Dự án **"Phân Tích Hành Vi Khách Hàng Dựa Trên Dữ Liệu Mua Hàng Của Công ty DataSquad Retail"**

Nguyễn Đức Anh

Mục lục

[Giới thiệu dự án 1](#_Toc191925962)

[1.1 Giới thiệu 1](#_Toc191925963)

[1.2 Yêu cầu của công ty 1](#_Toc191925964)

[1.3 Lập kế hoạch dự án 1](#_Toc191925965)

[2 Phân tích yêu cầu công ty 2](#_Toc191925966)

[2.1 Phân tích yêu cầu 2](#_Toc191925967)

[2.2 Câu chuyện dữ liệu 5](#_Toc191925968)

[2.2.1 Đặt vấn đề 5](#_Toc191925969)

[2.2.2 Xác định vấn đề 5](#_Toc191925970)

[2.2.3 Xác định rõ đối tượng 5](#_Toc191925971)

[2.3 Kiến trúc hệ thống 6](#_Toc191925972)

[2.3.1 Kiến trúc 6](#_Toc191925973)

[2.3.2 Giải thích 7](#_Toc191925974)

[2.4 Giải thích về bộ dữ liệu khách hàng 7](#_Toc191925975)

[2.4.1 Các khái niệm 7](#_Toc191925976)

[2.4.2 Các trường dữ liệu 8](#_Toc191925977)

[3 Làm sạch và chuyển đổi dữ liệu 9](#_Toc191925978)

[3.1 Chuẩn bị dữ liệu 9](#_Toc191925979)

[3.1.1 Giải pháp lưu trữ dữ liệu 9](#_Toc191925980)

[3.2 Làm sạch dữ liệu 13](#_Toc191925981)

[3.2.1 Các bước làm sạch dữ liệu 14](#_Toc191925982)

[3.3 Chuyển đổi dữ liệu 19](#_Toc191925983)

[3.3.1 Các trường hợp cần chuyển đổi 19](#_Toc191925984)

[3.3.2 Trình bày các phép chuyển đổi trong dự án 19](#_Toc191925985)

[4 Xử lý dữ liệu 24](#_Toc191925986)

[4.1 Chuẩn hóa dữ liệu 24](#_Toc191925987)

[4.1.1 Trình bày các bước chuẩn hóa trong dự án 24](#_Toc191925988)

[4.2 Mô hình hóa dữ liệu 35](#_Toc191925989)

[4.2.1 Trình bày các bước mô hình hóa 35](#_Toc191925990)

[5 Trực quan hóa dữ liệu 37](#_Toc191925991)

[5.1 Các kỹ thuật trực quan hóa 37](#_Toc191925992)

[5.2 Các nguyên tắc trực quan hóa 38](#_Toc191925993)

[5.3 Trình bày tạo các report cho dự án 39](#_Toc191925994)

[5.3.1 Tạo visual thống kê chi tiết 39](#_Toc191925995)

[5.3.2 Tạo visual dự báo 50](#_Toc191925996)

[6 Xây dựng báo cáo 53](#_Toc191925997)

[6.1 Dashboard và report 53](#_Toc191925998)

[6.2 Xây dựng báo cáo 54](#_Toc191925999)

[6.2.1 Dashboard vs Report 54](#_Toc191926000)

[6.2.2 Dashboard 54](#_Toc191926001)

[7 KẾT LUẬN 59](#_Toc191926002)

[7.1 Báo cáo 59](#_Toc191926003)

[7.1.1 Các bước viết báo cáo 59](#_Toc191926004)

[7.2 Khó khăn 59](#_Toc191926005)

[7.3 Thuận lợi 59](#_Toc191926006)

[7.4 Hướng phát triển 60](#_Toc191926007)

# Giới thiệu dự án

## Giới thiệu

Dự án này nhằm phân tích dữ liệu khách hàng của **Công ty** **DataSquad Retail** với mục tiêu chính là tìm hiểu hành vi mua sắm, phân tích hiệu quả chiến dịch marketing và đề xuất các chiến lược cải thiện trải nghiệm khách hàng. Thông qua việc phân tích dữ liệu, công ty DataSquad Retail có thể cá nhân hóa các chiến dịch tiếp thị, từ đó nâng cao tỷ lệ tương tác của khách hàng và gia tăng doanh thu.

## Yêu cầu của công ty

 Phân tích nhân khẩu học

 Phân tích hành vi mua sắm

 Kết hợp các yếu tố để xác định các phân khúc khách hàng

 Xác định hồ sơ và đặc điểm từng phân khúc

 Ứng dụng vào chiến lược marketing

 Đánh giá và tối ưu hóa

## Lập kế hoạch dự án

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **TT** | **Hạng mục** | **Bắt đầu** | **Kết thúc** | **Kết quả** |
| 1 | Giới thiệu dự án | 14/9/2024 | 15/9/2024 | Hoàn thành |
| 1.1 | Giới thiệu | 15/9/2024 | 15/9/2024 | Hoàn thành |
| 1.2 | Yêu cầu | 15/9/2024 | 16/9/2024 | Hoàn thành |
| 2 | Phân tích yêu cầu | 19/9/2024 | 21/9/2024 | Hoàn thành |
| 2.1 | Phân tích yêu cầu KH | 21/9/2024 | 21/9/2024 | Hoàn thành |
| 2.2 | Câu chuyện dữ liệu | 22/9/2024 | 23/9/2024 | Hoàn thành |
| 2.3 | Cấu trúc hệ thông | 23/9/2024 | 25/9/2024 | Hoàn thành |
| 2.4 | Giải thích về bộ dữ liệu khách hàng | 25/9/2024 | 26/9/2024 | Hoàn thành |
| 3 | Làm sạch và chuyển đổi dữ liệu | 26/9/2024 | 28/9/2024 | Hoàn thành |
| 3.1 | Chuẩn bị dữ liệu | 28/9/2024 | 30/9/2024 | Hoàn thành |
| 3.2 | Làm sạch dữ liệu | 30/9/2024 | 2/10/2024 | Hoàn thành |
| 3.3 | Chuyển đổi dữ liệu | 2/10/2024 | 2/10/2024 | Hoàn thành |
| 4 | Xử lý dữ liệu | 12/10/2024 | 12/10/2024 | Hoàn thành |
| 4.1 | Chuẩn hóa dữ liệu | 12/10/2024 | 13/10/2024 | Hoàn thành |
| 4.2 | Mô hình hoá dữ liệu | 13/10/2024 | 15/10/2024 | Hoàn thành |
| 5 | Trực quan hóa dữ liệu | 23/10/2024 | 25/10/2024 | Hoàn thành |
| 5.1 | Các kĩ thuật trực quan hóa | 25/10/2024 | 26/10/2024 | Hoàn thành |
| 5.2 | Các nguyên tắc trực quan hoá | 26/10/2024 | 28/10/2024 | Hoàn thành |
| 5.3 | Trình bày cách thêm visual mới | 28/10/2024 | 30/10/2024 | Hoàn thành |
| 6 | Xây dựng báo cáo | 3/11/2024 | 5/11/2024 | Hoàn thành |
| 6.1 | Dashboard và Report | 6/11/2024 | 9/11/2024 | Hoàn thành |
| 6.2 | Xây dựng báo cáo | 11/11/2024 | 13/11/2024 | Hoàn thành |
| 7 | Kết luận | 13/11/2024 | 16/11/2024 | Hoàn thành |
| 7.1 | Báo cáo | 17/11/2024 | 20/11/2024 | Hoàn thành |
| 7.2 | Khó khăn | 21/11/2024 | 23/11/2024 | Hoàn thành |
| 7.3 | Thuận lợi | 23/11/2024 | 24/11/2024 | Hoàn thành |
| 7.4 | Hướng phát triển | 24/11/2024 | 25/11/2024 | Hoàn thành |

# Phân tích yêu cầu công ty

## Phân tích yêu cầu

**1. Phân tích nhân khẩu học**

Phân tích nhân khẩu học giúp hiểu rõ hơn về các đặc điểm cơ bản của khách hàng. Các yếu tố nhân khẩu học phổ biến bao gồm:

* Tuổi: Nhóm tuổi có thể cho thấy sở thích tiêu dùng khác nhau.
* Giới tính: Giúp xác định các sản phẩm hoặc thông điệp tiếp thị phù hợp với từng giới.
* Thu nhập: Ảnh hưởng đến khả năng chi tiêu và quyết định mua hàng.
* Trình độ học vấn và nghề nghiệp: Giúp hiểu rõ hơn về phong cách sống và sở thích mua sắm.

Sau khi thu thập dữ liệu này, sử dụng các kỹ thuật thống kê như phân tích tần suất hoặc mô hình hồi quy để xác định các nhóm khách hàng tiềm năng dựa trên đặc điểm nhân khẩu học.

**2. Phân tích hành vi mua sắm**

Phân tích hành vi mua sắm giúp làm rõ hơn về sở thích, thói quen và hành vi chi tiêu của khách hàng. Các yếu tố hành vi quan trọng gồm:

* Tần suất mua hàng: Khách hàng mua thường xuyên có xu hướng trung thành hơn.
* Mức chi tiêu trung bình: Xác định mức độ tài chính và tiềm năng tiêu thụ của từng nhóm.
* Loại sản phẩm ưa thích: Giúp tìm hiểu sở thích sản phẩm của các nhóm khách hàng.
* Kênh mua sắm ưa thích: Có thể bao gồm trực tuyến, tại cửa hàng, qua ứng dụng di động, hoặc trên mạng xã hội.
* Thời điểm mua sắm: Các xu hướng về thời gian mua hàng (theo ngày, giờ hoặc mùa) có thể giúp xây dựng chiến lược khuyến mãi hợp lý.

**3. Kết hợp các yếu tố để xác định các phân khúc khách hàng**

Sau khi đã phân tích nhân khẩu học và hành vi mua sắm, kết hợp các yếu tố này để tạo ra các phân khúc khách hàng. Có thể thực hiện qua các phương pháp như:

* Phân tích cụm: Kỹ thuật này nhóm khách hàng dựa trên các yếu tố giống nhau để xác định các nhóm có đặc điểm tương đồng.
* Phân khúc dựa trên giá trị khách hàng: Phân chia khách hàng thành các nhóm dựa trên giá trị lâu dài (tính toán bằng tổng chi tiêu dự kiến).
* Phân khúc dựa trên lợi ích: Nhóm khách hàng theo lợi ích mà họ tìm kiếm từ sản phẩm, ví dụ: chất lượng cao, giá rẻ, dịch vụ tiện lợi.

**4. Xác định hồ sơ và đặc điểm từng phân khúc**

Dựa vào các phân khúc đã hình thành, tạo hồ sơ khách hàng (customer persona) cho từng nhóm. Mỗi hồ sơ cần bao gồm:

* Đặc điểm chung: Tuổi, giới tính, thu nhập, nghề nghiệp, nơi sống.
* Hành vi mua sắm: Các sản phẩm, dịch vụ ưa thích, tần suất mua hàng, mức chi tiêu trung bình.
* Động lực và nhu cầu chính: Những yếu tố ảnh hưởng đến quyết định mua hàng.
* Kênh tiếp cận ưa thích: Nhận diện các kênh và phương pháp tiếp cận phù hợp cho từng phân khúc.

**5. Ứng dụng vào chiến lược marketing**

Với từng phân khúc đã xác định, doanh nghiệp có thể:

* Tạo các chương trình khuyến mãi và ưu đãi phù hợp với từng nhóm.
* Phát triển nội dung tiếp thị cá nhân hoá.
* Tối ưu hoá kênh bán hàng và trải nghiệm khách hàng, ví dụ, ưu tiên nền tảng thương mại điện tử cho nhóm trẻ yêu thích mua sắm trực tuyến.

**6. Đánh giá và tối ưu hóa**

Công ty bán lẻ này cần thường xuyên đánh giá hiệu quả của từng phân khúc khách hàng để điều chỉnh các chiến lược tiếp thị và bán hàng phù hợp. Ví dụ, nếu nhóm khách hàng trẻ bắt đầu giảm tỷ lệ mua sắm, họ có thể cần cải thiện tính năng mua sắm trực tuyến hoặc tạo ra các chiến dịch tiếp thị hấp dẫn hơn trên mạng xã hội như sau:

**Nhóm khách hàng trẻ (18-30 tuổi)**:

* + Yêu thích thời trang mới nhất và thường xuyên theo dõi các xu hướng mới.
  + Chủ yếu mua hàng qua mạng xã hội và thường tìm kiếm các chương trình khuyến mãi online.
  + Có xu hướng tham gia vào các sự kiện giảm giá hoặc flash sale.

**Nhóm khách hàng gia đình (30-50 tuổi)**:

* + Quan tâm đến giá cả và chất lượng sản phẩm, thường mua sắm cho cả gia đình.
  + Thích mua sắm trực tiếp tại cửa hàng, nơi có thể xem và thử sản phẩm trước khi mua.
  + Có xu hướng chọn những sản phẩm phù hợp cho trẻ nhỏ và nhu cầu gia đình.

**Nhóm khách hàng trung niên (50 tuổi trở lên)**:

* + Thu nhập cao và ưa chuộng các sản phẩm cao cấp, thường tìm kiếm sự sang trọng và chất lượng trong mua sắm.
  + Chủ yếu mua hàng qua kênh trực tuyến của thương hiệu, thích có trải nghiệm mua sắm thuận tiện và an toàn.
  + Thích các chương trình khách hàng thân thiết và dịch vụ chăm sóc khách hàng tốt.

## Câu chuyện dữ liệu

### Đặt vấn đề

* **Mô tả thực trạng:**

**Hiện nay, Công ty DataSquad Retail** đang gặp phải nhiều khó khăn trong việc tăng cường tương tác khách hàng và cải thiện hiệu quả của các chiến dịch marketing. Cụ thể, mặc dù công ty đã tiến hành nhiều chiến dịch quảng cáo qua email và các chương trình khuyến mãi, nhưng **tỷ lệ phản hồi** từ phía khách hàng vẫn còn rất thấp. Điều này dẫn đến việc **doanh thu không tăng trưởng mạnh** và **mức độ hài lòng của khách hàng** vẫn chưa đạt như kỳ vọng.

Ngoài ra, dữ liệu về khách hàng, bao gồm thông tin nhân khẩu học và hành vi chi tiêu, đã được thu thập qua nhiều kênh khác nhau, nhưng **chưa được phân tích một cách hiệu quả** để tối ưu hóa các chiến lược marketing. Hiện tại, công ty DataSquad Retail đang tìm kiếm giải pháp để **tận dụng dữ liệu** này nhằm **cá nhân hóa các chiến dịch tiếp thị** và **nâng cao trải nghiệm khách hàng**, với hy vọng thúc đẩy tăng trưởng doanh thu.

### Xác định vấn đề

* Dựa trên các phân tích, đề xuất các chiến lược marketing cụ thể nhằm tối ưu hóa sự tương tác của khách hàng với sản phẩm/dịch vụ của doanh nghiệp, từ đó nâng cao doanh thu.
* Tăng cường việc phân tích và dự đoán hành vi của các nhóm khách hàng khác nhau, nhằm xây dựng chiến lược tiếp thị cá nhân hóa và phản hồi kịp thời cho các nhu cầu của khách hàng.

### Xác định rõ đối tượng

1. **Các nhà quản lý marketing**:
   * Chịu trách nhiệm phát triển và thực hiện các chiến lược marketing tổng thể. Cần có dữ liệu phân tích để đưa ra quyết định đúng đắn về cách tiếp cận thị trường và điều chỉnh các chiến dịch để tăng cường sự tương tác của khách hàng.
2. **Đội ngũ kinh doanh**:
   * Là những người trực tiếp bán hàng và hiểu rõ nhu cầu của khách hàng. Họ cần thông tin chi tiết về hành vi và phản hồi của khách hàng để tối ưu hóa quy trình bán hàng và cải thiện mối quan hệ với khách hàng.
3. **Bộ phận chăm sóc khách hàng**:
   * Đóng vai trò quan trọng trong việc tạo ra trải nghiệm khách hàng tốt. Cần nắm rõ các phản hồi và thắc mắc của khách hàng để cải thiện dịch vụ, từ đó nâng cao sự hài lòng và trung thành của khách hàng.

## Kiến trúc hệ thống

### Kiến trúc

A diagram of a computer program

Description automatically generated

### Giải thích

**CSV: Lấy dữ liệu**

* **Mục đích:** Sử dụng CSV để thu thập và lưu trữ dữ liệu thô từ các nguồn trước khi thực hiện các bước xử lý và phân tích.

**SQL Server: Làm file Backup để đề phòng dữ liệu có biến cố**

* **Mục đích:** Tạo ra các bản sao lưu (Backup) định kỳ của cơ sở dữ liệu để đề phòng các sự cố xảy ra như mất dữ liệu hoặc sự cố hệ thống. Nhờ đó, nếu có bất kỳ biến cố nào, nhóm có thể dễ dàng khôi phục lại dữ liệu mà không mất thông tin quan trọng.

**Python: Làm sạch dữ liệu các thông tin bị trùng lặp, bị mất, chuẩn hoá, …**

* **Mục đích:** Đảm bảo dữ liệu trước khi phân tích không bị nhiễu, thiếu sót hay trùng lặp thông tin. Python sẽ giúp tự động hóa quá trình này, giảm thiểu sai sót và đảm bảo chất lượng dữ liệu, từ đó giúp việc phân tích và trực quan hóa dữ liệu chính xác hơn.

**Tableau: Tạo nên câu chuyện và thiết kế các biểu đồ**

* **Mục đích:** Sử dụng Tableau để tạo các biểu đồ, báo cáo và dashboards nhằm cung cấp cái nhìn tổng quan, bao quát về dữ liệu. Điều này giúp người dùng dễ dàng nắm bắt các xu hướng, phát hiện các vấn đề và ra quyết định kinh doanh dựa trên dữ liệu.

## Giải thích về bộ dữ liệu khách hàng

### Các khái niệm

Một số thuật ngữ phổ biến trong ngành bán lẻ:

* **SKU (Stock Keeping Unit)**: Đơn vị lưu kho, thường được sử dụng để xác định và theo dõi hàng hóa trong kho. Mỗi SKU thường đại diện cho một sản phẩm duy nhất hoặc một biến thể của sản phẩm (như kích cỡ, màu sắc).
* **OMNI-channel**: Chiến lược bán hàng và tiếp thị tích hợp qua nhiều kênh (như cửa hàng vật lý, website, ứng dụng di động) để cung cấp trải nghiệm liền mạch cho khách hàng. Đây là thuật ngữ quan trọng trong tiếp thị hiện đại, nhưng không phải ai cũng biết rõ về nó.
* **Promotions and Discounts**: Các chương trình khuyến mãi và giảm giá thường được sử dụng trong bán lẻ để thu hút khách hàng, nhưng có thể cần giải thích chi tiết về các loại và cách thực hiện chúng.

### Các trường dữ liệu

Một số trường dữ liệu quan trọng:

|  |  |  |
| --- | --- | --- |
| **Stt** | **Tên trường dữ liệu** | **Nội dung** |
| 1 | ID | Mã định danh duy nhất của khách hàng |
| 2 | Year Birth | Năm sinh của khách hàng |
| 3 | Education | Trình độ học vấn của khách hàng |
| 4 | Marital Status | Tình trạng hôn nhân của khách hàng |
| 5 | Income | Thu nhập hàng năm của hộ gia đình khách hàng |
| 6 | Kidhome | Số lượng trẻ em trong hộ gia đình của khách hàng |
| 7 | Teenhome | Số lượng thanh thiếu niên trong hộ gia đình của khách hàng |
| 8 | Dt Customer | Ngày đăng ký của khách hàng với công ty |
| 9 | Recency | Số ngày kể từ lần mua hàng gần nhất của khách hàng |
| 10 | Complain | Giá trị 1 nếu khách hàng đã khiếu nại trong 2 năm qua, 0 nếu không |
| 11 | MntWines | Số tiền chi tiêu cho rượu vang trong 2 năm qua |
| 12 | MntFruits | Số tiền chi tiêu cho trái cây trong 2 năm qua |
| 13 | MntMeatProducts | Số tiền chi tiêu cho thịt trong 2 năm qua |
| 14 | MntFishProducts | Số tiền chi tiêu cho cá trong 2 năm qua |
| 15 | MntSweetProducts | Số tiền chi tiêu cho đồ ngọt trong 2 năm qua |
| 16 | MntGoldProds | Số tiền chi tiêu cho vàng trong 2 năm qua |
| 17 | NumDealsPurchases | Số lượng giao dịch mua hàng có áp dụng giảm giá |
| 18 | AcceptedCmp1 | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ nhất, 0 nếu không |
| 19 | AcceptedCmp2 | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ hai, 0 nếu không |
| 20 | AcceptedCmp3 | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ 3, 0 nếu không |
| 21 | AcceptedCmp4 | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ tư, 0 nếu không |
| 22 | AcceptedCmp5 | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch thứ năm, 0 nếu không |
| 23 | Response | Giá trị 1 nếu khách hàng chấp nhận ưu đãi trong chiến dịch cuối cùng, 0 nếu không |
| 24 | NumWebPurchases | Số lượng giao dịch mua hàng được thực hiện qua trang web của công ty |
| 25 | NumCatalogPurchases | Số lượng giao dịch mua hàng được thực hiện thông qua danh mục sản phẩm |
| 26 | NumStorePurchases | Số lượng giao dịch mua hàng được thực hiện trực tiếp tại cửa hàng |
| 27 | NumWebVisitsMonth | Số lần truy cập vào trang web của công ty trong tháng qua |

# Làm sạch và chuyển đổi dữ liệu

## Chuẩn bị dữ liệu

### Giải pháp lưu trữ dữ liệu

**Giải pháp lưu trữ dữ liệu:** Dự án sẽ sử dụng **SQL Server** để làm hệ quản trị cơ sở dữ liệu chính. SQL Server cung cấp một môi trường quản lý dữ liệu mạnh mẽ, hỗ trợ khả năng lưu trữ dữ liệu lớn, quản lý truy xuất, và bảo mật cao.

**Lý do chọn giải pháp:**

* **Khả năng quản lý dữ liệu lớn:** SQL Server có khả năng xử lý và quản lý lượng dữ liệu lớn, đồng thời hỗ trợ nhiều tính năng cần thiết cho dự án như sao lưu, khôi phục và nhân bản dữ liệu.
* **Tính bảo mật:** SQL Server cung cấp nhiều cơ chế bảo mật như mã hóa dữ liệu, kiểm soát quyền truy cập, và giám sát hoạt động, giúp bảo vệ dữ liệu của dự án.
* **Tích hợp dễ dàng:** SQL Server có thể tích hợp dễ dàng với nhiều công cụ khác trong hệ thống như Python (cho việc làm sạch dữ liệu) và Tableau (cho trực quan hóa dữ liệu), giúp tăng cường tính liên kết và hiệu quả trong toàn bộ quy trình xử lý dữ liệu.
* **Khả năng mở rộng:** SQL Server hỗ trợ khả năng mở rộng cơ sở dữ liệu linh hoạt, từ đó đáp ứng nhu cầu phát triển của dự án trong tương lai khi dữ liệu tăng lên.

Trình bày:

Tạo một database mới sau đó ta chọn Database và Import Flat File

A screenshot of a computer

Description automatically generated

Nhấn Next

A screenshot of a computer

Description automatically generated

Vào mục Browse > địa chỉ thư mục

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Nhấn Next

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Nhấn Import

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Như vậy là đã backup file thành công

## Làm sạch dữ liệu

### Các bước làm sạch dữ liệu

**Đầu tiên ta kiểm tra dữ liệu thiếu**

A screen shot of a computer code

Description automatically generated

Bước xử lý

A screenshot of a computer error

Description automatically generated

Sau khi xử lý

A computer code on a black background

Description automatically generated

**Tiếp theo là kiểm tra giá trị ngoại lệ**

A screenshot of a computer

Description automatically generated

Bước xử lý

A computer screen shot of a code

Description automatically generated

Sau khi xử lý

A screen shot of a graph

Description automatically generated

**Kiểm tra dữ liệu trùng lặp**

A screen shot of a computer

Description automatically generated

Bước xử lý

A screen shot of a computer

Description automatically generated

Sau khi xử lý

A screenshot of a computer program

Description automatically generated

**Kế tiếp là các cột có giá trị duy nhất**

A screen shot of a computer code

Description automatically generated

Bước xử lý

A black screen with white text

Description automatically generated

Sau khi xử lý

A screenshot of a computer

Description automatically generated

## Chuyển đổi dữ liệu

### Các trường hợp cần chuyển đổi

Ở bộ dữ liệu này, các trường dữ liệu cần chuyển đổi là:

* Education
* Martial Status
* Dt\_Customer
* AcceptedCmp1
* AcceptedCmp2
* AcceptedCmp3
* AcceptedCmp4
* AcceptedCmp5
* Complaint
* Response

### Trình bày các phép chuyển đổi trong dự án

**Chuẩn hoá cấp độ học vấn**

A computer screen shot of a program code

Description automatically generated

**Chuẩn hoá tình trạng hôn nhân**

**A computer screen shot of a computer code

Description automatically generated**

**Chuyển đổi kiểu dữ liệu**

**A screen shot of a computer code

Description automatically generated**

**Chuyển đổi định dạng cột đồng thời kiểm tra giá trị bị lỗi**

**A screen shot of a computer

Description automatically generated**

**Chuyển đổi kiểu dữ liệu**

**A screen shot of a computer program

Description automatically generated**

**A screen shot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**Tạo một số cột mới để dễ dàng cho việc phân tích**

**A screen shot of a computer

Description automatically generated**

**A screen shot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A black background with white text

Description automatically generated**

**A screen shot of a computer code

Description automatically generated**

**A screen shot of a computer program

Description automatically generated**

# Xử lý dữ liệu

## Chuẩn hóa dữ liệu

### Trình bày các bước chuẩn hóa trong dự án

Load dữ liệu từ file csv

Ta nhân bảng để tạo 1 bảng gốc (Root), chuột phải vào table và chọn Duplicate

A screenshot of a computer

Description automatically generated

Đổi tên

A screenshot of a computer

Description automatically generated

Bắt đầu chia dữ liệu thành các bảng có quan hệ với nhau, theo các nhóm liên quan về nội dung

Chúng ta tiếp tục nhân bảng bằng Reference

A screenshot of a computer

Description automatically generated

Đặt tên cho table này là ‘tbl\_KhachHang’, table này chứa các thông tin cơ bản về khách hàng:

A screenshot of a computer

Description automatically generated

Chọn các cột thích hợp

Chọn 1 cột ngẫu nhiên, sau đó chuột phải => Remove Other Columns

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Click chuột trái vào kí hiệu bánh răng và chọn lựa các cột phù hợp, sau đó nhấn OK.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Thực hiện tương tự để tạo ra các bản: Chi\_Tieu, Hanh\_Vi\_Mua\_Sam, Chien\_Dich\_Marketing, Phan\_Hoi\_Khieu\_Nai

Bảng Chi Tiêu

Chứa các thông tin chi tiết về chi tiêu của khách hàng, với mỗi mã khách hàng:

A screenshot of a computer

Description automatically generated

Bảng Hành Vi Mua Sắm

Chứa thông tin về cách thức mua hàng của khách hàng:

A screenshot of a computer

Description automatically generated

Bảng Chiến Dịch Marketing

Chứa thông tin về sự tham gia của khách hàng trong các chiến dịch marketing:

A screenshot of a computer

Description automatically generated

Bảng Phản Hồi và Khiếu Nại

Chứa thông tin về khiếu nại và phản hồi của khách hàng:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

2 bảng đầu tiên không cần Load vào Modeling, ta kích chuột phải vào và chọn Enable load.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Cuối cùng chúng ta chọn Close & Apply

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Mô hình hóa dữ liệu

### Trình bày các bước mô hình hóa

Mô Hình Quan Hệ Các Bảng:

**Ma\_KH (1) - (n) tbl\_ChiTieu:**

Một khách hàng có thể thực hiện nhiều lần chi tiêu, nhưng mỗi lần chi tiêu chỉ

thuộc về một khách hàng. Điều này giúp theo dõi chi tiêu của từng khách hàng

cụ thể.

**Ma\_KH (1) - (n) tbl\_Hanh ViMuaSam:**

Một khách hàng có thể có nhiều hành vi mua sắm khác nhau trong suốt thời

gian họ giao dịch với công ty. Mối quan hệ này nhằm ghi nhận các hành vi

mua sắm như tần suất, loại sản phẩm ưa thích và thời gian mua sắm.

**Ma\_KH (1) - (n) tbl\_ChienDichMarketing:**

Một khách hàng có thể tham gia hoặc tiếp nhận nhiều chiến dịch marketing

khác nhau, và dữ liệu này giúp phân tích mức độ ảnh hưởng của các chiến

dịch đối với hành vi của khách hàng.

**Ma\_KH (1) - (n) tbl\_PhanHoiKhieuNai:**

Một khách hàng có thể có nhiều phản hồi hoặc khiếu nại về các dịch vụ hoặc

sản phẩm của công ty. Mối quan hệ này giúp theo dõi và quản lý các phản hồi,

từ đó cải thiện chất lượng dịch vụ.

A screenshot of a computer

Description automatically generated

Mỗi khách hàng có thể có nhiều lần chi tiêu, hành vi mua sắm, và phản hồi khác nhau. Vì vậy, mỗi bảng có Mã\_KH làm khóa ngoại để liên kết dữ liệu về từng khách hàng.

# Trực quan hóa dữ liệu

## Các kỹ thuật trực quan hóa

Dự án phân tích dữ liệu khách hàng của công ty DataSquad Retail đang áp dụng các kỹ thuật trực quan hóa sau:

**1. Biểu đồ cột (Bar Chart):**

* Vì sao: Kỹ thuật này được sử dụng để so sánh mức chi tiêu giữa các nhóm khách hàng theo độ tuổi, thu nhập, hoặc tình trạng hôn nhân.
* Lợi ích: Biểu đồ cột giúp dễ dàng nhận diện sự khác biệt giữa các nhóm khách hàng, từ đó hỗ trợ việc tối ưu hóa chiến lược tiếp thị cho từng nhóm.

**2. Biểu đồ cây (Tree Map):**

* Vì sao: Tree Map được sử dụng để trực quan hóa sự đóng góp doanh thu của các nhóm khách hàng theo danh mục sản phẩm.
* Lợi ích: Kỹ thuật này giúp dễ dàng nhận diện các nhóm sản phẩm có doanh thu cao nhất hoặc đóng góp nhỏ, đồng thời cung cấp một cái nhìn tổng quan về tỷ lệ phân bố doanh thu giữa các danh mục. Tree Map rất hiệu quả khi cần hiển thị tỷ lệ tương quan trong một không gian cố định.

**3. Biểu đồ phân tán (Scatter Plot):**

* Vì sao: Để phân tích mối quan hệ giữa thu nhập và tần suất mua hàng của khách hàng.
* Lợi ích: Biểu đồ phân tán giúp xác định liệu có mối tương quan chặt chẽ giữa các yếu tố này hay không, từ đó hỗ trợ xây dựng chiến lược tiếp cận khách hàng có thu nhập cao.

**4. Bản đồ nhiệt (Heat Map):**

* Vì sao: Được sử dụng để phân tích mật độ mua sắm theo khu vực địa lý.
* Lợi ích: Điều này giúp công ty DataSquad Retail nhận diện các khu vực có khách hàng tiềm năng cao để triển khai các chiến dịch tiếp thị và quảng bá hiệu quả hơn.

**5. Biểu đồ tròn (Pie Chart):**

* Vì sao: Pie Chart được dùng để phân tích tỷ lệ phần trăm chi tiêu của các nhóm khách hàng theo từng danh mục sản phẩm.
* Lợi ích: Biểu đồ tròn mang lại cái nhìn rõ ràng về mức độ đóng góp của từng danh mục sản phẩm trong tổng doanh thu. Đây là công cụ phù hợp khi cần so sánh tỷ lệ giữa các nhóm trong một tổng thể.

Các kỹ thuật này được áp dụng vì chúng trực quan, dễ hiểu và cung cấp những thông tin cần thiết cho các bên liên quan trong việc phân tích và ra quyết định chiến lược, đặc biệt là tối ưu hóa các chiến dịch marketing và nâng cao tương tác khách hàng.

## Các nguyên tắc trực quan hóa

Trong phạm vi dự án phân tích dữ liệu của công ty bán lẻ DataSquad Retail, **nguyên tắc "tính đơn giản"** là quan trọng nhất. Dưới đây là lý do:

* **Phân tích dễ tiếp cận:** Người xem của báo cáo là các nhà quản lý marketing, đội ngũ bán hàng, và chăm sóc khách hàng, họ có thể không phải là chuyên gia về phân tích dữ liệu. Trình bày đơn giản giúp họ dễ dàng hiểu được các kết luận mà không cần giải thích phức tạp.
* **Tập trung vào thông tin cốt lõi:** Dự án này tập trung vào việc phân tích hành vi khách hàng và hiệu quả của các chiến dịch marketing. Việc trình bày dữ liệu một cách rõ ràng và tập trung vào các yếu tố chính như mức chi tiêu, tương tác khách hàng sẽ giúp ban quản lý nhanh chóng nắm bắt các điểm quan trọng để ra quyết định.

**Giảm thiểu sai sót và nhầm lẫn:** Khi dữ liệu được trình bày đơn giản, các biểu đồ dễ đọc và dễ hiểu, người xem sẽ ít có khả năng hiểu sai hoặc bỏ sót các thông tin quan trọng. Điều này đặc biệt quan trọng trong bối cảnh cạnh tranh của doanh nghiệp, nơi mà quyết định dựa trên dữ liệu có thể ảnh hưởng đến hiệu quả chiến dịch và doanh thu

## Trình bày tạo các report cho dự án

### Tạo visual thống kê chi tiết

**Viz 1: Phân bố khách hàng theo trình độ học vấn**

**A graph with blue squares

Description automatically generated**

Insight: Nhóm khách hàng có trình độ học vấn "Tốt nghiệp" chiếm tỷ lệ cao nhất (1.124 người). Điều này cho thấy đây là phân khúc thị trường có tiềm năng lớn, có thể được khai thác thông qua các chiến dịch tiếp thị tập trung vào sản phẩm/dịch vụ dành cho nhóm có nền tảng học vấn này.

**Viz 2: Phân bố tuổi khách hàng**

**A graph of blue bars with numbers

Description automatically generated**

Insight: Độ tuổi 49-56 có số lượng khách hàng lớn nhất (537 người), cho thấy đây là nhóm tuổi chủ chốt mà công ty cần ưu tiên. Các sản phẩm/dịch vụ nhắm đến nhu cầu sức khỏe, tài chính, hoặc nghỉ hưu có thể phù hợp với nhóm này.

**Viz 3: Tỷ lệ khách hàng theo tình trạng độc thân**

**A blue pie chart with a number of numbers

Description automatically generated**

Insight: Khách hàng "Sống chung" chiếm tỷ lệ lớn hơn (64.38%) so với "Độc thân" (35.62%), gợi ý rằng các chiến lược tiếp thị và sản phẩm nên tập trung vào các gia đình hoặc cặp đôi, thay vì chỉ nhắm vào cá nhân.

**Viz 4: Mối quan hệ giữa thu và tổng số con**

**A graph with numbers and a triangle

Description automatically generated**

Insight: Thu nhập cao hơn tập trung ở nhóm khách hàng không có con (thu nhập trung bình 52.879.132 VND). Điều này cho thấy tiềm năng cung cấp các sản phẩm/dịch vụ cao cấp, đặc biệt là dành cho nhóm khách hàng không có gánh nặng tài chính lớn từ con cái.

**Viz 5: Biểu đồ phân bố chi tiêu theo tình trạng hôn nhân và tổng số con**

**A green and white chart

Description automatically generated**

Insight: Nhóm "Sống chung" chi tiêu cao hơn ở tất cả các mức số con. Điều này cho thấy các gia đình hoặc cặp đôi là phân khúc khách hàng có nhu cầu chi tiêu mạnh mẽ hơn. Các sản phẩm/dịch vụ hướng đến nhu cầu gia đình nên được ưu tiên phát triển và quảng bá.

**Viz 6: Chi tiêu của từng nhóm thu nhập**

**A graph with green squares

Description automatically generated**

Insight: Nhóm thu nhập cao đóng góp phần lớn tổng chi tiêu (4.164.468 VND). Điều này gợi ý rằng việc phát triển các sản phẩm/dịch vụ cao cấp hoặc gói dịch vụ dành riêng cho nhóm khách hàng giàu có có thể đem lại doanh thu đáng kể.

**Viz 7: Biểu đồ chi tiêu theo thâm niên khách hàng**

**A graph with numbers and a number of people

Description automatically generated with medium confidence**

* Insight: Những khách hàng có thâm niên từ 140 đến 147 tháng duy trì mức chi tiêu ổn định, cho thấy họ là nhóm trung thành và đáng tin cậy. Công ty nên tập trung vào các chiến lược giữ chân khách hàng, chẳng hạn như chương trình khách hàng thân thiết, giảm giá độc quyền, hoặc ưu đãi dài hạn.

**Viz 8: Tổng chi tiêu danh mục sản phẩm theo nhóm tuổi**

**A graph of different colored bars

Description automatically generated with medium confidence**

Insight: Các danh mục sản phẩm như "Rượu", "Thịt", và "Trái cây" thu hút sự quan tâm đặc biệt từ nhóm khách hàng trung niên và cao tuổi. Đây là cơ hội để công ty cá nhân hóa chiến lược tiếp thị và cung cấp các sản phẩm phù hợp với sở thích và nhu cầu của từng nhóm tuổi.

**Viz 9: Biểu đồ chi tiêu theo độ tuổi khách hàng**

A graph of blue and black bars

AI-generated content may be incorrect.

Insight: Nhóm khách hàng trong độ tuổi từ 36-55 có xu hướng chi tiêu cao hơn, đặc biệt là nhóm trung niên và cao tuổi. Đây là phân khúc tiềm năng, cho thấy họ có khả năng chi tiêu mạnh mẽ và có thể là đối tượng mục tiêu cho các chiến dịch marketing hoặc sản phẩm cao cấp. Việc xác định và phát triển các sản phẩm và dịch vụ phù hợp cho nhóm này sẽ góp phần tăng trưởng doanh thu.

**Viz 10: Hành vi mua sắm theo kênh mua hàng**

**A graph of gray and black bars

AI-generated content may be incorrect.**

Insight: Kênh "Mua tại cửa hàng" và "Mua truy cập trực tuyến" đều có số lượt mua cao nhất, lần lượt là 12.964 và 11.900 lượt. Điều này chỉ ra rằng cả hai kênh đều quan trọng và cần được tối ưu để phục vụ khách hàng hiệu quả hơn. Các chiến lược marketing cần được điều chỉnh để tận dụng tối đa tiềm năng của cả hai kênh, đồng thời cải thiện trải nghiệm khách hàng trên mỗi kênh mua sắm.

**Viz 11: Tỷ lệ chi tiêu theo danh mục sản phẩm**

**A green and white chart

AI-generated content may be incorrect.**

Insight: Các sản phẩm như "Thịt", "Rượu", và "Vàng bạc" có tỷ lệ chi tiêu cao, đặc biệt là trong nhóm khách hàng trung niên và cao tuổi. Việc tập trung vào các sản phẩm này có thể gia tăng doanh thu, vì chúng đang thu hút sự chú ý lớn từ nhóm khách hàng có khả năng chi tiêu cao. Điều này gợi ý rằng các chiến lược quảng bá và phân phối các sản phẩm này nên được tăng cường để tận dụng tối đa nhu cầu của thị trường.

**Viz 12: Số lượng chấp nhận của các chiến dịch**

A graph with different colored bars

Description automatically generated

Insight:

Chiến dịch thứ 4 có số lượng chấp nhận cao nhất, cho thấy chiến dịch này thu hút được sự chú ý và phản hồi tích cực từ khách hàng. Ngược lại, Chiến dịch thứ 2 có số lượng chấp nhận thấp nhất, điều này phản ánh chiến dịch này có thể không đủ hấp dẫn hoặc hiệu quả trong việc tiếp cận đúng đối tượng mục tiêu. Đây là một chỉ báo cần cải tiến chiến lược cho chiến dịch này trong tương lai.

**Viz 13:So sánh giữa các chiến dịch theo nhóm thu nhập**

**A graph with blue and white text

Description automatically generated**

Insight: Phân tích cho thấy Chiến dịch 5 rất thành công với khách hàng cao cấp, mở ra cơ hội tăng trưởng doanh thu. Để mở rộng thị trường, chúng ta cần điều chỉnh chiến lược, đặc biệt là đối với nhóm thu nhập thấp, nhằm tăng tương tác và chuyển đổi.

**Viz 14: Mối quan hệ giữa Phản Hồi và Khiếu Nại**

**A blue squares with white text

Description automatically generated**

Insight: Phần lớn khách hàng không có khiếu nại, với 1880 trường hợp không có khiếu nại so với chỉ 18 trường hợp có khiếu nại. Điều này cho thấy rằng hầu hết khách hàng hài lòng với các chiến dịch được triển khai, hoặc ít có ý kiến tiêu cực. Đây là một dấu hiệu tích cực cho thấy chiến lược marketing đang có hiệu quả và khách hàng cảm thấy thoải mái với các dịch vụ hoặc sản phẩm hiện có.

**Viz 15: So sánh Tổng Chi Tiêu giữa Nhóm Thu Nhập và Khiếu Nại**

**A graph with green and purple squares

Description automatically generated**

Insight:

* Khách hàng thu nhập trung bình cao không chỉ chi tiêu nhiều hơn mà còn ít khiếu nại hơn so với các nhóm thu nhập thấp và cao. Điều này cho thấy rằng nhóm này có khả năng chi tiêu ổn định và có mức độ hài lòng cao hơn, điều này là một dấu hiệu quan trọng cho việc phát triển các chiến dịch marketing nhắm vào nhóm khách hàng này.
* Trong khi đó, nhóm thu nhập thấp có mức chi tiêu không đáng kể và một số ít khiếu nại, phản ánh rằng nhóm khách hàng này có thể không phải là đối tượng tiềm năng cho các chiến dịch cao cấp hoặc cần có sự điều chỉnh chiến lược để phù hợp hơn với nhu cầu của họ.

### Tạo visual dự báo

**Biểu đồ 1: Ma trận nhầm lẫn - Logistic Regression**

A blue and white squares with white text

Description automatically generated

* **Ý nghĩa của ma trận:**
  + **True negatives (373):** Mô hình dự đoán đúng rằng không có phản hồi.
  + **False positives (5):** Mô hình dự đoán sai rằng có phản hồi (nhưng thực tế không có phản hồi).
  + **False negatives (68):** Mô hình dự đoán sai rằng không có phản hồi (nhưng thực tế có phản hồi).
  + **True positives (1):** Mô hình dự đoán đúng rằng có phản hồi.
* **Insight:**
  + Mô hình dự đoán tốt lớp "Không phản hồi," với số lượng dự đoán chính xác rất cao (373).
  + Tuy nhiên, lớp "Phản hồi" bị đánh giá thấp đáng kể, với tỷ lệ **True Positive Rate (TPR)** thấp (1/69). Điều này có thể là do dữ liệu không cân bằng (lớp "Phản hồi" quá ít so với lớp "Không phản hồi").
  + Với chỉ một dự đoán đúng cho lớp "Phản hồi," Logistic Regression hiện không hiệu quả trong bài toán này.
* **Hướng cải thiện:**
  + Thực hiện cân bằng dữ liệu (oversampling lớp nhỏ, undersampling lớp lớn).
  + Sử dụng các mô hình phức tạp hơn như Random Forest, Gradient Boosting, hoặc SVM.
  + Điều chỉnh threshold dự đoán để cải thiện khả năng nhận diện lớp "Phản hồi."

**Biểu đồ 2: Residual Plot - Hồi quy tuyến tính**

A blue dotted diagram with a red line

Description automatically generated

* **Ý nghĩa:**
  + Phần dư (Residuals) là sai số giữa giá trị thực tế và giá trị dự đoán.
  + Trục y thể hiện phần dư, trục x thể hiện giá trị dự đoán. Đường đỏ là mức 0 (không có sai số).
* **Insight:**
  + Phần lớn các điểm dữ liệu phân bố gần đường Residual=0Residual = 0Residual=0, cho thấy mô hình tuyến tính có xu hướng không thiên lệch (bias).
  + Tuy nhiên:
    - Ở giá trị dự đoán lớn (bên phải biểu đồ), độ phân tán phần dư tăng cao. Điều này cho thấy mô hình tuyến tính không xử lý tốt các giá trị lớn, có thể do mối quan hệ giữa biến độc lập và biến phụ thuộc không hoàn toàn tuyến tính.
    - Một số điểm ngoại lệ xuất hiện (phần dư nằm xa trục trung tâm), ảnh hưởng đến hiệu suất của mô hình.
* **Hướng cải thiện:**
  + Xem xét lại dữ liệu và loại bỏ hoặc xử lý các outliers.
  + Sử dụng mô hình phi tuyến tính như Random Forest hoặc Gradient Boosting để cải thiện độ chính xác.

**Tổng kết chung từ hai biểu đồ:**

1. **Logistic Regression:**
   * Hiện tại mô hình không hiệu quả với lớp "Phản hồi." Cần điều chỉnh hoặc thử mô hình khác để cải thiện khả năng phân loại lớp này.
2. **Hồi quy tuyến tính:**
   * Hiệu quả trong các trường hợp cơ bản nhưng gặp khó khăn khi xử lý các giá trị lớn. Thử nghiệm các mô hình phi tuyến tính có thể là bước tiếp theo hợp lý.

# Xây dựng báo cáo

## Dashboard và report

**Cách tối ưu hóa:**

* **Dashboard:**
  + Thiết kế giao diện trực quan, dễ hiểu, nhấn mạnh các chỉ số chính (KPIs).
  + Sử dụng các biểu đồ phù hợp để truyền đạt thông tin rõ ràng và nhanh chóng.
  + Đảm bảo khả năng tương tác cao, cho phép người dùng chọn lọc dữ liệu theo nhu cầu.
* **Report:**
  + Tập trung chi tiết vào dữ liệu thô và phân tích chuyên sâu.
  + Định dạng rõ ràng, dễ đọc, với chú thích và giải thích đầy đủ.
  + Tối ưu hóa báo cáo để phù hợp với các mục tiêu cụ thể, như báo cáo tuần, tháng, hoặc theo lĩnh vực.

## Xây dựng báo cáo

### Dashboard vs Report

### Dashboard

**Trình bày sự khác nhau:**

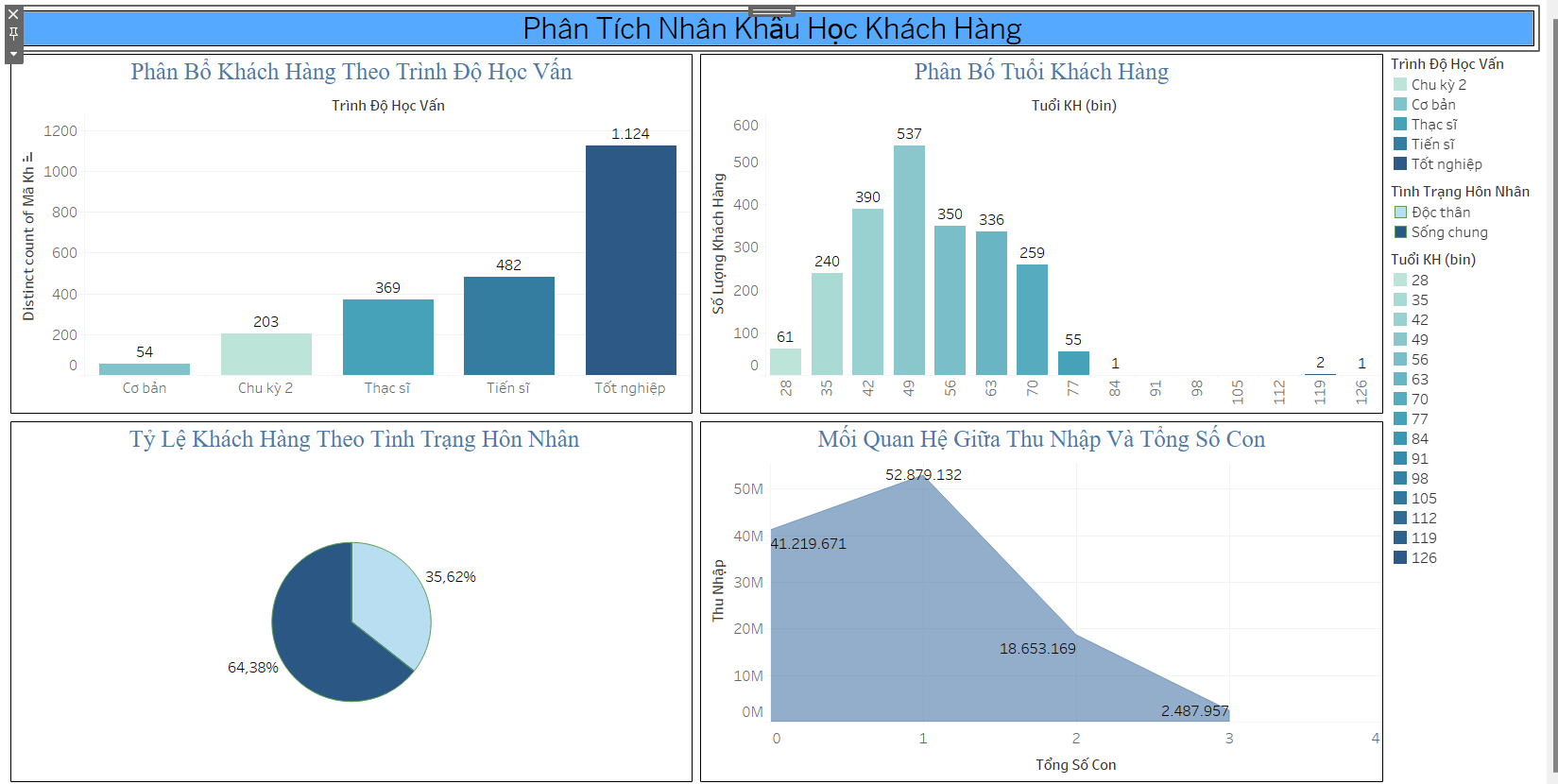
|  |  |  |
| --- | --- | --- |
| **Yếu tố** | **Dashboard** | **Report** |
| |  | | --- | | **Mục đích** |  |  | | --- | |  | | Tổng quan nhanh, ra quyết định tức thời | Phân tích chi tiết, hỗ trợ nghiên cứu sâu hơn |
| **Định dạng** | Trực quan, tương tác cao | Văn bản hoặc bảng biểu chi tiết, ít tương tác |
| **Người dùng mục tiêu** | Lãnh đạo, nhà quản lý | Nhà phân tích, chuyên gia dữ liệu |

**Lí do nhóm chọn làm Dashboard là:**

1. **Tăng cường tính tương tác và trải nghiệm người dùng:**
   * Dashboard cho phép người dùng tương tác trực tiếp với dữ liệu (ví dụ: lọc theo thời gian, khu vực, hoặc danh mục), giúp họ tìm kiếm thông tin cần thiết nhanh chóng mà không phải duyệt qua toàn bộ dữ liệu.
2. **Cung cấp thông tin tổng hợp dễ hiểu:**
   * Dashboard tập trung vào các biểu đồ, số liệu tổng quan, tránh làm người dùng bị "quá tải thông tin" so với việc xem các báo cáo chi tiết.
3. **Hỗ trợ đa nhiệm và phối hợp nhóm:**
   * Một Dashboard có thể được sử dụng đồng thời bởi nhiều nhóm như quản lý, marketing, và bán hàng, giúp phối hợp hiệu quả hơn khi mọi người đều dựa trên cùng một nguồn thông tin.
4. **Phát hiện xu hướng và cảnh báo bất thường:**
   * Dashboard có thể được thiết kế để hiển thị cảnh báo khi dữ liệu vượt ngưỡng (ví dụ: doanh thu giảm 20% so với tuần trước), giúp lãnh đạo nhanh chóng xác định và xử lý các vấn đề tiềm ẩn.
5. **Thúc đẩy việc ra quyết định dựa trên dữ liệu:**
   * Khi các quyết định cần dựa trên dữ liệu, việc có một giao diện trực quan sẽ khuyến khích lãnh đạo dựa vào thông tin từ Dashboard thay vì cảm tính.

Các loại Dashboard có trong dự án:

**Dashboard 1: Phân Tích Nhân Khẩu Học Khách Hàng**

****

**Chiến lược tiếp thị tổng hợp:**

* **Ưu tiên:** Phân khúc khách hàng có trình độ học vấn "Tốt nghiệp" và độ tuổi 49-56.
* **Nhắm vào:** Nhóm khách hàng không có con để tiếp thị các sản phẩm cao cấp.
* **Tập trung:** Tình trạng hôn nhân "Sống chung" với các giải pháp gia đình hoặc cặp đôi.

**Hướng đến hiệu quả lâu dài:**

* Thấu hiểu đặc điểm thị trường này giúp công ty tối ưu hóa chiến lược marketing, phân bổ nguồn lực hiệu quả và tăng cơ hội chuyển đổi khách hàng tiềm năng thành khách hàng trung thành.

**Dashboard 2: Phân tích chi tiêu khách hàng**

**A screenshot of a computer

Description automatically generated**

**Chiến lược tiếp thị dựa trên chi tiêu:**

* **Ưu tiên:** Tăng cường phát triển và quảng bá các sản phẩm/dịch vụ dành cho các gia đình sống chung và nhóm thu nhập cao.
* **Duy trì:** Tập trung vào nhóm khách hàng thâm niên để giữ sự ổn định trong doanh thu.
* **Đa dạng hóa:** Điều chỉnh danh mục sản phẩm theo nhu cầu đặc thù của từng nhóm tuổi để tăng khả năng tiếp cận thị trường.

**Tăng cường hiệu quả kinh doanh thông qua dữ liệu chi tiêu:**

* Bằng cách kết hợp các yếu tố như tình trạng hôn nhân, số con, thu nhập, thâm niên, và danh mục sản phẩm, công ty có thể xây dựng các chiến lược kinh doanh linh hoạt, phù hợp với đặc điểm của từng phân khúc khách hàng.

**Dashboard 3: Phân Khúc Khách Hàng**

**A screenshot of a computer screen

Description automatically generated**

**Tổng hợp chiến lược tiếp thị**

* Phân khúc chi tiêu cao: Nhóm tuổi 36-55 có xu hướng chi tiêu nhiều, là thị trường tiềm năng cần khai thác với sản phẩm/dịch vụ phù hợp.
* Tối ưu hóa kênh mua sắm: Kênh "Mua tại cửa hàng" và "Mua truy cập trực tuyến" cần được tối ưu để tăng trưởng doanh thu từ cả hai.
* Sản phẩm tiềm năng: Các sản phẩm "Thịt", "Rượu", "Vàng bạc" phổ biến ở nhóm trung niên và cao tuổi, cần tập trung vào nhóm này để gia tăng doanh thu.

**Định hướng chiến lược dài hạn**

* Công ty cần duy trì chiến lược marketing nhắm vào nhóm chi tiêu cao và tối ưu các kênh bán hàng để tăng trưởng doanh thu bền vững. Đồng thời, cần phát triển các sản phẩm phù hợp với nhóm khách hàng trung niên và cao tuổi, tận dụng các xu hướng tiêu dùng để xây dựng mối quan hệ lâu dài với khách hàng. Việc tối ưu hóa các sản phẩm và kênh bán hàng sẽ giúp công ty duy trì sự cạnh tranh trên thị trường trong dài hạn.

**Dashboard 4: Hiệu Quả Các Chiến Dịch Marketing**

A screenshot of a graph

Description automatically generated

**Tổng hợp và Định hướng chiến lược dài hạn:**

* Từ các phân tích trên, có thể thấy rằng **Chiến dịch 4** đang thành công nhất và cần được duy trì hoặc cải tiến để phát huy tối đa hiệu quả. Trong khi đó, **Chiến dịch 2** cần được điều chỉnh để nâng cao tỷ lệ chấp nhận. Mặt khác, nhóm khách hàng thu nhập trung bình cao là nhóm tiềm năng nhất với mức chi tiêu lớn và ít khiếu nại, cần tập trung tối đa vào họ.
* Về lâu dài, chiến lược tiếp thị cần chú trọng đến việc phát triển các chiến dịch nhắm vào các phân khúc thu nhập cao và trung bình, đồng thời điều chỉnh các chiến dịch để thu hút thêm nhóm thu nhập thấp. Ngoài ra, cần duy trì mối quan hệ tốt với khách hàng trung thành và giảm thiểu khiếu nại bằng cách cải tiến dịch vụ và chất lượng chiến dịch.

# KẾT LUẬN

## Báo cáo

### Các bước viết báo cáo

Việc viết báo cáo phân tích dữ liệu thường được thực hiện qua các bước sau:

1. **Xác định mục tiêu và đối tượng báo cáo**: Trước tiên, bạn cần xác định mục tiêu báo cáo và ai sẽ là người sử dụng báo cáo. Điều này giúp định hướng việc thu thập và phân tích dữ liệu sao cho phù hợp.
2. **Thu thập và xử lý dữ liệu**: Tập hợp các dữ liệu cần thiết, làm sạch dữ liệu, và chuẩn bị dữ liệu cho việc phân tích.
3. **Phân tích và trình bày dữ liệu**: Áp dụng các kỹ thuật phân tích như thống kê, phân nhóm, phân tích xu hướng để đưa ra những kết luận có giá trị. Trình bày kết quả thông qua các biểu đồ, bảng biểu để dễ dàng truyền đạt thông tin.
4. **Lập báo cáo**: Tạo bản báo cáo, trình bày rõ ràng các phân tích, kết quả và đưa ra những insight, khuyến nghị dựa trên dữ liệu. Báo cáo cần có sự kết hợp của các phần mở đầu, phương pháp, kết quả, và kết luận.
5. **Kiểm tra và chỉnh sửa**: Đọc lại báo cáo, kiểm tra tính logic, mạch lạc của báo cáo để đảm bảo rằng mọi thông tin đều chính xác và dễ hiểu.

## Khó khăn

* **Hạn chế kiến thức về Python Script**: Sử dụng Python để xử lý và làm sạch dữ liệu là một kỹ năng cần thiết, nhưng đôi khi còn gặp khó khăn khi sử dụng các thư viện phức tạp hoặc xử lý dữ liệu lớn.
* **Dữ liệu chưa đáp ứng đủ để khai thác**: Một số dữ liệu thiếu sót hoặc không đầy đủ, đặc biệt là dữ liệu thời gian, điều này hạn chế khả năng phân tích các xu hướng trong dài hạn.
* **Dữ liệu không nằm trong sự hiểu biết**: Có những loại dữ liệu phức tạp hoặc không quen thuộc, cần thêm thời gian để nghiên cứu và hiểu rõ hơn.

## Thuận lợi

* **Công cụ hỗ trợ**: Các công cụ như ChatGPT, Google, và các diễn đàn giúp giải quyết vấn đề nhanh chóng và cung cấp tài liệu tham khảo hữu ích.
* **Kinh nghiệm từ dự án mẫu**: Việc làm dự án mẫu giúp người tham gia dự án hiểu rõ hơn về quy trình phân tích dữ liệu và cách trình bày báo cáo hiệu quả.
* **Hỗ trợ từ giảng viên**: Giảng viên hướng dẫn đã cung cấp nhiều lời khuyên và hỗ trợ để giải quyết các vấn đề trong quá trình phân tích và báo cáo.

## Hướng phát triển

Hướng phát triển trong tương lai có thể bao gồm:

* **Nâng cao kiến thức về Python và các công cụ phân tích dữ liệu khác**: Học thêm về các thư viện Python như Pandas, Numpy, Matplotlib và Seaborn để làm việc hiệu quả hơn với dữ liệu lớn và phức tạp.
* **Tập trung vào việc cải thiện kỹ năng báo cáo**: Học cách trình bày thông tin một cách mạch lạc và dễ hiểu, giúp người đọc nắm bắt nhanh chóng những insight quan trọng.
* **Thu thập và phân tích dữ liệu theo thời gian**: Phát triển khả năng khai thác dữ liệu thời gian để nhận diện các xu hướng và mô hình trong hành vi của khách hàng, giúp đưa ra các chiến lược dài hạn.