ tuổi, ngành nghề, thu nhập,…
- Dữ liệu về lịch sử tương tác của khách hàng trên website của doanh nghiệp: Khách hàng tương tác với dòng sản phẩm nào? Thời gian bao lâu?
- Dữ liệu về lịch sử giao dịch của khách hàng đối với những sản phẩm của doanh nghiệp: Khách hàng đã mua sản phẩm gì? Vào thời điểm nào? Giá trị là bao nhiêu?

Nếu dữ liệu không có sẵn, thì sẽ phải lập kế hoạch tổng hợp những dữ liệu này hoặc sẽ phải xem xét giải pháp khác. Dữ liệu có thể được thu thập từ nhiều nguồn khác nhau, có thể từ nguồn nội bộ hay các nguồn bên ngoài khác theo các nhóm thông tin yêu cầu.

Bước 3: Chuẩn bị dữ liệu (Data Preparation)

Đây là bước chuyển đổi và tổng hợp các dữ liệu liên quan để dùng cho việc xây dựng mô hình ở bước tiếp theo. Bước chuẩn bị dữ liệu thông thường sẽ bao gồm các bước nhỏ hơn, ví dụ như sau:

- Làm sạch dữ liệu, ví dụ như xử lý dữ liệu bị thiếu và các điểm dị biệt
- Tạo thêm các thuộc tính mới
- Tổng hợp các nguồn dữ liệu với nhau

Ở bước chuẩn bị dữ liệu còn phụ thuộc vào thuật toán được áp dụng (thuật toán khác nhau có thể cần cách xử lý dữ liệu đầu vào khác nhau) nên bước này liên hệ chặt chẽ với bước xây dựng mô hình tiếp theo.

Bước 4: Xây dựng mô hình (Modeling)

Đây là bước áp dụng các thuật toán lên dữ liệu (được chuẩn bị ở bước trước đó) để xây dựng mô hình phục vụ cho mục tiêu đặt ra.

Do mỗi thuật toán lại có ưu nhược điểm khác nhau, do đó cần thử nghiệm xây dựng mô hình với nhiều thuật toán để tìm ra mô hình phù hợp nhất. Dưới đây là một số tiêu chí hay được áp dụng cho việc lựa chọn mô hình:

- Độ chính xác
- Khả năng diễn giải
- Thời gian huấn luyện mô hình

Bước 5: Đánh giá chất lượng mô hình (Evaluation)

Ở bước này, mô hình được xây dựng sẽ được đánh giá xem có đạt được mục tiêu đề ra ban đầu hay không. Liên hệ đến mục tiêu ở trong bước 1: Mục tiêu là tăng tỷ lệ chuyển đổi khách hàng tiềm năng từ 5% lên 10%. Mục tiêu này đòi hỏi đánh giá hiệu quả kinh doanh từ thực tế bằng cách thử nghiệm tiếp thị đến những khách hàng tiềm năng nhất (có chỉ số dự đoán khả năng mua sản phẩm cao) và đánh giá tỷ lệ chuyển đổi của phương pháp mới này. Nếu chỉ số này chưa đạt 10%, sẽ quay lại các bước trước đó để tìm cách nâng cao chất lượng mô hình dự đoán cho đến khi đạt được kết quả như mong muốn.

Bước 6: Triển khai giải pháp (Deployment)

Sau khi thực hiện đủ các bước trên và đạt được kết quả khả quan thì mô hình sẽ được coi là sẵn sàng cho việc triển khai vào thực tế. Quy trình phân tích dữ liệu theo chuẩn quy trình CRISP – DM không chỉ giúp doanh nghiệp nâng cao hiệu quả hoạt động và đưa ra quyết định chiến lược mà còn tạo ra giá trị cạnh tranh dài hạn thông qua việc hiểu rõ khách hàng, quản lý rủi ro và thúc đẩy đổi mới.

Phạm vi phân tích dữ liệu trong bán hàng là rất lớn. Nhưng tất cả đều tập trung vào việc lựa chọn phân tích và số liệu phù hợp giúp đạt được mục tiêu kinh doanh đặt ra. Phân tích bán hàng được chia thành bốn loại chính:

1. **Phân tích mô tả:** Phân tích mô tả đòi hỏi phải theo dõi lịch sử dữ liệu bán hàng như: doanh thu, số lượng người dùng,... Để có thể so sánh và hiểu rõ hơn những gì đang xảy ra. Phân tích này trả lời cho câu hỏi chuyện gì đã xảy ra?
2. **Phân tích chẩn đoán:** Phân tích chẩn đoán đang kiểm tra và đi sâu vào dữ liệu để xác định chính xác lý do tại sao điều gì đó xảy ra. Phân tích này trả lời cho câu hỏi tại sao nó lại xảy ra?
3. **Phân tích dự đoán:** Phân tích dự đoán là sử dụng những dữ liệu đã có đã về doanh số bán hàng trong quá khứ để đánh giá các mô hình và xu hướng. Điều này cho phép đưa ra những dự đoán có cơ sở. Phân tích này trả lời cho câu hỏi chuyện gì sẽ xảy ra?
4. **Phân tích quyết định:** Phân tích theo quy định bao gồm việc đánh giá tất cả dữ liệu và đề xuất kế hoạch hành động tốt nhất. Phương pháp này trả lời cho câu hỏi giải pháp hoặc chiến lược tốt nhất là gì?

1.1.3. Xu hướng và thách thức

Thị trường giải pháp phân tích bán hàng đang có sự tăng trưởng đáng kể do một số yếu tố. Thứ nhất, việc ra quyết định bán hàng dựa trên dữ liệu ngày càng được chú trọng vì các tổ chức mong muốn đạt được lợi thế cạnh tranh bằng cách tận dụng dữ liệu bán hàng của họ. Thứ hai, việc áp dụng ngày càng nhiều các giải pháp dựa trên đám mây và sự sẵn có của khả năng phân tích dữ liệu lớn đang thúc đẩy nhu cầu về các giải pháp phân tích bán hàng. Thứ ba, sự phổ biến ngày càng tăng của hệ thống tự động hóa lực lượng bán hàng và quản lý quan hệ khách hàng (CRM) đang thúc đẩy nhu cầu về các giải pháp phân tích nâng cao để rút ra những hiểu biết có ý nghĩa từ dữ liệu bán hàng.

Trước bối cảnh cạnh tranh gay gắt, cùng với những thay đổi mạnh mẽ của dữ liệu lớn dựa trên nền tảng khoa học kỹ thuật, đặc biệt là công nghệ từ cuộc Cách mạng công nghiệp lần thứ tư, các nhà bán nước ta đang đứng trước nhiều thách thức lớn:

- Đầu tiên, việc các doanh nghiệp bán lẻ nước ngoài ồ ạt đổ vốn vào thị trường Việt Nam cũng gây nên sức ép nhất định đến thị phần cho doanh nghiệp trong nước [2]. Doanh nghiệp FDI có lợi thế về năng lực cạnh tranh, về quy mô. Họ có lợi thế chuỗi liên kết toàn cầu với các nhà sản xuất. Trong khi đó, doanh nghiệp nội địa có quy mô nhỏ, năng lực yếu; đồng thời gánh các chi phí từ logistics, thuê mặt bằng cao.
- Khó khăn trong việc tìm hiểu nhu cầu khách hàng do nhu cầu của khách hàng luôn biến động và thay đổi theo từng giai đoạn. Việc nắm bắt thị hiếu của khách hàng đòi hỏi các doankh nghiệp bán lẻ cần tìm ra các phần mềm công nghệ phù hợp để phân tích, khai thác dữ liệu hiệu quả để hiểu rõ thay đổi của khách hàng.
- Phối hợp chưa chặt chẽ trong chuỗi cung ứng, lực lượng trong chuỗi cung ứng gồm nhà sản xuất, nhà phân phối và nhà bán lẻ. Lực lượng này sẽ phối hợp và làm việc theo những tiêu chuẩn thống nhất nhằm phục vụ khách hàng, nhưng đối với các doanh nghiệp vừa và nhỏ, chưa có nhiều kinh nghiệm thì sự phối hợp thiếu tính chặt chẽ, gắn kết giảm hiểu quả trong kinh doanh bán lẻ.
- Hình thức bán lẻ truyền thống vẫn chiếm tỉ lệ khá lớn, nhiều doanh nghiệp hiện nay vẫn tiếp tục duy trì hình thức bán hàng truyền thống, từ đó dẫn đến nguồn hàng hóa chưa phong phú, thiếu tính cạnh tranh cả về hàng hóa và giá cả. Quá trình kiểm soát chất lượng chưa đảm bảo và chưa đáp ứng được nhu cầu của khách hàng.

1.2. Phát biểu bài toán

Với đặc điểm là thị trường công nghệ đang phát triển mạnh mẽ thì công việc phân tích dữ liệu trong doanh nghiệp ở Việt Nam được đánh giá là có vai trò vô cùng cấp thiết. Mặc dù, hệ thống phân tích dữ liệu tự động đang được đưa vào sử dụng trong nhiều doanh nghiệp. Tuy nhiên, nó vẫn chưa thể đáp ứng hoàn toàn nhu cầu của người sử dụng bởi khả năng hiểu ngữ cảnh và tinh tế còn hạn chế,

chất lượng dữ liệu không đồng nhất, thiếu khả năng tùy chỉnh theo nhu cầu cụ thể, khó xử lý các tình huống bất ngờ, và khả năng tương tác, giao tiếp kém linh hoạt. Hơn nữa, vấn đề bảo mật và quyền riêng tư cũng đặt ra thách thức lớn.

Trong quá trình tìm hiểu và thực hiện đề tài với mục tiêu triển khai ứng dụng web được tích hợp báo cáo phân tích dữ liệu bằng Power BI đã rút ra được bài toán đặt ra gồm hai thách thức. Thứ nhất, là việc triển khai công nghệ và kỹ thuật để tạo ra một giao diện web linh hoạt, cho phép tương tác với người dùng và có khả năng nhúng báo cáo một cách dễ dàng và hiệu quả. Thứ hai, là việc nắm vững kiến thức về lĩnh vực ngành bán lẻ, bao gồm sự hiểu biết về dữ liệu bán lẻ, các bài toán quan trọng trong phân tích bán lẻ mà doanh nghiệp quan tâm, xu hướng thị trường và yếu tố ảnh hưởng đến hoạt động kinh doanh. Từ đó xác định và tiến hành triển khai bài toán phân tích quan trọng có tính ứng dụng cao trong các doanh nghiệp hiện nay.

Đối với phần công nghệ, cần xác định giải pháp để tích hợp báo cáo Power BI vào một ứng dụng web. Nắm được các khái niệm về REST API và dữ liệu JSON là cần thiết để tương tác với dữ liệu từ Power BI và hiển thị chúng trên trang web.

Về phần lĩnh vực ngành bán lẻ, hiểu biết sâu sắc về dữ liệu bán lẻ là rất quan trọng để tạo ra các báo cáo phân tích có ý nghĩa và hữu ích. Bao gồm nắm được lý thuyết của mỗi bài toán để ứng dụng trong phần thực nghiệm.

Vì vậy việc xây dựng một ứng dụng web được nhúng báo cáo phân tích dữ liệu bán lẻ là bài toán không chỉ là về công nghệ mà còn là về hiểu biết sâu sắc về phân tích dữ liệu trong ngành bán lẻ. Bằng cách kết hợp cả hai khía cạnh này, có thể tìm hiểu và áp dụng các giải pháp phân tích dữ liệu mạnh mẽ và mang lại giá trị lớn cho doanh nghiệp.

1.3. Giải pháp và công nghệ sử dụng

Đưa ra những đánh giá sáng suốt là rất quan trọng trong bối cảnh doanh nghiệp dựa trên dữ liệu ngày nay, đặc biệt là trong lĩnh vực bán lẻ. Dựa vào bài

toán đặt ra trong (phần 1.2) trong phần này sẽ xác định giải pháp để giải quyết thách thức mà bài toán đặt ra. Đầu tiên sẽ xác định giải pháp cho việc nhúng Power BI vào ứng dụng web MVC. Sau đó, sẽ xác định bài toán phổ biến trong ngành bán lẻ và chi tiết phương pháp giải quyết bài toán kết hợp với công cụ Power BI.

1.3.1. Giải pháp phân tích nhúng

Phân tích dữ liệu không còn xa lạ gì với các doanh nghiệp hiện nay, nó đã trở thành một phần không thể thiếu trong quá trình định hình chiến lược và tối ưu hóa hoạt động kinh doanh. Do đó, các doanh nghiệp không ngừng liên tục cập nhật những công nghệ phân tích dữ liệu tiên tiến nhất.

Business Intelligence (BI) là tập hợp các công nghệ, quy trình và công cụ giúp tổ chức thu thập, tích hợp, phân tích và trình bày dữ liệu kinh doanh. BI hỗ trợ các doanh nghiệp ra quyết định hiệu quả thông qua việc cung cấp báo cáo, bảng điều khiển (dashboards) và phân tích dữ liệu chi tiết. Đây là một phương pháp tiếp cận toàn diện, nhằm nâng cao hiệu quả và hiệu suất của quy trình ra quyết định, khi mà dữ liệu đã trở thành một yếu tố quen thuộc và quan trọng đối với các doanh nghiệp ngày nay.

Chủ đề phân tích dữ liệu đã được nghiên cứu rộng rãi trong những năm gần đây, như trong nghiên cứu của TS. Nguyễn Thị Hội & ThS. Bùi Quang Trường về "Một số công cụ phân tích dữ liệu thông minh và ứng dụng trong các bài toán kinh tế" được đăng tại Tạp chí Công Thương. Số 3, tháng 2 năm 2025 [2]. Nghiên cứu này so sánh một số công cụ phân tích dữ liệu thông minh ứng dụng trong các hoạt động sản xuất và kinh doanh của doanh nghiệp như: Tableau Public, Power BI, FineReport, R, và Python. Kết quả cho thấy mỗi công cụ có mức độ ảnh hưởng và ứng dụng khác nhau trong các bài toán cụ thể. Tuy nhiên, khi xem xét trên cùng giao diện và thay đổi các tham số như thị trường (theo tỉnh/thành phố), Power BI cho kết quả trực quan, sinh động, dễ dàng cập nhật và có giao diện kết nối tốt hơn các công cụ khác. Ngoài ra, Power BI cũng được sử dụng để phân tích và thiết kế trực quan hóa cơ sở dữ liệu cho các tổ chức giáo dục [4] .

Nhiều nghiên cứu khác cũng chỉ ra rằng Power BI là công cụ dẫn đầu trong nhóm các công cụ Business Intelligence (BI), được hầu hết các doanh nghiệp sử dụng để đơn giản hóa và nâng cao chất lượng hiển thị dữ liệu trên các báo cáo tài chính, doanh số, và các báo cáo kinh doanh khác. Điều này nhấn mạnh tầm quan trọng của giải pháp BI trong hoạt động doanh nghiệp.

Với sự phát triển mạnh mẽ của công nghệ và nhu cầu ngày càng cao về khả năng phân tích dữ liệu trực tiếp trong các ứng dụng doanh nghiệp, giải pháp Phân tích nhúng (Embedded Analytics) đã ra đời như một bước tiến tất yếu. Power BI, với những ưu điểm vượt trội đã được chứng minh qua nhiều nghiên cứu, là một lựa chọn hợp lý cho việc nhúng công cụ phân tích này vào các ứng dụng web tùy chỉnh.

Giải pháp phân tích nhúng (Embedded Analytics) là một cách tích hợp các công cụ và khả năng phân tích dữ liệu trực tiếp vào các ứng dụng cá nhân hoặc doanh nghiệp hiện có. Thay vì phải sử dụng một công cụ phân tích dữ liệu riêng biệt, người dùng có thể truy cập và tương tác với dữ liệu ngay trong ứng dụng mà họ đang sử dụng hàng ngày. Đây là một bước tiến quan trọng trong việc đưa ra quyết định dựa trên dữ liệu một cách hiệu quả và liền mạch.

Giải pháp này được sử dụng nhằm nhúng Power BI vào ứng dụng Web MVC sử dụng nền tảng .NET5 kết hợp viết mã C# phía máy chủ và mã JavaScript phía máy khách để nhúng báo cáo Power BI vào website tùy chỉnh cũng như thực hiện tương tác người dùng. Thông qua đó có thể giải quyết được vấn đề đầu tiên liên quan đến các nền tảng công nghệ trong bước đầu xây dựng ứng dụng để nhúng báo cáo Power BI.

Giải pháp phân tích nhúng Power BI(Power BI embedded analytics) là một dịch vụ của Azure Active Directory (Azure AD) cho phép nhúng nội dung Power BI như báo cáo, bảng thông tin vào một trang web hoặc ứng dụng web. Nó cung cấp trải nghiệm phong phú, được tích hợp và hỗ trợ API đầy đủ.

Ngoài ra, phân tích nhúng cho phép tự động hóa việc giám sát, quản lý và triển khai phân tích, đồng thời có toàn quyền kiểm soát các tính năng Power BI và phân tích thông minh. Phân tích nhúng Power BI hiện có hai phương pháp:

Bảng 1. 1 So sánh 2 phương pháp phân tích nhúng

Phương pháp	**Embed for your customers**	**Embed for your organization**
Định nghĩa	Còn được gọi là ứng dụng sở hữu dữ liệu	Còn được gọi là người dùng sở hữu dữ liệu
Dữ liệu	Ứng dụng sở hữu dữ liệu hướng đến người dùng bên ngoài, người dùng ứng dụng không cần đăng nhập bằng thông tin xác thực Power BI	Người dùng sở hữu dữ liệu hướng đến người dùng nội bộ, người dùng ứng dụng phải đăng nhập bằng thông tin xác thực Power BI (xác thực Azure AD)
Giấy phép	Người dùng ứng dụng không cần giấy phép	Người dùng ứng dụng cần có giấy phép Power BI Pro
Xác thực	Xác thực không tương tác. Ứng dụng sử dụng dịch vụ chính (service principal) hoặc người dùng chính (người dùng chính) để xác thực	Xác thực tương tác. Ứng dụng sử dụng thông tin đăng nhập của người dùng ứng dụng để xác thực

Trong (bảng 1.1) ở trên vừa so sánh những tính năng của hai phương pháp nhúng dành cho khách hàng với giải pháp nhúng dành cho tổ chức (embed for customers vs embed for organization). Phương pháp nhúng cho tổ chức nội bộ yêu cầu tất cả người dùng phải có tài khoản Power BI Pro hoặc Power BI Premium Per User (PPU) mới có thể truy cập báo cáo.

Còn phương pháp nhúng cho khách hàng thì khách hàng không cần có tài khoản Power BI riêng mà vẫn có thể tương tác với báo cáo và dữ liệu nhúng. Nhưng tài khoản của quản trị viên cần phải có quyền truy cập Power BI Pro hoặc PPU để sử dụng các tính năng nhúng và cho phép quản lý và chia sẻ nội dung Power BI với khách hàng mà không yêu cầu họ có tài khoản riêng. Do đó để người dùng có thể là khách hàng của một doanh nghiệp không cần tài khoản Power BI của riêng họ mà vẫn có thể truy cập được báo cáo. Quản trị viên sẽ phân quyền sử dụng báo cáo theo nhu cầu của mỗi người dùng. Vì vậy phương pháp phân tích nhúng cho khách hàng rất phù hợp với mục đích của đồ án này.

Việc đăng ký license Microst Power BI sẽ quyết định số lượng tính năng có thể sử dụng trên Power BI Service. Do đó, việc chọn đúng License Power BI phù hợp với nhu cầu của là rất quan trọng.

Hiện tại Microsoft Power BI cung cấp cả tài khoản miễn phí và trả phí. Với tài khoản miễn phí thì có thể kết nối nguồn dữ liệu, phân tích dữ liệu và tạo báo cáo. Phù hợp cho người dùng cá nhân muốn tạo báo cáo cho nhu cầu riêng và không cần chia sẻ lên đám mây.

Còn tài khoản Power BI Pro sẽ có phí là $10/người dùng/tháng. Cho phép chia sẻ dữ liệu và báo cáo với những người dùng khác cũng có License Pro. Lưu trữ dữ liệu tối đa 10GB cho mỗi người dùng. Xuất báo cáo lên website hoặc trang SharePoint. License tính theo người dùng, nghĩa là phải trả phí cho từng người sử dụng. Thích hợp cho các công ty vừa và nhỏ cần tạo, chia sẻ và xem báo cáo của người dùng khác.

Ngoài ra còn tài khoản PPU với mức phí là $20/người dùng/tháng. Kết hợp các lợi ích của License Power BI Pro với gói đăng ký Power BI Premium. Người dùng được cung cấp các tính năng cao cấp như mô hình dữ liệu lớn (tối đa 100GB), cập nhật thường xuyên (lên đến 48 lần/ngày), báo cáo phân tích trang (Paginated Reports) và trí tuệ nhân tạo (AI).

Tiếp theo, sẽ tìm hiểu hai bài toán lớn hiện nay được hầu như các doanh nghiệp sử dụng trong phân tích dữ liệu của mình. Nội dung này sẽ là cơ sở để có thể áp dụng trên bộ dữ liệu cụ thể được triển khai trong (chương 3) của đề tài này.

1.3.2 Giải pháp phân tích bài toán bán lẻ

Thông thường khi phân tích dữ liệu báo cáo sẽ chia làm 2 loại: Thứ nhất là báo cáo tổng quan với mục đích phục vụ cho việc theo dõi số liệu. Thứ hai là báo cáo phân tích là loại báo cáo đi sâu vào việc phân tích dữ liệu và thông tin để xác định nguyên nhân, xu hướng, và mối quan hệ giữa các yếu tố. Báo cáo này thường chứa các biểu đồ, đồ thị, và các công cụ phân tích khác để làm rõ các chỉ số và cung cấp cái nhìn chi tiết về vấn đề đang được xem xét. Hai loại báo cáo này đều sẽ được triển khai để thấy rõ được lợi ích của báo cáo phân tích so với báo cáo tổng quan.

Trong môi trường cạnh tranh gay gắt của ngành bán lẻ, việc hiểu rõ khách hàng và tận dụng thông tin từ dữ liệu là một yếu tố quyết định sự thành công của các doanh nghiệp. Trong phần này sẽ tìm hiểu phương pháp để giải bài toán phân khúc khách hàng và phân tích bán chéo. Và trình bày báo cáo theo phương pháp DAR (Dashboard, Analysis, Reporting). Phương pháp gồm 3 bước chính:

- Xây dựng trang tổng quan (dashboard): đây là trang đầu tiên chứa những thông tin quan trọng và cần thiết. Trang này sẽ cung cấp thông tin theo hướng tổng quan và ít thao tác nhất có thể.
- Trang tiếp theo là trang phân tích (analysis), trang này giúp người xem tìm hiểu kỹ hơn về dữ liệu cũng như đi tìm các câu trả lời mà họ thắc mắc khi xem dashboard. Các trang này cần đưa ra nhiều tùy chọn về bộ lọc, chú thích (tooltips), dấu trang (bookmark) để người dùng có thể dễ dàng tương tác.
- Cuối cùng là trang report: trang này cung cấp các thông tin ở mức độ chi tiết nhất, đa phần được thể hiện bằng nhiều bảng biểu (matrix table) để người xem có thể đào sâu phân tích bằng cách sắp xếp hoặc lọc dữ liệu.

Đồng thời, đây cũng là giai đoạn nhận thức đầy đủ về dữ liệu và ra quyết định cụ thể.

Dựa trên phương pháp DAR báo cáo để nhúng vào ứng dụng Power BI Sales sẽ gồm ba trang: Đầu tiên là trang báo cáo tổng quan, trang thứ hai là báo cáo phân tích phân khúc khách hàng, trang cuối cùng sẽ là phân tích bán chéo.

Với trang báo cáo tổng quan, đây là trang đầu tiên nên chỉ đưa ra những thông tin/KPIs quan trọng và tổng quan. Không nên sử dụng quá nhiều các bộ lọc nhằm giảm thiểu các thao tác động trên dashboard. Bố cục nên tập trung vào các thông tin quan trọng, các chỉ tiêu chính nên có kích thước lớn hơn và thu hút người xem. Trang báo cáo tổng quan chứa những chỉ số dễ tính và khá đơn giản hầu như chưa cần áp dụng những bài toán phức tạp cũng do mục đích của trang này chủ yêu là để theo dõi số liệu. Còn bài toán phân khúc và bán chéo có phạm vi lớn hơn và có ứng dụng thực tế rất cao, do đó những chỉ số tính toán cũng phức tạp hơn.

Tuy nhiên, trong phần này tôi sẽ trình bày mỗi bài toán theo hướng là một giải pháp hoàn chỉnh để dễ hình dung các công đoạn. Khung dàn ý chung các vấn đề cần làm rõ trong mỗi bài toán gồm :

1. Xác định mục tiêu bài toán
2. Phân tích và đặc tả bài toán trong bối cảnh hiện tại
3. Nghiên cứu đề xuất lựa chọn các giải pháp
4. Phát triển 1 giải pháp hoặc cải tiến giải pháp đã lựa chọn
5. Thử nghiệm và đánh giá

❖ **Bài toán phân khúc khách hàng**

Phân khúc khách hàng được hiểu là quá trình mà doanh nghiệp chia nhỏ đối tượng khách hàng của mình thành các nhóm nhỏ dựa trên các đặc điểm tương đồng như sở thích, tính cách, độ tuổi, hành vi, nhu cầu,… Trong quá trình kinh doanh, khách hàng thường có những mong muốn và nhu cầu khác nhau đối với sản phẩm và dịch vụ.

Các phương pháp phân loại khách hàng:

Có nhiều phương pháp để phân loại khách hàng dựa trên các yếu tố như vị trí địa lý, nhân khẩu học, tâm lý học,... Tuy nhiên, dưới tác động của sự toàn cầu hóa thông qua công nghệ, các yếu tố này bị xóa nhòa, mọi khách hàng đều được tiếp cận và truyền thông như nhau. Khi đó, các chỉ số về vị trí địa lý, độ tuổi,... trở nên kém quan trọng. Thay vào đó, hành vi tiêu dùng của khách hàng trở thành một yếu tố quyết định hàng đầu trong việc phân loại họ.

Hơn nữa, việc chia phân khúc khách hàng cũng tùy thuộc vào hành vi của từng lĩnh vực. Phương pháp phổ biến trong phân khúc nhất có lẽ là sử dụng mô hình RFM(Recency - Frequency - Monetary). Cách phân khúc theo mô hình RFM phù hợp với các công ty bán lẻ khách hàng cá nhân (B2C) và kinh doanh nhiều loại sản phẩm khác nhau.

Mô tả mục tiêu:

Mục đích của bài toán phân khúc khách hàng là khi nào cần tập trung khai thác những khách hàng cũ, hoặc tập trung vào tìm kiếm những khách hàng mới . Bài toán chia nhóm khách hàng theo những hành vi khác nhau dựa theo 3 tham số đầu vào là Recency, Frequency, Monetary (lý do chỉ có 3 tham số này là vì bài toán đã được tính toán và đo đạc qua nhiều lần và có độ tương quan tốt nhất trong việc chia khúc khách hàng)

Mô hình phân loại khách hàng chủ yếu chia khách hàng thành các nhóm đồng nhất, hiểu rõ đặc điểm của từng nhóm và triển khai các chiến dịch phù hợp theo hành vi của khách hàng. Điều này không chỉ giúp doanh nghiệp kinh doanh hiệu quả hơn mà còn làm cho khách hàng cảm thấy được quan tâm và phục vụ tốt hơn, từ đó tăng cường tình cảm và sự tín nhiệm vào thương hiệu.

Các nhà quản lý khách hàng cần hiểu tầm quan trọng của khách hàng và sử dụng dữ liệu lịch sử để nâng cao hiểu biết về khách hàng. Thay vì chỉ tập trung vào việc giới thiệu sản phẩm mới, họ cần duy trì và xây dựng mối quan hệ với khách hàng hiện có.

Cách giải quyết bài toán:

Đầu tiên, sẽ tạo bảng biến lấy ra những khách hàng đã mua (không lấy những khách hàng mới) bởi vì những khách hàng mới họ không có những hành vi nên không phân tích và sẽ chỉ phân tích những khách hàng đã mua hàng trong bảng Sale. Vì bài toán đã được chứng minh, đo đạc và rút ra 3 biến. Còn các bài toán lớn hơn như credit scorecard thì sẽ phải tạo ra hàng nghìn biến.

Các biến này là các nhân tố có thể ảnh hưởng tới biến đích, rồi sau đó rút bớt đi và thường được làm ở trên SQL hay gọi là tạo cơ sở dữ liệu các đặc trưng chung (feature store) của doanh nghiệp . Doanh nghiệp họ sẽ thường tính toán sẵn feature store chứa hàng nghìn biến để phục vụ cho nhiều bài toán khác nhau và sau này khi phân tích không cần phải chuẩn hóa lại mà sẽ dùng luôn những bảng biến đó. Các bảng biến trong phân tích dữ liệu được quy về một dạng biến theo khách hàng (lấy khách hàng làm gốc) là dạng customer centric (một chiến lược kinh doanh tập trung vào việc đáp ứng nhu cầu của khách hàng) và tất cả những biến số đó sẽ quy về theo đầu khách hàng.

Lý thuyết áp dụng:

Mô hình RFM là một kĩ thuật phân khúc khách hàng dựa trên hành vi giao dịch của khách hàng trong quá khứ. Mô hình này rất phổ biến vì tính khách quan từ việc mô tả ngắn gọn về khách hàng, đơn giản, dễ sử dụng, trực quan và dễ giải thích. Mô hình RFM [5], đánh giá giá trị của một khách hàng dựa trên ba yếu tố:

- Recency (Gần đây): Khoảng thời gian từ lần mua hàng gần nhất đến thời điểm hiện tại. Chỉ số này cho biết khách hàng có đang hoạt động không và cảnh báo xu hướng rời bỏ nếu chỉ số này lớn.
- Frequency (Tần suất): Số lần khách hàng mua hàng. Tần suất cao cho thấy giá trị doanh số lớn, nhưng cần xem xét cùng với giá trị đơn hàng.
- Monetary (Giá trị): Tổng số tiền khách hàng đã chi tiêu. Đây là yếu tố trực quan nhất ảnh hưởng tới doanh số.

Hiện nay, nhiều nghiên cứu hiện tại đã cung cấp một cái nhìn tổng quan về phân khúc khách hàng như nghiên cứu sử dụng mô hình RFM kết hợp với thuật toán K-means [6], hay nghiên cứu dựa trên kỹ thuật định tính và định lượng kết hợp thuật toán phân cụm K-means [7] để xác định các phân khúc khách hàng dựa trên các yếu tố nhân khẩu học và hành vi. Tuy nhiên, những nghiên cứu này hầu như đều gặp hạn chế trong việc tìm được nhóm khách hàng quan trọng. Lý do là khi mô hình RFM kết hợp với thuật toán K-means sẽ tập trung vào việc phân chia khách hàng thành các nhóm dựa trên dữ liệu lịch sử mua hàng, nhưng sẽ bỏ qua tầm quan trọng tương đối của từng nhóm khách hàng đối với tổng doanh thu của doanh nghiệp.

Hướng phân tích bài toán:

Theo nguyên lý pareto hay còn gọi là quy luật (80/20), 20% khách hàng sẽ mang lại 80% doanh số [8]. Do đó doanh nghiệp cần xác định được những nhóm khách hàng quan trọng nhất để chăm sóc đặc biệt.

- Nếu 20% khách hàng tốt nhất đem lại < 80% tổng doanh thu, doanh nghiệp nên tập trung vào chăm sóc khách hàng hiện hữu.
- Nếu 20% khách hàng tốt nhất đem lại > 80% tổng doanh thu, doanh nghiệp nên tìm kiếm thêm khách hàng mới.

Cách đánh giá khách hàng:

Để đánh giá khách hàng, sử dụng ba chỉ số RFM và thuật toán K-Mean hỗ trợ phân loại, tính điểm gãy WCSS (Elbow) để xác định số phân khúc hợp lý. Tuy nhiên, bản đồ án này tôi đang mong muốn đề ra chiến lược phân khúc khách hàng theo phân tích Pareto thì việc áp dụng thuật toán học máy gây khó khăn cho việc việc áp dụng, do đó với bài này các biến RFM sẽ chủ yếu được phân tích theo khoảng phân vị. Trong phân tích Pareto, thường chia mỗi biến thành 5 khoảng phân vị, có thể chia 4 hoặc 10 khoảng tùy theo cách phân tích.

Để phân loại khách hàng theo mô hình RFM, mỗi tiêu chí Recency, Frequency và Monetary được chia thành 5 nhóm phân vị. Điều này dẫn đến tổng

cộng 125 (5x5x5) phân khúc khách hàng khác nhau. Ví dụ khách hàng được đánh giá theo mô hình RFM sẽ được chia thành các nhóm phân vị cho từng tiêu chí:

- Recency (R): Khoảng thời gian từ lần mua hàng gần đây nhất.
- Frequency (F): Tần suất mua hàng.
- Monetary (M): Tổng số tiền chi tiêu.

Mỗi tiêu chí được chia thành 5 nhóm phân vị như (bảng 1.2) dưới đây:

Bảng 1. 2 Ví dụ chia phân vị

Recency (R)	**Frequency (F)**	**Monetary (M)**
Phân vị 1	Phân vị 1	Phân vị 1
Phân vị 2	Phân vị 2	Phân vị 2
Phân vị 3	Phân vị 3	Phân vị 3
Phân vị 4	Phân vị 4	Phân vị 4
Phân vị 5	Phân vị 5	Phân vị 5

Theo phân vị được chia như trong bảng trên và giả sử chấm điểm ví dụ một khách hàng:

- Khách hàng thuộc nhóm 20% người mua gần đây nhất (R = 5).
- Khách hàng thuộc nhóm 20% người mua số lượng nhiều nhất (F = 5).
- Khách hàng thuộc nhóm 20% người chi tiêu nhiều nhất (M = 5).

Phân khúc RFM của khách hàng này là 5-5-5, tức là thuộc nhóm khách hàng tốt nhất.

Triển khai và áp dụng kết quả:

Sau khi mô hình RFM đáp ứng các tiêu chí kỹ thuật, triển khai thử nghiệm A/B để đánh giá hiệu quả thực tế và rút ra kinh nghiệm về cách tiếp cận từng nhóm khách hàng. Các nhóm khách hàng có thể có sự ưa thích khác nhau đối với các hình thức ưu đãi và kênh truyền thông.

Thử nghiệm A/B sẽ giúp xác định những nhóm khách hàng nhạy cảm nhất với các chiến lược tiếp thị hoặc kênh truyền thông cụ thể. Các yếu tố cần xem xét

khi thiết kế chương trình tiếp cận bao gồm: Đặc điểm của các phân khúc khách hàng được tác động, thời gian tiến hành A/B testing, nội dung ưu đãi (nếu có), thông điệp truyền tải, kênh tiếp cận (online/offline), phương tiện sử dụng (email, tin nhắn, thông báo qua ứng dụng, quảng cáo mục tiêu, v.v.), phương án tiếp cận (cá nhân hoặc gửi hàng loạt), thời điểm và tần suất tương tác.

❖ **Bài toán phân tích bán chéo**

Bài toán phân tích dữ liệu gợi ý sản phẩm bán chéo là một bài toán quen thuộc và được áp dụng rộng rãi. Nhiệm vụ của bài toán này là dự đoán và gợi ý các sản phẩm có khả năng được mua cùng nhau bởi các khách hàng. Ví dụ, khi một khách hàng mua sản phẩm A, hệ thống phân tích dữ liệu sẽ gợi ý những sản phẩm B, C, D... mà thường được mua kèm với sản phẩm A.

Ứng dụng trong lĩnh vữc:

Thương mại điện tử như gợi ý sản phẩm trên các sàn thương mại điện tử. Hoặc trong ngành ngân hàng như gợi ý bán chéo các sản phẩm bảo hiểm, mở thẻ tín dụng, hoặc mở tài khoản tiết kiệm. Ngoài ra trong ngành giải trí cũng được ứng dụng như gợi ý các bài báo, bộ phim liên quan trên các trang web trực tuyến. Các lĩnh vực khác như ứng dụng trong các trang web bán lẻ, dịch vụ trực tuyến,...

Các phương pháp phổ biến:

Phương pháp phổ biến để thực hiện bài toán gợi ý là xây dựng mô hình đề xuất gợi ý (recommender system). Đây là hệ thống lọc thông tin, giúp đưa ra đề xuất sản phẩm hoặc nội dung cho khách hàng. Có một số cách tiếp cận phổ biến để thực hiện bài toán này:

- Sử dụng luật kết hợp (Association rule): Tìm ra mối liên hệ giữa các cặp sản phẩm thường được bán kèm với nhau.
- Collaborative filtering : Mô hình lọc ra các khách hàng có đặc điểm và xu hướng tương tự nhau, từ đó đề xuất chéo các ưu đãi cho khách hàng.
- Sử dụng phương pháp Deep Learning: Phương pháp này phức tạp và trong phần này sẽ không cập nhật chi tiết về phương pháp này.

Bán chéo là bài toán có tính ứng dụng rất cao trong các doanh nghiệp trong ngành bán lẻ. Các hệ thống gợi ý sử dụng luật kết hợp có thể áp dụng trong nhiều lĩnh vực khác nhau để tăng doanh thu và nâng cao trải nghiệm khách hàng. Để giải quyết bài toán này, sẽ sử dụng quy tắc kết hợp sản phẩm (association rule). Do đó, trong phần này sẽ chỉ mô tả hướng giải quyết bài toán áp dụng quy tắc kết hợp. Còn phương pháp collaborative filtering sẽ được trình bày trong phần (phụ lục1).

Phương pháp sử dụng luật kết hợp (Association rule)

Mỗi quy tắc kết hợp sản phẩm (association rule)[9] bao gồm danh sách sản phẩm tiền đề (antecedent) và danh sách sản phẩm hệ quả (consequent). Các quy tắc được đánh giá dựa trên ba hệ số đo lượng bao gồm support, confidence, lift. Công thức cụ thể cho association rule như sau:

Xét một quy tắc kết hợp $\{X\} \rightarrow \{Y\}$, X và Y là các hạng mục (itemset) với X là danh sách sản phẩm tiền đề và Y là danh sách sản phẩm hệ quả. Quy tắc này có thể hiểu theo hướng nếu khách hàng mua các sản phẩm trong X, thì họ có khả năng sẽ mua các sản phẩm trong Y. Vậy 3 hệ số được tính như sau:

Hệ số Support: Thể hiện tần suất xuất hiện của itemset trong dữ liệu quan sát, được tính theo công thức như sau.

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Tổng\ số\ bản\ ghi\ chứa\ X\ và\ Y}{Tổng\ số\ bản\ ghi} \qquad (1.1)$$

Hệ số Support có giá trị nằm trong đoạn $[0, 1]$. Hệ số càng cao thể hiện số lượng quan sát đối có chứa bộ quy tắc càng lớn, thể hiện mức độ tin cậy của bộ quy tắc càng lớn và ngược lại (nếu hệ số này quá nhỏ, thì sẽ có rất ít bản ghi chứa bộ quy tắc đang xét, và khả năng bộ quy tắc được xuất hiện do ngẫu nhiên càng cao, tính tin cậy thấp). Ví dụ tính hệ số Support: Nếu 50 trong số 1000 giao dịch chứa sản phẩm X và Y, dựa theo công thức (1.1) thì hệ số support của X -> Y là 0.05 (5%).

Hệ số Confidence: Thể hiện tần suất xuất hiện của sản phẩm Y trong itemset, được tính toán với công thức như sau.

$$Confidence(\{X\} \rightarrow \{Y\}) = \frac{Tổng\ số\ bản\ ghi\ chứa\ X\ và\ Y}{Tổng\ số\ bản\ ghi\ chứa\ X} \qquad (1.2)$$

Hệ số Confidence có giá trị nằm trong đoạn $[0, 1]$, giá trị Confidence càng cao thể hiện mức độ liên kết giữa X và Y càng lớn. Trong đó, nếu giá trị này là 1 thì 100% khi khách hàng sử dụng các sản phẩm X, sẽ sử dụng đồng thời sản phẩm Y. Ví dụ tính hệ số Confidence: nếu 50 trong số 200 giao dịch chứa sản phẩm X cũng chứa sản phẩm Y, dựa theo công thức (1.2) thì confidence của X -> Y là 0.25 (25%).

Hệ số Lift: Hệ số này so sánh khả năng mua Y kèm điều kiện đã mua X với khả năng mua Y không kèm điều kiện, tức việc mua X khiến khả năng mua Y thay đổi như thế nào.

$$\text{Lift}(\{X\} \rightarrow \{Y\}) = \frac{\text{Số lượng bản ghi chứa X và Y} / \text{Số lượng bản ghi chứa X}}{Số\ lượng\ bản\ ghi\ chứa\ Y / Tổng\ số\ bản\ ghi} \qquad (1.3)$$

Hệ số Lift có giá trị nằm trong khoảng $[0, +\infty)$. Giá trị Lift bằng 1 thể hiện X và Y không có ảnh hưởng đến nhau, nếu:

- Giá trị Lift nhỏ hơn 1 thể hiện ảnh hưởng tiêu cực của hai nhóm hàng, khi khách hàng mua X thì xu hướng mua Y giảm đi.
- Giá trị Lift lớn hơn 1 thể hiện ảnh hưởng tích cực của hai nhóm hàng, khi mua X thì xu hướng mua Y tăng lên.

Ví dụ tính hệ số Lift: Dựa theo công thức (1.3) ở trên Nếu confidence của X -> Y là 0.25 và tỉ lệ giao dịch chứa sản phẩm Y là 0.1 (10%), thì lift của X -> Y là 2.5.

Kết quả đầu ra mô hình là một số các bộ quy tắc kết hợp giữa sản phẩm tiền đề và sản phẩm hệ quả tốt nhất, thỏa mãn các tiêu chí sau:

- Các quy tắc được chọn phải có giá trị support, confidence và lift đủ cao để đảm bảo tính hữu ích và sự liên kết chặt chẽ giữa các sản phẩm.
- Số lượng sản phẩm vừa phải: Quy tắc bao gồm 2 - 3 sản phẩm để dễ dàng triển khai trong các chiến dịch Marketing.

- Sử dụng rộng rãi và liên kết tốt: Các sản phẩm trong quy tắc phải được một lượng lớn người dùng sử dụng để đảm bảo tính phổ biến, đồng thời các sản phẩm phải có sự liên kết tốt để xây dựng các chương trình Marketing hiệu quả.

Sử dụng kết quả đầu ra bằng cách đánh giá các quy tắc này bởi các chuyên có chuyên môn xây dựng thông điệp Marketing tới khách hàng. Các quy tắc có thể được triển khai dưới dạng: Thông báo trên Web và App, gửi email hoặc tin nhắn, giới thiệu trực tiếp bởi nhân viên kinh doanh. Sau khi xác nhận được nhóm các sản phẩm có tính liên kết mạnh và thỏa mãn các tiêu chí trên. Đơn vị kinh doanh sẽ thực hiện các chiến dịch Marketing để quảng cáo sản phẩm tới khách hàng. Mục tiêu là cải thiện hiệu quả trong việc thu hút khách hàng thực hiện giao dịch.

Phương án tiếp cận

Trên thực tế dự án bán chéo sản phẩm sẽ là một quy trình tối ưu trải nghiệm trong một hành trình khách hàng (Customer journey) và sẽ thường kết hợp với các bài toán khác để tối ưu hiệu quả, ví dụ :

Bước 1: Phân tích dữ liệu khách hàng để xây dựng các quy tắc bán chéo (association rule), tìm kiếm mối liên kết giữa các sản phẩm, từ đó đề xuất các chương trình tiếp thị, quảng cáo bán chéo sản phẩm tới khách hàng dựa trên các quy tắc này.

Bước 2: Chạy chiến dịch các chương trình tiếp thị trong khoảng 6 tháng, gán nhãn khách hàng sau đó dựa vào nhãn khách hàng phản hồi từ chương trình tiếp thị bán chéo, xây dựng tiếp mô hình dự đoán khả năng khách hàng chấp nhận tiếp thị quảng cáo. Bài toán kết hợp xây dựng mô hình dự đoán khách hàng chấp nhận tiếp thị quảng cáo sản phẩm được mô tả trong (phụ lục 2).

Bước 3: Chạy tiếp mô hình dự đoán khách hàng chấp nhận tiếp thị khoảng 6 tháng sau đó lại gán nhãn, xây dựng bài toán tối ưu nên tiếp thị quảng cáo sản

phẩm nào tới các khách hàng như thế nào. Bài toán này được mô tả chi tiết trong (phụ lục 3).

Bước 4: Xây dựng bài toán dự đoán các khách hàng nào sẽ có nhu cầu mua hàng trong vòng 6 tháng tới để có thể tiếp thị quảng cáo chính xác sản phẩm tới các khách hàng có nhu cầu.

Việc triển khai bán chéo kết hợp các bài toán tối ưu trải nghiệm của khách hàng để tránh tình trạng Spam quảng cáo và tiếp thị quảng cáo bán chéo tới các khách hàng không có nhu cầu, gây khó chịu và tốn nguồn lực triển khai mà không đạt hiệu quả cao. Phần lớn ở đây đều là bài toán phân loại quen thuộc.

Phương án đánh giá

Phương án đánh giá hiệu quả sẽ được thực hiện thông qua A/B testing, với tiêu chí chính là tỷ lệ chuyển đổi từ khách hàng được tiếp thị, quảng cáo bán chéo sang khách hàng mua sản phẩm đó. Để triển khai A/B testing chiến dịch Marketing bán chéo, quá trình được thực hiện trên hai tập khách hàng khác nhau:

- Tập khách hàng được chỉ định sẵn bởi bộ phận phân tích: Tập này gồm các khách hàng đã được xác định có khả năng cao chấp nhận tiếp thị quảng cáo dựa trên phân tích dữ liệu.
- Tập khách hàng ngẫu nhiên: Tập này gồm các khách hàng được chọn ngẫu nhiên mà không dựa trên phân tích trước.

Tỷ lệ chuyển đổi của hai tập khách hàng được đo cụ thể như sau:

$$Tỷ\ lệ\ chuyển\ đổi = \frac{Số\ lượng\ KH\ sử\ dụng\ làm\ sản\ phẩm}{Số\ lượng\ KH\ nhận\ quảng\ cáo/marketing} \quad (1.4)$$

Việc đánh giá hiệu quả của mô hình sẽ được thực hiện bằng cách so sánh tỷ lệ chuyển đổi giữa hai tập khách hàng này.

Vậy trong phần này đã xác định được các chỉ số chính của 3 trang báo cáo. Ngoài ra, mỗi bài toán đều được trình bày theo hướng là một giải pháp hoàn chỉnh. Qua đó có thể làm rõ được vấn đề của mỗi bài toán.

1.3.3 Giải pháp bảo mật và nâng cao hiệu xuất xử lý dữ liệu

1. Giải pháp bảo mật dữ liệu

- Xác thực và ủy quyền: Sử dụng thư viện Microsoft.Identity.Web để xác minh danh tính và phân quyền (Authentication and Authorizationcho người dùng. Thư viện này giúp quản lý và bảo vệ quá trình xác minh danh tính, đảm bảo chỉ những người dùng có quyền mới có thể truy cập vào dữ liệu.
- Mã thông báo chỉ dành cho ứng dụng (App-only Access Tokens): Sử dụng các mã thông báo truy cập chỉ dành cho ứng dụng để gọi Power BI Service API nhằm đảm bảo chỉ có ứng dụng đã được xác thực mới có thể truy cập vào Power BI Service, giảm nguy cơ truy cập trái phép.
- Azure Active Directory (Azure AD): Sử dụng Azure AD để quản lý người dùng và nhóm, cung cấp một lớp bảo mật bổ sung bằng cách giới hạn quyền truy cập vào các tài nguyên chỉ cho những người dùng và nhóm đã được xác định trước.

2. Giải pháp công nghệ nâng cao hiệu suất xử lý dữ liệu

Để nâng cao hiệu suất xử lý dữ liệu lướn trong Power BI có thể áp dụng các giải pháp công nghệ sau:

- Dataflow và Datamart: Sử dụng Dataflow và Datamart trong Power BI để chuyển đổi, làm sạch và chuẩn bị dữ liệu trước khi nhập vào mô hình giúp giảm tải công việc trên Power BI và cải thiện hiệu suất xử lý.
- DirectQuery và Import Mode: Sử dụng chế độ DirectQuery để truy vấn dữ liệu trực tiếp từ nguồn dữ liệu mà không cần tải về Power BI, giúp giảm tải dữ liệu lớn. Trong khi đó, chế độ Import có thể tải và nén dữ liệu vào Power BI để tăng tốc độ truy vấn.
- Azure Synapse Analytics: Kết hợp Power BI với Azure Synapse Analytics để xử lý dữ liệu lớn. Azure Synapse cung cấp một nền tảng phân tích tích hợp giúp xử lý và phân tích dữ liệu lớn nhanh chóng và hiệu quả.
- Sử dụng các phương pháp làm sạch dữ liệu lớn trước khi tiến hành trực quan hóa dữ liệu điển hình là sử dụng ngôn ngữ python và jupyter notebook.

1.3.4 Công nghệ sử dụng

Nền tảng .NET 5 và mô hình MVC

Ứng dụng được xây dựng trên nền tảng .NET 5, một nền tảng phát triển đa nền tảng của Microsoft, cho phép xây dựng các ứng dụng web, di động, máy tính để bàn và đám mây. Với .NET 5, ứng dụng có thể tận dụng hiệu năng cao, bảo mật và tính nhất quán trong phát triển và triển khai. Mô hình MVC (Model-View-Controller) giúp tách biệt rõ ràng giữa giao diện người dùng, logic xử lý và dữ liệu, làm cho ứng dụng dễ bảo trì và mở rộng.

Công nghệ phía máy khách: JavaScript, TypeScript và Node.js

Phía máy khách của ứng dụng sử dụng JavaScript và TypeScript. Ban đầu, mã JavaScript được sử dụng để nhúng báo cáo Power BI vào trang web. Tuy nhiên, để tận dụng lợi thế của IntelliSense (tính năng cung cấp gợi ý mã về các hàm, biến và kiểu dữ liệu) và kiểm tra loại thời gian biên dịch, mã sẽ được chuyển sang TypeScript. Node.js và webpack được sử dụng để xây dựng và quản lý các mô-đun TypeScript, đảm bảo mã nguồn được sắp xếp gọn gàngvà dễ dàng bảo trì.

Công cụ phát triển

Các công cụ phát triển như Visual Studio Code và Visual Studio 2019 cung cấp một môi trường phát triển mạnh mẽ và linh hoạt. Visual Studio Code, với khả năng mở rộng mạnh mẽ và hỗ trợ nhiều ngôn ngữ lập trình, là công cụ lý tưởng cho việc phát triển mã nguồn TypeScript và JavaScript. Visual Studio 2019, với tích hợp sâu với các công nghệ của Microsoft, là công cụ lý tưởng cho việc phát triển và gỡ lỗi các ứng dụng .NET.

1.4. Power BI

1.4.1. Giới thiệu công cụ Power BI

Power BI là một công cụ phân tích và trực quan hóa dữ liệu mạnh mẽ của Microsoft. Power BI Desktop là phiên bản cài đặt trên máy tính giúp người dùng thiết kế và tạo báo cáo, trong khi Power BI Service là nền tảng trực tuyến cho

phép chia sẻ và xem báo cáo. Power BI giúp doanh nghiệp dễ dàng kết nối với nhiều nguồn dữ liệu, tạo ra các báo cáo và dashboard tùy chỉnh.

Power BI hỗ trợ kết nối với nhiều nguồn dữ liệu khác nhau như Excel, SQL Server, SharePoint, Azure, Google Analytics, và nhiều nguồn dữ liệu khác. Điều này giúp doanh nghiệp dễ dàng tích hợp và phân tích dữ liệu từ nhiều nguồn khác nhau, tạo ra cái nhìn toàn diện về hoạt động kinh doanh.

1.4.2. Thành phần kiến trúc Power BI

Power BI là một bộ công cụ phân tích kinh doanh mạnh mẽ được phát triển bởi Microsoft, cho phép người dùng kết nối với nhiều nguồn dữ liệu khác nhau, xử lý và trực quan hóa dữ liệu. Các thành phần chính của Power BI bao gồm :

Power BI Desktop: Là ứng dụng cài đặt trên máy tính cá nhân, cho phép người dùng thiết kế, phát triển và tạo ra các báo cáo dữ liệu. Đây là công cụ mạnh mẽ để xử lý, phân tích và trực quan hóa dữ liệu với giao diện thân thiện và dễ sử dụng.

- Power BI Service: Là nền tảng trực tuyến (SaaS) cho phép người dùng lưu trữ, chia sẻ và cộng tác trên các báo cáo và dashboard. Power BI Service hỗ trợ truy cập báo cáo từ bất kỳ đâu, chỉ cần có kết nối internet, giúp tối ưu hóa việc chia sẻ và hợp tác giữa các thành viên trong tổ chức.
- Power BI Mobile: Ứng dụng di động cho phép người dùng truy cập và tương tác với các báo cáo và dashboard trên thiết bị di động.
- Power BI Report Server: Máy chủ tại chỗ (on-premises) cho phép lưu trữ và quản lý các báo cáo Power BI trong môi trường nội bộ của doanh nghiệp.
- Power BI Embedded: Dịch vụ Azure cho phép nhúng các báo cáo và dashboard Power BI vào các ứng dụng tùy chỉnh.

Power BI cung cấp các công cụ và tính năng phù hợp với nhiều vai trò khác nhau trong tổ chức. Đối với nhà quản lý, Power BI cung cấp các báo cáo và dashboard tổng quan giúp theo dõi hiệu suất kinh doanh. Đối với nhà phân tích dữ liệu, Power BI cung cấp các công cụ phân tích mạnh mẽ để khám phá dữ liệu

và phát hiện các mẫu hình. Đối với nhân viên, Power BI cung cấp các báo cáo chi tiết giúp hỗ trợ công việc hàng ngày.

Các luồng tác vụ chính trong Power BI như trực quan hóa dữ liệu, quản lý bộ dữ liệu, tạo báo cáo, và xây dựng dashboard. Trực quan hóa dữ liệu giúp biến đổi dữ liệu thành các biểu đồ và bảng biểu, bộ dữ liệu là nơi lưu trữ và quản lý dữ liệu, báo cáo là tài liệu chứa các phân tích chi tiết, và dashboard là bảng điều khiển tổng quan giúp theo dõi các chỉ số quan trọng.

Các luồng tác vụ chính trong phân tích dữ liệu cũng vậy, tuy nhiên sẽ có phần khác với các tác vụ trong Power BI như từ kết nối, khai phá, tính toán các chỉ số cần thiết cho báo cáo trực quan hóa, đồng thời sẽ có những biểu đồ để biểu diễn các chỉ số đó.

Để giúp giúp người dùng dễ dàng tạo và tùy chỉnh báo cáo theo nhu cầu Power BI cũng cung cấp nhiều mẫu báo cáo sẵn. Các mẫu báo cáo bao gồm báo cáo tài chính, báo cáo bán hàng, báo cáo tiếp thị, và nhiều loại báo cáo khác. Người dùng có thể sử dụng các mẫu này để nhanh chóng tạo ra các báo cáo chuyên nghiệp và hữu ích.

Ngoài ra, trong Power BI là một phần không thể thiếu để đảm bảo tính bảo mật và chính xác của dữ liệu trong quá trình phân tích và trình bày thông tin. Row-level Security (RLS) là tính năng cho phép kiểm soát quyền truy cập dựa trên hàng (row) trong bảng dữ liệu. Với RLS, quản trị viên có thể đảm bảo rằng người dùng chỉ có thể truy cập và thấy dữ liệu mà họ được phép.

1.4.3. Mô hình hóa dữ liệu trong Power Bi

Mô hình hóa dữ liệu trong Power BI là quá trình tạo ra các mô hình dữ liệu bằng cách xác định cấu trúc, mối quan hệ giữa các bảng dữ liệu, và các biện pháp tính toán nhằm tối ưu hóa việc phân tích và trực quan hóa dữ liệu. Quá trình này giúp tổ chức dữ liệu một cách hợp lý, từ đó người dùng có thể dễ dàng truy vấn, xử lý và phân tích dữ liệu để đưa ra các quyết định kinh doanh. Dưới đây là quy trình mô hình hóa dữ liệu:

Bước 1: Kết nối và nhập dữ từ các nguồn khác nhau vào Power BI. Các nguồn dữ liệu phổ biến bao gồm: Excel, Cơ sở dữ liệu SQL Server hoặc các dịch vụ web và API như Google Analytics, Azure, SharePoint, và nhiều nguồn khác

Bước 2: Xử lý và chuẩn hóa dữ liệu để đảm bảo tính nhất quán và độ tin cậy. Các bước xử lý bao gồm:

- Loại bỏ dữ liệu trùng lặp: Xác định và loại bỏ các bản ghi trùng lặp.
- Xử lý các giá trị thiếu: Điền vào các giá trị thiếu hoặc loại bỏ các bản ghi không đầy đủ.
- Chuẩn hóa định dạng dữ liệu: Đảm bảo rằng tất cả các dữ liệu có định dạng thống nhất.

Bước 3: Tạo mối quan hệ giữa các bảng dữ liệu, kết nối các bảng dữ liệu có liên quan để tạo ra một mô hình dữ liệu toàn diện. Cho phép Power BI tự động kết hợp dữ liệu từ các bảng liên quan khi thực hiện các truy vấn và phân tích.

Bước 4: Tạo các hàm column DAX và DAX Measure để tính toán các chỉ số quan trọng trong mô hình hóa dữ liệu, giúp thực hiện các phép tính phức tạp và tạo ra các chỉ số quan trọng

1.5. Tổng kết chương

Trong chương này, đã thảo luận về hai khía cạnh chính là phân tích dữ liệu bán hàng và công nghệ Power BI. Bằng cách tổng hợp thông tin từ hai lĩnh vực này đã tạo nên cơ sở lý luận và hiểu biết sâu sắc hơn về quy trình và công cụ cần thiết cho việc phân tích dữ liệu trong lĩnh vực bán hàng . Các kiến thức và thông tin thu thập từ chương này sẽ là cơ sở cho việc phân tích, thiết kế và xây dựng hệ thống, cũng như thực hiện các thử nghiệm và đánh giá kết quả để đảm bảo rằng ứng dụng đáp ứng được các yêu cầu và mục tiêu đặt ra.

Chương 2
PHÂN TÍCH THIẾT KẾ HỆ THỐNG

2.1. Đặc tả yêu cầu

Dựa trên mục tiêu của ứng dụng là cung cấp phân tích chuyên sâu làm nổi bật các số liệu và xu hướng kinh doanh quan trọng, xác định các khu vực cần tối ưu hóa và đưa ra quyết định sáng suốt thông qua trang báo cáo tổng quan sử dụng Power BI trong ngành bán hàng. Quyển đồ án này sẽ khai thác khả năng của Power BI để trao quyền cho các bên liên quan tức là những người sử dụng trang báo cáo tổng quan bằng một nền tảng động để phân tích dữ liệu bán hàng. Do đó sẽ gồm những yêu cầu về chức năng như sau:

2.1.1. Yêu cầu chức năng

- Kết nối dữ liệu: Ứng dụng cần có khả năng kết nối với các nguồn dữ liệu khác nhau như cơ sở dữ liệu, tệp tin Excel, hoặc dịch vụ web để lấy dữ liệu bán hàng.
- Biểu diễn dữ liệu: Ứng dụng cần có khả năng biểu diễn dữ liệu bán hàng dưới dạng các biểu đồ, bảng thống kê và bảng điều khiển tương tác.
- Tích hợp Power BI: Yêu cầu tích hợp công nghệ Power BI vào ứng dụng để hiển thị các báo cáo và bảng điều khiển được tạo trong Power BI Desktop.
- Tương tác người dùng: Ứng dụng cần cho phép người dùng tương tác với các biểu đồ và bảng điều khiển, bao gồm việc lọc, sắp xếp và phân tích chi tiết dữ liệu.

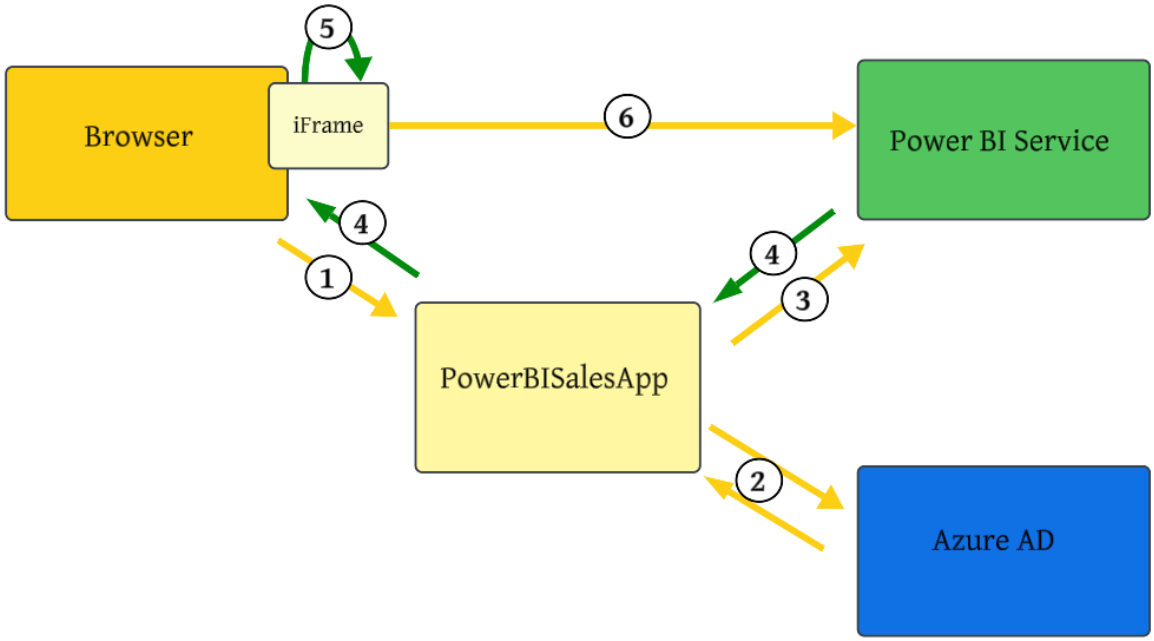
2.1.2. Yêu cầu phi chức năng

- Bảo mật dữ liệu: Yêu cầu ứng dụng đảm bảo bảo mật dữ liệu bán hàng và hạn chế quyền truy cập dựa trên vai trò của người dùng.
- Hiệu suất: Yêu cầu ứng dụng hoạt động một cách mượt mà và hiệu quả, đảm bảo thời gian phản hồi nhanh chóng khi người dùng tương tác.

- Dễ sử dụng: Yêu cầu ứng dụng có giao diện người dùng thân thiện và dễ sử dụng, đảm bảo người dùng có thể tương tác và thao tác dữ liệu một cách dễ dàng và thuận tiện.

2.1.3. Tổng quan hệ thống

Sơ đồ dưới đây cung cấp cái nhìn tổng quan của giải pháp phân tích nhúng sẽ hoạt động như thế nào khi triển khai nhúng Power BI vào ứng dụng web MVC với xác thực Azure.



Hình 2. 1 Quy trình hoạt động phương pháp phân tích nhúng

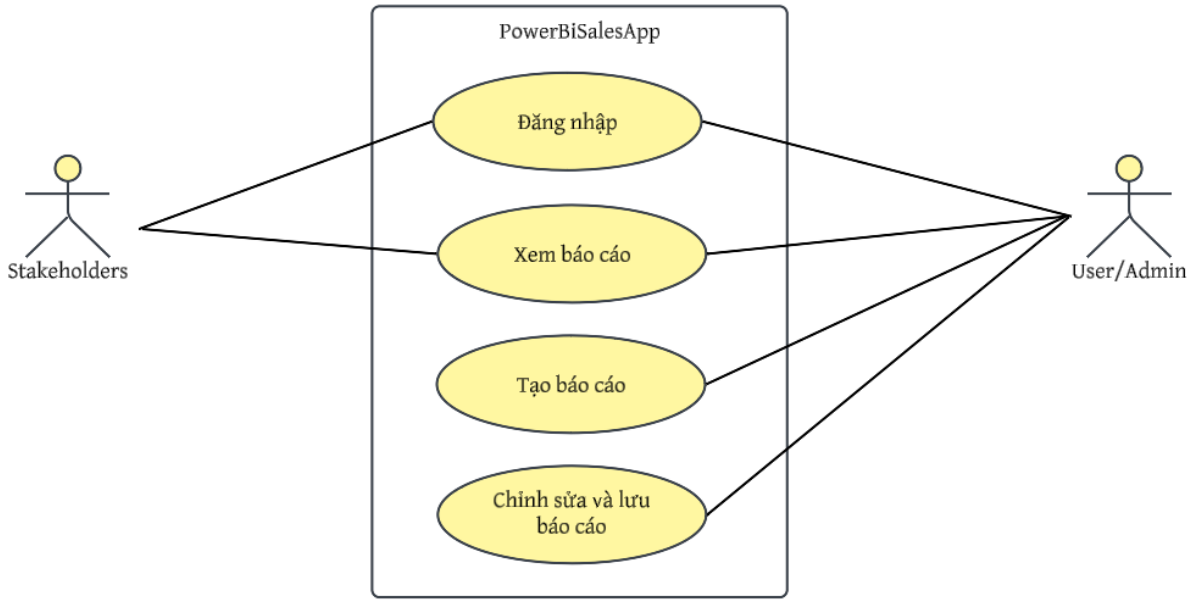
Khi ứng dụng web MVC (đặt tên là PowerBISalesApp) thực hiện nhúng Power BI để hiển thị và tương tác với báo cáo trong giao diện web của ứng dụng bằng giải pháp phân tích nhúng, các bước triển khai phải tuân theo sơ đồ trong hình 1.1 để đảm bảo tài nguyên Power BI được nhúng một cách an toàn, hiệu quả và cung cấp trải nghiệm tốt nhất cho người dùng. Các bước thực hiện trong sơ đồ bắt đầu từ :

1. Người dùng mở trình duyệt và truy cập vào ứng dụng PowerBISalesApp. Ứng dụng này được xây dựng trên nền tảng .NET 5 với kiến trúc MVC.
2. Ứng dụng xác thực người dùng bằng cách sử dụng Azure AD. Khi người dùng đăng nhập thành công, Azure AD cung cấp cho ứng dụng một access

token. Token này sẽ được sử dụng để xác thực các yêu cầu đến Power BI Service API.

3. Ứng dụng sử dụng access token đã nhận được để thực hiện các cuộc gọi đến Power BI Service API. Các cuộc gọi này có thể bao gồm việc lấy thông tin về báo cáo, dashboard, và các nội dung Power BI khác.
4. Ứng dụng lấy dữ liệu cần thiết cho các nội dung được nhúng từ Power BI Service và chuyển dữ liệu này về trình duyệt. Điều này đảm bảo rằng dữ liệu hiển thị trên ứng dụng là mới nhất và chính xác.
5. Mã phía máy khách (trong trình duyệt) sử dụng Power BI JavaScript API để tạo và hiển thị nội dung được nhúng, chẳng hạn như báo cáo hoặc dashboard, trực tiếp trong trang web của ứng dụng. API này cho phép tương tác trực tiếp với các tài nguyên Power BI từ phía máy khách.
6. Một iFrame (khung nội dung nhúng) chứa nội dung nhúng được tạo giữa trình duyệt của người dùng và dịch vụ Power BI Service. IFrame này cho phép truyền tải dữ liệu và tương tác thời gian thực, giúp người dùng có thể tương tác với báo cáo Power BI như nó đang được chạy trực tiếp trên Power BI Service.

2.2. Use case



Hình 2. 2 Biểu đồ ca sử dụng

2.2.1. Đặc tả use case đăng nhập

Bảng 2. 1 Đặc tả use case đăng nhập

Ca sử dụng	**Đăng nhập**
Tác nhân	Các bên liên quan (Stakeholders) hoặc người dùng/ quản trị viên (User/Admin).
Mô tả	Stakeholders hoặc user/admin nhập thông tin đăng nhập và ứng dụng xác thực thông tin với Azure AD.
Luồng sự kiện chính	1. Người dùng đăng nhập tên đăng nhập và mật khẩu vào giao diện đăng nhập 2. Ứng dụng xác thực thông tin người dùng với Azure AD 3. Nếu xác thực thành công, người dùng sẽ được đưa đến trang chủ của ứng dụng

2.2.2. Đặc tả use case xem báo cáo

Bảng 2. 2 Bảng đặc tả use case xem báo cáo

Ca sử dụng	**Xem báo cáo**
Tác nhân	Các bên liên quan (Stakeholders) hoặc người dùng/ quản trị viên (User/Admin).
Mô tả	Stakeholders hoặc user/admin chọn xem báo cáo Power Bi trên ứng dụng.
Luồng sự kiện chính	1. Người dùng chọn một báo cáo từ danh sách báo cáo có sẵn. 2. Ứng dụng gửi yêu cầu đến Power BI Service API để lấy dữ liệu của báo cáo. 3. Ứng dụng hiển thị báo cáo cho người dùng trên giao diện của ứng dụng.

2.2.3. Đặc tả use case tạo báo cáo

Bảng 2. 3 Đặc tả use case tạo báo cáo

Ca sử dụng	**Tạo báo cáo**
Tác nhân	User/admin
Mô tả	User/admin có thể tạo báo cáo mới Power BI trên ứng dụng
Luồng sự kiện chính	1. Người dùng chuyển đổi sang chế độ tạo báo cáo. 2. Ứng dụng cấp quyền tạo tùy chỉnh và hiển thị báo cáo trong chế độ tạo báo mới cho người dùng.

2.2.4. Đặc tả use case chỉnh sửa và lưu báo cáo

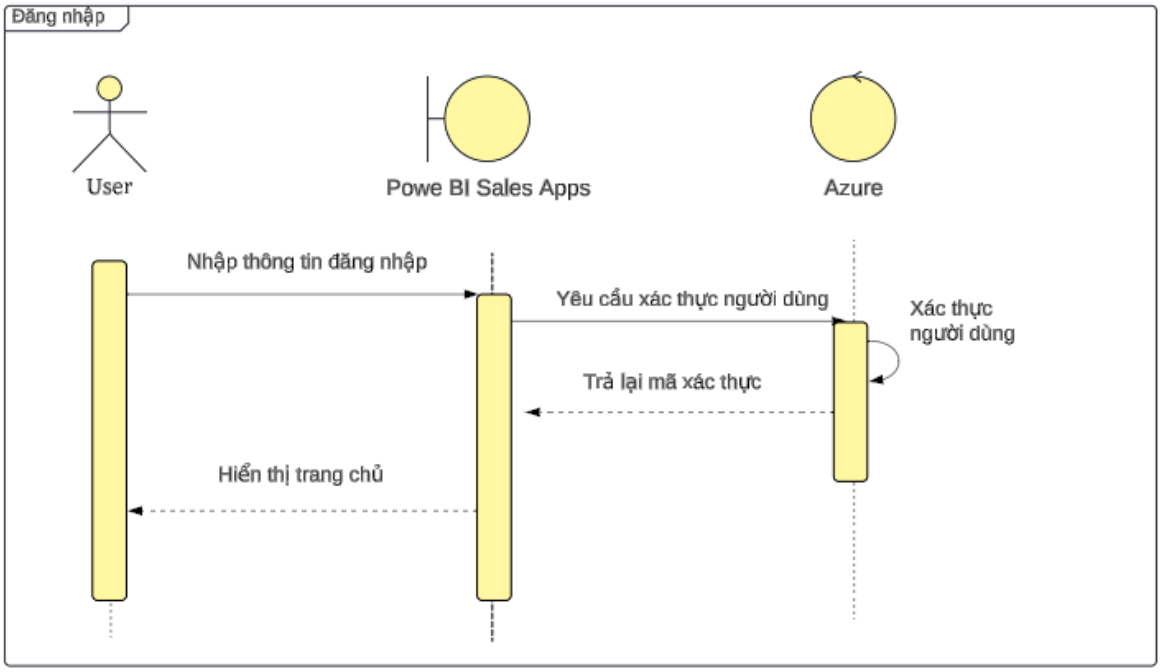
Bảng 2. 4 Đặc tả use case chỉnh sửa và lưu báo cáo

Ca sử dụng	**Chỉnh sửa báo cáo và lưu báo cáo**
Tác nhân	User/admin

Mô tả	User/admin có thể chỉnh sửa và lưu các thay đổi đã thực hiện trên báo cáo Power BI.
Luồng sự kiện chính	1. Người dùng chuyển đổi báo cáo sang chế độ chỉnh sửa. 2. Ứng dụng cấp quyền chỉnh sửa và hiển thị báo cáo trong chế độ chỉnh sửa cho người dùng. 3. Người dùng thực hiện các thay đổi trên báo cáo và nhấn nút lưu. 4. Ứng dụng gửi yêu cầu đến Power BI Service API để lưu báo cáo. 5. Ứng dụng thông báo cho người dùng biết về kết quả lưu.

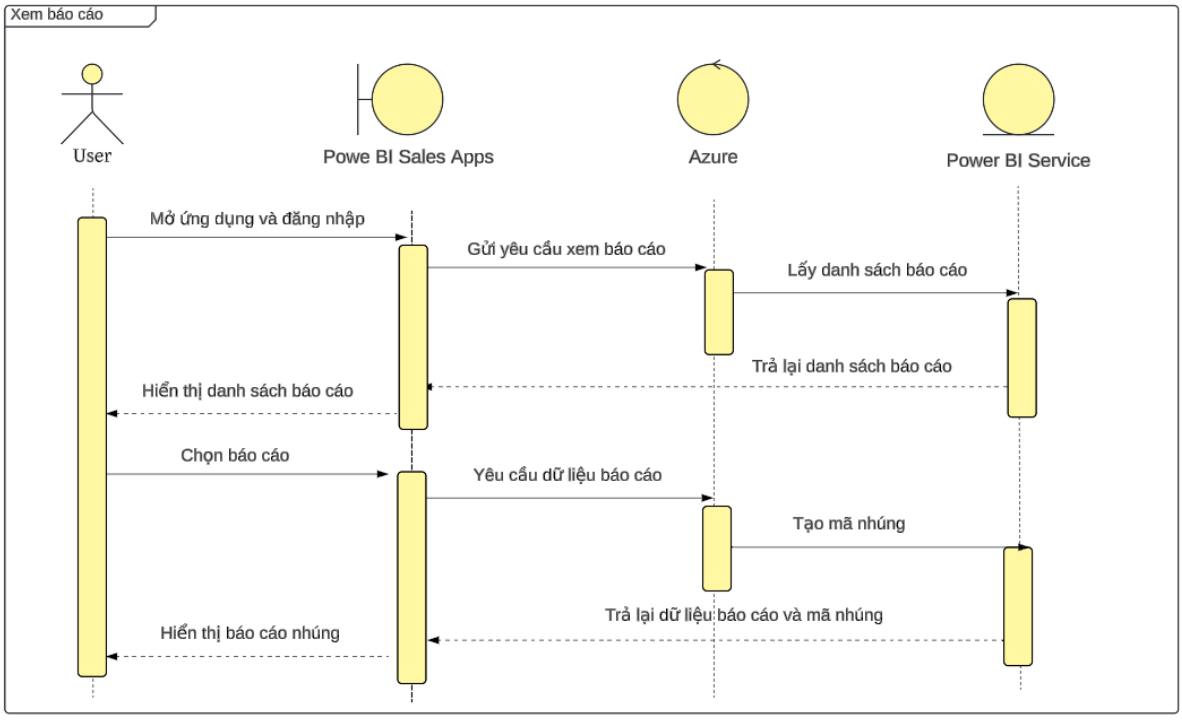
2.3. Sequence/Activity

2.3.1 Sequence chức năng đăng nhập



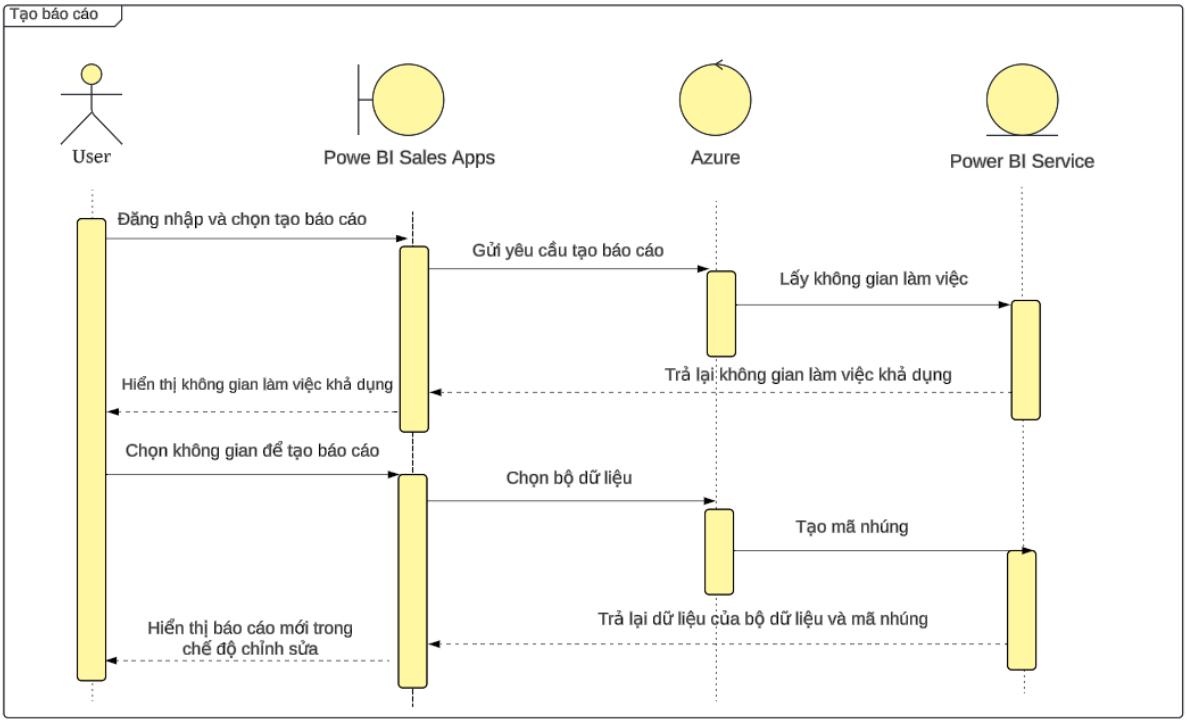
Hình 2. 3 Sequence chức năng đăng nhập

2.3.2 Sequence chức năng xem báo cáo



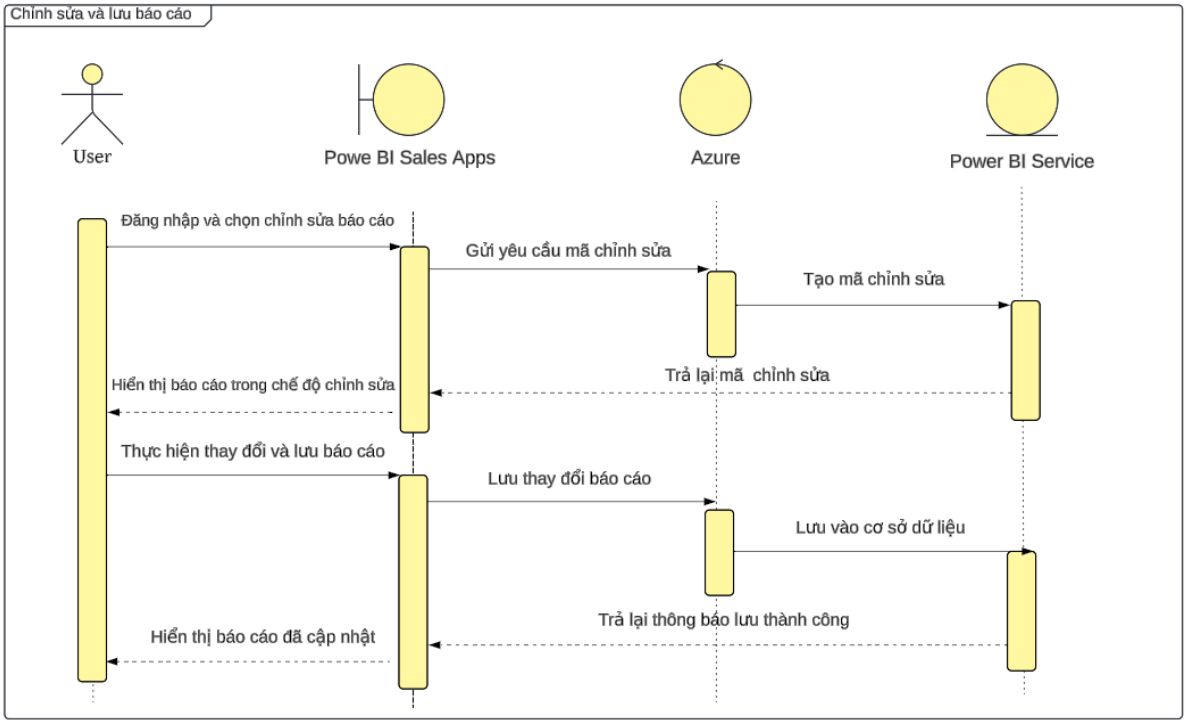
Hình 2. 4 Sequence chức năng xem báo cáo

2.3.3 Sequence chức năng tạo báo cáo



Hình 2. 5 Sequence chức năng tạo báo cáo

2.3.4 Sequence chức năng chỉnh sửa và lưu báo cáo



Hình 2. 6 Sequence chức năng chỉnh sửa và lưu báo cáo

2.4. Tổng kết chương

Trong chương này đã tìm hiểu và đi sâu vào quá trình phân tích và thiết kế hệ thống. Những điểm chính được thảo luận bao gồm: Phân tích yêu cầu, chức năng, thiết kế mô hình hệ thống và mô tả quy trình hoạt động của các thành phần trong hệ thống. Quá trình này không chỉ giúp làm rõ yêu cầu bài toán đặt ra mà còn đảm bảo rằng hệ thống được xây dựng theo một cấu trúc logic và có thể mở rộng. Sự kỹ lưỡng và chính xác trong giai đoạn này sẽ giảm thiểu rủi ro và sai sót trong chương tiếp theo.

Chương 3
XÂY DỰNG VÀ THỬ NGHIỆM HỆ THỐNG

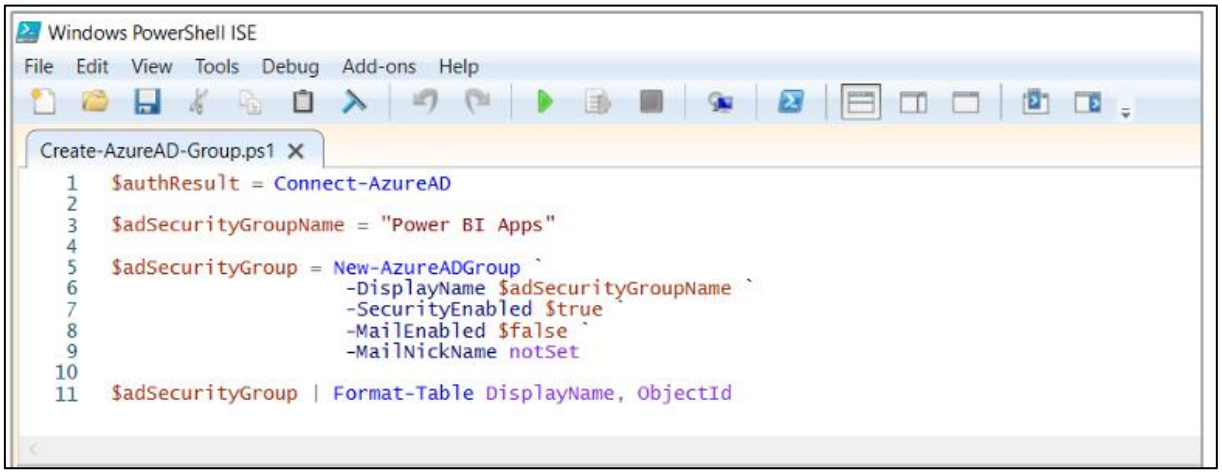
3.1. Xây dựng ứng dụng web MVC

Để tạo ứng dụng web MVC.NET5 cho phép nhúng báo cáo Power BI cần thiết lập môi trường phát triển, cấu hình xác thực và viết mã cả phía máy chủ và máy khách. Đầu tiên là những thiết lập và công cụ phần mềm cần thiết để xây dựng ứng cần phải đáp ứng những yêu cầu sau:

1. Cài đặt phần mềm và công cụ cần thiết:
    - Thư viện tập lệnh PowerShell cho AzureAD
    - .NET 5 SDK
    - Visual Studio Code
    - Node.js
    - Visual Studio 2019 (tùy chọn)
2. Microsoft 365 Tenant: Với quyền quản trị viên Power BI.
3. Quyền Azure AD: Có khả năng tạo nhóm và ứng dụng mới trong Azure AD.

3.1.1 Tạo ứng dụng web MVC với xác thực Azure AD

Ứng dụng sẽ được tạo theo quy trình tạo nhóm và sau đó thêm ứng dụng vào nhóm của quản lý quyền truy cập và bảo mật trong Azure AD. Đầu tiên là tạo một nhóm mới trong Azure AD sử dụng Power Shell ISE. Tệp này sẽ thực hiện các thao tác kết nối với Azure AD và tạo một nhóm bảo mật Azure AD mới đặt tên là Power BI Apps như (Hình 3. 1) dưới đây.



Hình 3. 1 Tạo nhóm bảo mật trong Azure AD

Tiếp theo, sau khi nhóm đã được tạo thành công, cần tạo thêm tệp có tên Create_AzureAD_Application.ps1.Tệp này sẽ thực hiện thao tác kết nối với Azure AD và tạo một ứng dụng mới trong Azure AD và cấu hình các thiết lập như tên hiển thị, URL phản hồi (Redirect URL) và mã bảo mật ứng dụng. Ngoài ra, còn lấy thông tin về service principal của ứng dụng mới tạo và thêm service principal của ứng dụng vào nhóm Azure AD đã tạo trước đó. Vậy ứng dụng web MVC với tính năng xác thực Azure AD để sử dụng trong các bước tiếp theo của quá trình xây dựng ứng dụng bước đầu được tạo thành công.

3.1.2 Tạo dự án .NET5 cho ứng dụng web MVC

Tạo dự án .NET 5 cho ứng dụng web MVC với hỗ trợ xác thực Azure AD cần sử dụng công cụ .NET CLI (Command Line Interface). Đặt tên dự án là PowerBISalesApps. Sử dụng cửa sổ terminal trong visual studio code nhập và thực thi lệnh “dotnet new mvc” --auth SingleOrg. Tham số --auth SingleOrg chỉ định cho .NET 5 CLI tạo ứng dụng web mới với mã nguồn bổ sung hỗ trợ xác thực sử dụng thư viện xác thực của Microsoft có tên là Microsoft.Identity.Web.

Khi chạy lệnh dotnet new, khá nhiều thư mục và tệp mới sẽ được thêm vào dự án. Cài đặt phần mở rộng Omnisharp để phát triển C# với .NET 5. Tạo các tệp cần thiết để xây dựng dự án. Chạy và kiểm tra ứng dụng web PowerBISalesApps bằng Visual Studio Code và trình gỡ lỗi .NET 5. Sau khi chạy ứng dụng sẽ có giao diện như (Hình 3. 2) dưới đây.

Hình 3. 2 Tạo dự án .NET5 cho ứng dụng MVC với xác thực Azure AD

3.1.3 Triển khai giải pháp bảo mật và hiệu xuất xử lý dữ liệu

Như đã nhắc ở phần yêu cầu phi chức năng mục 2.1.2 về vấn đề bảo mật dữ. Yêu cầu ứng dụng đảm bảo vấn đề bảo mật dữ liệu. Do đó, sau khi tạo được dự án .NET5 cho ứng dụng web MVC thì sẽ triển khai yêu cầu về bảo mật dữ liệu và áp dụng công nghệ nâng cao hiểu suất xử lý dữ liệu. Dưới đây là các bước thực hiện bảo mật và áp dụng giải pháp nâng cao hiệu suất xử lý dữ liệu:

1. **Triển khai giải pháp bảo mật**

- Sử dụng thư viện Microsoft.Identity.Web để thiết lập đăng nhập cho người dùng, bảo mật dữ liệu bán hàng và hạn chế quyền truy cập dựa trên vai trò của người dùng. Hình 3.3 dưới đây là đoạn mã thêm thư viện vào ứng dụng.

AppOwnsData.csproj

```
1  <Project Sdk="Microsoft.NET.Sdk.Web">
2
3    <PropertyGroup>
4      <TargetFramework>net5.0</TargetFramework>
5      <UserSecretsId>aspnet-AppOwnsData-74548193-D24B-4C99-9C55-AF543E13A630</UserSecretsId>
6    </PropertyGroup>
7
8    <ItemGroup>
9      <PackageReference Include="Microsoft.AspNetCore.Authentication.JwtBearer" Version="5.0.1" NoWarn="NU1605" />
10     <PackageReference Include="Microsoft.AspNetCore.Authentication.OpenIdConnect" Version="5.0.1" NoWarn="NU1605"
11     <PackageReference Include="Microsoft.Identity.Web" Version="1.1.0" />
12     <PackageReference Include="Microsoft.Identity.Web.UI" Version="1.1.0" />
13     <PackageReference Include="Microsoft.PowerBi.Api" Version="3.20.1" />
14   </ItemGroup>
15
16   <Target Name="PostBuild" AfterTargets="PostBuildEvent">
17     <Exec Command="npm install" />
18     <Exec Command="npm run build" />
19   </Target>
20
21 </Project>
```

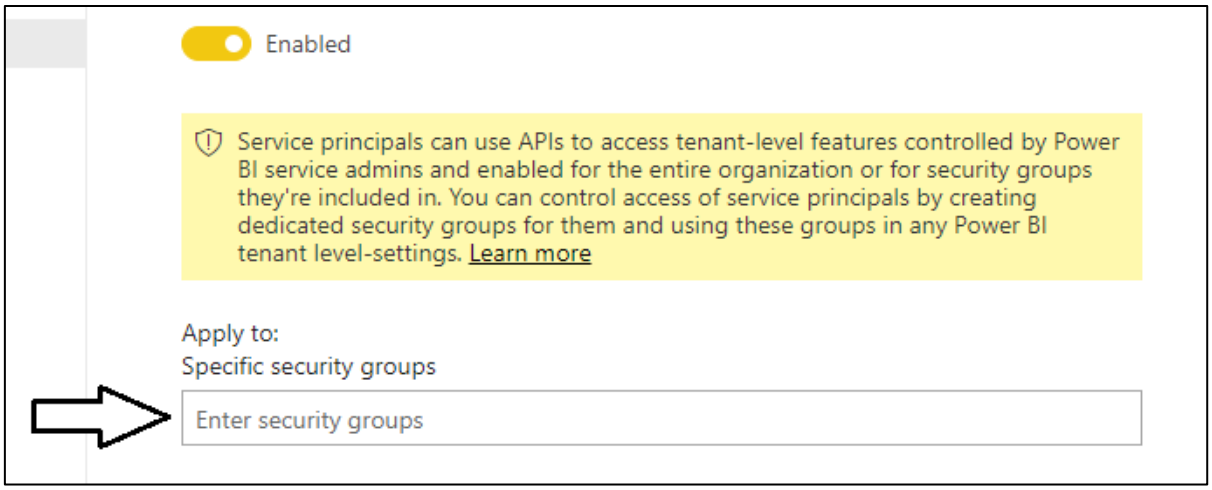
Hình 3. 3 Thư viện Microsoft.Identity.Web được thêm vào ứng dụng

- Sử dụng các token truy cập chỉ dành cho ứng dụng để gọi Power BI Service API như trong (hình 3.4) dưới đây.

```
21  public void ConfigureServices(IServiceCollection services) {
22
23      services.AddMicrosoftIdentityWebAppAuthentication(Configuration)
24              .EnableTokenAcquisitionToCallDownstreamApi()
25              .AddInMemoryTokenCaches();
26
27      services.AddScoped(typeof(PowerBiServiceApi));
28
29      services.AddControllersWithViews(options => {
30        var policy = new AuthorizationPolicyBuilder()
31            .RequireAuthenticatedUser()
32            .Build();
33        options.Filters.Add(new AuthorizeFilter(policy));
34      });
35      services.AddRazorPages()
36              .AddMicrosoftIdentityUI();
37  }
```

Hình 3. 4 Cấu hình token và Power BI Service

- Sử dụng Azure AD để quản lý người dùng và nhóm, cung cấp một lớp bảo mật bổ sung bằng cách giới hạn quyền truy cập vào các tài nguyên chỉ cho những người dùng và nhóm đã được xác định trước.



Hình 3. 5 Đặt bảo mật áp dụng cho nhóm cụ thể

Sau khi đã triển khai được các yêu cầu về bảo mật dữ liệu và hạn chế quyền truy cập dựa trên vai trò của người dùng. Sẽ tiếp tục sử dụng một số giải pháp nâng cao hiệu suất xử lý dữ liệu.

2. Triển khai giải pháp nâng cao hiệu suất xử lý dữ liệu

Các phương pháp nêu trong phần 1.3.3 đều hướng đến tối ưu nguồn dữ liệu lớn tuy nhiên trong phạm vi dữ liệu sử dụng của đồ án này sẽ triển khai phương pháp làm sạch dữ liệu sử dụng ngôn ngữ lập trình Python và DirectQuery.

Để thấy được sự phân hóa của dữ liệu khi trực quan nên đồ án sẽ sử dụng lượng dữ liệu lớn cụ thể là dữ liệu trong ba năm của bộ AdventureWorks. Power BI không thể xử lý được lượng dữ liệu lớn này do đó để chuyển đổi, làm sạch dữ liệu trước khi trực quan thì sẽ sử dụng Python và jupyter notebook.

TẢI DỮ LIỆU THÔ

```
In [ ]: customer = pd.read_excel('Raw_data.xlsx' , sheet_name='dventureWorks_Customers.xlsx')
```

CHUYỂN SANG DATA FRAME

```
In [ ]: customer_df =pd.DataFrame(customer)
        prod_cat_df =pd.DataFrame(product_categories)
        subcat_df = pd.DataFrame(subcategories)
        product_df = pd.DataFrame(products)
        returns_df = pd.DataFrame(returns)
        territ_df = pd.DataFrame(territories)
```

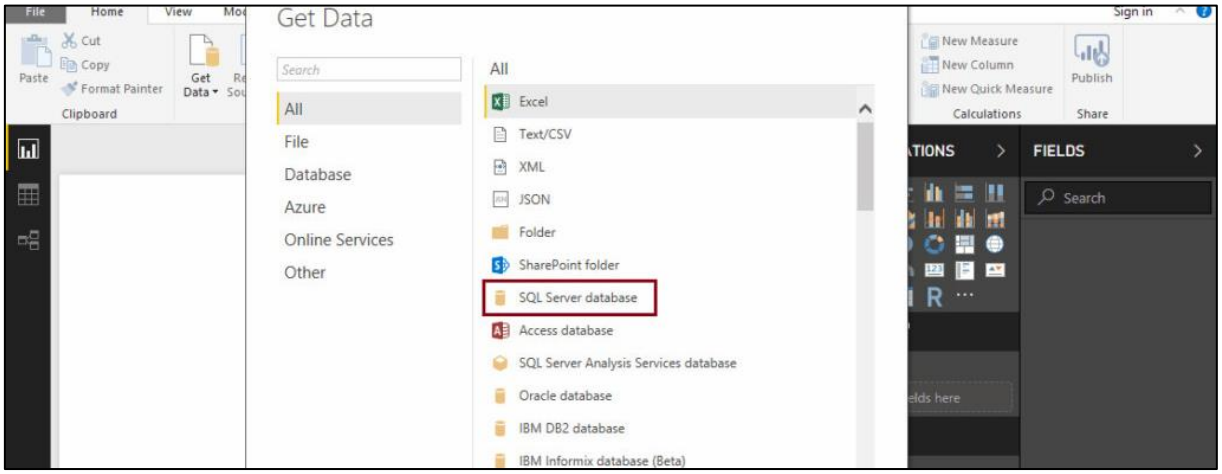
GỘP BẢNG SALES 2015 - 2017

```
In [ ]: sales=pd.concat([sales2015,sales2016,sales2017],ignore_index=True)
```

```
In [ ]: sales_df = pd.DataFrame(sales)
        type(sales_df)
```

Hình 3. 6 Xử lý dữ liệu xử dụng Python trong jupyter notebook

Sau khi thực hiện sử lý dữ liệu xong sẽ lưu dữ liệu ở dạng “csv” và lưu trong SQL Server. Đồng thời để có thể truy vấn trực tiếp dữ liệu thì sẽ kết nối SQL Server với Power BI Desktop.



Hình 3. 7 Kết nối SQL Server database với Power BI Desktop

3.2. Tạo báo cáo Power BI

Như đã trình bày trong (phần 1.3.2) thì bài toán cần phân tích là phân khúc khách hàng và phân tích bán chéo sản phẩm thực nghiệm trên bộ CSDL Adventure Works.

3.2.1. Chuẩn bị dữ liệu

Bộ dữ liệu AdventureWorks sẽ được sử dụng làm cơ sở dữ liệu (CSDL) để phân tích bởi, Adventure Works được thiết kế như một trường hợp thực tế giống như một công ty có thật trong ngành bán lẻ. Adventure Works cung cấp cơ sở dữ liệu bao gồm quy trình kinh doanh từ bán hàng, quản lý vật tư, sản xuất, tài chính và quản lý vốn nhân lực. Bộ CSDL lấy bối cảnh của một công ty hư cấu có tên AdventureWorksCycles. Công ty có 97 các nhãn hiệu xe đạp khác khau được nhóm theo ba loại: xe đạp leo núi(mountain bikes), xe đạp đường bộ (road bikes), xe đạp du lịch (touring bikes). Adventure Works không chỉ bán xe đạp mà còn cung cấp phụ kiện, quần áo và linh kiện. Nhiều thứ trong số đó được sản xuất bởi các nhà cung cấp, vì vậy Adventure Works đóng vai trò là đại lý bán lẻ.

Adventure Works phục vụ khách hàng trên toàn cầu, bao gồm cả khu vực Bắc Mỹ, Châu Âu và Châu Á. Có 2 mô hình kinh doanh ở Adventure Works là cửa hàng bán lẻ bán xe đạp và bán hàng qua internet phục vụ khách hàng cá nhân. Để điều hành các hoạt động kinh doanh, Adventure Works có tổng cộng 290 nhân viên đảm nhiệm một số chức năng như bán hàng, sản xuất, mua hàng, kỹ thuật,

tài chính, dịch vụ thông tin, tiếp thị, vận chuyển và nhận hàng cũng như R&D. Khách hàng của Adventure Works bao gồm hơn 700 cửa hàng và hơn 19000 cá nhân trên toàn thế giới và các nhà cung cấp của nó được ước tính khoảng 100 công ty cung cấp nguyên liệu thô, phụ kiện, quần áo và linh kiện.

Để hiệu quả phân tích được tối ưu nhất, sữ liệu sẽ chuẩn bị để phân tích gồm 3 năm 2015, 2016, 2017 của bộ dữ liệu Adventure Works.

Sử dụng công cụ SQL Server Management Studio (SSMS) để kết nối và truy vấn dữ liệu từ cơ sở dữ liệu từ Power BI Desktop. Các bảng dữ liệu cần thiết đã được chuẩn bị gồm : Sales, Customer, Product, Calendar, Returns, Territoties…

Trong quá trình chuẩn bị dữ liệu thông thường sẽ chỉnh sửa dữ liệu được thực hiện ở trong Tab Transform data gồm ba loại dữ liệu ở dạng:

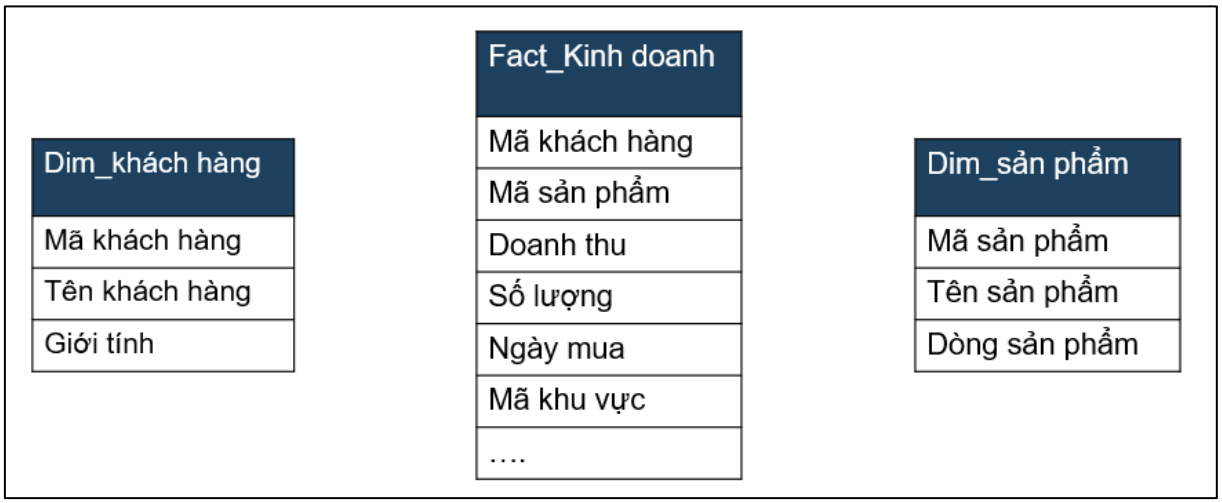
- **Dạng Text:** Thực hiện tách chữ, gom chữ, thay thế giá trị cần thiết.
- **Dạng Number**: Dạng số thì sẽ có các kiểu cộng, trừ, nhân, chia, và làm tròn số sau 2 dấu phẩy để số liệu dễ theo dõi.
- **Dạng Date:** Thay đổi kiểu dữ dữ liệu đúng định dạng , có thể tạo thêm cột mới là Year, Month, Day trong add Tab Colums để áp dụng khi trực quan sẽ dễ hiểu và dễ theo dõi hơn.

Trong quá trình chuẩn bị dữ liệu, vì bảng Sales gồm có 3 bảng (Sales_2015, Sales_2016, Sales_2016) nên sẽ nối các bảng Sales này vào thành một bảng chung có thể dùng truy vấn trong SQL để nối. Hoặc nếu dữ liệu được tải lên Power BI ở định dạng .csv thì gom lại tạo một thư mục (Folder) tên là Sales. Sau khi gom thành một bảng cần vào Transform data để bung dữ liệu nối theo chiều dọc.

3.2.2. Tạo mô hình dữ liệu

Sau khi chuẩn bị tất cả bảng cần thiết, bước tiếp theo là gán mối quan hệ giữa chúng. Mối quan hệ rất cần thiết trong mô hình kho dữ liệu vì nó sẽ kiểm soát cách các dữ liệu liên quan với nhau. Hình 3.8 dưới đây là ví dụ cho một mô hình dữ liệu chỉ có một bảng Fact (chứa những dữ liệu quan trọng như bảng Sales

và thường sẽ ở định dạng khóa chính (Key, ID) thì nó sẽ là bảng Fact) và nhiều bảng Dimension (Dim) có nhiệm vụ giải nghĩa cho bảng Fact.



Hình 3. 8 Ví dụ các thành phần trong mô hình dữ liệu

Khi thiết kế mô hình dữ liệu trong Power BI thì phải xác định được đâu là bảng Fact và đâu là bảng Dim, vì những liên kết sẽ dựa vào tính chất mỗi bảng để thực hiện kết nối. Bảng 3.1 dưới đây là giải nghĩa cho các thành phần của bảng Fact và bảng Dim.

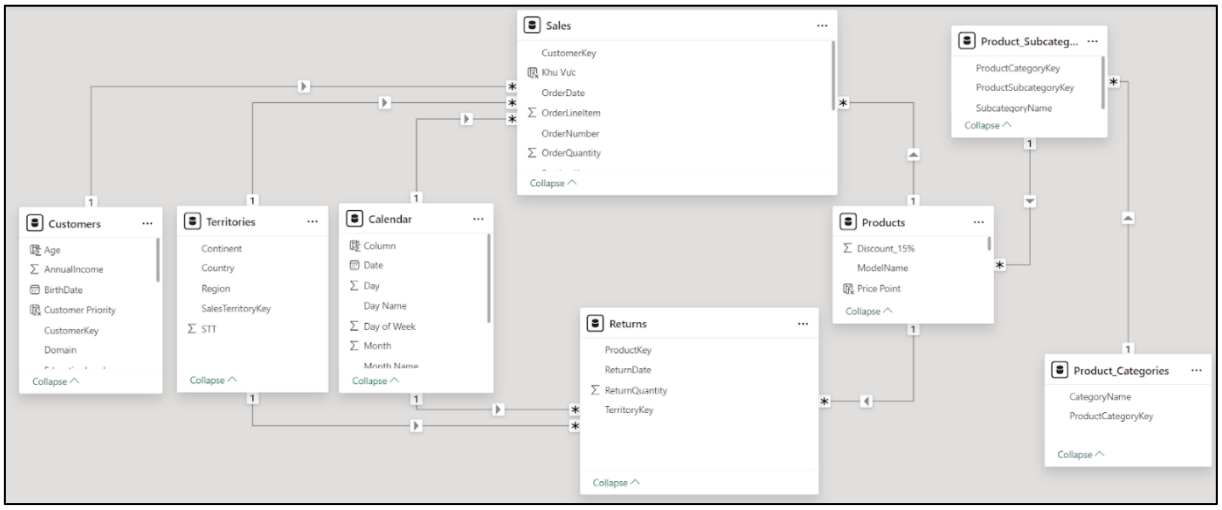
Bảng 3. 1 Ý nghĩa thành phần trong bảng Fact và Dimension

	Bảng Fact	**Bảng Dimension**
Khái niệm	Các phép đo, chỉ số, dữ liệu sinh ra	Bảng chứa các dữ kiện thuộc tính mô tả được sử dụng làm ràng buộc truy vấn
Tác dụng	Đo lường các dữ liệu các thông tin trong bảng Dimension	Giải thích ý nghĩa các trường dữ liệu trong bảng Fact
Loại dữ liệu	Chủ yếu dữ liệu lượng hóa và là các thông tin cơ bản nhất	Có thể chứa dạng text , mô tả đầy đủ

Phân cấp	Không phân cấp	Có phân cấp (vd :dòng sp có nhiều loại sp,…)
Vị trí	Trung tâm của mô hình với các bảng Dimension xung quanh	Ở xung quanh bổ xung ý nghĩa cho bảng Fact

Như đã nói thì một mô hình dữ liệu thì chỉ có một bảng Fact và nhiều bảng Dim. Nhưng có thể có 2 mô hình dữ liệu ví dụ như bảng Fact Sales và bảng Fact Return trong mô hình dữ liệu (hình 3.9) dưới đây (hai bảng này là 2 mô hình dữ liệu khác nhau nhưng có thể dùng chung bảng Dim). Và khi một cơ sở dữ liệu có nhiều bảng Fact gọi là mô hình galaxy schema. Trong phần này, mô hình được tạo ra sử dụng phương pháp lược đồ galaxy schema.

Thực hiện nối bảng: Thiết lập các mối quan hệ giữa các bảng dữ liệu dựa trên các khóa chính và khóa ngoại trong Power BI Desktop. Sau khi tạo liên kết có mô hình dữ liệu như dưới đây.



Hình 3. 9 Mô hình dữ liệu

Gồm có ba loại mối quan hệ trong mô hình dữ liệu:

- **Quan hệ 1-1 (one to one):** Đây là loại quan hệ mà một giá trị duy nhất ở 1 cột của 1 bảng tương ứng với 1 giá trị duy nhất trong 1 cột của bảng liên quan.

- **Quan hệ 1-n (one to many) hoặc n-1 (many to one):** Đây là loại quan hệ phổ biến và được sử dụng nhất trong Power BI. Một giá trị duy nhất trong cột của bảng gốc tương ứng với nhiều giá trị trong cột của bảng liên quan.
- **Quan hệ n-n (many to many):** Đây là loại quan hệ mà nhiều giá trị trong cột của 1 bảng tương ứng với nhiều giá trị trong cột của bảng liên quan. Đây là mối quan hệ cần phải đặc biệt lưu ý cẩn thận. Ví dụ điển hình là một đơn hàng có nhiều đợt giao hàng và mỗi đợt giao hàng giao nhiều kiện hàng. Bảng 3.2 dưới đây là tất cả liên kết của mô hình dữ liệu

Bảng 3. 2 Mối quan hệ trong mô hình dữ liệu Power BI

Từ bảng(cột)	**Mối quan hệ**	**Đến bảng(cột)**
Product_Subcategories (ProductCategoryKey)	Many to one	Product_Categories (ProductCategoryKey)
Products (ProductSubcategoryKey)	Many to one	Product_Subcategories (ProductSubcategoryKey)
Returns (ProductKey)	Many to one	Products (ProductKey)
Returns (ReturnDate)	Many to one	Calendar (Date)
Returns (TerritoryKey)	Many to one	Territories (SalesTerritoryKey)
RFM (RFM_Score)	Many to one	Sheet1 (Scores)
Sales (CustomerKey)	Many to one	Customers (CustomerKey)
Sales (OrderDate)	Many to one	Calendar (Date)
Sales (ProductKey)	Many to one	Products (ProductKey)
Sales (TerritoryKey)	Many to one	Territories (SalesTerritoryKey)

3.2.3 Tính toán các chỉ số

Sử dụng các hàm DAX (Data Analysis Expressions) là tập hợp các hàm, toán tử và hằng số, sử dụng trong công thức hoặc biểu thức tính để giải quyết những bài toán từ cơ bản tới phức tạp, kết quả trả về là một hoặc nhiều giá trị từ dữ liệu có sẵn. Có 2 cách dùng chính.

Cách 1: Dax column

Đây là công thức DAX được sử dụng để thêm một cột vào bảng. Công thức phải trả về giá trị vô hướng và nó được đánh giá cho mỗi hàng trong bảng.

Ngữ cảnh tính toán (Row context) lặp qua các dòng trong bảng. DAX sẽ kích hoạt Row context khi sử dụng Calculated column (tạo cột mới trong bảng bằng DAX).

Một số hàm vòng lặp (Hàm X : SUMX, COUNTX,…) có thể tính toán trực tiếp trên Row context (tính toán trên cột ảo). Row context không tự động lọc dữ liệu qua các bảng, cần sử dụng 1 số hàm liên kết : Related, lookupvalue,…Hàm này được sử dụng khi mà các trường hợp khó quá không giải quyết được bằng Power Query. Thì mới ra ngoài Tab transform để thực hiện các hàm DAX.

Cách 2: DAX Measure

Một measure tạo một biểu thức, một logic tính toán không phải là một cột hay một hàm cụ thể. Do đó, chỉ có thể thấy được công thức của nó và không thấy kết quả trả về. Và tất cả các giá trị của biểu thức DAX Measure này sẽ chỉ hiển thị khi mà biểu diễn nó bằng thao tác kéo thả vào trang trực quan.

Ngữ cảnh bộ lọc (Filter context) lặp qua các bảng trong mô hình dữ liệu (các bảng đã được liên kết với nhau). DAX sẽ kích hoạt Filter context khi sử dụng các measure được biểu diễn trên báo cáo. Hàm CALCULATE có thể thay đổi được bộ lọc khi tính toán, filter context luôn luôn sử dụng mối liên kết one- to- many không và không sử dụng many- to- many [10].

Dưới đây là cách tính các chỉ số của mỗi trang báo cáo:

❖ **Chỉ số trang báo cáo tổng quan:**

Để xác định được các chỉ số cần tính, phần này lấy ví dụ một doanh nghiệp nhỏ mong muốn tăng doanh thu ngành bán lẻ .Thì các yếu tố có thể ảnh hưởng tới doanh thu như là:

- Các yếu tố có thể lượng hóa được (có thể lấy được dữ liệu) như: Số lượng khách hàng, địa điểm, giới tính, chi phí.
- Các yếu tố khó lượng hóa được (khó lấy được dữ liệu): Lòng trung thành, sự hài lòng, sự trải nghiệm…

Để nắm được hiệu suất kinh doanh của doanh nghiệp và chuẩn bị chiến lược đưa ra những chương trình kinh doanh mới. Thông thường, thì báo cáo tổng quan có thể gồm một số chỉ số như: Tổng doanh thu, tổng giao dịch thay đổi theo thời gian(cùng kỳ năm trước, cùng kỳ tháng trước…), thay đổi theo vị trí, loại hàng …. Ngoài ra, còn thay đổi theo giới tính.. tùy thuộc vào số liệu mong mong muốn của doanh nghiệp. Bảng 3.3 dưới đây Dưới đây là các chỉ số được chọn để trực quan (sẽ sử dụng thêm một số bộ lọc vê thời gian. vị trí, sản phẩm để có cái nhìn đa chiều) trong nội dung báo cáo của tôi.

Bảng 3. 3 Tính toán các chỉ số của trang tổng quan

Chỉ số	**Hàm DAX**
Tổng doanh thu của bảng Sales	Tổng doanh thu =SUM(Sales[Revenue])
Tổng doanh thu cùng kỳ năm trước: Tức là tính doanh thu của năm trước với năm nay	*Tổng doanh thu_Last year = CALCULATE([Tổng doanh thu], SAMEPERIODLASTYEAR('Calendar[Date]))*
Tổng doanh thu cùng kỳ tháng trước: Tức là tính so sánh doanh thu của tháng này so với tháng trước:	*Tổng doanh thu_Last month = CALCULATE([Tổng doanh thu], DATEADD('Calendar[Date], -1,MONTH))*

Tổng doanh thu kì trước	*Revenue_Last period =* *varfirst_date = FIRSTDATE('Calander'[Date])* *var last_date= LASTDATE('Calandar'[Date])* *var khoang = DATEDIFF* *(first_date, last_date, DAY) +1* *return CALCULATE ([Tổng doanh thu], DATEADD('Calandar'[Date], - khoang, DAY))*
Tổng lũy kế(mục tiêu năm hoặc tháng đến nay đạt bao nhiêu phầnn trăm)	*TOTALYTD=* *([Tổng doanh thu], 'Calandar'[Date])*

- **Chỉ số trang báo cáo phân khúc khách hàng**

Dựa vào giải pháp được đề xuất trong phần (1.3.2) thì mục tiêu bài toán phân khúc khách hàng không chỉ đơn giản là chia tập dữ liệu khách hàng thành những tập có đặc điểm tương đồng như những nghiên cứu cũ. Mà còn liên hệ đến thực tế trong phân tích doanh nghiệp thường sẽ hướng đến chỉ số 20%, 80% . Ví dụ họ mong muốn 20% khách hàng vip sẽ mang lại 80% doanh thu dựa theo quy luật pareto [8]. Do đó các chỉ số biểu diễn báo cáo phân khúc khách hàng thể hiện được hai mục đích sau:

- 20% khách hàng tốt nhất đem lại bao nhiêu % doanh thu ?(trả lời cho câu hỏi nên tập trung vào KH cũ hay KH mới).
- 80% doanh thu tới từ các phân khúc khách hàng nào?(trả lời cho câu hỏi nên tập trung vào những nhóm KH nào ?).

Để báo cáo thể hiện được điều đó trong phần này sẽ tạo một bảng mới tên là RFM, sử dụng hàm SUMMARIZE để tạo một bảng tổng hợp mới với các cột được liệt kê trong (bảng 3.4) dưới đây với những chỉ số sau: CustomerKey, Recency, Frequency, Monetary, R_Score, F_Score, M_Score, RFM_Score, Segment.

Bảng 3. 4 Thành phần trong bảng RFM

Tên cột	Giải nghĩa
CustomerKey	Mã khách hàng
Recency	Lần cuối cùng khách hàng mua hàng cách nay bao lâu?
Frequency	Khách hàng đã mua tổng cộng bao nhiêu lần ?
Monetary	Tổng số tiền khách hàng đã tiêu là bao nhiêu?
R_Score	Chấm điểm cột Recency
F_Score	Chấm điểm cột Frequency
M_Score	Chấm điểm cột Monetary
RFM_Score	Gom điểm theo khách hàng
Segment	Phân khúc khách hàng

Dưới đây là chi tiết các bước tính các chỉ số này:

Bước 1: Tính ba tham số đầu vào là Recency, Frequency, Monetary.

```
RFM =
SUMMARIZE(Sales, Sales[CustomerKey],
        "Recency", DATEDIFF(MAX(Sales[OrderDate]),TODAY(),DAY),
        "Frequency", DISTINCTCOUNT(Sales[OrderNumber]),
        "Monetary", SUM(Sales[Revenue]))
```

CustomerKey	Recency	Frequency	Monetary
12646	2841	1	9.98
12837	2852	1	9.98
14404	2840	1	9.98
14629	2827	1	9.98
14909	2838	1	9.98
14920	2784	1	9.98

Hình 3. 10 Tính chỉ số Recency, Frequency, Monetary

Bước 2: Chấm điểm cho từng hành vi của khách hàng

Sau khi tạo bảng gồm 3 chỉ số thì sẽ tiến hành chấm điểm cho hành vi của từng khách hàng. Chia cột Recency thành 5 phần 20/40/60/80/100 (theo phân vị

theo 20/80). Thì 20% số lượng KH có Recency thấp nhất (phân vị nhỏ nhất) sẽ được 5 điểm , tiếp theo sẽ được 4, 3, 2, 1 điểm. Gọi cột điểm mới tạo tên là R_Score trong bảng RFM với công thức trong (hình 3.11) dưới đây:

```
R_Score =
var percent_20 = PERCENTILE.EXC(RFM[Recency], 0.2)
var percent_40 = PERCENTILE.EXC(RFM[Recency], 0.4)
var percent_60 = PERCENTILE.EXC(RFM[Recency], 0.6)
var percent_80 = PERCENTILE.EXC(RFM[Recency], 0.8)
return
SWITCH(TRUE(),
        RFM[Recency] >= percent_80 , 1,
        RFM[Recency] >= percent_60 , 2,
        RFM[Recency] >= percent_40 , 3,
        RFM[Recency] >= percent_20 , 4, 5)
```

CustomerKey	Recency	Frequency	Monetary	R_Score
12646	2841	1	9.98	1
12837	2852	1	9.98	1
14404	2840	1	9.98	1
14629	2827	1	9.98	1
14909	2838	1	9.98	1
14920	2784	1	9.98	2
15315	2795	1	9.98	1

Hình 3. 11 Tính điểm cột Recency

Điểm R_Score này sẽ chia cột Recency thành 5 khúc với số lượng KH và mỗi nhóm này chứa 20% số lượng khách hàng(vì phân vị là 20). Thông thường thì 20% nhóm KH tốt nhất nó sẽ là 25 nhóm đầu tiên nên sẽ rơi vào khoảng 3 khúc (segment) đầu tiên. Có thể trong 20% nhóm KH trong 3 khúc đầu tiên này

không thể chiếm được 20% tổng số lượng người hoặc chiếm hơn 20% tổng số lượng (xét trong điều kiện lý tưởng).

Cột Frequency tính tương tự như trên nhưng cách đánh giá sẽ khác một chút như (hình 3.12) dưới đây. Với KH mua càng nhiều càng tốt nên KH nào phân vị càng lớn thì điểm sẽ là 5.

```
F_Score =
var percent_20 = PERCENTILE.EXC(RFM[Frequency], 0.2)
var percent_40 = PERCENTILE.EXC(RFM[Frequency], 0.4)
var percent_60 = PERCENTILE.EXC(RFM[Frequency], 0.6)
var percent_80 = PERCENTILE.EXC(RFM[Frequency], 0.8)
return
SWITCH(TRUE(),
        RFM[Frequency] >= percent_80 , 5,
        RFM[Frequency] >= percent_60 , 4,
        RFM[Frequency] >= percent_40 , 3,
        RFM[Frequency] >= percent_20 , 2, 1)
```

CustomerKey	Recency	Frequency	Monetary	R_Score	F_Score
12646	2841	1	9.98	1	4
12837	2852	1	9.98	1	4
14404	2840	1	9.98	1	4
14629	2827	1	9.98	1	4
14909	2838	1	9.98	1	4
14920	2784	1	9.98	2	4

Hình 3. 12 Tính điểm cột Frequency

Tuy nhiên khi chia phân vị thì cột R_Score có thể chia được 5 nhóm, còn F_Score thì chỉ chia được 2 nhóm. Bởi vì là dữ liệu ở Recency có rất nhiều dữ liệu khác nhau và nó phân hóa rõ ràng, còn Frequency thì số lượng mua hàng rất ít nghĩa là hành vi số lần mua hàng của công ty này chưa đủ phân hóa để chia thành các phân khúc khách hàng. Vậy nên dữ liệu càng lớn , càng lâu năm thì sẽ

càng dễ nhận biết được hành vi bên trong. Ngược lại nếu dữ liệu quá ít thì nó không đủ để phân hóa thành các nhóm khác nhau rõ ràng.

Tiếp tục tính cột Monetory như (hình 3.13):

```
M_Score =
var percent_20 = PERCENTILE.EXC(RFM[Monetary], 0.2)
var percent_40 = PERCENTILE.EXC(RFM[Monetary], 0.4)
var percent_60 = PERCENTILE.EXC(RFM[Monetary], 0.6)
var percent_80 = PERCENTILE.EXC(RFM[Monetary], 0.8)
return
SWITCH(TRUE(),
        RFM[Monetary] >= percent_80 , 5,
        RFM[Monetary] >= percent_60 , 4,
        RFM[Monetary] >= percent_40 , 3,
        RFM[Monetary] >= percent_20 , 2, 1)
```

CustomerKey	Recency	Frequency	Monetary	R_Score	F_Score	M_Score
12646	2841	1	9.98	1	4	1
12837	2852	1	9.98	1	4	1
14404	2840	1	9.98	1	4	1
14629	2827	1	9.98	1	4	1
14909	2838	1	9.98	1	4	1

Hình 3. 13 Tính điểm cột Monetory

Sau khi chấm điểm xong 3 tham số, sẽ tiến hành gom điểm lại thành cột RFM_Score và import bảng RankScore vào để tiến hành lọc và xếp hạng khách hàng theo nhóm điểm dựa trên bảng RankScore trong (hình 3.14). Trước khi lọc,cần biến đổi xếp hạng RFM từ dạng liệt kê sang từng dòng để có thể tự động so sánh với cột RFM_Score bằng edit query trong Power BI Desktop.

Segment	Scores
Champions	555, 554, 544, 545, 454, 455, ...
Loyal	543, 444, 435, 355, 354, 345, ...
Potential Loyalist	553, 551, 552, 541, 542, 533, ...
Promising	525, 524, 523, 522, 521, 515, ...
New Customers	512, 511, 422, 421, 412, 411, ...
Need Attention	535, 534, 443, 434, 343, 334, ...
About To Sleep	331, 321, 312, 221, 213, 231, ...
At Risk	255, 254, 245, 244, 253, 252, ...
Cannot Lose Them	155, 154, 144, 214,215,115, 1...
Hibernating customers	332, 322, 233, 232, 223, 222, ...
Lost customers	111, 112, 121, 131,141,151

Segment	Scores	Index
Champions	555	1
Champions	554	1
Champions	544	1
Champions	545	1
Champions	454	1
Champions	455	1
Champions	445	1
Loyal	543	2
Loyal	444	2
Loyal	435	2
Loyal	355	2
Loyal	354	2
Loyal	345	2

Hình 3. 14 Nhóm điểm khách hàng

Điểm số RFM sau khi chấm đại diện cho hành vi mua sắm của khách hàng, (bảng 3.5) dựa vào điểm số RFM để chia những phân khúc gồm các hành vi mua sắm:

Bảng 3. 5 Phân khúc hành vi khách hàng

Segment	Activity
Champion	Đã mua gần đây, đặt hàng thường xuyên và chi tiêu nhiều nhất.
Loyal	Đặt hàng thường xuyên và tương tác chương trình khuyến mãi.
Potential Loyalist	Đã mua gần đây, chi tiêu tốt.
Promising	Từng là khách trung thành. Chi tiêu thường xuyên và khá nhiều tiền. Nhưng lần mua cuối cùng cách đây đã lâu.
New Customers	Khách hàng mới
Needs attention	Khách hàng có lần mua gần đây nhất đã xảy ra hơn một tháng trước
About to sleep	Khách phổ biến, mua hàng cách đây không lâu.
At Risk	Có thể nói không thể mất họ nhưng lượng tiền và tần suất thấp hơn
Can’t lose them	Thực hiện các đơn đặt hàng lớn nhất và thường xuyên. Nhưng đã lâu không trở lại.
Hibernating customers	Đã thực hiện lần mua hàng cuối cùng của họ cách đây khá lâu.
Lost	Đã thực hiện mua hàng gần đây nhất cách đây khá lâu , mua ít và tiêu ít.

Sử dụng hàm RELATED (trả về giá trị liên quan) để xếp hạng khách hàng. Ngoài ra có thể dùng hàm LOOKUPVALUE(tìm kiếm giá trị so sánh) hoặc hàm

SUMMARIZE(bảng tóm tắt) để lọc điểm. Nhưng hàm RELATED (trả về giá trị liên quan) sẽ tối ưu hơn, vì dùng liên kết đúng dạng một mô hình dữ liệu (cột Score ở bảng Rank_Score sẽ liên kết với cột RFM_Score của bảng RFM. Sau khi thực hiện chấm điểm và xếp hạng điểm thì dữ liệu phân loại khách hàng theo từng phân khúc sẽ như dưới (hình 3.15) dưới đây.

```
Segment =
-- LOOKUPVALUE(Sheet1[Segment], Sheet1[Scores], RFM[RFM_Score])
-- SUMMARIZE(FILTER(Sheet1, RFM[RFM_Score] = Sheet1[Scores]), Sheet1[Segment])
RELATED(Sheet1[Segment])
```

CustomerKey	Recency	Frequency	Monetary	R_Score	F_Score	M_Score	RFM_Score	Segment	FullName
12646	2841	1	9.98	1	4	1	141	Lost customers	Kenneth Xu
12837	2852	1	9.98	1	4	1	141	Lost customers	Nicole Anderson
14404	2840	1	9.98	1	4	1	141	Lost customers	Rachel Davis
14629	2827	1	9.98	1	4	1	141	Lost customers	Blake Hernandez
14909	2838	1	9.98	1	4	1	141	Lost customers	Benjamin Bryant
14920	2784	1	9.98	2	4	1	241	About To Sleep	Jason Hill
15315	2795	1	9.98	1	4	1	141	Lost customers	Katherine Bennett
15344	2786	1	9.98	2	4	1	241	About To Sleep	Luke Russell
15400	2772	1	9.98	2	4	1	241	About To Sleep	Samuel Roberts

Hình 3. 15 Xếp hạng điểm phân khúc

- **Chỉ số trang báo cáo phân tích bán chéo:**

Bán chéo là bài toán có tính ứng dụng rất cao trong các doanh nghiệp trong ngành bán lẻ. Các hệ thống gợi ý sử dụng luật kết hợp có thể áp dụng trong nhiều lĩnh vực khác nhau để tăng doanh thu và nâng cao trải nghiệm khách hàng. Bài toán bán chéo sẽ sử dụng quy tắc kết hợp sản phẩm (association rule). Do đó, (bảng 3.6) dưới đây là cách tính ba chỉ số của quy luật kết hợp:

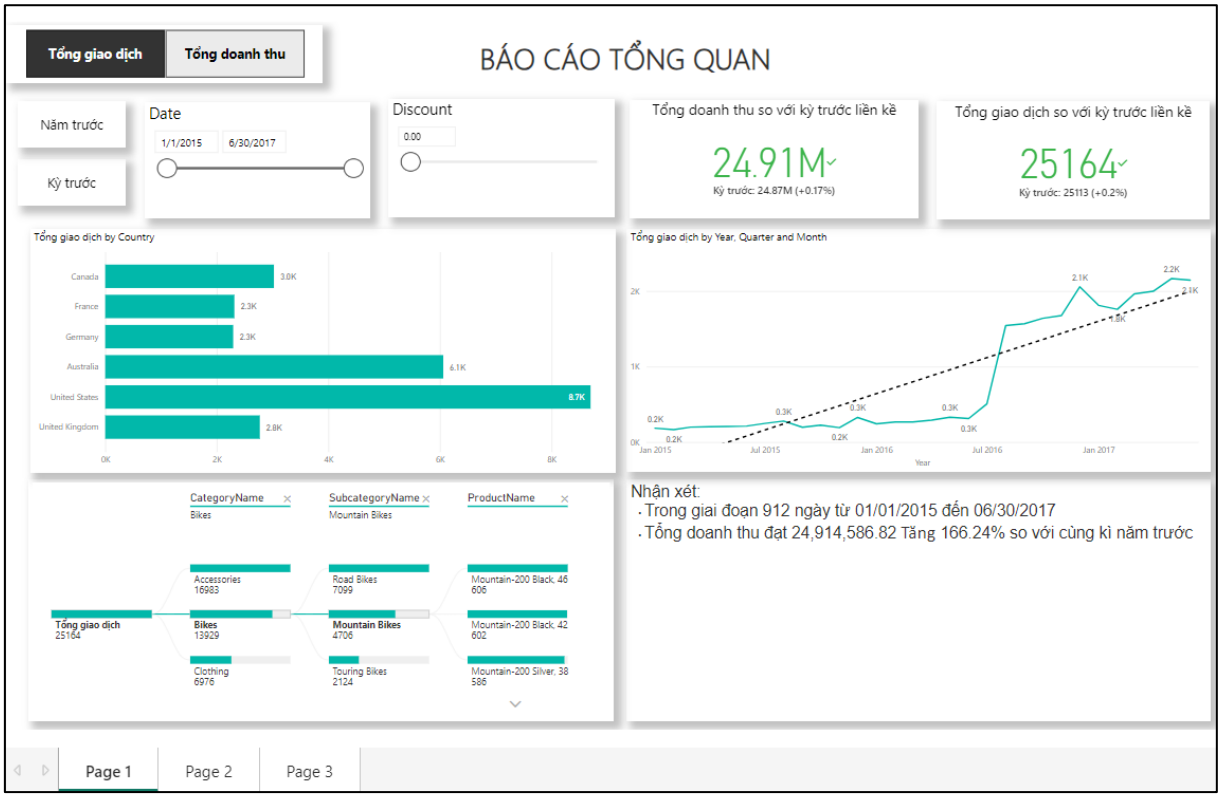
Bảng 3. 6 Các chỉ số trang phân tích bán chéo

Các sản phẩm tiền đề	Order(Tiền đề) = DISTINCTCOUNT(Sales[OrderNumber])
Các sản phẩm hệ quả	Order(Hệquả) = CALCULATE(DISTINCTCOUNT (Sales[OrderNumber]), REMOVEFILTERS(Products), USERELATIONSHIP(Sales[ProductKey], 'Products_Kèm'[ProductKey]))

Các sản phẩm kèm	Order(Kèm) = var cungdonhang = CALCULATETABLE(SUMMARIZE (Sales, Sales[OrderNumber]), REMOVEFILTERS(Products), REMOVEFILTERS(Sales[ProductKey]), USERELATIONSHIP(Sales[ProductKey], 'Products_Kèm'[ProductKey])) var soluong = CALCULATE(DISTINCTCOUNT (Sales[OrderNumber]),KEEPFILTERS(cungdonhang)) return soluong
Các sản phẩm kèm lọc trùng	Order(Kèm lọc trùng) = IF(ISEMPTY(INTERSECT (DISTINCT(Products[ProductKey]), DISTINCT('Products_Kèm'[ProductKey]))), [Order(Kèm)])
Tổng sản phẩm	Order(Tổng) = CALCULATE(DISTINCTCOUNT (Sales[OrderNumber]),REMOVEFILTERS(Sales))
Chỉ số Support	Support = [Order(Kèm lọc trùng)] / [Order(Tổng)]
Chỉ số Confidence	Confidence = [Order(Kèm lọc trùng)]/[Order(Tiền đề)]
Chỉ số Lift	Lift = ([Order(Kèm lọc trùng)]/[Order(Tiền đề)]) / ([Order(Hệ quả)]/[Order(Tổng)])

3.2.4 Trực quan hóa dữ liệu

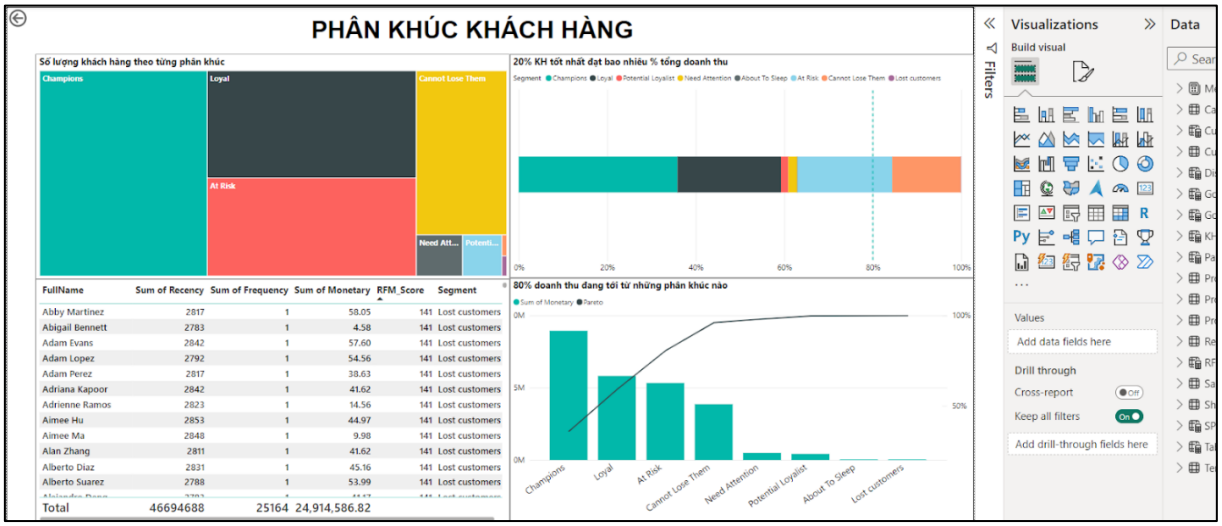
Trang báo cáo tổng quan: Trang báo cáo này sẽ sử dụng biểu đồ cột, biểu đồ đường và biểu đồ tròn để thể hiện các chỉ số gồm doanh thu, lợi nhuận và tỷ lệ tăng trưởng qua các kỳ.



Hình 3. 16 Trực quan dữ liệu trang báo cáo tổng quan

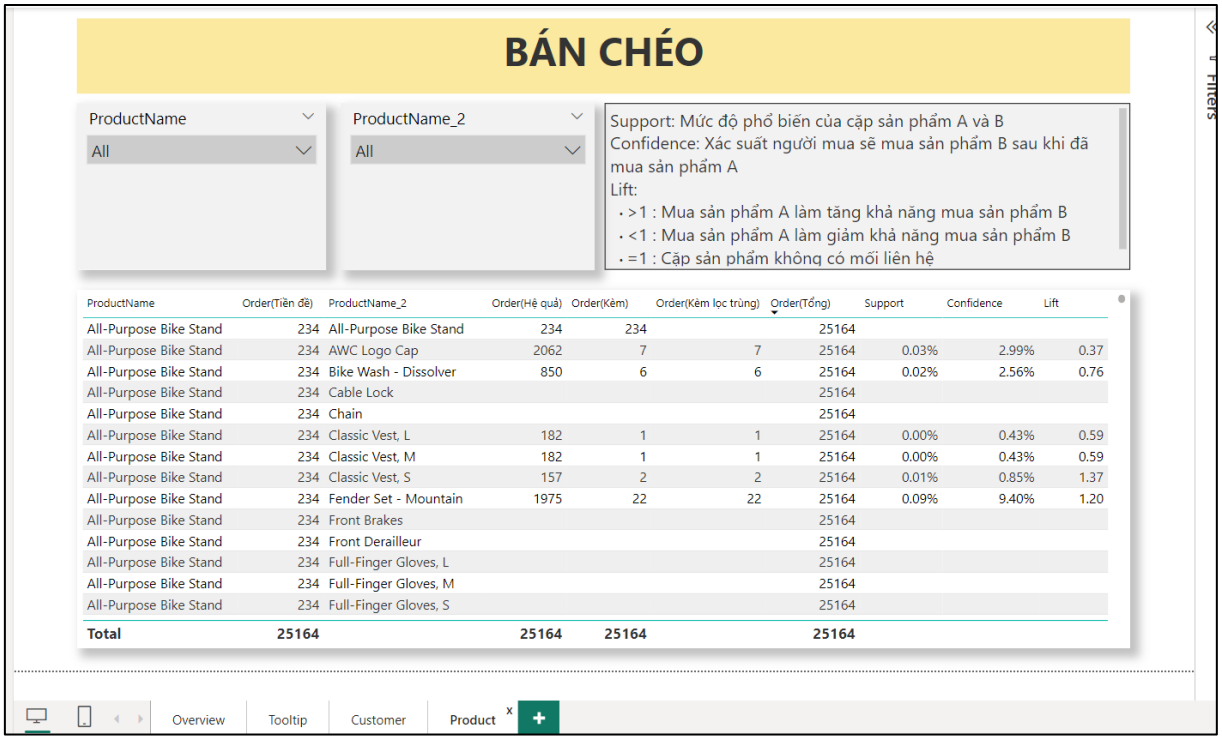
Trang báo cáo phân khúc khách hàng

Trang báo cáo phân khúc khách hàng tập trung vào việc phân tích và so sánh các phân khúc khác nhau của khách hàng. Biểu đồ phù hợp gồm các biểu đồ thanh chồng (stacked), biểu đồ biểu đồ cây (Treemap chart) và biểu đồ đường được sử dụng để minh họa sự phân bổ khách hàng theo độ tuổi, giới tính, địa lý và hành vi mua sắm.



Hình 3. 17 Trực quan hóa dữ liệu trang phân khúc khách hàng

Trang báo cáo bán chéo:



Hình 3. 18 Trực quan hóa trang phân tích bán chéo

3.3. Nhúng báo cáo Power BI vào ứng dụng web MVC

Sau khi tải báo cáo Power BI lên workspace Power BI service. Để mã của ứng dụng có thể gọi API Power BI service để truy xuất dữ liệu về một báo cáo cần thiết cho việc nhúng thì cần thêm hỗ trợ vào ứng dụng web PowerBISalesApps để lấy token truy cập chỉ dành cho ứng dụng từ Azure AD và gọi API Power

service. Đây là quá trình yêu cầu và nhận một mã thông báo truy cập (access token) từ Azure AD để có thể truy cập vào các API bảo mật. Mã thông báo này là cần thiết để chứng thực và ủy quyền cho các yêu cầu mà ứng dụng sẽ gửi đến API Power BI service. Chạy ứng dụng web trong trình gỡ lỗi Visual studio code để kiểm tra trang nhúng (Embed). Khi đăng nhập thành công thì trang chủ của ứng dụng sẽ như (hình 3.19) dưới đây.

Hình 3. 19 Trang chủ sau khi đăng nhập thành công

Tạo tệp PowerBiServiceApi chứa mã phía server để gọi API Dịch Vụ Power BI và truy xuất dữ liệu cần thiết để nhúng một báo cáo. Tiếp theo, để hoàn thiện việc nhúng Power BI bằng cách thêm mã JavaScript phía client để tương tác các thành phần Power BI được nhúng trên trang web. Kiểm tra Ứng dụng Web trong Visual Studio Code, Ứng dụng web sẽ hiển thị trang chủ như đối với người dùng ẩn danh. Nhấp vào liên kết Embed trong menu điều hướng trên để điều hướng đến trang Embed và đăng nhập khi được yêu cầu.

Bây giờ người dùng có thể điều hướng đến trang em báo cáo và xem báo cáo Power BI hiển thị trên trang. Có thể điều hướng đến cả trang chỉnh sửa báo cáo và thay đổi kích thước cửa sổ trình duyệt. Báo cáo được nhúng sẽ liên tục thích ứng với kích thước của cửa sổ.

Tiếp theo, thêm hỗ trợ vào dự án mã phía client sử dụng TypeScript thay vì JavaScript. Bằng cách chuyển sang TypeScript, mang lại nhiều lợi ích cho quá trình phát triển ứng dụng web. Như là lập trình kiểu mạnh (Strongly-typed programming). TypeScript là một ngôn ngữ lập trình có kiểu mạnh, nghĩa là phải khai báo kiểu dữ liệu cho các biến, hàm, và đối tượng giúp tránh các lỗi liên quan

đến kiểu dữ liệu trong quá trình phát triển. Một lợi thế nữa là kiểm tra kiểu trong thời gian biên dịch (Compile-time type checking) là kiểm tra kiểu dữ liệu của mã nguồn trong quá trình biên dịch, trước khi chạy chương trình giúp phát hiện lỗi sớm hơn trong quá trình phát triển.

Ngoài ra còn có IntelliSense tốt hơn (Better IntelliSense) là tính năng của các IDE và trình soạn thảo mã nguồn cung cấp gợi ý mã, thông tin về kiểu dữ liệu, và các tài liệu liên quan trực tiếp trong quá trình viết mã. TypeScript cung cấp IntelliSense tốt hơn nhờ vào việc khai báo kiểu rõ ràng.

3.4. Thực nghiệm và đánh giá kết quả

3.4.1. Thực nghiệm

Để đánh giá hiệu quả của ứng dụng phân tích dữ liệu PowerBISalesApps với khả năng phân tích tự động, chỉnh sửa báo cáo trực tiếp. Dưới đây là các bước thực nghiệm hệ thống ứng dụng

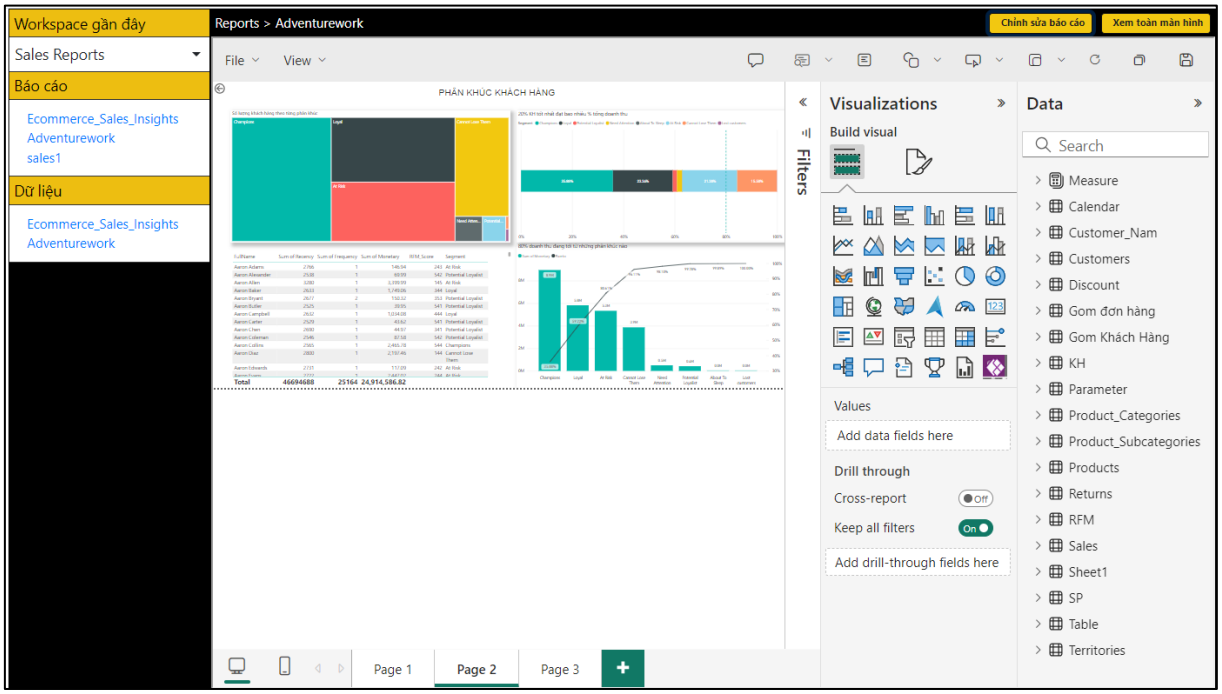
Dưới đây là các bước tiến hành thực nghiệm hệ thống ứng dụng:

Bước 1: Triển khai và kiểm ra chức năng xem báo báo:

- Thực hiện thao tác thêm phương thức GetEmbeddedViewModel vào lớp PowerBiServiceApi.
- Kiểm tra xem phương thức có thể lấy được dữ liệu JSON đúng từ Power BI Service API hay không.
- Thêm phương thức GetFirstWorkspace để hỗ trợ kiểm tra các trường hợp không có workspace nào khả dụng.

Bước 2: Cập nhật giao diện người dùng và kiểm tra tương tác:

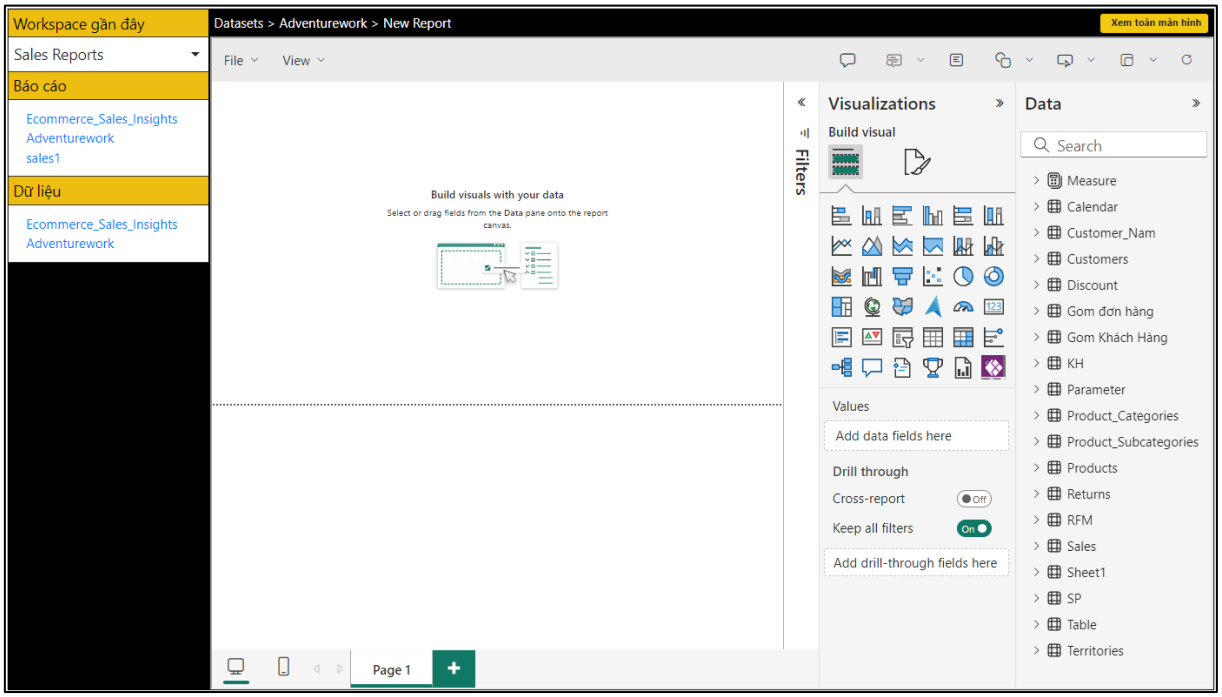
- Kiểm tra tính tương tác và giao diện người dùng
- Kiểm tra chức năng điều hướng và chuyển đổi giữa các workspace.



Hình 3. 20 Thực nghiệm chức năng chỉnh sửa và lưu báo cáo

Bước 3: Chạy ứng dụng và kiểm tra:

- Chạy ứng dụng trong chế độ debug để kiểm tra các chức năng.
- Kiểm tra việc mở và chỉnh sửa báo cáo, chuyển đổi giữa các chế độ xem và chế độ chỉnh sửa.
- Thử nghiệm tạo báo cáo mới từ dataset và lưu báo cáo.



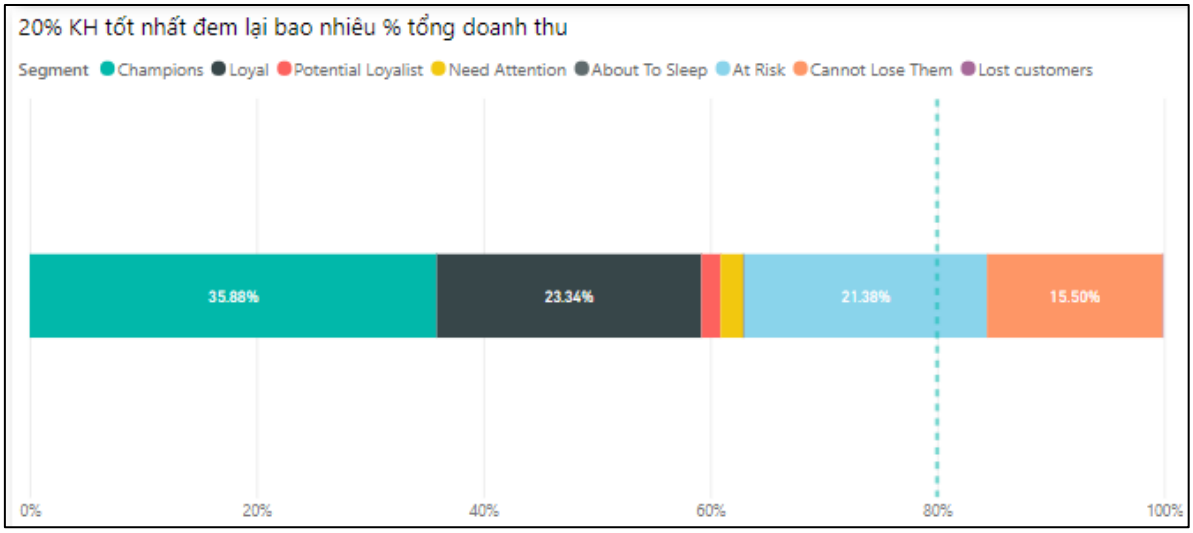
Hình 3. 21 Thực nghiệm chức năng tạo báo cáo mới

Bước 4: Đánh giá nội dung báo cáo được nhúng

- **Đối với trang báo cáo phân khúc khách hàng:** Liên hệ đến thực tế trong các doanh nghiệp bán lẻ hiện nay để đánh giá giải pháp sử dụng quy luật pareto [8] trong bài toán phân tích phân khúc khách hàng có những cải tiến nào so với những phương pháp cũ hiện nay. Nội dung đánh giá theo những ý sau:
    - 20% khách hàng tốt nhất đem lại bao nhiêu % doanh thu ?(trả lời cho câu hỏi nên tập trung vào KH cũ hay KH mới).
    - Câu 2: 80% doanh thu tới từ các phân khúc khách hàng nào?(trả lời cho câu hỏi nên tập trung vào những nhóm KH nào ?).

Như đã đề cập trong (phần 3.2.3). Để biết được 20% khách hàng(KH) sẽ đem lại bao nhiêu % doanh thu thì sẽ biểu diễn dữ liệu. Số nhóm KH là 5*5*5=125 nhóm (chia theo phân vị 20/80. Thông thường thì 20% nhóm KH tốt nhất nó sẽ là 25 nhóm đầu tiên nên sẽ rơi vào khoảng 3 khúc (segment)đầu tiên.

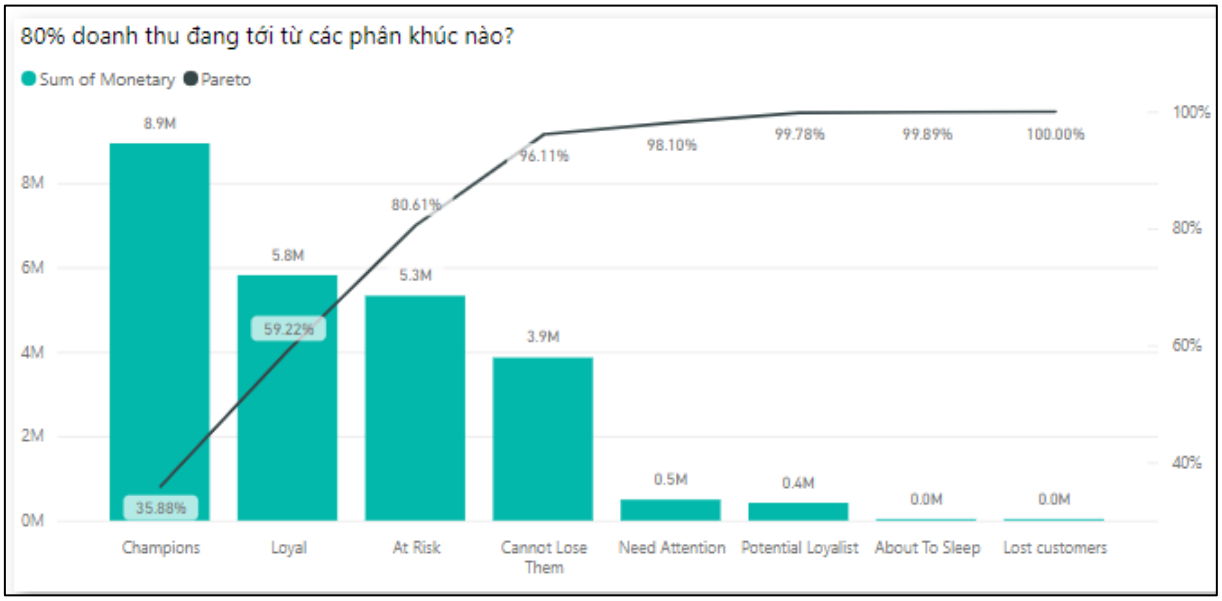
Mục đích là làm thế nào để biểu diễn được 3 khúc này đang chiếm tổng doanh thu là bao nhiêu.



Hình 3. 22 Xem 20% KH tốt nhất đem lại bao nhiêu % tổng doanh thu

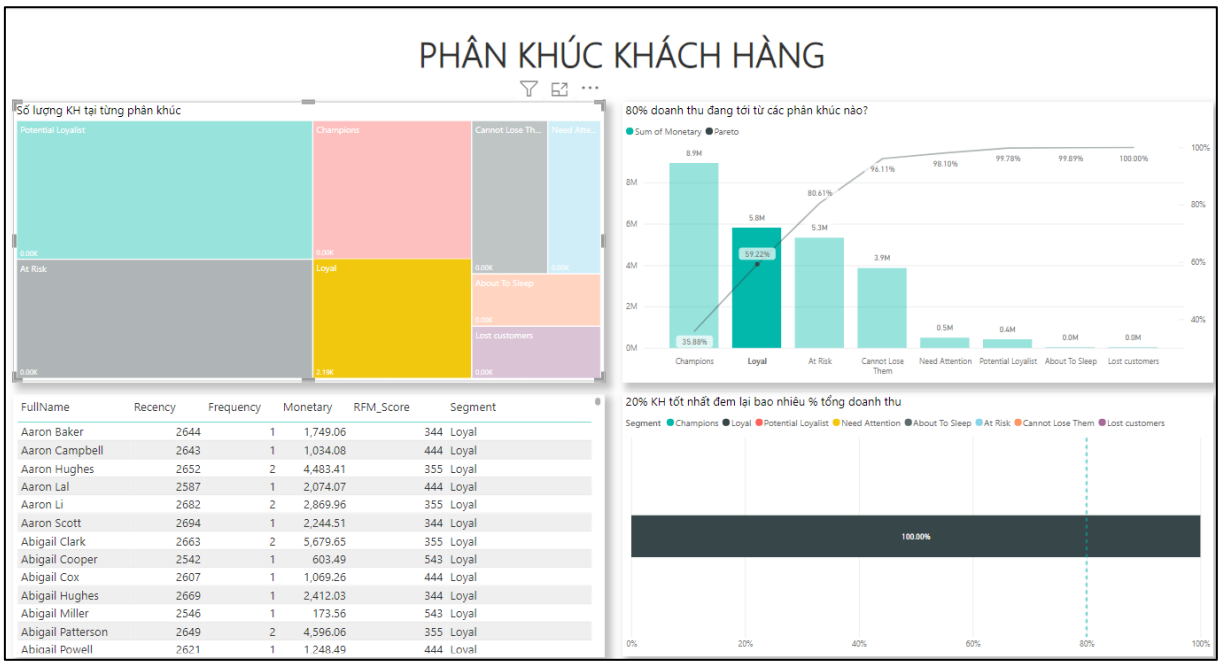
Vậy theo quan sát (hình 3.22) thì 3 nhóm đầu chỉ đạt 60% doanh thu, đang không đạt được 80% . Vậy nên doanh nghiệp này nên tập trung vào những tệp khách hàng cũ(tức là khách hàng trong 3 khúc) chưa cần phải tìm kiếm khách hàng mới.

Thông thường khi phân tích pareto thì sẽ sử dụng một số dạng biểu đồ (chart) gọi là pareto chart. Biểu đồ này sẽ ưu tiên cộng phần trăm từ phân khúc lớn nhất cộng dồn vào phần trăm phân khúc thứ 2 sau đó cộng dồn các phân khúc còn lại. Các phân khúc này sắp xếp theo chiều từ lớn đến nhỏ. Sau đó, với mong muốn 80% nó dừng ở đâu thì sẽ biết được nó đang dừng ở nhóm phân khúc nào. Khi đó, sẽ biết được 80% doanh thu của công ty đến từ những tệp khách hàng nào. Đường thể hiện sự cộng dồn của doanh thu chính là công thức để tính lũy kế như (hình 3.23).



Hình 3. 23 Xem 80% doanh thu đang tới từ nhóm phân khúc nào

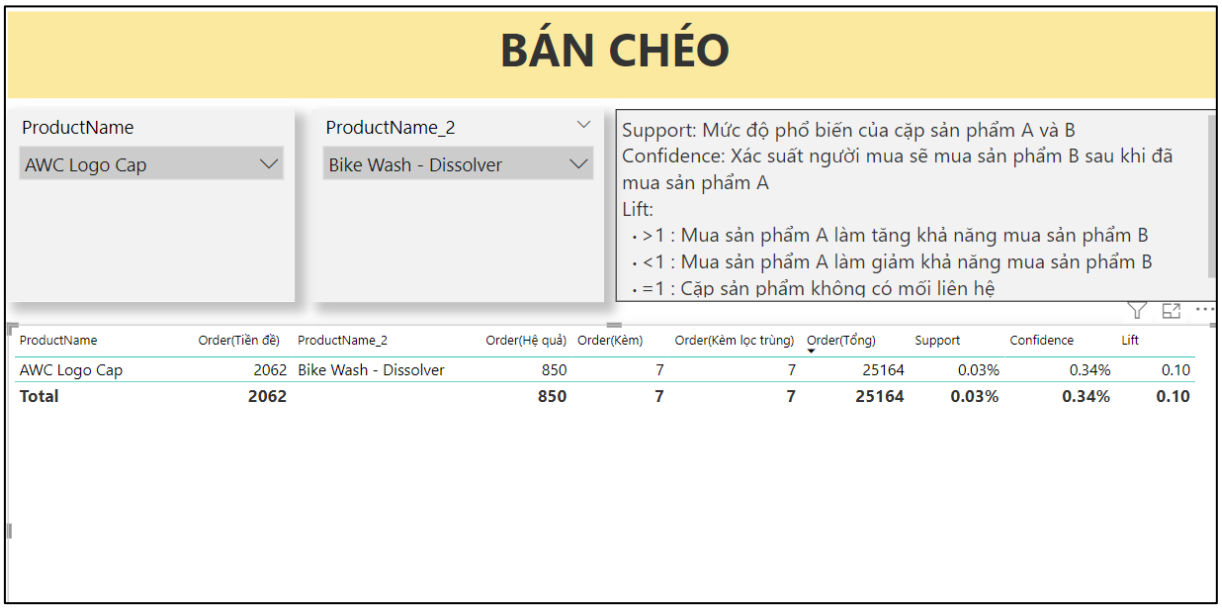
Để lọc ra những danh sách KH cho nhân viên Sales có thể dễ dàng lấy được sẽ tạo một bảng để lọc ra những danh sách KH đó. Và bảng cuối cùng là treemap . Bảng treemap này có thể xem số lượng KH của từng phân khúc. Như (hình 3.24) dưới đây là chọn xem số lượng khách hàng của phân khúc Loyal



Hình 3. 24 Xem số lượng khách hàng phân khúc Loyal

- **Đối với trang trực quan của phân tích bán chéo**. Thực hiện thao tác trực tiếp báo cáo, chọn hiển thị báo cáo theo bộ lọc của trang bán chéo. Dưới

(hình 3.25) đây là thực hiện sử dụng bộ lọc xem sự tương quan của hai sản phẩm AWC Logo Cap và Bike Wash – Dissolver .



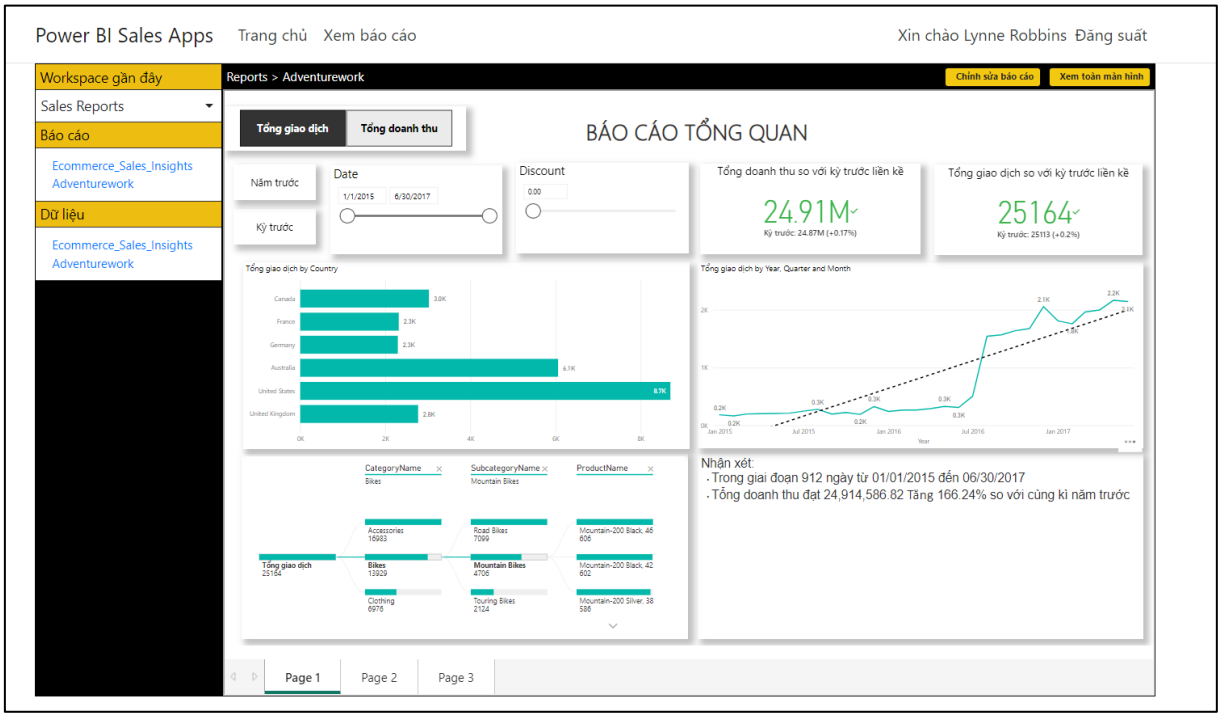
ProductName	Order(Tiền đề)	ProductName_2	Order(Hệ quả)	Order(Kèm)	Order(Kèm lọc trùng)	Order(Tổng)	Support	Confidence	Lift
AWC Logo Cap	2062	Bike Wash - Dissolver	850	7	7	25164	0.03%	0.34%	0.10
Total	**2062**		**850**	**7**	**7**	**25164**	**0.03%**	**0.34%**	**0.10**

Hình 3. 25 Thao tác bộ lọc

3.4.2 Đánh giá kết quả

Sau khi thực hiện các bước trong (phần 3.4.1) ở trên, cơ bản thu được những kết quả sau:

1. **Nhúng và chỉnh sửa báo cáo:**
   - Các báo cáo được nhúng thành công và hiển thị đúng chế độ (chỉ đọc hoặc chỉnh sửa) dựa trên trạng thái hiện tại.
   - Chế độ chỉnh sửa và lưu báo cáo hoạt động ổn định, cho phép người dùng tạo và chỉnh sửa báo cáo trực tiếp từ giao diện ứng dụng.



Hình 3. 26 Chức năng xem báo cáo

2. **Tương tác người dùng:**
    - Giao diện người dùng được cải thiện với tính tương tác cao hơn, đặc biệt là khả năng chuyển đổi giữa các workspaces và báo cáo.
    - Tính năng toàn màn hình và chế độ chỉnh sửa giúp nâng cao trải nghiệm người dùng khi làm việc với các báo cáo.
3. **Hiệu suất và độ ổn định:**
    - Ứng dụng hoạt động ổn định, không gặp phải lỗi nghiêm trọng trong quá trình kiểm tra.
    - Thời gian phản hồi của ứng dụng khi chuyển đổi giữa các workspaces và báo cáo là nhanh chóng, cải thiện hiệu suất tổng thể của ứng dụng.
4. **Đánh giá nội dung báo cáo được nhúng**

Có thể thấy được khi sử dụng quy luật Pareto có những cải tiến hơn so với nghiên cứu trước đây sử dụng sử dụng mô hình RFM kết hợp với thuật toán K-means[6], sử dụng kỹ thuật định tính và định lượng kết hợp thuật toán phân cụm K-means [7]. Bởi những nghiên cứu này đều gặp hạn chế trong việc tìm được nhóm khách hàng quan trọng. Lý do là khi mô hình RFM kết hợp với thuật toán

K-means sẽ tập trung vào việc phân chia khách hàng thành các nhóm dựa trên dữ liệu lịch sử mua hàng, nhưng sẽ bỏ qua tầm quan trọng tương đối của từng nhóm khách hàng đối với tổng doanh thu của doanh nghiệp. Đồng thời, khi hoàn thành phân khúc thì không 20% khách hàng tốt nhất đem lại bao nhiêu % doanh thu ?(trả lời cho câu hỏi biết nên tập trung vào KH cũ hay KH mới, 80% doanh thu tới từ các phân khúc khách hàng nào.

Về báo cáo phân tích sản phẩm bán chéo, đây là dạng báo cáo được ứng dụng trong thực tế rất cao, những bộ lọc kết hợp bảng biểu đã thực hiện lọc được những cặp sản phẩm nên được bán cùng nhau. Từ đó, định hình lại chiến lược kinh doanh của doanh nghiệp, góp phần vào mục đích tăng doanh thu.

3.5. Tổng kết chương

Qua việc nghiên cứu và phân tích trong chương này nhận thấy sự quan trọng của việc sử dụng công nghệ phân tích dữ liệu Power BI trong quá trình kinh doanh cụ thể là trong ngành bán lẻ. Bằng một loạt các triển khai xây dựng ứng dụng web cho phép tích hợp báo cáo và thực hiện các thao tác tùy chỉnh, cũng như mô tả chi tiết quá trình từ kết nối dữ liệu cho đến khai phá và trực quan dữ liệu thành những thông tin có ý nghĩa. Từ đó, tiến hành triển khai ứng dụng và đánh ứng dụng so với mục tiêu đã đặt ra.

KẾT LUẬN

Bài báo cáo đồ án tốt nghiệp này tập trung vào việc xây dựng trang web được nhúng báo cáo phân tích dữ liệu bán hàng sử dụng công nghệ Power BI. Ba chương của báo cáo đã thể hiện được quá trình nghiên cứu và phát triển ứng dụng này, từ lý thuyết đến ứng dụng thực tiễn và đề xuất hướng phát triển trong tương lai.

Chương 1: Tổng quan về phân tích dữ liệu bán hàng và Power BI. Chương này giới thiệu về tầm quan trọng của phân tích dữ liệu bán hàng trong kinh doanh và công nghệ Power BI. Chương 2: Phân tích và thiết kế hệ thống. Chương này tập trung vào quá trình phân tích và thiết kế trang báo cáo phân tích dữ liệu bán hàng. Chương 3: Xây dựng và thử nghiệm hệ thống. Trong chương này, mô tả quá trình xây dựng và thử nghiệm trang website phân tích dữ liệu bán hàng sử dụng Power BI. Từ xây dựng website, chuẩn bị dữ liệu và thiết kế mô hình, đến tích hợp dữ liệu và phân tích dữ liệu, tất cả được thực hiện và đánh giá kết quả. Người dùng có thể dễ dàng điều hướng và làm việc với các báo cáo và không gian làm việc (workspaces) khác nhau, nâng cao hiệu quả công việc và trải nghiệm tổng thể khi sử dụng ứng dụng.

Trong tương lai, ứng dụng có thể được mở rộng thêm với các tính năng như: Hỗ trợ nhiều loại báo cáo và bộ dữ liệu (datasets) khác nhau, tích hợp các công cụ phân tích và trực quan hóa dữ liệu nâng cao, tối ưu hóa hiệu suất cho việc xử lý các báo cáo lớn và phức tạp. Những cải tiến này sẽ tiếp tục nâng cao giá trị và khả năng của ứng dụng, đồng thời đáp ứng tốt hơn nhu cầu của người dùng trong việc quản lý và phân tích dữ liệu.

TÀI LIỆU THAM KHẢO

[1] TS. Đinh Tiên Minh &ThS. Lê Vũ Lan Oanh, "Phân khúc khách hàng mua sắm dựa trên thuộc tính của các trung tâm thương mại tại Thành phố Hồ Chí Minh," Tạp Chí Công Thương, HCM, 2020.

[2] TS.Nguyễn Thị Hội & ThS.Bùi Quang Trường, "Một số công cụ phân tích dữ liệu thông minh và ứng dụng trong các bài toán kinh tế," Tạp Chí công Thương, 2022.

[3] S. N. D. Thành H T, "Một nghiên cứu liên ngành giữa phân tích phân khúc khách hàng trong marketing và phương pháp học máy," Tạp chí Phát triển Khoa học và Công nghệ, 2015.

[4] T. N. T. Hải, "Giải pháp phát triển thị trường bán lẻ Việt Nam," Tạp chí Công Thương, Khoa Thương mại,Trường Đại học Kinh tế, 2022.

[5] "Competing in a data-driven world," McKinsey Global Institute, 2016.

[6] G. B. Mandava, "Analysis and Design of Visualization of Educational Institution Database using Power BI Tool," *Global Journal of Computer Science and Technology. Software & Data Engineering,* vol. Vol 1(4), 2018.

[7] Jo-Ting Wei, Shih-Yen Lin, and Hsin-Hung Wu, "A Review of the Application of RFM Model," *African Journal of Business Management,* vol. 4, no. 19, p. 4199–4206, 2010.

[8] J. Juran, "What is 80/20 Pareto’s Law," [Online]. Available: http://ww38.80-20presentationrule.com/. [Accessed 2008].

[9] Agrawal, R. and Srikant, "Fast algorithms for mining association rules," *VLDB'94,* pp. 487-499, 1994.

[10] "RelationalDBDesign," 1 March 2024. [Online]. Available: https://www.relationaldbdesign.com/.

PHỤ LỤC

1. Phương pháp Collaborative Filtering

Đánh giá phương pháp Collaborative Filtering

Collaborative Filtering là một kỹ thuật quan trọng trong lĩnh vực hệ thống gợi ý (recommendation systems), với mục tiêu cung cấp cho người dùng các đề xuất về sản phẩm, dịch vụ hoặc nội dung mà họ có thể quan tâm dựa trên dữ liệu về hành vi hoặc sở thích của họ. Collaborative Filtering có hai hình thức chính:

Hình thức 1: User-Based Collaborative Filtering**:** Kỹ thuật này dựa trên sự tương đồng giữa các người dùng, xác định mức độ tương tự dựa trên lịch sử đánh giá hoặc sở thích. Sau đó, nó sẽ gợi ý các sản phẩm mà người dùng khác (có sở thích tương tự) đã thích nhưng người dùng đang xem xét.

Ưu điểm:

- Phát hiện và ghi nhận thay đổi trong sở thích và hành vi của khách hàng.
- Cung cấp những gợi ý đa dạng, hữu ích, và có khả năng thích ứng cao.
- Tăng khả năng bán chéo sản phẩm.
- Phát hiện và ghi nhận điểm tương đồng của nhiều khách hàng khác nhau.

Nhược điểm:

- Cần thời gian để thấu hiểu dữ liệu từ khách hàng mới hoặc sản phẩm mới.
- Cần có tệp dữ liệu lớn để trích xuất thông tin.
- Gặp khó khăn trong việc đề xuất sản phẩm mới nếu chưa có khách hàng nào khác đã mua hoặc đánh giá sản phẩm đó.

Hình thức 2: Item-Based Collaborative Filtering: Kỹ thuật này tập trung vào sự tương đồng giữa các sản phẩm, xác định mức độ tương tự dựa trên cách người dùng đã tương tác với chúng. Hệ thống sau đó gợi ý các sản phẩm tương tự với những gì người dùng đang xem xét.

Ưu điểm:

- Có thể áp dụng trong trường hợp không có nhiều dữ liệu.
- Đưa ra đề xuất dựa trên lịch sử của khách hàng.
- Tốc độ xử lý nhanh.

Nhược điểm:

- Chỉ có thể đề xuất các sản phẩm tương tự, không đa dạng hóa được đề xuất.
- Chất lượng đề xuất phụ thuộc vào việc gắn nhãn cho sản phẩm.

Một kỹ thuật phổ biến trong Collaborative Filtering là phân rã ma trận (Matrix Factorization). Ý tưởng cơ bản của Matrix Factorization là phân rã ma trận đánh giá thành hai ma trận nhỏ hơn, một ma trận cho người dùng và một ma trận cho các sản phẩm, sao cho tích của hai ma trận này xấp xỉ ma trận đánh giá ban đầu.

Đánh giá hiệu quả hệ thống recommendation

PWC cung cấp báo cáo chi tiết về đánh giá hiệu quả của hệ thống gợi ý sản phẩm thông qua các chỉ số như:

- Tỷ lệ nhấp chuột (clickthrough rates)
- Tỷ lệ chuyển đổi (conversion rates)
- Mức độ tương tác của người dùng (user engagement)
- Sự thay đổi trong khối lượng bán hàng (change in sales volume)
- Tác động đến phân phối bán hàng (impact on sales distribution)
- Trải nghiệm khách hàng với các gợi ý (customer experience with recommendations)
- Mức độ cá nhân hóa (level of personalisation)
- Giảm thiểu quá tải thông tin (reduced information overload)
- Tạo doanh thu (revenue generation)
- Độ bao phủ - số lượng người dùng và sản phẩm được phân tích (coverage)
- Đa dạng hóa - mức độ bán chéo và gợi ý đa dạng (diversity)

- Khả năng tổng quát hóa - gợi ý yếu/mạnh (generalisation)
- Xử lý người dùng mới (handling new users)

2. **Phương án xây dựng chiến dịch Marketing**

Nếu mô hình được đánh giá là hiệu quả, nó sẽ được sử dụng để chọn các sản phẩm để quảng cáo và truyền thông đến khách hàng. Do hành vi giao dịch và thị hiếu của khách hàng thường thay đổi, các quy tắc về liên kết sản phẩm cũng cần được cập nhật định kỳ, khoảng 3 đến 6 tháng một lần. Quy trình thực hiện sẽ như sau:

- Đơn vị kinh doanh nêu nhu cầu hỗ trợ đưa các sản phẩm vào nội dung thông điệp quảng cáo.
- Đơn vị phân tích dữ liệu thực hiện phân tích mối liên kết các sản phẩm trong một khung thời gian gần đây.
- Đơn vị phân tích dữ liệu đưa ra danh sách các quy tắc liên kết sản phẩm.
- Đơn vị phân tích và đơn vị kinh doanh review các quy tắc này, chọn một quy tắc phù hợp nhất cho mục tiêu xây dựng chương trình.
- Thông điệp quảng cáo được đơn vị kinh doanh xây dựng dựa trên danh sách các sản phẩm được chọn.
- Triển khai thông điệp quảng cáo qua Web, App, Email, hoặc marketing trực tiếp tới khách hàng.

3. **Kết hợp xây dựng mô hình dự đoán khách hàng chấp nhận quảng cáo tiếp thị sản phẩm.**

Sau khi triển khai gửi chương trình tiếp thị quảng cao theo kết quả của quy luật kết hợp (association rules) trong khoảng 6 tháng, việc ghi nhận và theo dõi phản hồi của khách hàng sẽ được thực hiện. Các phản hồi này (chấp nhận/không chấp nhận) sẽ được gán nhãn: 0 cho khách hàng chấp nhận tiếp thị và 1 cho khách hàng không chấp nhận tiếp thị. Dữ liệu này sẽ được sử dụng để xây dựng mô hình dự đoán khả năng khách hàng chấp nhận tiếp thị.

Mục tiêu: Dự đoán nhu cầu khách hàng, giúp tối ưu hóa Marketing tới những khách hàng có nhu cầu thực sự, tránh spam quảng cáo và tiết kiệm chi phí.

Thuật toán áp dụng: Các thuật toán phân loại truyền thống như hồi quy logistic, decision tree, XGBoost, v.v. Có thể sử dụng các phương pháp học máy khác nhau để đạt hiệu quả tối ưu.

Kết quả đầu ra: Mô hình sẽ trả về điểm/xác suất chấp nhận tiếp thị của khách hàng. Điểm càng cao thì khả năng khách hàng chấp nhận tiếp thị sản phẩm càng lớn.

Phương pháp đánh giá hiệu suất mô hình: Chỉ số đánh giá hiệu suất phổ biến là hệ số Gini. Mô hình có hệ số Gini lớn hơn 30% là có khả năng sử dụng được, và nếu lớn hơn 50% thì mô hình được coi là có khả năng phân biệt tốt.