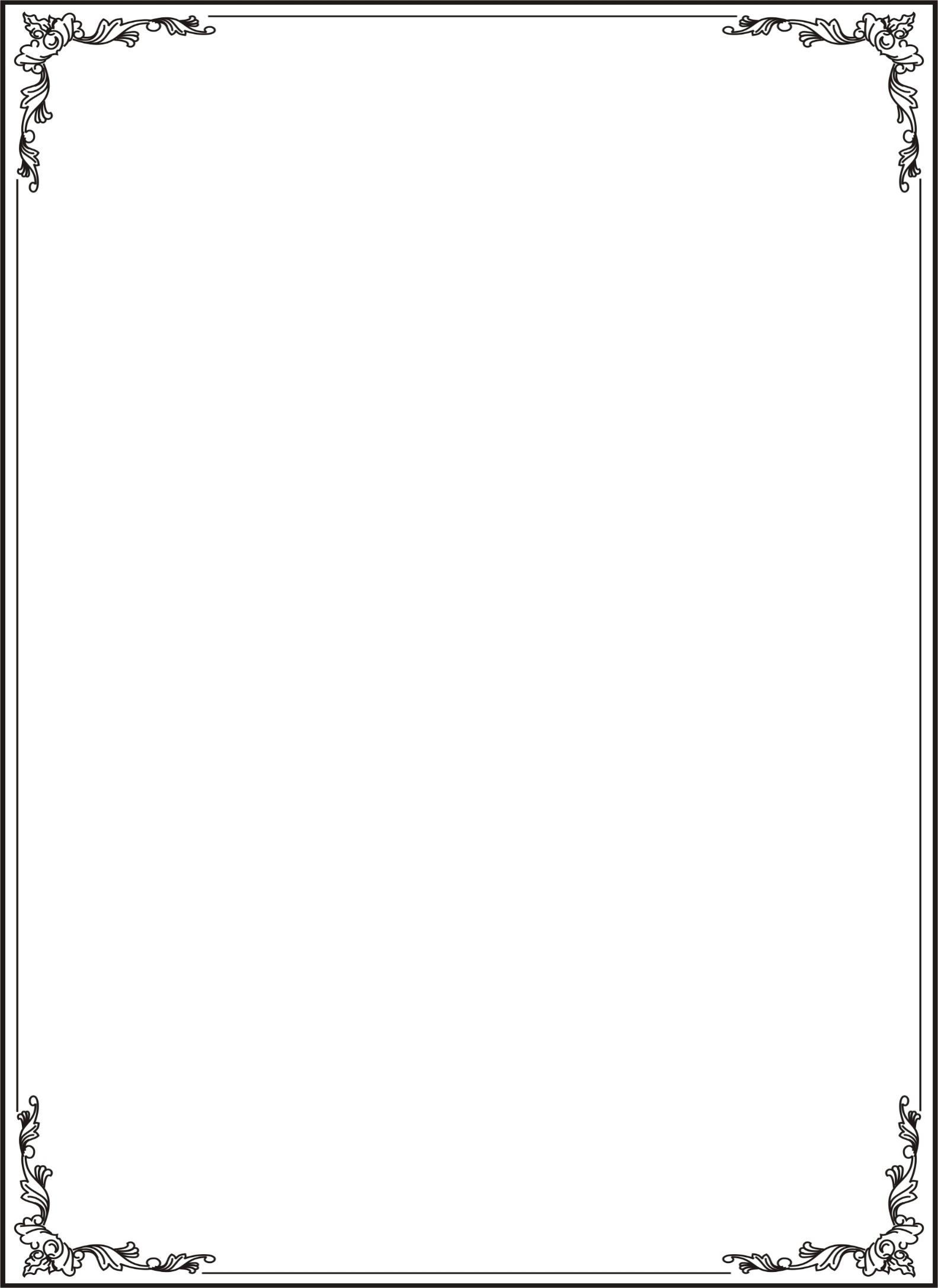
A logo with a ball and a circle

Description automatically generated with medium confidence

**GVHD:** ThS.Nguyễn Văn Thành

**Lớp HP:** DAWH430784\_23\_2\_02

**Nhóm thực hiện:** Nhóm 8

**Học kỳ:** 2

**Năm học:** 2023 - 2024

**MÔN: KHO DỮ LIỆU**

**ĐỀ TÀI: TÌM HIỂU VÀ XÂY DỰNG KHO DỮ LIỆU CAR - STORE**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**DANH SÁCH SINH VIÊN THỰC HIỆN**

**HỌC KÌ 2 NĂM HỌC 2023 – 2034**

**NHÓM 8**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Đào Quang Duy | Mai Trần Bách | Ngô Minh Lợi | Đỗ Phú Hiếu |
| **Thiết kế DW** |  |  |  |  |
| * DimDate * DimCar |  |  |  | **x** |
| * DimSalePerson * DimEmployee * DimBrand |  |  |  | **x** |
| * FactSale, BrandFact | **x** |  |  |  |
| **ETL** |  |  |  |  |
| * Thực hiện đưa dữ liệu vào StageDate và DimDate |  |  |  | **x** |
| * Thực hiện đưa dữ liệu vào StageSalesPerson và DimSalePerson |  |  | **x** |  |
| * Thực hiện đưa dữ liệu vào StageBrand và StageSale | **x** |  |  |  |
| * Thực hiện đưa dữ liệu vào FactBrand và FactSale | **x** |  |  |  |
| * Nhập dữ liệu vào SSAS, tạo datasource, view, cube |  | **x** |  |  |
| **Phân cấp dữ liệu** |  |  |  |  |
| * DimDate * DimCar |  | **x** |  |  |
| * DimSalePerson * DimEmployee |  | **x** |  |  |
| **Thực hiện trả lời câu hỏi cho Sale Analysis** |  |  |  |  |
| * Trả lời câu hỏi cho Power PI |  |  | **x** |  |
| * Trả lời câu hỏi cho SSAS |  | **x** |  |  |
| * Cho biết xếp hạng KPI * Xếp hạng khách hàng |  |  | **x** |  |
| * Xếp hạng hãng xe, loại xe * Doanh thu theo tháng, quý, năm |  |  | **x** |  |

***Nhận xét của giảng viên:***……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………….………………………………………………………

*Tháng 05 năm 2024*  
**Giảng viên chấm điểm**

**LỜI CẢM ƠN**

Lời đầu tiên, chúng em xin gửi lời cảm ơn chân thành đến Thầy đã hướng dẫn, giúp đỡ và chỉ dẫn chúng em trong quá trình thực hiện đề tài này. Thầy đã tận tình chia sẻ kiến thức và kinh nghiệm của mình, giúp chúng em hiểu sâu hơn về đề tài. Thầy cũng đã giúp đỡ chúng em giải quyết những khó khăn và trở ngại trong quá trình nghiên cứu, giúp chúng em hoàn thành đề tài này. Không có sự hỗ trợ đó, chúng em không thể hoàn thành đề tài một cách thành công.

Chúng em cũng xin gửi lời cảm ơn đến các thành viên trong nhóm. Mỗi thành viên đã đóng góp một phần công sức, tài năng và thời gian để hoàn thành đề tài này. Chúng em đã học hỏi và trau dồi thêm nhiều kỹ năng, kiến thức mới từ nhau. Đặc biệt, chúng em cảm ơn vì mối quan hệ tình cảm và sự đoàn kết của nhóm trong suốt quá trình làm việc.

Một lần nữa, chúng em xin chân thành cảm ơn tất cả những người đã giúp đỡ chúng em trong quá trình nghiên cứu và thực hiện đề tài này. Sự hỗ trợ của các bạn đã làm cho đề tài của chúng em trở nên hoàn hảo hơn

# Giới thiệu tổng quát về dữ liệu

* **Nguồn dữ liệu:**
* <https://www.kaggle.com/datasets/bartoszpieniak/poland-cars-for-sale-dataset/code>
* <https://www.kaggle.com/datasets/suraj520/car-sales-data>
* **Mô tả tập dữ liệu:**
* Tập dữ liệu 1: Yêu cầu mua ô tô từ một trang quảng cáo ở Ba Lan.

Tập dữ liệu này được tạo bằng cách thu thập dữ liệu trên web từ hơn 200.000 yêu cầu mua ô tô từ một trong những trang quảng cáo ô tô lớn ở Ba Lan.

* Tập dữ liệu 2:

Tập dữ liệu này có thể được sử dụng để phân tích các xu hướng bán ô tô theo thời gian, đánh giá hiệu suất của từng nhân viên bán hàng trong đại lý, và phân tích tác động của các yếu tố như hãng sản xuất ô tô, mẫu mã, năm sản xuất và hiệu suất của nhân viên bán hàng đến doanh số bán ô tô. Nó cũng hữu ích cho việc dự đoán hiệu quả bán hàng trong tương lai và đưa ra các chiến lược kinh doanh phù hợp.

# Chi tiết thuộc tính

* **Tập dữ liệu 1:**

| **Thuộc tính** | **Mô tả** | **Ví dụ** |
| --- | --- | --- |
| ID | Mã định danh của yêu cầu | 12345 |
| Price | Giá của sản phẩm | 25000 |
| Currency | Loại tiền tệ sử dụng | PLN |
| Condition | Trạng thái của sản phẩm, mới hoặc đã qua sử dụng | Used |
| Vehicle\_brand | Thương hiệu của xe | Abarth |
| Vehicle\_model | Mẫu xe | 595 |
| Vehicle\_Generation | Thế hệ xe | 1.4 MultiAir Scorpione |
| Vehicle\_version | Phiên bản xe | SE |
| Production\_year | Năm sản xuất của xe | 2020 |
| Mileage\_km | Quãng đường xe đã đi được, tính bằng km | 50000 |
| Power\_HP | Công suất động cơ, tính bằng mã lực | 185 |
| Displacement\_cm3 | Dung tích động cơ, tính bằng cm³ | 2500 |
| Fuel\_type | Loại nhiên liệu sử dụng | Hunt |
| CO2\_emissions | Lượng khí thải CO2, tính bằng g/km | 150 |
| Drive | Loại dẫn động | Dẫn động cầu trước |
| Transmission | Loại hộp số | Auto |
| Type | Kiểu dáng thân xe | Sedan |
| Doors\_number | Số cửa xe | 4 |
| Colour | Màu sơn thân xe | Silver |
| Origin\_country | Quốc gia xuất xứ của xe | Japan |
| First\_owner | Chủ sở hữu có phải là chủ sở hữu đầu tiên không | No |
| First\_registration\_date | Ngày đăng ký đầu tiên | 2020-06-15 |
| Offer\_publication\_date | Ngày xuất bản ưu đãi | 2024-05-01 |
| Offer\_location | Địa chỉ do nhà phát hành cung cấp | Japan |
| Features | Các tính năng xe được liệt kê | ABS, Túi khí, Cảm biến đỗ xe |

* **Tập dữ liệu 2:**

|  |  |  |
| --- | --- | --- |
| **Thuộc tính** | **Mô tả** | **Ví dụ** |
| Date: | Ngày xe được bán | 2022-08-01, 2023-03-15, … |
| Salesperson | Tên của người bán xe | Monica Moore MD, Roberto Rose, … |
| Customer Name | Tên của người mua xe | Mary Butler, Richard Pierce, … |
| Car Make | Hãng sản xuất của xe | Honda, Ford, ... |
| Car Model | Mẫu của chiếc xe | F-150, Civic, … |
| Car Year: | Năm của xe ra mắt | 2010 to 2022 |
| Sale Price | Giá của xe (Theo USD) | 10 000 to 50 000 USD |
| Commission Rate | Phần trăm hoa hồng phải trả cho saleperson | 0.05 to 0.15 |
| Commission Earned: | Số tiền hoa hồng mà saleperson nhận được từ việc bán xe. (Sale Price \* Commission Rate) | 500 to 8 000 USD |

# 3. Giới thiệu công cụ sử dụng

## 3.1. SSMS (SQL Server Management Studio)

SQL Server Management Studio (SSMS) là một công cụ quản lý cơ sở dữ liệu, một công cụ mạnh mẽ và toàn diện được thiết kế để quản lý, phát triển, và điều hành SQL Server được Microsoft phát triển. SSMS cung cấp một môi trường trực quan và tiện lợi cho người dùng để làm việc với cơ sở dữ liệu SQL Server và Azure SQL Database.

Các chức năng chính của SSMS bao gồm các thao tác tạo, sửa đổi, và xóa các cơ sở dữ liệu. Quản lý bảo mật và quyền truy cập cho người dùng và vai trò. Theo dõi và quản lý các phiên và tiến trình đang chạy. SSMS cung cấp giao diện đồ họa và một loạt các công cụ để thực hiện các tác vụ quản lý cơ sở dữ liệu. Điều này cho phép người dùng truy vấn cơ sở dữ liệu, tạo và chỉnh sửa các đối tượng cơ sở dữ liệu như bảng, chế độ xem, thủ tục lưu trữ, hàm, và các đối tượng khác. Nó cũng hỗ trợ quản lý phiên bản và quản lý bảo mật cho cơ sở dữ liệu.

SQL Server Management Studio là một công cụ không thể thiếu đối với các nhà phát triển và quản trị viên cơ sở dữ liệu, giúp họ thực hiện các nhiệm vụ quản lý và phát triển cơ sở dữ liệu một cách hiệu quả và chính xác.

## 3.2. Visual Studio

Visual Studio là một môi trường phát triển tích hợp (IDE) toàn diện được Microsoft phát triển, được sử dụng để xây dựng các ứng dụng trên nhiều nền tảng khác nhau, bao gồm Windows, web, điện thoại di động, và đám mây. Visual Studio cung cấp một bộ công cụ mạnh mẽ và các tính năng hỗ trợ cho các nhà phát triển trong việc viết, gỡ lỗi, kiểm thử, và triển khai phần mềm.

Visual Studio là một môi trường phát triển tích hợp (Integrated Development Environment - IDE) được phát triển bởi Microsoft, được sử dụng để xây dựng các ứng dụng trên nhiều nền tảng khác nhau, bao gồm Windows, web, điện thoại di động, và đám mây. Visual Studio cung cấp một bộ công cụ mạnh mẽ và các tính năng hỗ trợ cho các nhà phát triển trong việc viết, gỡ lỗi, kiểm thử, và triển khai phần mềm trên nhiều nền tảng khác nhau, bao gồm Windows, macOS và Linux.

Visual Studio hỗ trợ nhiều ngôn ngữ lập trình như C++, C#, Visual Basic, Python, JavaScript, và nhiều ngôn ngữ khác. IDE này cung cấp cho nhà phát triển một loạt các tính năng và công cụ để tăng năng suất và hỗ trợ quá trình phát triển phần mềm.

Các tính năng chính của Visual Studio bao gồm hỗ trợ nhiều ngôn ngữ lập trình như C#, VB.NET, F#, C++, Python, JavaScript, TypeScript, và nhiều ngôn ngữ khác. Giao diện người dùng thân thiện và tùy biến cao, cho phép các nhà phát triển thiết lập môi trường làm việc theo ý muốn.

## 3.3. Ngôn ngữ truy vấn SQL

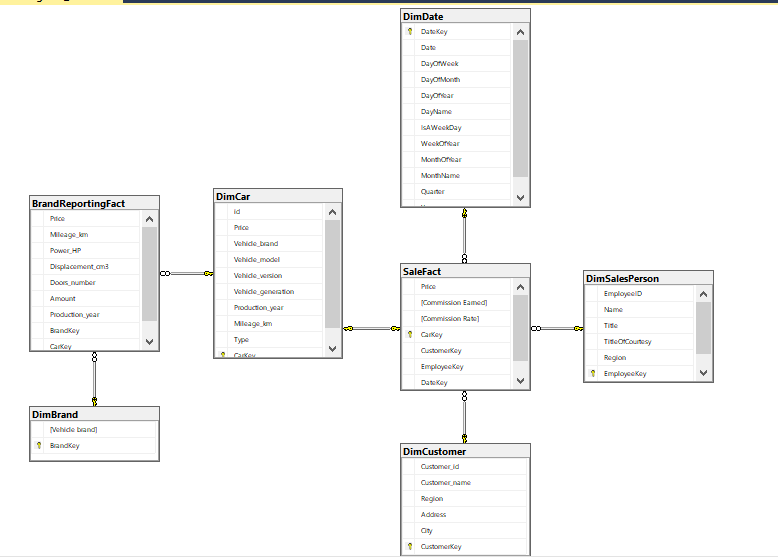
Ngôn ngữ truy vấn SQL (Structured Query Language)là một ngôn ngữ truy vấn chuẩn hóa được sử dụng để quản lý và thao tác các cơ sở dữ liệu quan hệ. SQL là một ngôn ngữ chuẩn được chấp nhận rộng rãi trong ngành công nghiệp và là một phần quan trọng của hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System - RDBMS). SQL được phát triển vào những năm 1970 và từ đó trở thành một ngôn ngữ chuẩn để làm việc với hầu hết các hệ quản lý cơ sở dữ liệu quan hệ (RDBMS) phổ biến như MySQL, Oracle, Microsoft SQL Server, PostgreSQL và SQLite.

SQL cho phép bạn truy vấn dữ liệu từ cơ sở dữ liệu. Bằng cách sử dụng các câu lệnh SELECT, bạn có thể lấy ra các bản ghi từ bảng dữ liệu dựa trên các điều kiện nhất định. SQL cũng cung cấp các toán tử và hàm để thực hiện các phép tính và xử lý dữ liệu trước khi trả về kết quả. Tạo, sửa đổi và xóa cấu trúc cơ sở dữ liệu bằng cách sử dụng các câu lệnh CREATE, ALTER và DROP, bạn có thể tạo bảng, chỉnh sửa cấu trúc bảng và xóa bảng khỏi cơ sở dữ liệu. SQL cũng cung cấp các câu lệnh để tạo, sửa đổi và xóa các ràng buộc, chỉ mục và quyền truy cập. Cung cấp các câu lệnh INSERT, UPDATE và DELETE để thêm, cập nhật và xóa dữ liệu trong cơ sở dữ liệu. Bằng cách sử dụng các câu lệnh này, bạn có thể thay đổi nội dung của bảng dữ liệu.

# Câu hỏi Business Process

* Các hãng xe bán chạy nhất và có doanh thu cao nhất
* Các loại xe bán chạy nhất doanh thu cao nhất
* Nhân viên bán hàng nhiều nhất và đem lại nhiều doanh thu
* Số lượng xe bán ra và doanh thu của công ty trong từng tháng của năm hiện tại?
* Doanh số của từng nhân viên bán hàng theo từng loại hãng xe trong năm gần đây là bao nhiêu?

# Xây dựng kho dữ liệu



# 6. Quá trình SSIS

## 6.1. Staging

A screenshot of a computer

Description automatically generated

* StageCustomer:

A screenshot of a computer

Description automatically generated

* StageSalesPerson:

A screen shot of a computer screen

Description automatically generated

* StageCar:

A screen shot of a computer

Description automatically generated

* StageDate:

A screenshot of a computer

Description automatically generated

* StageBrand:

A screen shot of a computer

Description automatically generated

* StageSales:

A screenshot of a computer

Description automatically generated

* CarBrandReportStage:

A screen shot of a computer

Description automatically generated

## 6.2. Đưa dữ liệu vào Data Warehouse

A screenshot of a computer

Description automatically generated

* DimCar:

A screen shot of a computer script component

Description automatically generated

* DimCustomer:

A screenshot of a computer screen

Description automatically generated

* DimSalePerson:

A screen shot of a computer

Description automatically generated

* DimDate:

A computer screen shot of a diagram

Description automatically generated

* DimBrand:

A blue line with a white dot

Description automatically generated

* SaleFact:

A computer screen shot of a computer

Description automatically generated

* CarBrandReport:

A computer screen with a black background

Description automatically generated

# 7. Quá trình SSAS

## 7.1. Tạo kết nối Data Sources

- Kết nối với SQL Server:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A close up of a screen

Description automatically generated

## 7.2. Tạo và kết nối Data Sources Views

- Chọn tất cả các bảng Dim và Fact trong database:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer flow chart

Description automatically generated

A screenshot of a computer

Description automatically generated

## 7.3. Tạo Cubes

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

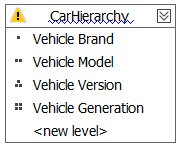
A screenshot of a computer

Description automatically generated

## 7.4. Các Dimensions



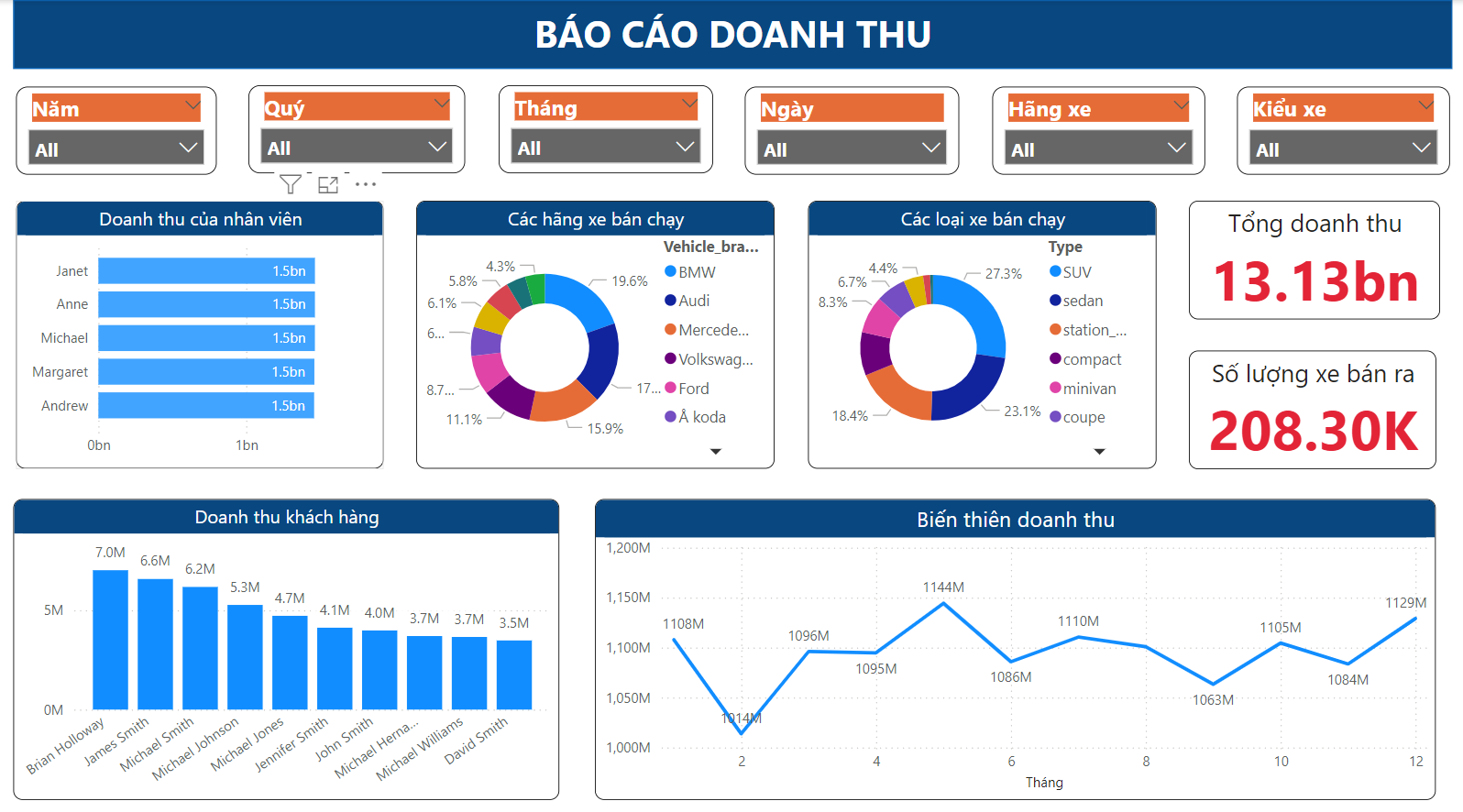






# 8. Phân tích dữ liệu trên Power Bi

## 8.1. Dashboard báo cáo doanh thu



## 8.2. Tổng doanh thu và số lượng xe bán ra của chuỗi cửa hàng trong giai đoạn quý 2 năm 2022 đến quý 2 năm 2023



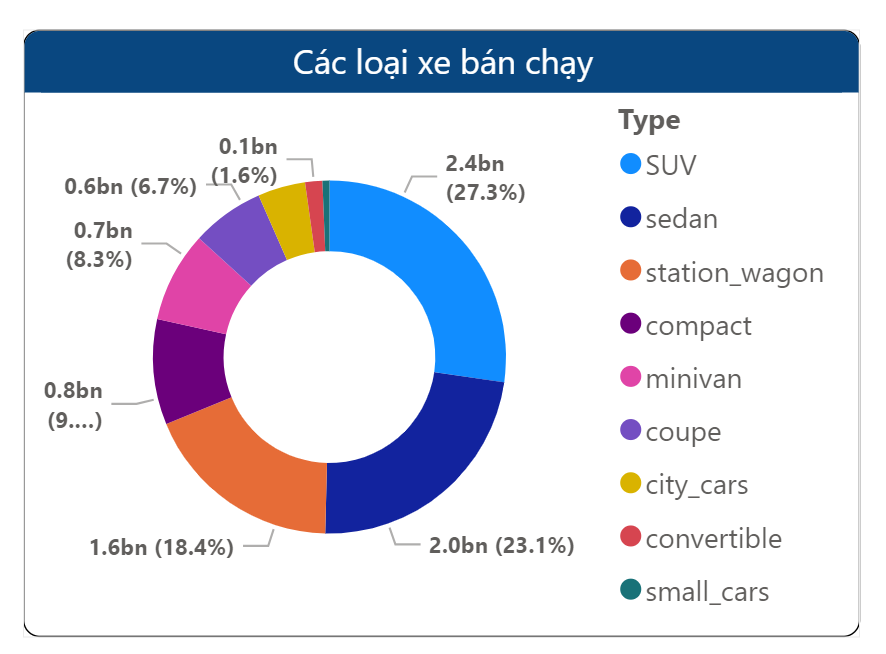
***Nhận xét:*** *Trong vòng một năm (từ quý 2 năm 2022 đến quý 2 năm 2023) Tổng doanh thu là 1,13 tỷ USD và số lượng xe bán ra là 17,87 nghìn chiếc* có thể thấy rằng công ty đã đạt được một mức doanh thu đáng kể. Đây là kết quả của những đóng góp cá nhân cho công ty trong việc phát triển sản phẩm và dịch vụ, tiếp cận khách hàng và tăng cường doanh số bán hàng. Ngoài ra, công ty cũng cần tăng cường quản lý chi phí để tối ưu hóa lợi nhuận và tiếp tục đổi mới và cải tiến các sản phẩm và dịch vụ của mình.

## 8.3. Các hãng xe bán chạy và có doanh thu cao nhất



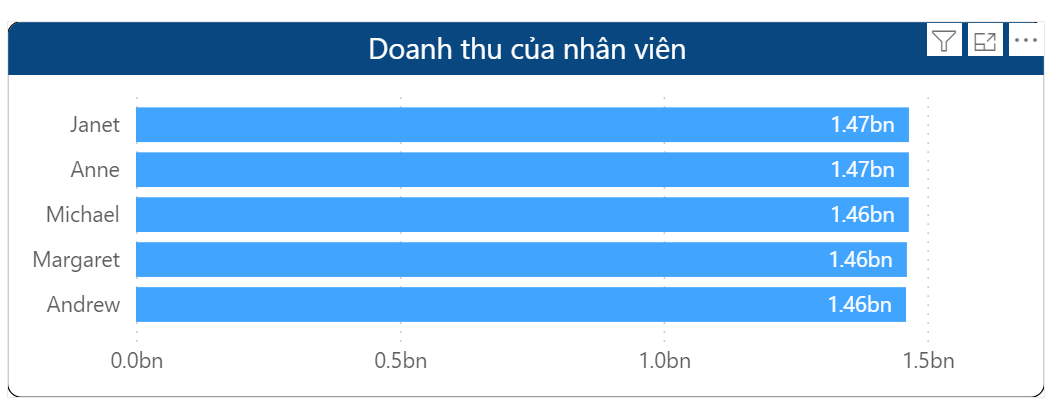
***Nhận xét:*** *Có thể thấy rằng những hãng xe nổi tiếng và được biết đến nhiều nhất có thể kể đến là BMW, Audio, Mercedes-Benz, Volkswagen, v.v.* Mỗi mẫu xe đều có tính năng và ưu điểm riêng, đáp ứng nhu cầu của các đối tượng khách hàng khác nhau.

## 8.4. Các loại xe bán chạy nhất và doanh thu của nó



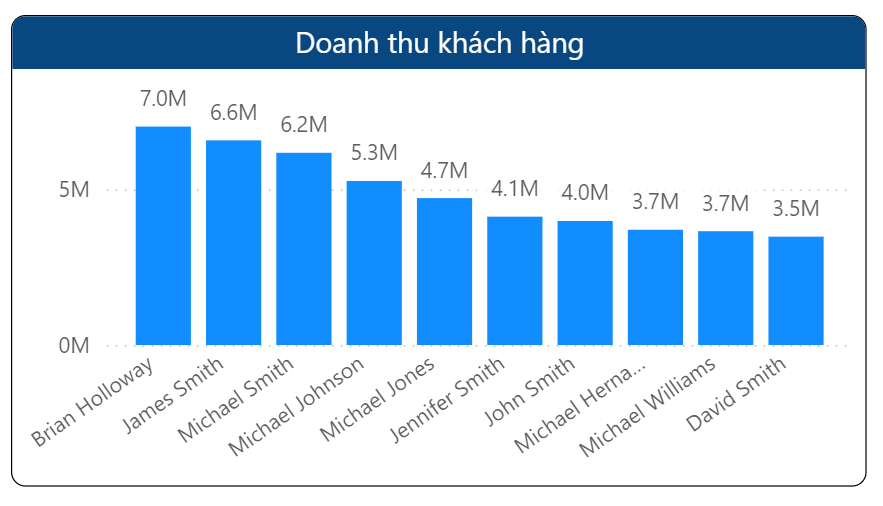
***Nhận xét:*** *Có thể thấy rằng những loại xe được bán ra là SUV, sedan, station\_wagon, compact, minivan, v.v.* Mỗi loại xe đều có tính năng và ưu điểm riêng, đáp ứng nhu cầu của các đối tượng khách hàng khác nhau, nhưng phổ biến nhất đáp ứng cho các nhu cầu đi lại.

## 8.5 Nhân viên bán hàng nhiều nhất và mang lại thành tích doanh số cho công ty



***Nhận xét:*** *Có thể thấy rằng các nhân viên có thành tích khá gần nhau và đều mang lại cho công ty những doanh thu kỷ lục, chứng tỏ nhân viên của công ty ai cũng có năng lực. Và Top 5 người có doanh thu cao nhất là: Janet, Anne, Michael, Margaret, Andrew, có thể kết luận rằng đây là những nhân viên có kỹ năng giao tiếp tốt, khả năng giải thích sản phẩm một cách rõ ràng và hiểu được nhu cầu của khách hàng.*

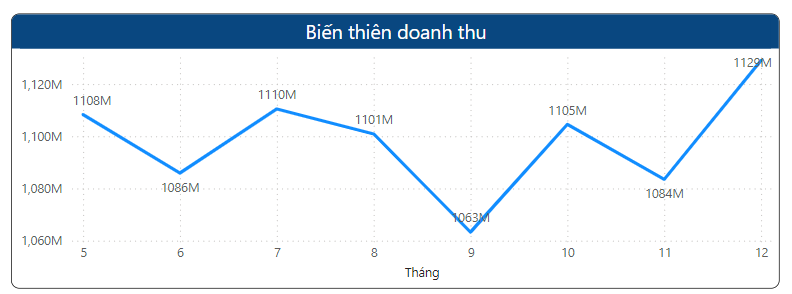
## 8.6. Thông tin những khách hàng VIP

**

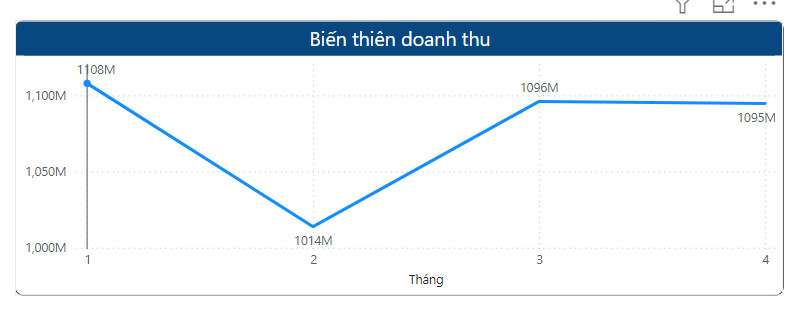
***Nhận xét:*** *Đây là những khách hàng thân thiết, mua số lượng xe lớn và là những khách hàng quan trọng của công ty, do đó cần đặc biệt quan tâm đến những khách hàng này.*

## 8.7. Tình hình bán hàng qua các tháng của các năm

*-* ***Năm 2022***

**

*-* ***Năm 2023***

******

***Nhận xét:*** *Nhìn chung doanh thu của công ty qua các tháng đều vượt 1000 triệu, song lại có những tháng lao dốc xuống như tháng 9 năm 2022 (từ 1101 triệu giảm xuống 1063 triệu) giảm tận 50 triệu và tháng 2 năm 2023 (từ 1108 triệu giảm xuống 1014 triệu) giảm gần 100 triệu, công ty cần đặc biệt chú tâm vào tại những thời điểm này, cần phân tích thêm và điều tra về các yếu tố nội bộ lẫn lên ngoài như điều kiện kinh tế thế giới, dịch bệnh, công ty gặp vấn đề về thương hiệu, ban lãnh đạo, v,v từ đó điều chỉnh chiến lược cho phù hợp hơn với từng hoàn cảnh.*

**Tài liệu tham khảo**

1. [[4 hours]] SSIS complete tutorial - {end to end} full course - SQL server integration services (2022) YouTube. Available at: <https://www.youtube.com/watch?v=G_wG-bzTCZY&amp;t=13844s>

(Accessed: 04 May 2023).

1. Thabresh (2023) Car sales data - EDA, Kaggle. Available at: <https://www.kaggle.com/code/thabresh/car-sales-data-eda>

(Accessed: 01 May 2023).

1. ETL project from Excel data source to star schema with SSIS (2021) YouTube. Available at: <https://www.youtube.com/watch?v=Yp8fXLnVCp8&amp;t=596s>

(Accessed: 08 May 2023).

1. SSAS tutorial (4/15) - what is Star schema | why it is important in modelling | dimension modelling (2020) YouTube. Available at: <https://www.youtube.com/watch?v=rIir3Wsz2-g>

(Accessed: 08 May 2023).

1. Power Bi Nâng Cao | Tạo dashboard quản Lý Bán Hàng (kèm file Thực Hành) (2021) YouTube. Available at: <https://www.youtube.com/watch?v=k9dmVlPqGDM&amp;t=1726s>

(Accessed: 12 May 2023).