

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM



HCMUTE

BÁO CÁO CUỐI KÌ

ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU CAR-STORE

Môn: Kho Dữ Liệu

GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện : Nhóm 6

Thành Phố Hồ Chí Minh, tháng 5 năm 2023

DANH SÁCH NHÓM LÀM ĐỒ ÁN CUỐI KỲ

MÔN HỌC: KHO DỮ LIỆU

HỌC KỲ II – NĂM HỌC: 2022-2023

- 1. Mã môn học: DAWH430784_22_2_02**
- 2. Giảng viên hướng dẫn: ThS. Nguyễn Văn Thành**
- 3. Tên đề tài: Xây dựng kho dữ liệu Car - Store**
- 4. Danh sách nhóm 6 làm đồ án cuối kỳ**

STT	Họ và tên sinh viên	MSSV	% Hoàn Thành	Kí Tên
1	Phan Quốc Lưu	20133065	100%	
2	Cao Trọng Nghĩa	20133071	100%	
3	Mai Thành Trung	20133105	100%	

- Tỷ lệ hoàn thành: 100%**
- Trưởng nhóm: Phan Quốc Lưu**

Nhận xét của giảng viên:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Tháng 05 năm 2023
Giảng viên chấm điểm

Mục Lục

LỜI MỞ ĐẦU	5
ĐỀ TÀI: CAR STORE	6
1. Giới thiệu tổng quát về dữ liệu.....	6
2. Chi tiết thuộc tính.....	6
3. Giới thiệu công cụ sử dụng.....	7
3.1 Giới thiệu về SSMS (SQL Server Management Studio)	7
3.2 Giới thiệu Visual Studio	7
3.3 Ngôn ngữ truy vấn sql	7
4. Câu hỏi Business Process.....	8
5. Xây dựng kho dữ liệu.....	8
5.1 Làm sạch dữ liệu	8
5.2 Car_Sales_Store_DW	9
5.3 Thông tin các bảng Car_Sales_Store_Stage.....	10
6. Quá trình SSIS.....	10
6.1 Kết nối SQL	10
6.2 Dataflow – StageSales.....	10
6.3 Dataflow – Fact_Sale	11
7. Quá trình SSAS	12
7.1 Tạo Kết nối Data Sources.....	12
7.2 Tạo và Kết nối Data Source Views	13
7.3 Tạo Cubes	13
7.4 Dimensions.....	14
8. Phân tích dữ liệu trên Power BI	14
8.1 Dashboard quản lý doanh số	14
8.2 Tổng doanh thu của năm qua các lần bán xe.....	15
8.3 Các mẫu xe bán chạy nhất và có doanh thu cao nhất	16
8.4 Các hãng xe bán chạy nhất doanh thu cao nhất	16
8.5 Nhân viên bán hàng nhiều nhất mang về hoa hồng nhiều nhất	17
8.6 Sự khác biệt giữa việc lựa chọn năm sản xuất xe và giá bán không	17
8.7 Thông tin tên khách hàng và số tiền dành cho việc mua xe	18
TÀI LIỆU THAM KHẢO	19

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

LỜI MỞ ĐẦU

Trong đề tài này, chúng ta sẽ sử dụng dữ liệu thị trường được cung cấp bởi Kaggle để thực hiện phân tích và đánh giá. Dữ liệu này bao gồm thông tin về khách hàng, sản phẩm và chiến dịch tiếp thị của một doanh nghiệp. Chúng ta sẽ áp dụng các kỹ thuật phân tích dữ liệu như trực quan hóa dữ liệu, phân tích hồi quy và phân tích nhân tố để tìm hiểu và giải thích mối quan hệ giữa các biến trong dữ liệu.

Mục đích của đề tài này là tìm hiểu cách thức hoạt động của thị trường và các yếu tố ảnh hưởng đến hoạt động kinh doanh. Kết quả của đề tài này sẽ giúp các doanh nghiệp có thể đưa ra các quyết định thông minh và hiệu quả trong hoạt động của mình. Chúng ta cũng hy vọng rằng đề tài này sẽ giúp các bạn hiểu thêm về phân tích dữ liệu và ứng dụng của nó trong thực tiễn.

ĐỀ TÀI: CAR STORE

1. Giới thiệu tổng quát về dữ liệu

Tập dữ liệu "Car Sales Data" là một tập dữ liệu bảng chứa thông tin về việc bán xe từ một đại lý bán xe trong vòng một năm. Tập dữ liệu này bao gồm 2.500.000 hàng và 9 cột dữ liệu cho mỗi lần bán xe.

Tập dữ liệu này hữu ích để phân tích xu hướng bán xe qua thời gian và đánh giá hiệu suất của từng nhân viên bán hàng trong đại lý. Nó cũng có thể được sử dụng để dự đoán hiệu suất bán hàng và phân tích tác động của các yếu tố khác nhau đến doanh số bán hàng, chẳng hạn như hãng xe, mẫu xe, năm sản xuất và hiệu suất của nhân viên bán hàng.

Nguồn dữ liệu: <https://www.kaggle.com/datasets/suraj520/car-sales-data>

2. Chi tiết thuộc tính

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 Giới thiệu về SSMS (SQL Server Management Studio)

SSMS là một công cụ quản lý SQL Server phổ biến được phát triển bởi Microsoft. SSMS được sử dụng để quản lý và điều khiển SQL Server, cho phép người dùng thực hiện nhiều tác vụ khác nhau như thiết lập cấu hình SQL Server, tạo, xóa hoặc sửa đổi cơ sở dữ liệu, tạo và chạy các truy vấn SQL, xem và quản lý các đối tượng trong cơ sở dữ liệu (bảng, thủ tục lưu trữ, chức năng, chỉ mục, v.v.).

SSMS có giao diện trực quan và đơn giản, giúp người dùng dễ dàng tương tác với SQL Server. Nó hỗ trợ nhiều phiên bản SQL Server khác nhau và cung cấp nhiều tính năng mạnh mẽ.

Tóm lại, SSMS là một công cụ hữu ích và cần thiết cho những người quản lý SQL Server hoặc làm việc với cơ sở dữ liệu. Hiện nay đã ra mắt bản 2022.

3.2 Giới thiệu Visual Studio

Microsoft Visual Studio là một môi trường phát triển tích hợp (IDE) từ Microsoft. Nó được sử dụng để phát triển chương trình máy tính cho Microsoft Windows, cũng như các trang web, các ứng dụng web và các dịch vụ web. Visual Studio sử dụng nền tảng phát triển phần mềm của Microsoft như Windows API, Windows Forms, Windows Presentation Foundation, Windows Store và Microsoft Silverlight. Nó có thể sản xuất cả hai ngôn ngữ máy và mã số quản lý.

3.3 Ngôn ngữ truy vấn sql

SQL (Structured Query Language) là một ngôn ngữ lập trình dùng để truy vấn và quản lý cơ sở dữ liệu quan hệ (RDBMS - Relational Database Management System). SQL là một ngôn ngữ đơn giản và dễ học, nhưng cũng rất mạnh mẽ và linh hoạt.

SQL cho phép người dùng truy vấn và tương tác với các cơ sở dữ liệu quan hệ, như truy xuất dữ liệu, thêm mới, sửa đổi và xóa bỏ dữ liệu. Nó cũng cho phép người dùng tạo và quản lý bảng, chỉ định ràng buộc, tạo truy vấn phức tạp và thực hiện các hoạt động quản lý cơ sở dữ liệu.

SQL được sử dụng rộng rãi trong các hệ thống quản lý cơ sở dữ liệu như MySQL, Oracle, SQL Server và PostgreSQL. Nó là một phần quan trọng trong việc xử lý dữ liệu và đóng vai trò quan trọng trong việc phân tích dữ liệu và quản lý dữ liệu cho các ứng dụng và website.

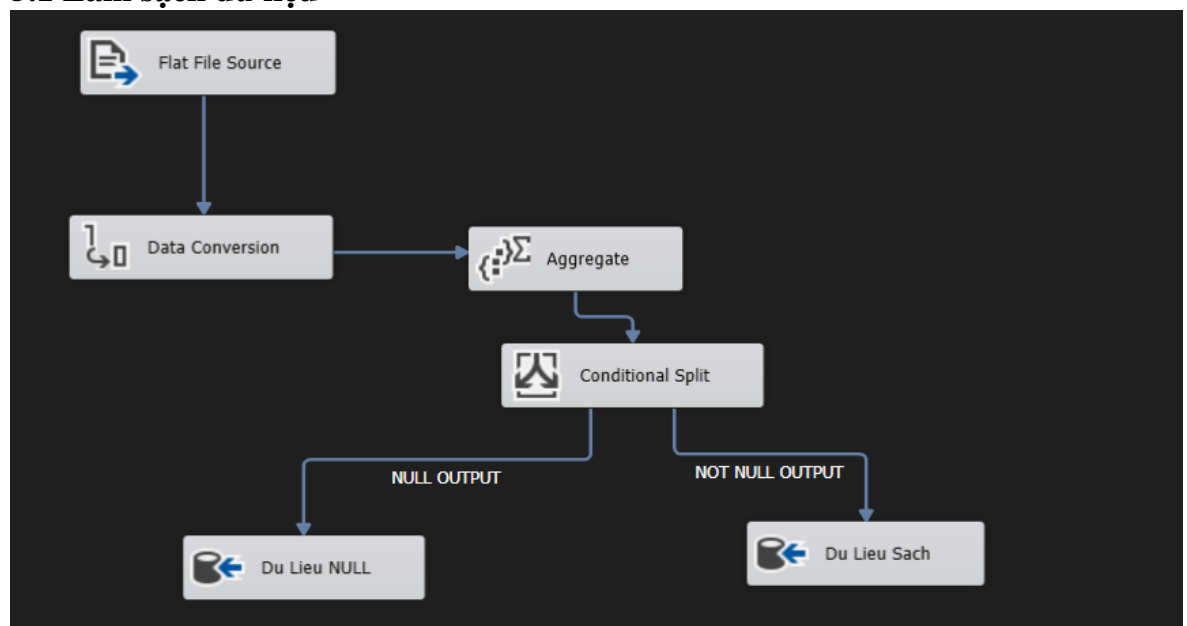
4. Câu hỏi Business Process

Báo cáo doanh số (sales reporting): người quản lý muốn theo dõi doanh số bán ô tô, bao gồm mẫu ô tô, hãng ô tô, nhân viên bán hàng, giá bán, tỷ lệ hoa hồng, tiền hoa hồng kiếm được và năm bán ô tô.

- Các mẫu xe bán chạy nhất và có doanh thu cao nhất
- Các hãng xe bán chạy nhất doanh thu cao nhất
- Nhân viên bán hàng nhiều nhất mang về hoa hồng nhiều nhất
- 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 mẫu xe trong năm gần đây là bao nhiêu?
- Có sự khác biệt giữa việc chọn hãng xe và giá bán không

5. Xây dựng kho dữ liệu

5.1 Làm sạch dữ liệu

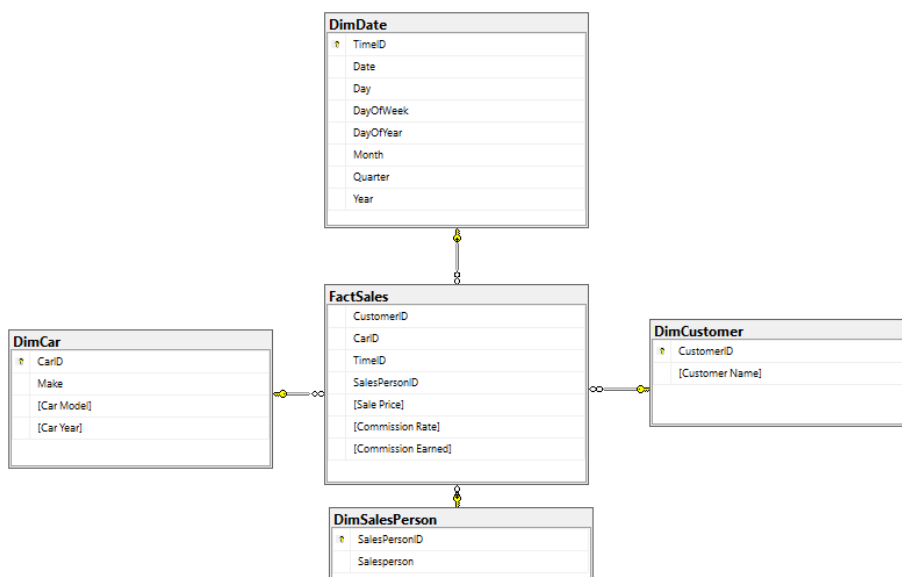


Thông tin dữ liệu khi làm sạch

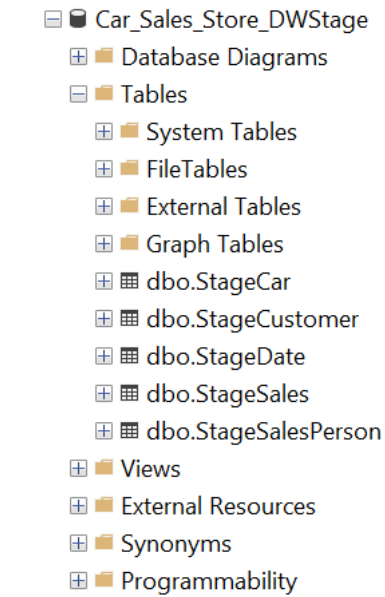
DuLieuSach
Date
Salesperson
[Customer Name]
Make
[Car Model]
[Car Year]
[Sale Price]
[Commission Rate]
[Commission Earned]

DuLieuNULL
Date
Salesperson
[Customer Name]
Make
[Car Model]
[Car Year]
[Sale Price]
[Commission Rate]
[Commission Earned]

5.2 Car_Sales_Store_DW

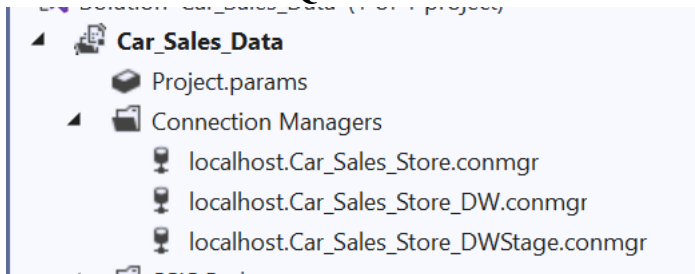


5.3 Thông tin các bảng Car_Sales_Store_Stage

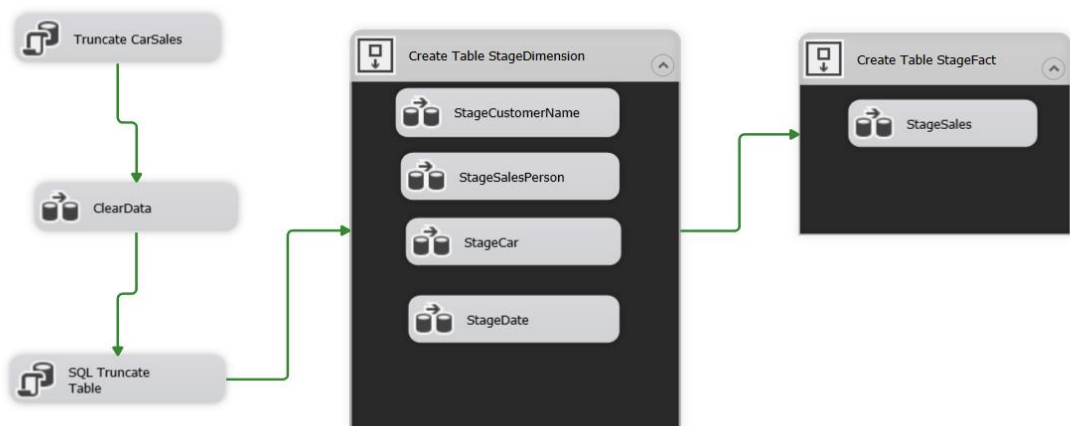


6. Quá trình SSIS

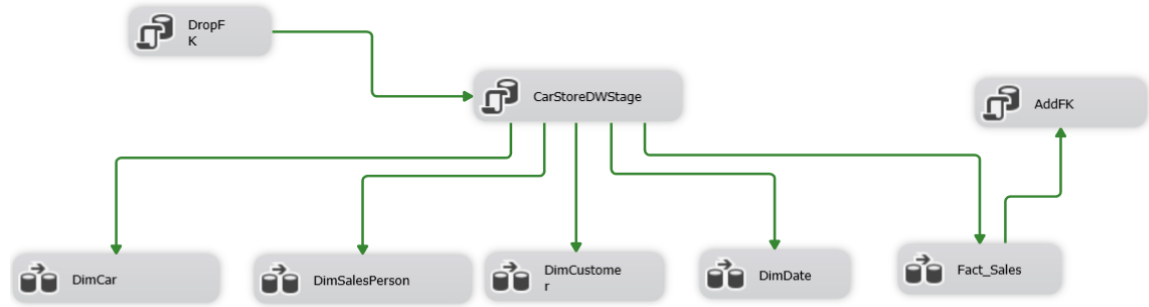
6.1 Kết nối SQL



6.2 Dataflow – StageSales



6.3 Dataflow – Fact_Sale



7. Quá trình SSAS

7.1 Tạo Kết nối Data Sources

Data Source Designer

General Impersonation Information

Data source name: Car Sales Store DW

Provider:

Connection string: Provider=SQLNCLI11.1;Data Source=localhost;Integrated Security=SSPI; Edit...

Data source references

☐ Maintain a reference to another object in the solution

Create a data source based on an existing data source

Isolation: ReadCommitted

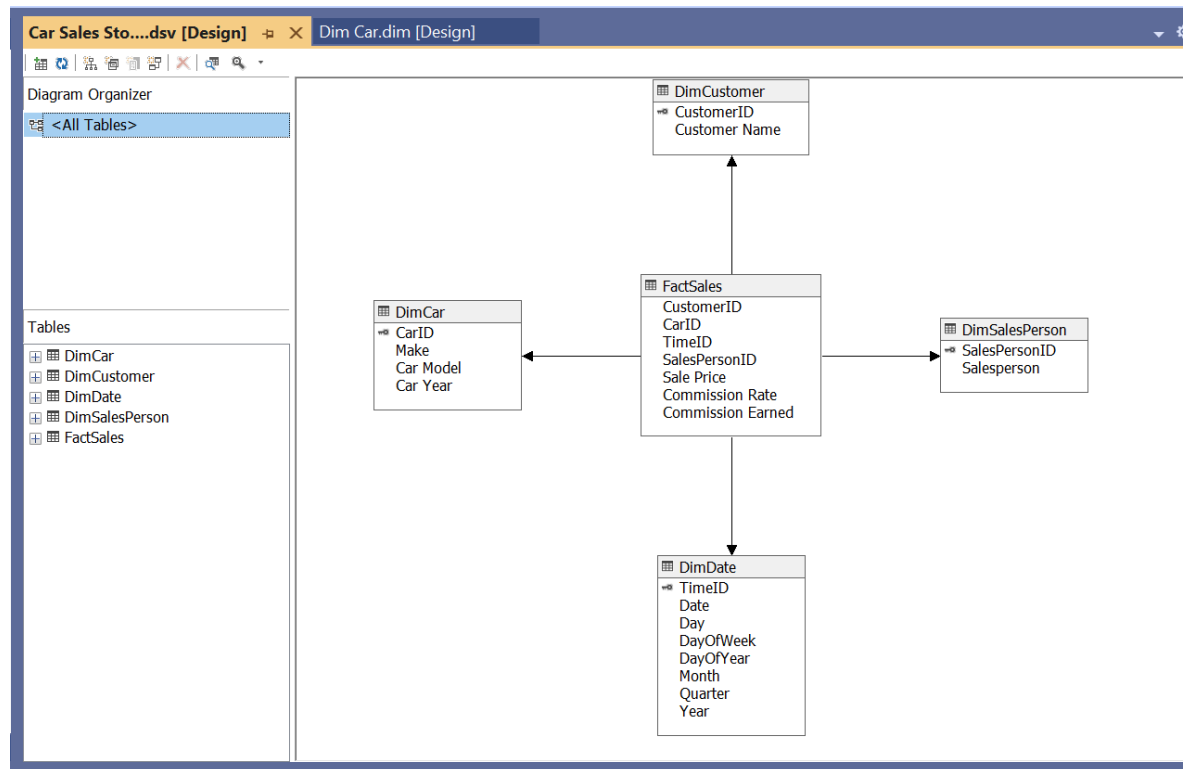
Query timeout (in seconds): 0

Maximum number of connections: 10

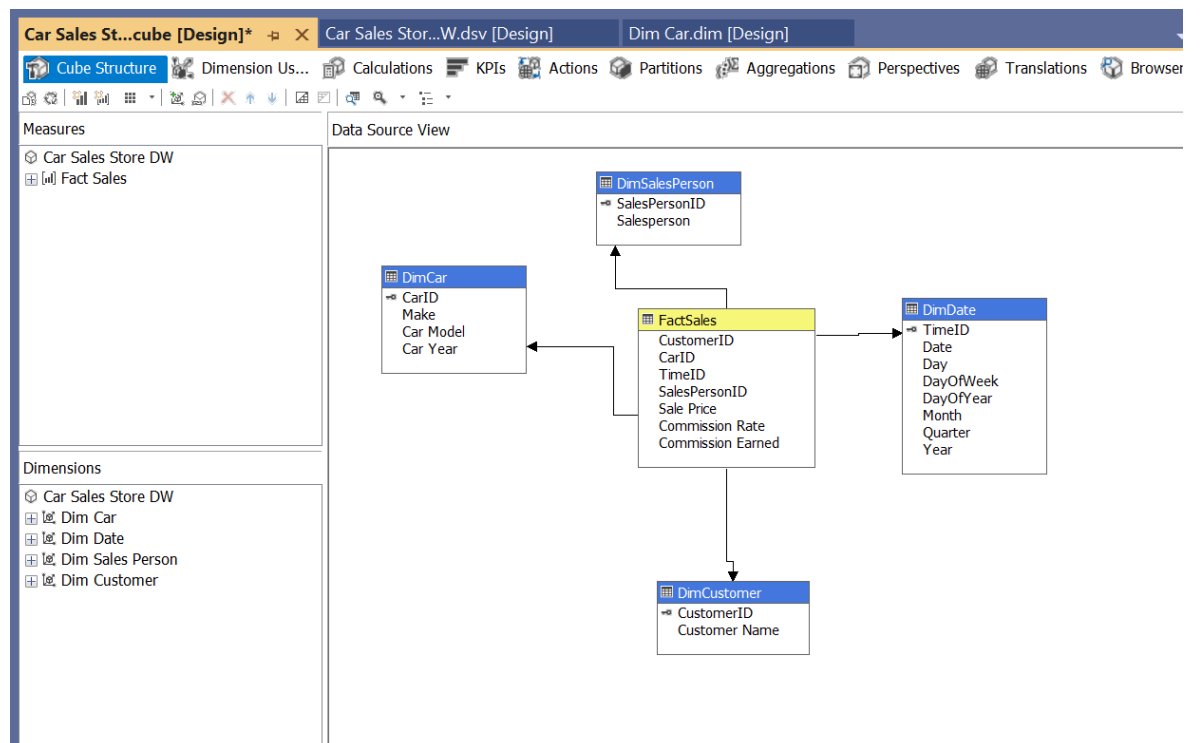
Data source description:

OK Cancel Help

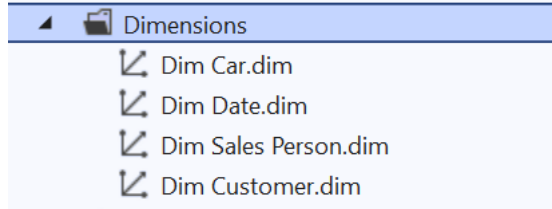
7.2 Tạo và Kết nối Data Source Views



7.3 Tạo Cubes

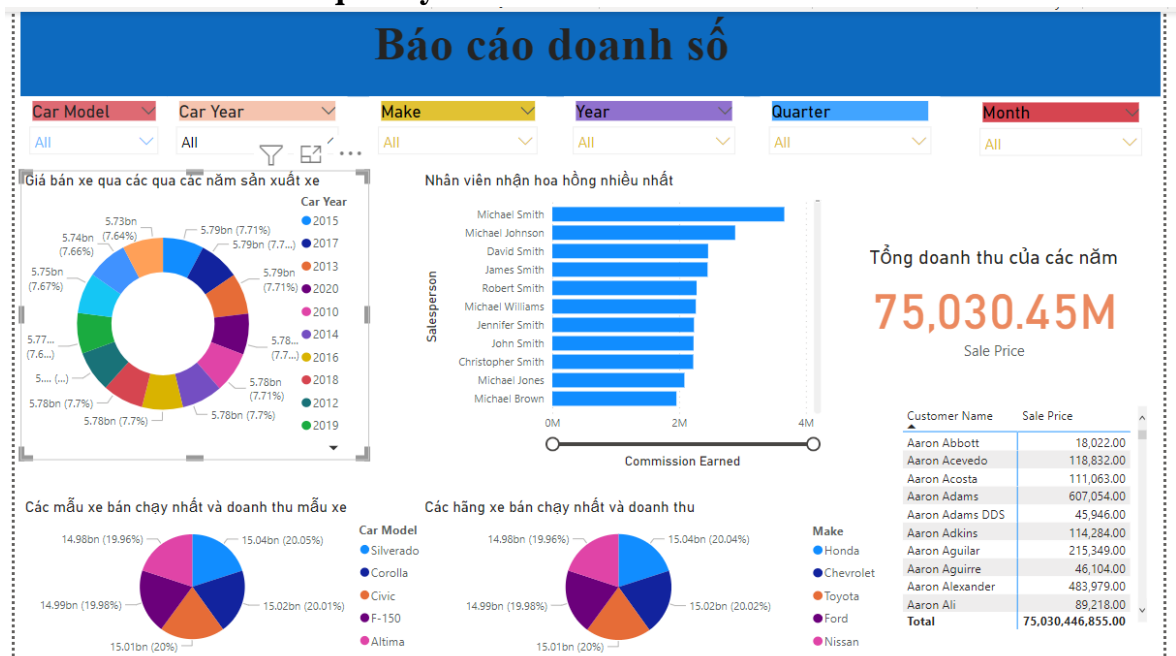


7.4 Dimensions



8. Phân tích dữ liệu trên Power BI

8.1 Dashboard quản lý doanh số



8.2 Tổng doanh thu của năm qua các lần bán xe

Tổng doanh thu của các năm

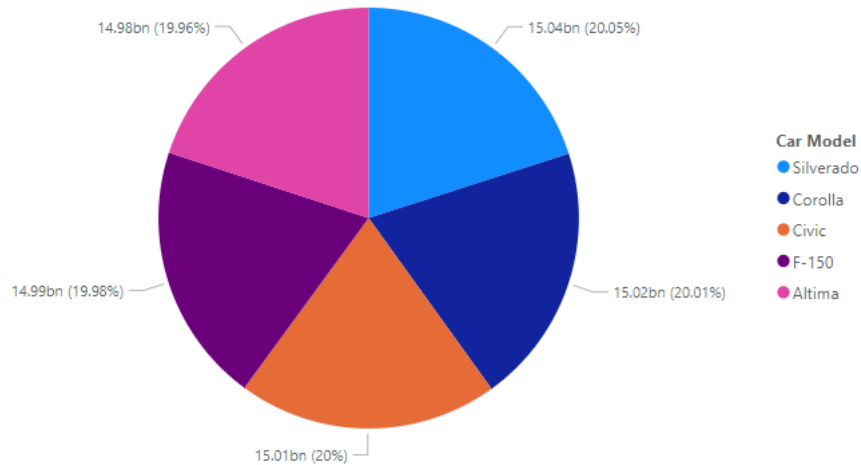
75,030.45M

Sale Price

Nhận xét: Trong 2 năm 2022 và 2023 Tổng doanh thu là 75M 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 mẫu xe bán chạy nhất và có doanh thu cao nhất

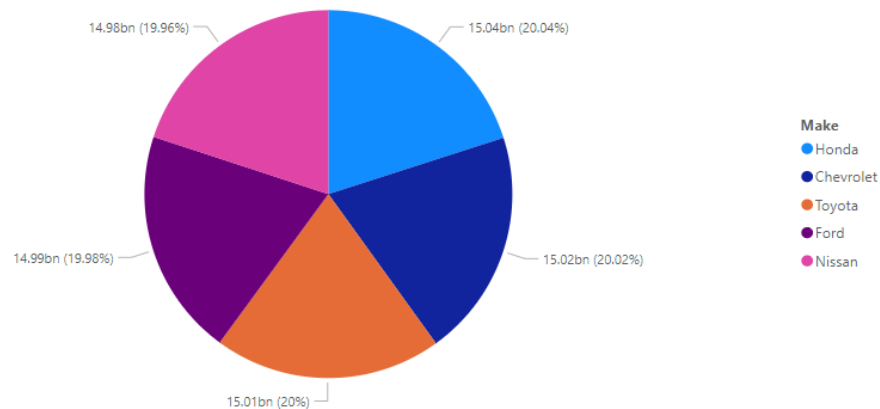
Các mẫu xe bán chạy nhất và doanh thu mẫu xe



Nhận xét: Có thể thấy rằng những mẫu xe phổ biến nhất được bán ra là Silverado, Civic, Corolla, F-150 và Altima. Các mẫu xe này đều thuộc loại xe phổ thông và được sản xuất bởi các hãng xe nổi tiếng. 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 hãng xe bán chạy nhất doanh thu cao nhất

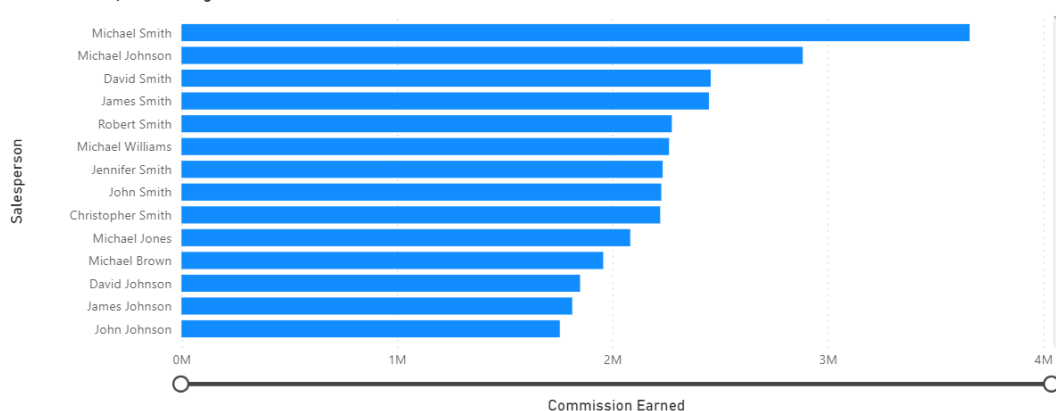
Các hãng xe bán chạy nhất và doanh thu



Nhận xét: Có thể thấy sự phổ biến của các hãng xe Honda, Chevrolet, Toyota, Ford và Nissan trên thị trường ô tô hiện nay. Những hãng xe này đều là những nhà sản xuất ô tô lớn và uy tín, có rất nhiều khách hàng trung thành trên toàn thế giới, và những hãng xe đã chứng tỏ được chất lượng sản phẩm và dịch vụ của mình trên thị trường ô tô.

8.5 Nhân viên bán hàng nhiều nhất mang về hoa hồng nhiều nhất

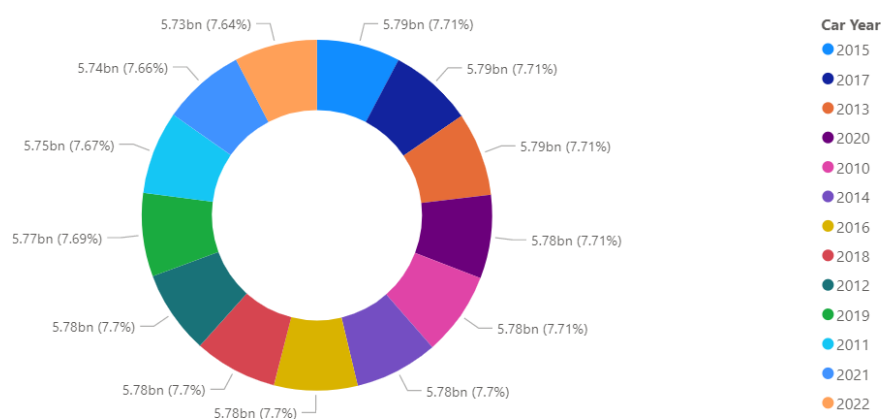
Nhân viên nhận hoa hồng nhiều nhất



Nhận xét: Có thể thấy rằng những nhân viên bán hàng phát huy hiệu quả là Michael Smith, Michael Johnson, David Smith, James Smith và Robert Smith. Những người này đã mang về hoa hồng nhiều nhất trong danh sách top 5. Là những nhân viên bán hàng nổi bật và thành công trong việc tăng doanh số bán hàng và mang lại lợi nhuận cho công ty các nhân viên bán hàng này có thể 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 Sự khác biệt giữa việc lựa chọn năm sản xuất xe và giá bán không

Giá bán xe qua các năm sản xuất xe



Nhận xét: Các thông tin về giá bán xe và các thuộc tính của các chiếc xe như năm sản xuất, loại xe, động cơ, màu sắc, số chỗ ngồi, v.v. ta có thể khám phá sự khác biệt giữa việc lựa chọn năm sản xuất xe và giá bán. Giá bán xe có xu hướng tăng theo năm sản xuất xe. Những chiếc xe mới hơn thường có giá bán cao hơn so với những

chiếc xe cũ hơn, bởi sự tiến bộ trong công nghệ và thiết kế của các chiếc xe mới hơn, điều này khiến cho giá bán tăng lên. Tuy nhiên sự khác biệt trong giá bán còn phụ thuộc vào nhiều yếu tố khác nhau. Do đó, khi quan tâm đến giá bán của một chiếc xe, ta nên xem xét nhiều yếu tố khác nhau, không chỉ riêng năm sản xuất.

8.7 Thông tin tên khách hàng và số tiền dành cho việc mua xe

Thông tin khách hàng và số tiền chi
chi việc mua xe

Customer Name	Sale Price
Aaron Abbott	18,022.00
Aaron Acevedo	118,832.00
Aaron Acosta	111,063.00
Aaron Adams	607,054.00
Aaron Adams DDS	45,946.00
Aaron Adkins	114,284.00
Aaron Aguilar	215,349.00
Aaron Aguirre	46,104.00
Aaron Alexander	483,979.00
Aaron Ali	89,218.00
Aaron Ali MD	15,575.00
Aaron Allen	1,008,733.00
Aaron Allen MD	42,632.00
Aaron Allison	149,404.00
Aaron Alvarado	285,425.00
Total	75,030,446,855.00

Nhận xét: Dữ liệu cung cấp thông tin về các khách hàng và số tiền họ đã dành cho việc mua xe. Điều này hữu ích cho các công ty ô tô để hiểu được nhu cầu của khách hàng và phát triển chiến lược bán hàng phù hợp, hiểu về hành vi mua sắm của khách hàng. Chẳng hạn, chúng ta có thể biết được khách hàng nào đã chi tiêu nhiều tiền hơn cho việc mua xe. Nhờ đó, chúng ta có thể đưa ra các giải pháp bán hàng phù hợp để thu hút và duy trì những khách hàng này.

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&t=13844s
(Accessed: 04 May 2023).
2. Thabresh (2023) Car sales data - EDA, Kaggle. Available at:
<https://www.kaggle.com/code/thabresh/car-sales-data-eda>
(Accessed: 01 May 2023).
3. ETL project from Excel data source to star schema with SSIS (2021) YouTube. Available at:
<https://www.youtube.com/watch?v=Yp8fXLnVCp8&t=596s>
(Accessed: 08 May 2023).
4. 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).
5. Power Bi Nâng Cao | Tạo dashboard quản Lý Bán Hàng (kèm file Thực Hành) (2021) YouTube. Available at:
<https://www.youtube.com/watch?v=k9dmVIPqGDM&t=1726s>
(Accessed: 12 May 2023).