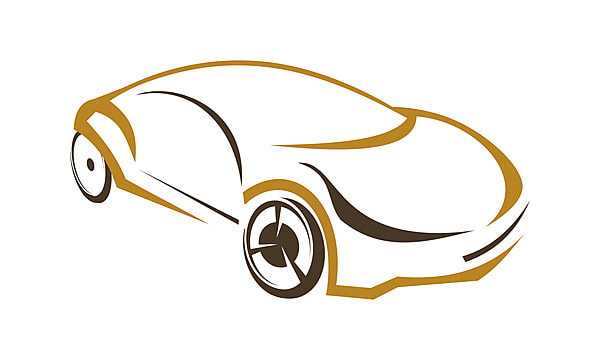


**TÀI LIỆU DỰ ÁN**

**Phân tích bán xe ô tô cũ - Hope**



| **Giảng viên** | **n : TRẦN TUẤN PHONG** |
| --- | --- |
| **Nhóm : 5** |  |
| **Lớp/Kỳ** | **: DP19302/3** |

MỤC LỤC

[1](#_heading=h.gjdgxs) Giới thiệu dự án 6

[1.1](#_heading=h.30j0zll) Giới thiệu công ty Hope 6

[1.2](#_heading=h.1fob9te) Yêu cầu của công ty 6

[1.3](#_heading=h.3znysh7) Lập kế hoạch dự án 7

[2](#_heading=h.2et92p0) Phân tích yêu cầu khách hàng 8

[2.1](#_heading=h.tyjcwt) Phân tích yêu cầu 8

[2.2](#_heading=h.3dy6vkm) Câu chuyện dữ liệu 13

[2.2.1](#_heading=h.1t3h5sf) Đặt vấn đề 13

[2.2.2](#_heading=h.4d34og8) Xác định câu chuyện 13

[2.2.3](#_heading=h.2s8eyo1) Xác định rõ đối tượng 13

[2.2.4](#_heading=h.17dp8vu) Xác định câu chuyện chi tiết 13

[2.2.5](#_heading=h.3rdcrjn) Trình bày dữ liệu 14

[2.2.6](#_heading=h.26in1rg) Những điều cần lưu ý 15

[2.3](#_heading=h.lnxbz9) Kiến trúc hệ thống 17

[2.3.1](#_heading=h.35nkun2) Kiến trúc 17

[2.3.2](#_heading=h.1ksv4uv) Giải thích 18

[2.4](#_heading=h.44sinio) Giải thích về bộ dữ liệu khách hàng 18

[2.4.1](#_heading=h.2jxsxqh) Các khái niệm 18

[2.4.2](#_heading=h.z337ya) Các trường dữ liệu 21

[3](#_heading=h.3j2qqm3) Làm sạch và chuyển đổi dữ liệu 22

[3.1](#_heading=h.1y810tw) Chuẩn bị dữ liệu 22

[3.1.1](#_heading=h.4i7ojhp) Giải pháp lưu trữ dữ liệu 23

[3.1.2](#_heading=h.2xcytpi) Giải pháp phân bố dữ liệu 24

[3.2](#_heading=h.1ci93xb) Làm sạch dữ liệu 25

[3.2.1](#_heading=h.3whwml4) Các vấn đề ảnh hưởng tới dữ liệu 25

[3.2.2](#_heading=h.2bn6wsx) Các tiêu chí đánh giá chất lượng dữ liệu 25

[3.2.3](#_heading=h.qsh70q) Các bước làm sạch dữ liệu 27

[3.3](#_heading=h.1pxezwc) Chuyển đổi dữ liệu 29

[3.3.1](#_heading=h.49x2ik5) Các trường hợp cần chuyển đổi 29

[3.3.2](#_heading=h.2p2csry) Các kỹ thuật chuyển đổi 30

[3.3.3](#_heading=h.147n2zr) Trình bày các phép chuyển đổi trong dự án 30

[4](#_heading=h.3o7alnk) Xử lý dữ liệu 31

[4.1](#_heading=h.23ckvvd) Chuẩn hóa dữ liệu 31

[4.1.1](#_heading=h.ihv636) Trình bày các bước chuẩn hóa trong dự án 31

[4.2](#_heading=h.32hioqz) Mô hình hóa dữ liệu 32

[4.2.1](#_heading=h.1hmsyys) Các loại mô hình hóa 32

[4.2.2](#_heading=h.41mghml) Các tiêu chí đánh giá mô hình dữ liệu 33

[4.2.3](#_heading=h.2grqrue) Trình bày các bước mô hình hóa 33

[4.2.4](#_heading=h.vx1227) Trình bày các bước tạo bảng dữ liệu 36

[4.3](#_heading=h.3fwokq0) Xử lý dữ liệu DAX 38

[4.3.1](#_heading=h.1v1yuxt) Measure 38

[4.3.2](#_heading=h.4f1mdlm) Calculated column 42

[4.3.3](#_heading=h.2u6wntf) Filter 43

[5](#_heading=h.19c6y18) Trực quan hóa dữ liệu 44

[5.1](#_heading=h.3tbugp1) Các kỹ thuật trực quan hóa 44

[5.2](#_heading=h.28h4qwu) Các nguyên tắc trực quan hóa 46

[5.3](#_heading=h.nmf14n) Trình bày cách thêm visual mới 49

[5.4](#_heading=h.37m2jsg) Trình bày tạo các report cho dự án 52

[6](#_heading=h.2lwamvv) Xây dựng báo cáo 55

[6.1](#_heading=h.111kx3o) Dashboard và report 55

[6.2](#_heading=h.3l18frh) Xây dựng báo cáo 57

[6.2.1](#_heading=h.206ipza) Dashboard vs Report 57

[6.2.2](#_heading=h.4k668n3) Dashboard 59

[6.2.3](#_heading=h.2zbgiuw) Report 60

[6.2.4](#_heading=h.1egqt2p) Bookmark 61

[7](#_heading=h.3ygebqi) KẾT LUẬN 66

[7.1](#_heading=h.2dlolyb) Báo cáo 66

[7.1.1](#_heading=h.sqyw64) Các bước viết báo cáo 66

[7.1.2](#_heading=h.3cqmetx) Tổng hợp 67

[7.2](#_heading=h.1rvwp1q) Khó khăn 69

[7.3](#_heading=h.4bvk7pj) Thuận lợi 70

[7.4](#_heading=h.2r0uhxc) Hướng phát triển 70

**Danh sách thành viên nhóm:**

| **STT** | **Họ tên** | **Mã sinh viên** | **Email** |
| --- | --- | --- | --- |
| 1 | *Nguyễn Văn Minh* | PH54780 | [nvanminhdata211105@gmail.com](mailto:nvanminhdata211105@gmail.com) |
| 2 | *Nguyễn Văn Khải* | PH53468 | [khainvph53468@gmail.com](mailto:khainvph53468@gmail.com) |
| 3 | *Hà Vũ Đức* | PH54334 | [Duchv612@gmail.com](mailto:Duchv612@gmail.com) |
| 4 | *Nguyễn Văn Thành* | PH52035 | [thanhph52035@gmail.com](mailto:thanhph52035@gmail.com) |
| 5 | *Lê Huy Hoàng* | PH53268 | [hoanglhph53268@gmai.com](mailto:hoanglhph53268@gmai.com) |
| 6 | *Nguyễn Việt Anh* | PH52136 | [anhnvph52136@gmail.com](mailto:anhnvph52136@gmail.com) |

# Giới thiệu dự án

## Giới thiệu công ty Hope

Công ty Hope hiện đang gặp phải một vấn đề phổ biến trong thời đại số hóa: không phải là thiếu dữ liệu mà là làm sao để quản lý và sử dụng dữ liệu một cách hiệu quả. Ngày nay, dữ liệu có thể dễ dàng được thu thập, mua bán hoặc trao đổi từ nhiều nguồn khác nhau, dẫn đến việc sở hữu một lượng lớn thông tin không đồng nhất về cấu trúc và định dạng. Nhiều dữ liệu chỉ ở dạng thô, không tuân theo bất kỳ tiêu chuẩn nào, làm cho việc xử lý và phân tích trở nên khó khăn. Hơn nữa, việc có quá nhiều thông tin nhưng thiếu sự diễn giải cụ thể khiến cho quá trình ra quyết định trở nên phức tạp. Để cải thiện tình hình, Hope cần tập trung vào việc chuẩn hóa và tối ưu hóa dữ liệu để biến chúng thành thông tin hữu ích, hỗ trợ quá trình ra quyết định một cách hiệu quả hơn.

Tập dữ liệu này liên quan đến các phiên giao dịch trên thị trường xe ô tô cũ, được thu thập từ nhiều website khác nhau, cả có trả phí và miễn phí, trong một khoảng thời gian nhất định. Dữ liệu chủ yếu là những con số khô khan và mang tính phẳng, không theo cấu trúc. Đây là nguồn thông tin quý giá để phân tích và hiểu rõ hơn về thị trường xe ô tô cũ.

## Yêu cầu của công ty

***1. Truyền đạt và diễn giải dữ liệu***

* *Trình bày thông tin phức tạp một cách rõ ràng và hấp dẫn: Mục tiêu là đảm bảo rằng thông điệp được truyền tải dễ hiểu và lôi cuốn, giúp mọi người nắm bắt nhanh chóng.*
* *Tạo Data Insights sống động, dễ hiểu và thuyết phục: Đảm bảo rằng các vấn đề được hiểu rõ ràng và chính xác nhất, từ đó nâng cao hiệu quả làm việc và hỗ trợ việc đưa ra các quyết định kinh doanh chiến lược.*

**2. Quản lý và Lưu trữ Dữ liệu:**

* *Giảm không gian lưu trữ và quản lý tài liệu khoa học: Tối ưu hóa việc lưu trữ dữ liệu để tiết kiệm thời gian tìm kiếm và giảm không gian lưu trữ.*
* *Tăng cường bảo mật thông tin: Bảo vệ dữ liệu quan trọng bằng cách nâng cao khả năng bảo mật thông tin.*
* *Chia sẻ thông tin nhanh chóng và hiệu quả: Đảm bảo rằng thông tin được chia sẻ kịp thời để nâng cao hiệu suất làm việc.*
* *Giảm chi phí vận hành và quản lý: Đạt được hiệu quả cao trong quản lý dữ liệu mà vẫn giữ chi phí ở mức thấp, tối ưu hóa hoạt động kinh doanh và nâng cao khả năng cạnh tranh.*

**3. Năng lực :**

| **Công việc chính** | **Công nghệ sử dụng** |
| --- | --- |
| Thu thập làm sạch dữ liệu | SQL, Excel |
| Thiết kế cơ sở dữ liệu và nhập liệu | SQL |
| Phân tích dữ liệu cơ bản và thống kê | Excel |
| Xây dựng báo cáo trực quan | Power Bi |
| Viết báo cáo phân tích | Word |
| Kiểm tra hiệu chỉnh báo cáo cuối cùng | Word, excel |

## 1.3 Lập kế hoạch dự án

| 1 | Giới thiệu dự án | 25/06/2024 | 25/06/2024 | Hoàn thành |
| --- | --- | --- | --- | --- |
| 1.1 | Giới thiệu công ty | 27/06/2024 | 27/06/2024 | Hoàn thành |
| 1.2 | Yêu cầu công ty | 29/06/2024 | 29/06/2024 | Hoàn thành |
| 1.3 | Lập dự án kế hoạch | 09/07/2024 | 09/07/2024 | Hoàn thành |
| 2 | Phân tích yêu cầu | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 2.1 | Phân tích yêu cầu KH | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 2.2 | Câu chuyện dữ liệu | 06/08/2024 | 06/08/2024 | Hoàn thành |
| 2.3 | Kiến trúc hệ thống | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 2.3.1 | Kiến trúc | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 2.3.2 | Giải thích | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 2.4 | Giải thích về bộ dữ liệu khách hàng | 10/07/2024 | 10/07/2024 | Hoàn thành |
| 3 | Làm sạch và chuyển đổi dữ liệu | 11/07/2024 | 11/07/2024 | Hoàn thành |
| 3.1 | Chuẩn bị dữ liệu | 11/07/2024 | 11/07/2024 | Hoàn thành |
| 3.1.1 | Giải pháp lưu trữ dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.1.2 | Giải pháp phân bố dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.2 | Làm sạch dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.2.1 | Các vấn đề ảnh hưởng đến dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.2.2 | Các tiêu chí đánh giá chất lượng dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.2.3 | Các bước làm sạch dữ liệu | 12/07/2024 | 12/07/2024 | Hoàn thành |
| 3.3 | Chuyển đổi dữ liệu | 13/07/2024 | 13/07/2024 | Hoàn thành |
| 3.3.1 | Các trường hợp cần chuyển đổi dữ liệu | 13/07/2024 | 13/07/2024 | Hoàn thành |
| 3.3.2 | Các kỹ thuật chuyển đổi | 13/07/2024 | 13/07/2024 | Hoàn thành |
| 3.3.3 | Trình bày các phép chuyển đổi trong dự án | 13/07/2024 | 13/07/2024 | Hoàn thành |
| 4 | Xử lý dữ liệu | 20/07/2024 | 20/07/2024 | Hoàn thành |
| 4.1 | Chuẩn hóa dữ liệu | 20/07/2024 | 20/07/2024 | Hoàn thành |
| 4.2 | Mô hình hóa dữ liệu | 20/07/2024 | 20/07/2024 | Hoàn thành |
| 4.3 | Xử lý dữ liệu DAX | 30/7/2024 | 1/8/2024 | Hoàn thành |
| 5 | Trực quan hóa dữ liệu | 3/8/2024 | 6/8/2024 | Hoàn thành |
| 5.1 | Các kĩ thuật trực quan hóa | 3/8/2024 | 3/8/2024 | Hoàn thành |
| 5.2 | Các nguyên tắc trực quan hóa | 4/8/2024 | 4/8/2024 | Hoàn thành |
| 5.3 | Trình bày cách thêm visual mới | 4/8/2024 | 4/8/2024 | Hoàn thành |
| 5.4 | Trình bày tạo các report cho dự án | 5/8/2024 | 5/8/2024 | Hoàn thành |
| 6 | Xây dựng báo cáo | 5/8/2024 | 5/8/2024 | Hoàn thành |
| 6.1 | Dashboard và Report | 5/8/2024 | 5/8/2024 | Hoàn thành |
| 6.2 | Xây dựng báo cáo | 5/8/2024 | 5/8/2024 | Hoàn thành |

# Phân tích yêu cầu khách hàng

## 2.1Phân tích yêu cầu

**Dữ liệu**

| ***Danh mục*** | ***Chi tiết*** |
| --- | --- |
| **Dữ liệu** |  |
| * Loại dữ liệu | Các phiên giao dịch trên thị trường xe ô tô cũ |
| * Nguồn dữ liệu | Trong một khoảng thời gian nhất định |
| * Thời gian thu thập | Những con số khô khan, thông tin phẳng, không theo cấu trúc |
| * Đặc điểm dữ liệu | Cần có phương pháp truyền đạt và diễn giải thông tin dữ liệu một cách hiệu quả đến nhân viên và sếp |
| **Truyền đạt và diễn giải** |  |
| * Phương pháp | Cần có phương pháp truyền đạt và diễn giải thông tin dữ liệu một cách hiệu quả đến nhân viên và sếp |
| * Thông điệp | Thông điệp cần được trình bày rõ ràng và hấp dẫn |
| * Data Insights | Data Insights phải sống động, dễ hiểu và thuyết phục |
| **Insight** |  |
| * Tình trạng xe và giá bán | Phân tích tình trạng xe và giá bán |
| * Ảnh hưởng của thương hiệu và model | Đánh giá ảnh hưởng của thương hiệu và model xe tới giá bán và sự ưa chuộng |
| * Tỷ lệ khấu hao xe ô trung bình | Tính toán tỷ lệ khấu hao trung bình của xe ô tô |
| * Dòng đời xe ô | Tính toán tỷ lệ khấu hao trung bình của xe ô tô |
| * Đặc điểm khách hàng | Nghiên cứu các đặc điểm của khách hàng |
| * Yếu tố ảnh hưởng đến doanh số | Phân tích các yếu tố ảnh hưởng đến doanh số |
| * Phân khúc xe phổ biến | Xác định phân khúc xe nào phổ biến nhất |
| * Giá ước tính và giá thực tế | Điều gì ảnh hưởng tới sự chênh lệch giữa giá ước tính trên thị trường và giá bán thực tế |
| * Trung bình giá xe cũ | Tính toán giá trung bình của xe cũ |
| * Hãng xe được yêu thích | Xác định hãng xe cũ được khách hàng yêu thích |
| * Màu sắc và loại hộp số ưa chuộng | Xác định màu sắc và loại hộp số nào được người mua ưa chuộng |
| * Xu hướng giá bán và lượng xe bán | Phân tích xu hướng giá bán và lượng xe bán ra theo thời gian |

### Quản lý và Lưu trữ

| **Danh mục** | **Phương pháp** | **Công cụ** |
| --- | --- | --- |
| **Quản lý và Lưu trữ** |  |  |
| * Giảm không gian lưu trữ và tiết kiệm thời gian tìm kiếm | Số hóa tài liệu, sử dụng hệ thống quản lý tài liệu (DMS). | Google Drive, Dropbox, hoặc SharePoint. |
| * Tăng cường bảo mật thông tin | Sử dụng mã hóa dữ liệu, quyền truy cập người dùng. | Azure Information Protection, BitLocker, hoặc các phần mềm mã hóa khác. |
| * Chia sẻ thông tin nhanh chóng | Sử dụng nền tảng chia sẻ đám mây. | Google Drive, OneDrive, hoặc Dropbox. |
| * Chi phí vận hành và quản lý thấp nhưng hiệu quả cao | Tối ưu hóa quy trình, sử dụng dịch vụ đám mây. | Trello, AWS, Google Cloud Platform, hoặc Microsoft Azure. |

### Công nghệ

| Danh mục | Phương pháp | Công cụ |
| --- | --- | --- |
| Công nghệ phổ biến | Áp dụng các công nghệ đã được công nhận rộng rãi | - Microsoft Office Suite, Slack, Jira |
| Chi phí hợp lý | Chọn các công nghệ có chi phí phù hợp với ngân sách công ty | - Google Workspace, Office 365, Asana |
| Có thể mở rộng nhanh chóng | Sử dụng các nền tảng linh hoạt và dễ dàng mở rộng | - Kubernetes, Docker, Azure |

**Lựa chọn công nghệ**

| **Danh mục** | **Chi tiết** |
| --- | --- |
| ***Dữ liệu*** |  |
| Công nghệ lựa chọn | Power BI |
| Lý do lựa chọn | * Power BI là công cụ mạnh mẽ và dễ sử dụng để tạo các báo cáo trực quan, biểu đồ và dashboard. |
| * Hỗ trợ biểu đồ tương tác và storytelling với dữ liệu, giúp data insights sống động, dễ hiểu và thuyết phục. |
| * Power BI cung cấp khả năng tóm tắt dữ liệu thành các điểm chính và cung cấp dashboard dữ liệu thời gian thực. |
| Truyền đạt và diễn giải thông tin | * Truyền đạt và diễn giải thông tin tới nhân viên và sếp với thông điệp rõ ràng và hấp dẫn. |
| * Data insights sống động, dễ hiểu, thuyết phục, giúp hiểu vấn đề nhanh và chính xác. |
| ***Quản lý và Lưu trữ*** |  |
| Công nghệ chọn | SQL Server |
| Lý do chọn | * Hiệu suất cao: SQL Server cung cấp hiệu suất xử lý tốt với khả năng tối ưu hóa truy vấn và xử lý khối lượng dữ liệu lớn. |
| * Khả năng mở rộng: Hỗ trợ mở rộng cả về chiều ngang và chiều dọc, phù hợp với các doanh nghiệp có nhu cầu tăng trưởng dữ liệu. |
| * Tính năng bảo mật mạnh mẽ: Cung cấp các tính năng bảo mật nâng cao như mã hóa dữ liệu, kiểm soát quyền truy cập, và các công cụ phát hiện và phản ứng đối với các mối đe dọa. |
| * Hỗ trợ báo cáo và phân tích: Tích hợp tốt với các công cụ báo cáo và phân tích với Power Bi |
| * Tính năng sao lưu và phục hồi: Cung cấp các giải pháp sao lưu và phục hồi mạnh mẽ để bảo vệ dữ liệu và đảm bảo khả năng khôi phục sau sự cố. |
| ***Công nghệ*** | Google Workspace |
| Lý do chọn | * Google Workspace là bộ công cụ văn phòng phổ biến, bao gồm Gmail, Google Docs, Google Sheets, và Google Slides. |
| * Đáp ứng các nhu cầu công nghệ phổ biến với chi phí hợp lý, phù hợp với ngân sách công ty. |
| * Dễ dàng mở rộng khi cần thiết, cung cấp các dịch vụ linh hoạt và tích hợp với nhiều công cụ khác. |

## 

## 2.2 Câu chuyện dữ liệu

### 2.2.1 Giới thiệu:

**Bối cảnh**

Hope là một công ty mới thành lập trong lĩnh vực mua bán xe cũ với sứ mệnh mang lại những chiếc xe chất lượng tốt, giá cả phải chăng cho người dân. Để hiểu rõ thị trường và giá cả xe cũ hiện tại, Hope quyết định phân tích xem liệu có nên đầu tư vào thị trường này hay không.

**Dữ liệu**

Hope sử dụng một bộ dữ liệu chi tiết về các xe bao gồm các thông tin như năm sản xuất, hãng sản xuất, mẫu xe, phiên bản, kiểu dáng thân xe, loại hộp số, số VIN, bang nơi xe được bán hoặc đăng ký, tình trạng xe, số km đã đi, màu sắc bên ngoài, nội thất, người bán, giá trị thị trường ước tính (MMR), giá bán thực tế và ngày bán xe.

**Mục tiêu**

Công ty sử dụng bộ dữ liệu này để:

Đánh giá giá trị xe.

Quản lý tồn kho.

Phân tích xu hướng mua bán.

Cung cấp dịch vụ khách hàng tốt hơn.

Điều này giúp khách hàng dễ dàng tìm kiếm và mua xe với đầy đủ thông tin chi tiết, đồng thời giúp công ty tối ưu hóa quy trình kinh doanh, tăng hiệu quả bán hàng và cải thiện dịch vụ khách hàng. Hope mong muốn mở rộng thị trường và tiếp tục cải thiện dịch vụ thông qua việc sử dụng dữ liệu và phân tích để đưa ra các quyết định kinh doanh chiến lược.

**Đối tượng**

Ban giám đốc và cấp quản lý, bao gồm Hội đồng quản trị, Ban cố vấn, Ban kiểm soát nội bộ, Quản lý điều hành cấp cao, Quản lý chi nhánh, Trưởng phòng các bộ phận.

### 2.4.2 Hành động gia tăng

**Xung đột**

Có một tranh luận phổ biến trong ngành ô tô về việc yếu tố nào ảnh hưởng nhiều nhất đến giá bán xe cũ: tình trạng xe (Condition) hay số km đã đi (Odometer). Những người cho rằng tình trạng xe là yếu tố quan trọng nhất lập luận rằng một chiếc xe trong tình trạng tốt, dù có số dặm cao, vẫn có giá trị cao hơn vì đã được bảo dưỡng tốt. Ngược lại, những người tin rằng số km đã đi là yếu tố quan trọng nhất cho rằng số dặm là chỉ số quan trọng về mức độ hao mòn của xe.

**Dự đoán và tầm quan trọng**

Câu hỏi đặt ra là: "Yếu tố nào giữa tình trạng xe (Condition) và số km đã đi (Odometer) ảnh hưởng nhiều nhất đến giá trị bán lại của xe ô tô cũ?"

### 2.4.3 Cao trào

Phân tích dữ liệu cho thấy rằng cả hai yếu tố tình trạng xe và số km đã đi đều ảnh hưởng đến giá bán xe:

Giá bán xe có xu hướng tăng khi tình trạng xe được cải thiện.

Giá bán xe cũng tăng khi số km đã đi giảm.

Điều này cho thấy cả tình trạng và số dặm đều đóng vai trò quan trọng trong việc định giá xe, và cả hai yếu tố cần được xem xét đồng thời để đưa ra một đánh giá chính xác và công bằng.

### 2.4.3 Hành động rơi

**Sự biến động giá xe ô tô cũ theo thời gian:**

Mô tả sự thay đổi giá trung bình theo các năm, tháng hoặc quý.

Giải thích các yếu tố ảnh hưởng đến sự biến động giá.

**Xác định các yếu tố chính ảnh hưởng đến giá bán xe:**

Phân tích các yếu tố như số km đã đi, hãng xe.

So sánh sự biến động giá giữa các loại xe.

Tìm hiểu sự ảnh hưởng của các yếu tố bên ngoài.

**Lợi nhuận trung bình thu được từ việc bán xe ô tô cũ:**

So sánh lợi nhuận giữa các loại xe khác nhau.

Tính toán trung bình lợi nhuận từ việc bán ô tô cũ.

Phân tích các yếu tố ảnh hưởng tới lợi nhuận như số km đã đi, loại hộp số.

**Sự thay đổi trong nhu cầu mua xe ô tô cũ theo thời gian và các yếu tố ảnh hưởng:**

Phân tích số lượng xe được bán qua các tháng hoặc năm để xác định nhu cầu của khách hàng.

Phân tích sự khác biệt trong nhu cầu mua xe giữa các tháng hay các năm khác nhau.

### 2.4.5 Giải quyết

**Tóm tắt kết quả chính:**

Giá bán xe ô tô cũ bị ảnh hưởng mạnh bởi năm sản xuất, số km đã đi, tình trạng xe, và yếu tố mùa vụ.

Lợi nhuận cao hơn thường thấy ở các loại xe có tình trạng tốt và số km ít.

Nhu cầu mua xe ô tô cũ thay đổi theo mùa vụ và chịu ảnh hưởng của nhiều yếu tố kinh tế.

**Tầm quan trọng của kết quả:**

Hiểu rõ xu hướng giá và yếu tố ảnh hưởng giúp doanh nghiệp đưa ra quyết định mua vào và bán ra hợp lý, tối ưu hóa lợi nhuận.

Phân tích chi tiết giúp xác định thời điểm tốt nhất để bán xe và định giá hợp lý, tăng cường chiến lược tiếp thị và quảng cáo.

Kết quả phân tích giúp doanh nghiệp quản lý tồn kho hiệu quả hơn, đảm bảo tồn kho luôn ở mức hợp lý và không bị ứ đọng xe cũ.

**Quá trình hành động:**

Điều chỉnh giá bán: Sử dụng kết quả phân tích để điều chỉnh giá bán xe theo tình trạng, số km đã đi và thời điểm bán để tối ưu hóa lợi nhuận.

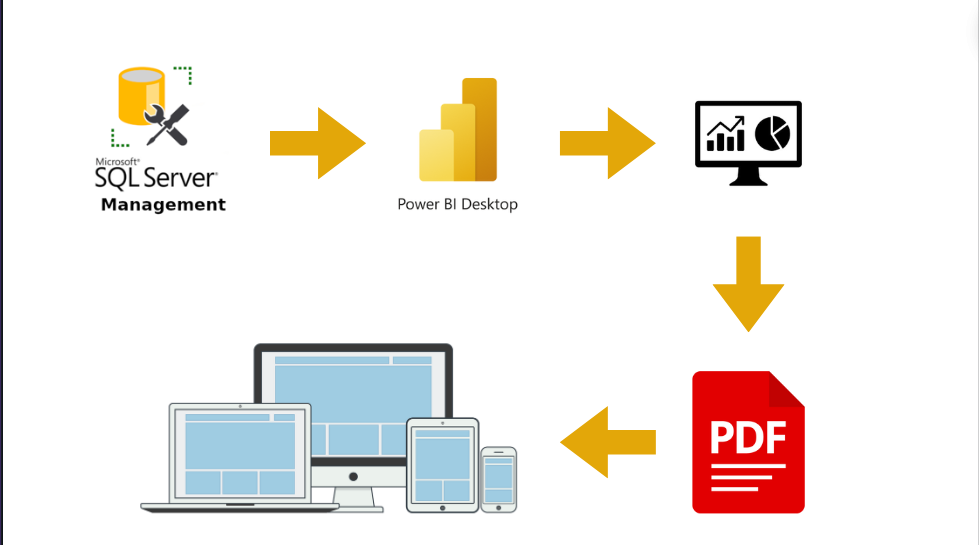
Tối ưu hóa chiến lược tiếp thị: Tập trung vào các kênh tiếp thị và thời điểm có nhu cầu cao, như mùa xuân và mùa thu, để tăng cường hiệu quả bán hàng.

**Kết luận:**

Kết quả phân tích này cung cấp những hiểu biết quan trọng giúp doanh nghiệp nắm bắt xu hướng thị trường, điều chỉnh chiến lược và cải thiện hiệu quả hoạt động kinh doanh trong lĩnh vực mua bán xe ô tô cũ.

## 2.3 Kiến trúc hệ thống

### 2.3.1 Kiến trúc



### 2.3.2 Giải thích

**Máy chủ lưu trữ (SQL Server):**

Đảm bảo lưu trữ an toàn, hiệu quả và tiết kiệm với bộ dữ liệu thống kê mua bán xe cũ

**Power BI Desktop:**  
Kết nối, xử lý, phân tích và trực quan hóa dữ liệu từ bộ dữ liệu đã thu thập

Làm sạch, mô hình dữ liệu, tạo bảng và đưa ra các báo cáo có ích

Xuất báo và gửi về khách hàng.

**Khách hàng:**

Truy cập và sử dụng các báo cáo, dashboard bằng bản tải về trên thiết bị di động cá nhân.

Đánh giá chất lượng và phản hồi chất lượng.

## 2.4 Giải thích về bộ dữ liệu khách hàng

### 2.4.1 Các khái niệm

| ***Trường dữ liệu*** | ***Khái niệm*** |
| --- | --- |
| Year: Năm sản xuất của xe | Đây là năm mà chiếc xe được sản xuất. Thông tin này giúp xác định tuổi của xe, từ đó có thể suy ra một phần giá trị hiện tại và mức độ khấu hao của xe. Thường các xe mới hơn sẽ có giá trị cao hơn. |
| Make: Hãng sản xuất xe | Tên của công ty sản xuất xe, ví dụ như Toyota, Ford, Honda. Hãng sản xuất có thể ảnh hưởng đến độ tin cậy, chất lượng, và giá trị thị trường của xe. Một số hãng nổi tiếng về chất lượng và bền bỉ, trong khi những hãng khác có thể có giá trị bán lại thấp hơn. |
| Model: Mẫu xe | Tên của mẫu xe cụ thể được sản xuất bởi hãng xe, ví dụ như Camry, Mustang, Civic. Mẫu xe cung cấp thông tin chi tiết hơn về dòng sản phẩm cụ thể. Một số mẫu xe có thể được ưa chuộng hơn do tính năng, hiệu suất hoặc thiết kế. |
| Trim: Phiên bản hoặc biến thể của mẫu xe | Các phiên bản hoặc biến thể khác nhau của cùng một mẫu xe, thường có sự khác biệt về tính năng, trang bị, và giá cả. Trim có thể ảnh hưởng lớn đến giá trị và sự hấp dẫn của xe. |
| Body: Loại thân xe | Loại thân xe, ví dụ như SUV, Sedan, Coupe, Hatchback, Convertible, Truck, Minivan. Loại thân xe ảnh hưởng đến mục đích sử dụng và sở thích của người dùng. |
| Transmission: Loại hộp số của xe | Loại hộp số của xe, có thể là tự động (Automatic) hoặc số sàn (Manual). Loại hộp số có thể ảnh hưởng đến hiệu suất, trải nghiệm lái và tiêu thụ nhiên liệu của xe. |
| VIN: Mã số nhận dạng xe (Vehicle Identification Number - VIN) | VIN là mã số duy nhất cho mỗi chiếc xe, chứa thông tin về nhà sản xuất, đặc điểm của xe và năm sản xuất. VIN rất quan trọng cho việc tra cứu lịch sử xe và xác minh tính hợp pháp. |
| State: Bang nơi xe được đăng ký | Bang hoặc khu vực nơi chiếc xe được đăng ký. Thông tin này có thể liên quan đến điều kiện thời tiết, luật pháp và yêu cầu bảo trì của khu vực đó. |
| Condition: Tình trạng của xe | Tình trạng của xe, thường được đánh giá theo thang điểm, ví dụ từ 1 đến 5 hoặc từ 1 đến 10, để xác định mức độ mới/cũ và chất lượng của xe. Xe ở tình trạng tốt hơn thường có giá trị cao hơn và ít cần sửa chữa hơn. |
| Odometer: Số km đã đi được | Số km mà xe đã đi được, đo bởi công tơ mét. Thông tin này giúp xác định mức độ sử dụng của xe. Xe đã đi ít km thường có giá trị cao hơn và có thể ít bị mài mòn hơn. |
| Color: Màu sắc bên ngoài của xe | Màu sơn bên ngoài của xe, ví dụ như đỏ, đen, trắng, xanh. Màu sắc của xe có thể ảnh hưởng đến sở thích cá nhân và giá trị thẩm mỹ. |
| Interior: Màu sắc nội thất của xe | Màu sắc của nội thất bên trong xe, ví dụ như màu be, đen, xám. Sự kết hợp màu sắc nội thất và ngoại thất có thể ảnh hưởng đến giá trị xe. |
| Seller: Tên người bán hoặc đại lý bán xe | Tên của người bán hoặc đại lý bán xe. Thông tin này có thể hữu ích cho việc liên hệ và xác minh thông tin xe. |
| MMR: Giá thị trường hiện tại của xe (Manheim Market Report) | Giá trị thị trường hiện tại của xe dựa trên báo cáo từ Manheim Market Report, một trong những nguồn thông tin đáng tin cậy về giá xe đã qua sử dụng. |
| Selling Price: Giá bán của xe | Giá mà xe được bán. Thông tin này cho biết mức giá giao dịch thực tế của xe. |
| Sale Date: Ngày bán xe | Ngày mà xe được bán. Thông tin này giúp xác định thời điểm giao dịch xe diễn ra. |

### 2.4.2 Các trường dữ liệu

| ***Các trường dữ liệu*** | ***Giải thích*** |
| --- | --- |
| Year | Năm sản xuất xe |
| Make | Hãng sản xuất hoặc thương hiệu của xe |
| Model | Mẫu xe cụ thể |
| Trim | Phiên bản hoặc biến thể của mẫu xe |
| Body | Kiểu dáng xe |
| Transmission | Loại hộp số |
| Vin | Mã số nhận dạng xe, là mã duy nhất cho mỗi chiếc xe |
| State | Nơi xe được đăng ký hoặc định vị. |
| Condition | Mức độ tình trạng của xe được biểu thị bằng số. |
| Odometer | Số km đã đi của xe |
| Color | Màu ngoại thất của xe |
| Interior | Màu nội thất của xe |
| Seller | Cửa hàng hoặc người đã bán xe |
| Mmr | Giá trị ước tính của xe theo báo cáo thị trường Manheim |
| Selling Price | Giá bán thực tế của xe |
| Saledate | Ngày và giờ xe được bán |

# 3 Làm sạch và chuyển đổi dữ liệu

## 3.1 Chuẩn bị dữ liệu

| ***Khái niệm*** | ***Chi tiết*** |
| --- | --- |
| Chuẩn bị dữ liệu | * Quá trình làm sạch và chuyển đổi dữ liệu thô trước khi xử lý và phân tích. |
| Quá trình chuẩn bị dữ liệu | * Chuẩn hóa các định dạng dữ liệu * Kết hợp các bộ dữ liệu để gia tăng độ phong phú * Loại bỏ giá trị ngoại lai (outliers) |
| Mục tiêu | * Phát hiện các lỗi trước khi xử lí * Gia tăng chất lượng dữ liệu * Đẩy nhanh tốc độ và hiệu quả phân tích dữ liệu * Hỗ trợ quyết định quan trọng |

### 3.1.1 Giải pháp lưu trữ dữ liệu

**Lưu trữ trên SQL Server**

| ***Ưu điểm*** | ***Nhược điểm*** |
| --- | --- |
| Hiệu suất cao | **Yêu cầu phần cứng mạnh**: Cần phần cứng đủ mạnh để tối ưu hóa hiệu suất, đặc biệt với khối lượng dữ liệu lớn. |
| Khả năng mở rộng | **Quản lý phức tạp**: Cần quản lý và bảo trì cơ sở dữ liệu định kỳ, điều này có thể phức tạp và tốn thời gian. |
| Tính năng bảo mật mạnh mẽ | **Tương thích phần mềm**: Một số ứng dụng và công cụ bên ngoài có thể không hoàn toàn tương thích với SQL Server. |
| Hỗ trợ báo cáo và phân tích | **Khả năng tùy biến**: Mặc dù mạnh mẽ, nhưng một số tính năng có thể không dễ tùy biến theo nhu cầu đặc thù của doanh nghiệp. |
| Khả năng tích hợp | **Cần kiến thức chuyên môn**: Để tận dụng tối đa các tính năng và khả năng của SQL Server, cần có kiến thức chuyên môn. |
| Tính năng sao lưu và phục hồi | **Yêu cầu cập nhật thường xuyên:** Cần phải cập nhật thường xuyên để bảo mật và hiệu suất tối ưu. |

### 3.1.2 Giải pháp phân bố dữ liệu

#### 3.1.2.1 Ý nghĩa việc phân bố dữ liệu

| ***Ý nghĩa của việc phân bố dữ liệu*** | ***Chi tiết*** |
| --- | --- |
| Gia tăng độ phong phú của dữ liệu | Cải thiện định dạng và giá trị của dữ liệu, kết nối với thông tin liên quan khác để cung cấp hiểu biết sâu sắc hơn. |
| Lưu trữ dữ liệu | Sau khi chuẩn bị, dữ liệu được lưu trữ hoặc chuyển vào các ứng dụng như Power BI để xử lý và phân tích. |
| Giảm tải cho việc báo cáo | Chạy các truy vấn trên các cơ sở dữ liệu nhân bản giúp giải phóng tài nguyên của cơ sở dữ liệu chính, làm tăng hiệu năng hệ thống. |
| Bảo vệ dữ liệu khỏi các lỗi quản trị | Sử dụng cơ sở dữ liệu nhân bản để khôi phục dữ liệu trong trường hợp xảy ra lỗi trong quá trình thao tác. |
| Bảo vệ dữ liệu trước lỗi người dùng | Giảm thiểu các lỗi do người dùng gây ra, như xóa bảng hoặc cập nhật dữ liệu sai. |
| Quản lý các dữ liệu thử nghiệm | Tạo và quản lý cơ sở dữ liệu thử nghiệm hữu ích cho các giao thức thử nghiệm liên tục. |
| Tăng độ tin cậy cho hệ thống | Sử dụng bản sao dữ liệu để duy trì tính sẵn sàng của hệ thống trong trường hợp dữ liệu chính bị lỗi. |

#### 3.1.2.2 Trình bày cách phân bố dữ liệu

| ***Cách thực hiện*** | ***Các bước*** | ***Chi tiết*** |
| --- | --- | --- |
| ***Sử Dụng Trình Wizard*** | 1. Chọn cơ sở dữ liệu SAMPLE | Mở SQL Server Management Studio (SSMS), kết nối với server chứa cơ sở dữ liệu SAMPLE. |
| 2. Chọn server chứa SAMPLE và server đích | Nhấp chuột phải vào cơ sở dữ liệu SAMPLE, chọn "Tasks" > "Copy Database..." để mở trình Wizard. |
| 3. Chọn phương thức chuyển dữ liệu | Chọn giữa phương thức **Online** hoặc **Offline** tùy theo yêu cầu. |
| 4. Đặt tên cho cơ sở dữ liệu mới | Nhập tên "Staging" cho cơ sở dữ liệu đích |
| 5. Chọn các tùy chọn khác | Cấu hình tùy chọn lưu trữ log, thời điểm bắt đầu và hoàn tất quá trình sao chép. |
| 6. Hoàn thành quá trình và kiểm tra kết quả | Nhấp "Finish" để bắt đầu sao chép, sau đó kiểm tra cơ sở dữ liệu Staging để đảm bảo dữ liệu đã được sao chép thành công. |
| ***Sử Dụng Script*** | 1. Viết script để sao lưu SAMPLE database | Viết và thực thi script sao lưu cơ sở dữ liệu SAMPLE:  BACKUP DATABASE SAMPLE TODISK='C:\Backup\SAMPLE.bak' |
| 2. Điều chỉnh đường dẫn sao lưu và cài đặt SQL Server | Đảm bảo đường dẫn sao lưu tồn tại và điều chỉnh theo cấu hình của hệ thống và SQL Server. |
| 3. Thực thi script để sao lưu và khôi phục cơ sở dữ liệu với tên Staging | Viết và thực thi script khôi phục cơ sở dữ liệu từ bản sao lưu:  RESTORE DATABASE Staging FROM DISK = 'C:\Backup\SAMPLE.bak' |

## 3.2 Làm sạch dữ liệu

### 3.2.1 Các vấn đề ảnh hưởng tới dữ liệu

| ***Vấn đề ảnh hưởng tới dữ liệu*** | ***Chi tiết*** |
| --- | --- |
| Dữ liệu bị trùng lặp (duplicate data) | Có hai hoặc nhiều record có thông tin giống hệt nhau. Điều này có thể gây ra sai lệch về số lượng hàng tồn kho, trùng lặp tài sản thế chấp |
| Dữ liệu bị mâu thuẫn (conflict data) | Khi cùng một record nhưng lại tạo ra hai dòng dữ liệu với hai thuộc tính khác nhau, dữ liệu lúc này đang ở trạng thái bị mâu thuẫn |
| Dữ liệu không đầy đủ (incomplete data) | Dữ liệu bị thiếu một số thuộc tính |
| Dữ liệu không hợp lệ (invalid data) | Một thuộc tính dữ liệu không phù hợp với các tiêu chuẩn dữ liệu đã đặt ra. Ví dụ: số điện thoại được lưu với 9 chữ số thay vì 10 chữ số |

### 3.2.2 Các tiêu chí đánh giá chất lượng dữ liệu

| **Tiêu chí đánh giá** | **Áp dụng trong dự án(Có/Không)** |
| --- | --- |
| *Tính chuẩn xác (validity)* | *Có* |
| *Tính chính xác (accuracy):* | *Có* |
| *Tính đầy đủ (completeness)* | *Có* |
| *Tính nhất quán (consistency)* | *Có* |
| *Tính đồng nhất (uniformity)* | *Có* |
| *Tính minh bạch của nguồn dữ liệu (traceability)* | *Không* |
| *Tính kịp thời (timeliness):* | *Không* |

### 3.2.3 Các bước làm sạch dữ liệu

#### 3.2.3.1 Trình bày các bước làm sạch

| ***Bước xác định dữ liệu*** | ***Sửa chữa các mục bị lỗi*** |
| --- | --- |
| Dữ liệu trùng lặp | => Loại bỏ thông tin trùng lặp |
| Dữ liệu không liên quan | => Xác định các trường quan trọng đối với phân tích và loại bỏ dữ liệu không liên quan khỏi phân tích |
| Dữ liệu ngoại lai | => Dữ liệu ngoại lai có thể ảnh hưởng đáng kể đến hiệu suất của mô hình, vậy nên cần phải xác định các dữ liệu ngoại lai và tiến hành biện pháp thích hợp |
| Dữ liệu bị thiếu | => Gắn cờ và loại bỏ hoặc thay thế dữ liệu bị thiếu |
| Lỗi cấu trúc | => Sửa lỗi đánh máy và các điểm không nhất quán khác, đồng thời khiến dữ liệu tuân theo mẫu hoặc quy ước chung |

#### 3.2.3.2 Trình bày các bước làm sạch trong phạm vi dự án

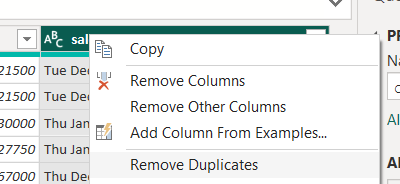
**Xử lý các dữ liệu bị thiếu:**

Thay thế các giá trị bị thiếu trong các cột: make, model, trim, body, transmission, condition, odometer, color, interior, mmr, và sellingprice thành “Chưa cập nhật.

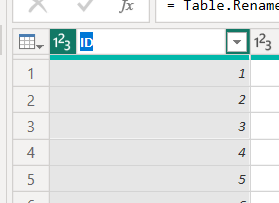
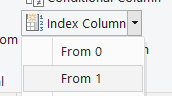


**Loại bỏ dữ liệu trùng lặp:**

Loại bỏ dữ liệu trùng lặp ở các cột vin , saledate



Thêm cột mới ID để đánh Số thứ tự cho từng dòng trong bảng.



## 3.3 Chuyển đổi dữ liệu

### 3.3.1 Các trường hợp cần chuyển đổi

| **Các trường cần chuyển đổi** | **Chi tiết** |
| --- | --- |
| Price Different | Chênh lệch giữa giá bán và MMR |
| Đánh giá Hiệu Quả Kinh Doanh | Lợi nhuận: Chênh lệch giữa giá bán và MMR có thể giúp xác định lợi nhuận của từng giao dịch. Nếu giá bán cao hơn MMR, giao dịch có khả năng mang lại lợi nhuận cao hơn cho người bán.  Hiệu quả chiến lược giá: Giúp đánh giá liệu chiến lược định giá của bạn có hiệu quả hay không. |
| Phân Tích Thị Trường | Xu hướng giá: Phân tích chênh lệch giữa giá bán và MMR trên nhiều giao dịch có thể giúp nhận ra các xu hướng giá cả trên thị trường. Ví dụ, nếu giá bán thường cao hơn MMR, có thể cho thấy một thị trường mạnh.  Định giá theo khu vực: Có thể phân tích dữ liệu theo khu vực địa lý để xác định nơi nào có giá bán cao hơn hoặc thấp hơn MMR. |
| Đánh Giá Giá Trị Xe | Nếu một chiếc xe bán với giá cao hơn nhiều so với MMR, điều này có thể chỉ ra rằng chiếc xe đó có các đặc điểm hoặc tình trạng tốt hơn so với các xe tương tự trên thị trường. |

### 3.3.2 Các kỹ thuật chuyển đổi

**Sử dụng Power bi để tiến hành chuyển đổi dưới đấy là các bước để thực hiện.**

Cột tính toán về giá trị chênh lệch giữa giá bán và MMR:

Tên cột mới: Price Difference

Công thức: Price Difference = [sellingprice] - [mmr]

Chuẩn hóa dữ liệu ngày (Date)

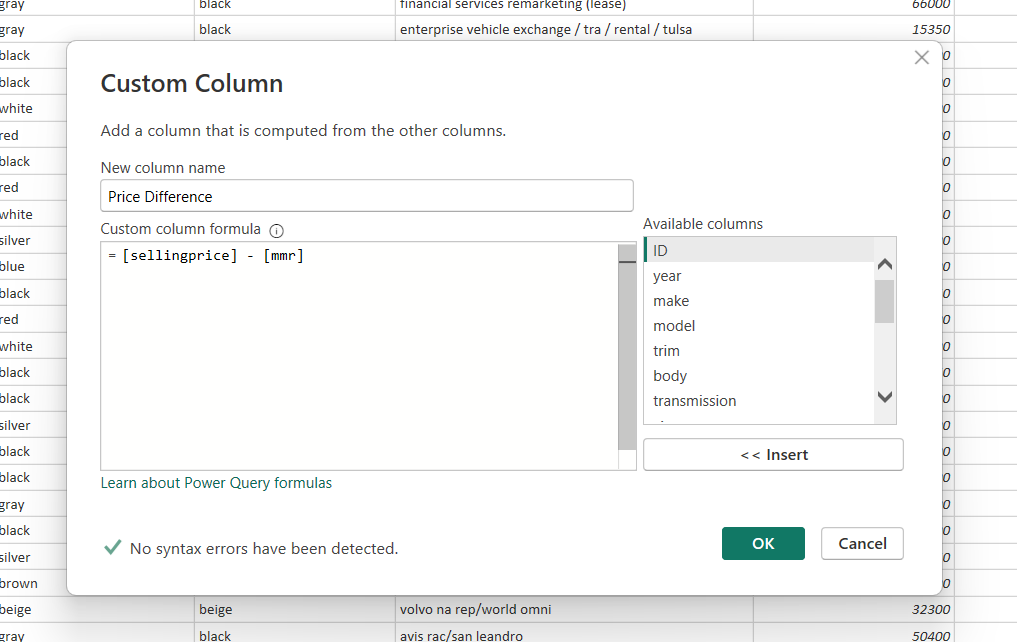
Chuyển đổi cột ngày sang định dạng ngày chuẩn:

Chọn cột saledate.

Trong tab Transform, chọn Data Type > Date/Time.

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

Tạo cột tính toán mới Price Difference



# 4 Xử lý dữ liệu

## 4.1 Chuẩn hóa dữ liệu

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

Chọn Load table car\_prices để bắt đầu xử lý.

Rename car\_prices thành [NGUON]

Chuyển DataType [mmr], [sellingprice] thành fixed decimal number.

| ***Trước*** | ***Sau*** |
| --- | --- |
|  |  |

Chuyển DataType [condition], [odometer] thành decimal number

| ***Trước*** | ***Sau*** |
| --- | --- |
|  |  |

Đổi tất cả các giá trị null thành “chưa cập nhật

| ***Trước*** | ***Sau*** |
| --- | --- |
|  |  |

## 4.2 Mô hình hóa dữ liệu

### 4.2.1 Các loại mô hình hóa

| **Các loại mô hình** | **Khái niệm** | **Sử dụng trong dự án(có/không)** |
| --- | --- | --- |
| *Conceptual data models (Mô hình Dữ liệu Khái niệm)* | *Mô hình dữ liệu khái niệm ghi nhận các mối quan hệ cấp cao nhất giữa các thực thể khác nhau.* | *Không* |
| *Logical Data Model (Mô hình Dữ liệu Logic)* | *Logical Data Model xác định thông tin theo cấu trúc mà không cần quan tâm về mặt vật lý trong cơ sở dữ liệu.* | *Có* |
| *Physical data models (Mô hình Dữ liệu Vật lý)* | *Mô hình dữ liệu này mô tả cách hệ thống sẽ được triển khai bằng cách sử dụng một hệ thống quản lý cơ sở dữ liệu cụ thể.* | *Có* |

### 4.2.2 Các tiêu chí đánh giá mô hình dữ liệu

| **Tiêu chí đánh giá** | **Áp dụng được trong dự án không(có/không)** | **Ghi chú** |
| --- | --- | --- |
| *Clearness (Tính rõ ràng)* | *Có* | *Mô hình dễ hiểu cho người sử dụng và nhà phát triển.* |
| *Flexibility (Tính linh hoạt)* | *Có* | *Có thể dễ dàng chỉnh sửa và mở rộng khi có yêu cầu mới.* |
| *Performance (Hiệu suất)* | *Có* | *Thiết kế đúng đắn giúp cải thiện tốc độ đọc và ghi dữ liệu.* |
| *Productivity (Năng suất)* | *Có* | *Mô hình dễ làm việc, không tốn nhiều thời gian.* |
| *Traceability (Khả năng truy xuất nguồn gốc)* | *Có* | *Có thể truy xuất nguồn gốc dữ liệu và thông tin hệ thống.* |

### 4.2.3 Trình bày các bước mô hình hóa

***Mô hình hóa dữ liệu car***

| Nhóm dữ liệu theo cột make |  |
| --- | --- |
| Thêm cột mới, đánh index, đặt tên là [make\_ID] |  |
| Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích. |  |

***Mô hình hóa dữ liệu Seller***

| Nhóm dữ liệu theo cột seller |  |
| --- | --- |
| Thêm cột mới, đánh index, đặt tên là [seller\_ID] |  |
| Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích. |  |

***Mô hình hóa dữ liệu Transaction***

| Nhóm dữ liệu theo cột vin |  |
| --- | --- |
| Thêm cột mới, đánh index, đặt tên là [vin\_ID] |  |
| Trả lại trạng thái ban đầu cho cột để tiếp tục phân tích. |  |

### 4.2.4 Trình bày các bước tạo bảng dữ liệu

***Tạo bảng car***

| * Tạo bảng Reference từ NGUON * Rename thành car * Giữ lại các cột ([make\_ID], [year], [make], [model], [trim], [body],[transmission], [vin], [condition], [odometer], [color], [interior]). |  |
| --- | --- |
| * *Remove duplicate data.* |  |

***Tạo bảng seller***

| * Tạo bảng Reference từ NGUON * Rename thành seller * Giữ lại các cột ([seller\_ID], [seller], [state]). |  |
| --- | --- |
| * *Remove duplicate data.* |  |

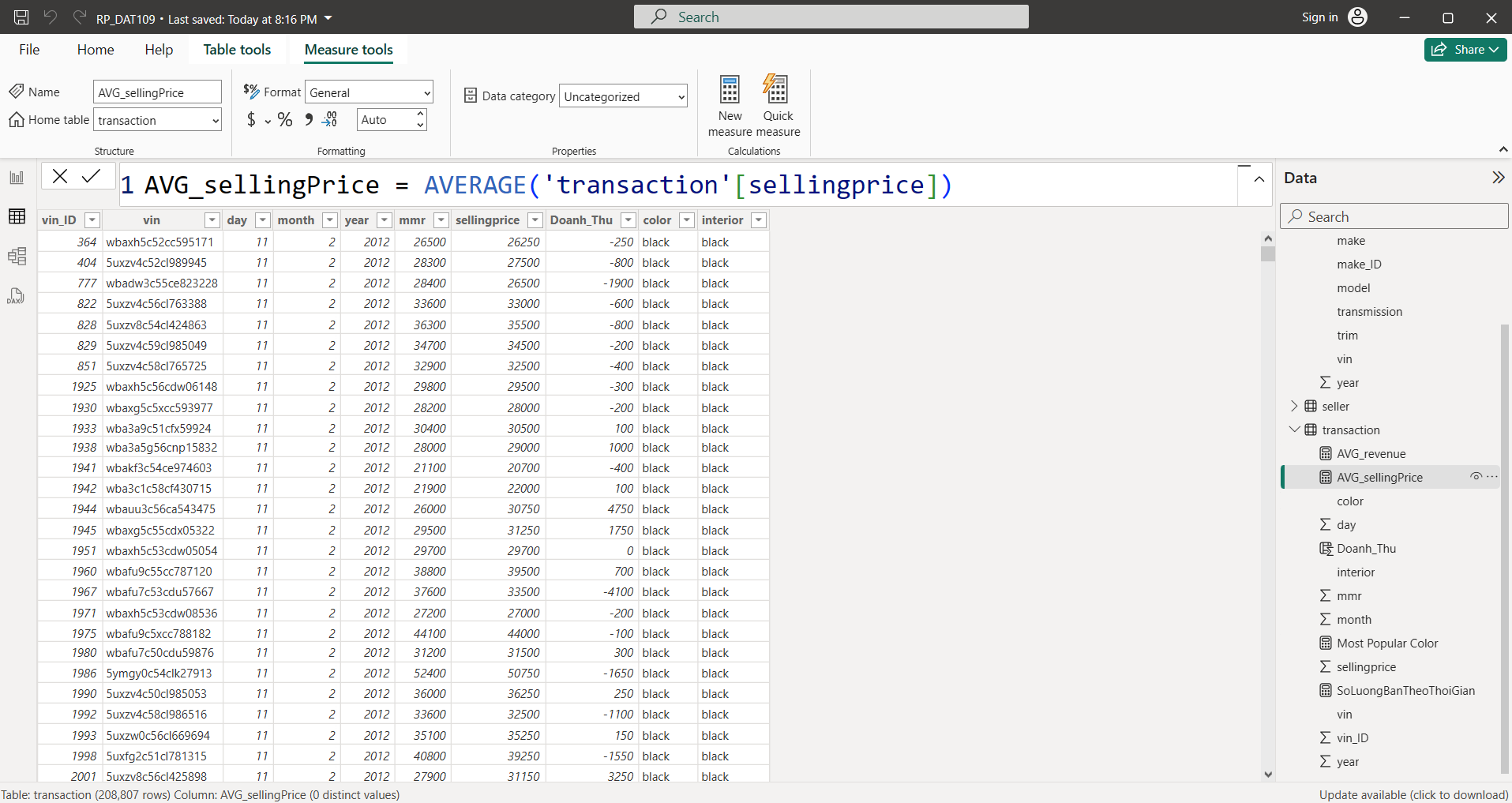
***Tạo bảng Transaction***

| * Tạo bảng Reference từ NGUON * Rename thành Transaction * Giữ lại các cột ([vin\_ID], [vin], [mmr], [sellingprice], [year]). |  |
| --- | --- |
| * *Remove duplicate data.* |  |

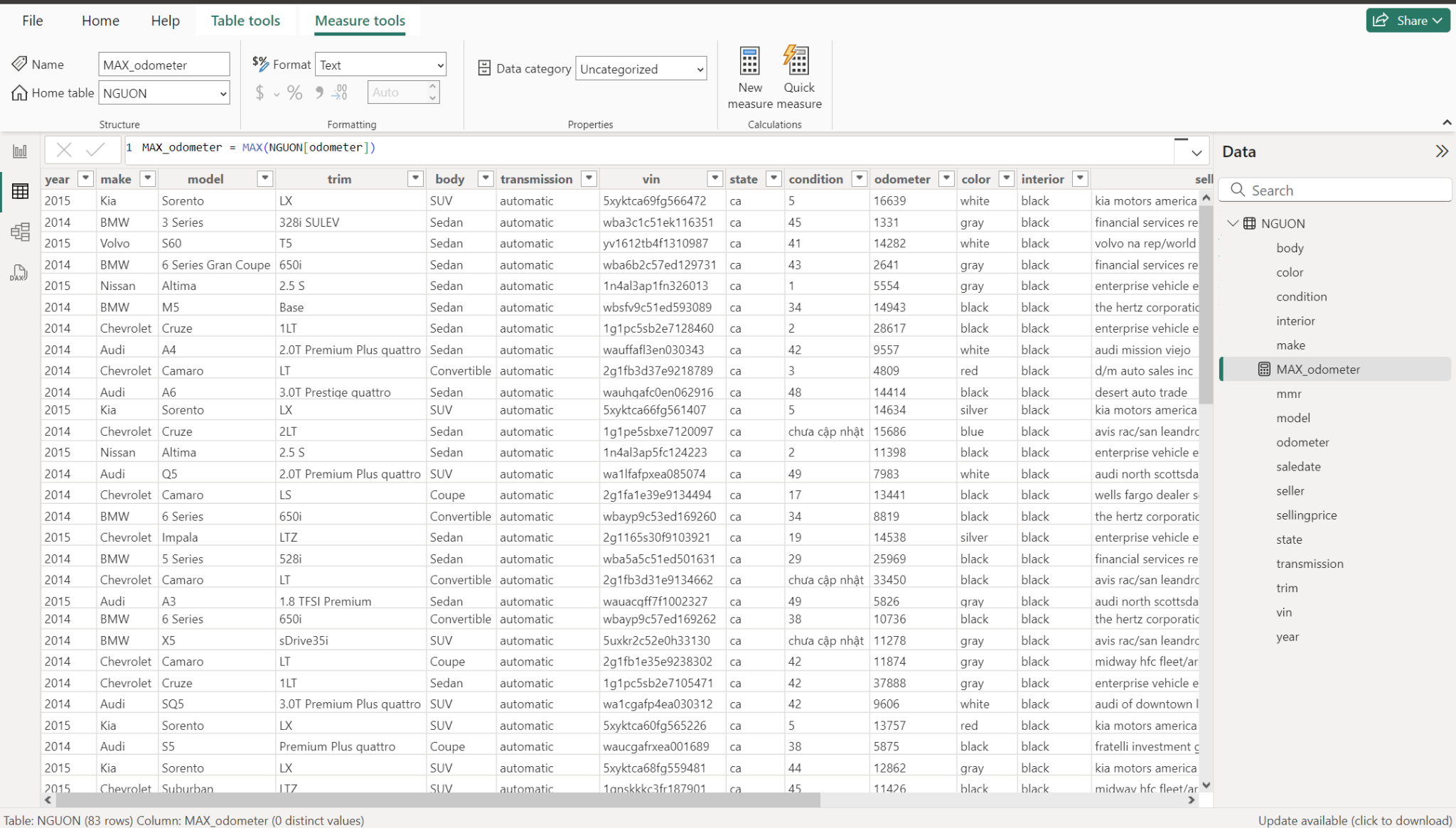
## 4.3 Xử lý dữ liệu DAX

### 4.3.1 Measure

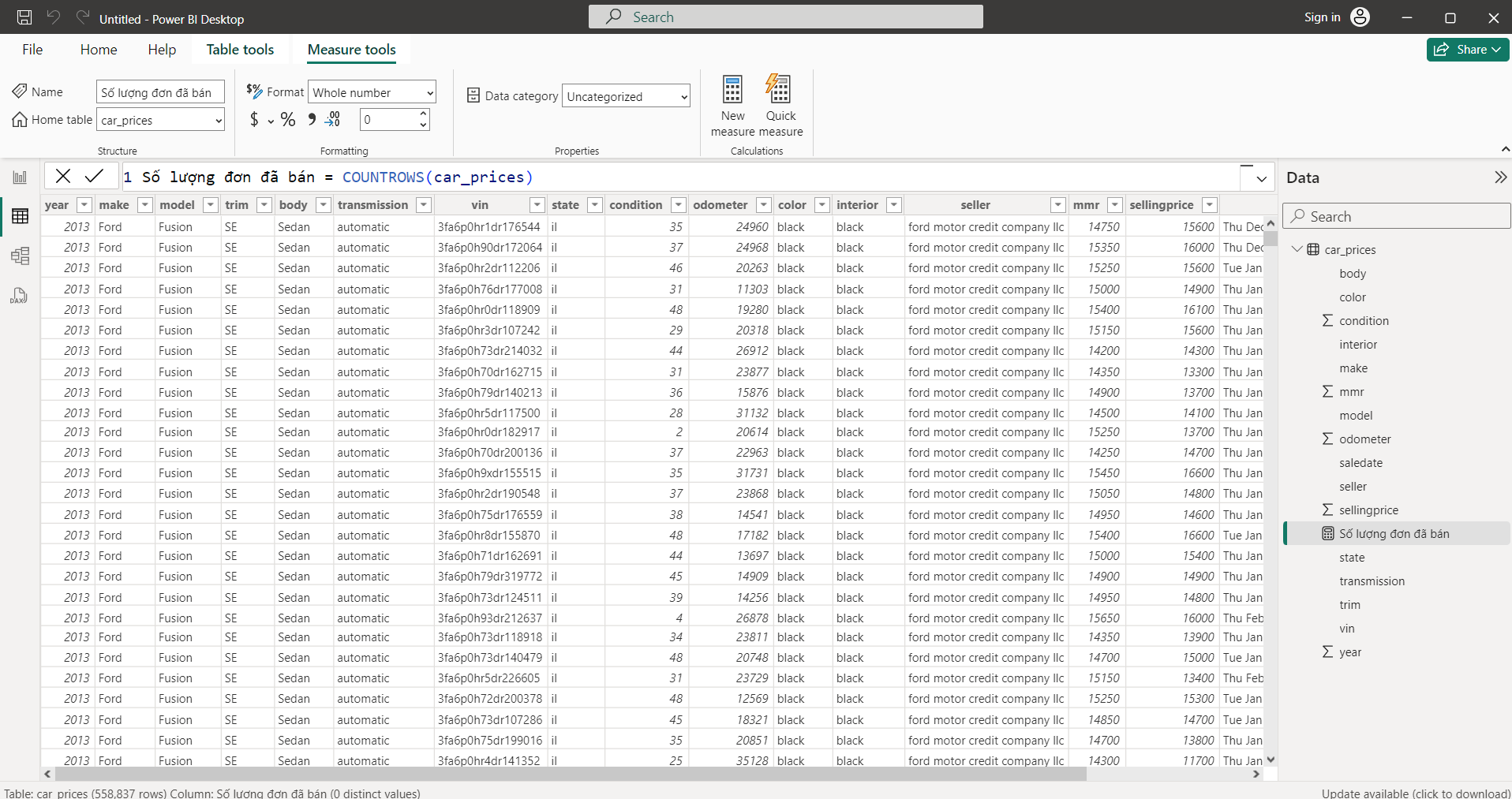
#### 4.3.1.1 Tạo measure cho giá bán trung bình:



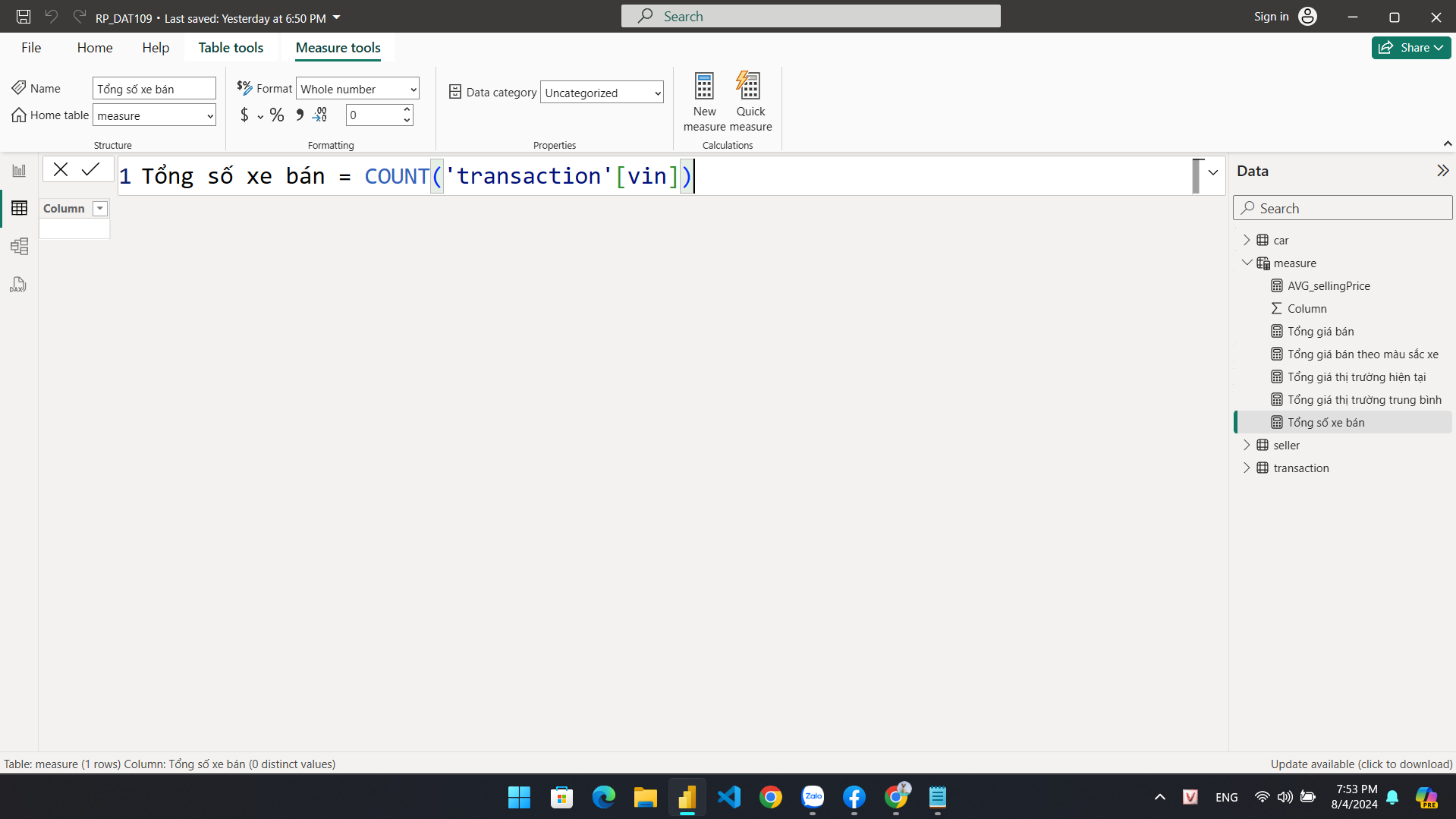
#### 4.3.1.2 Tạo measure cho quãng đường đi lớn nhất:



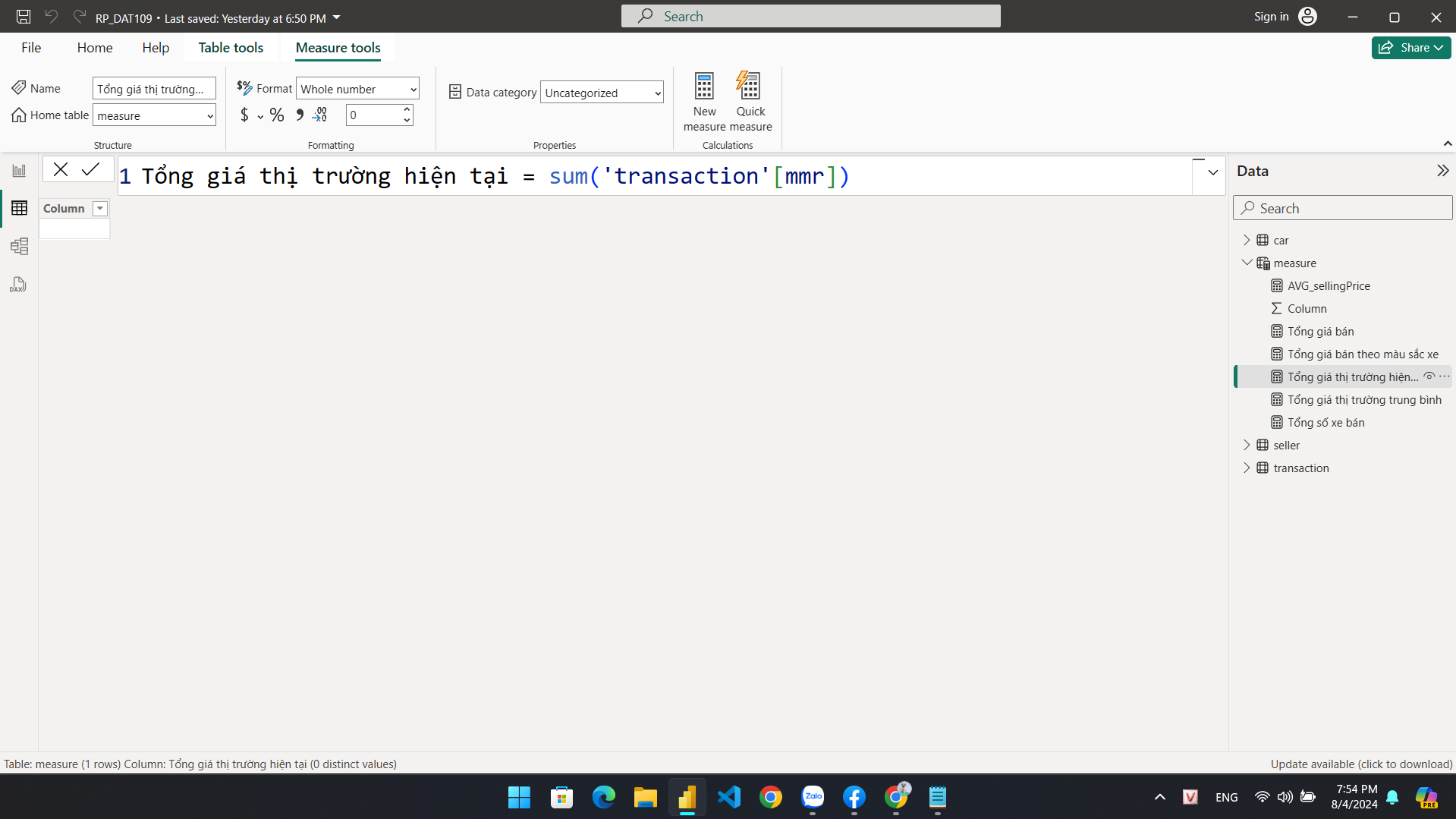
#### 4.3.1.3 Tạo measure cho số lượng đơn đã bán:



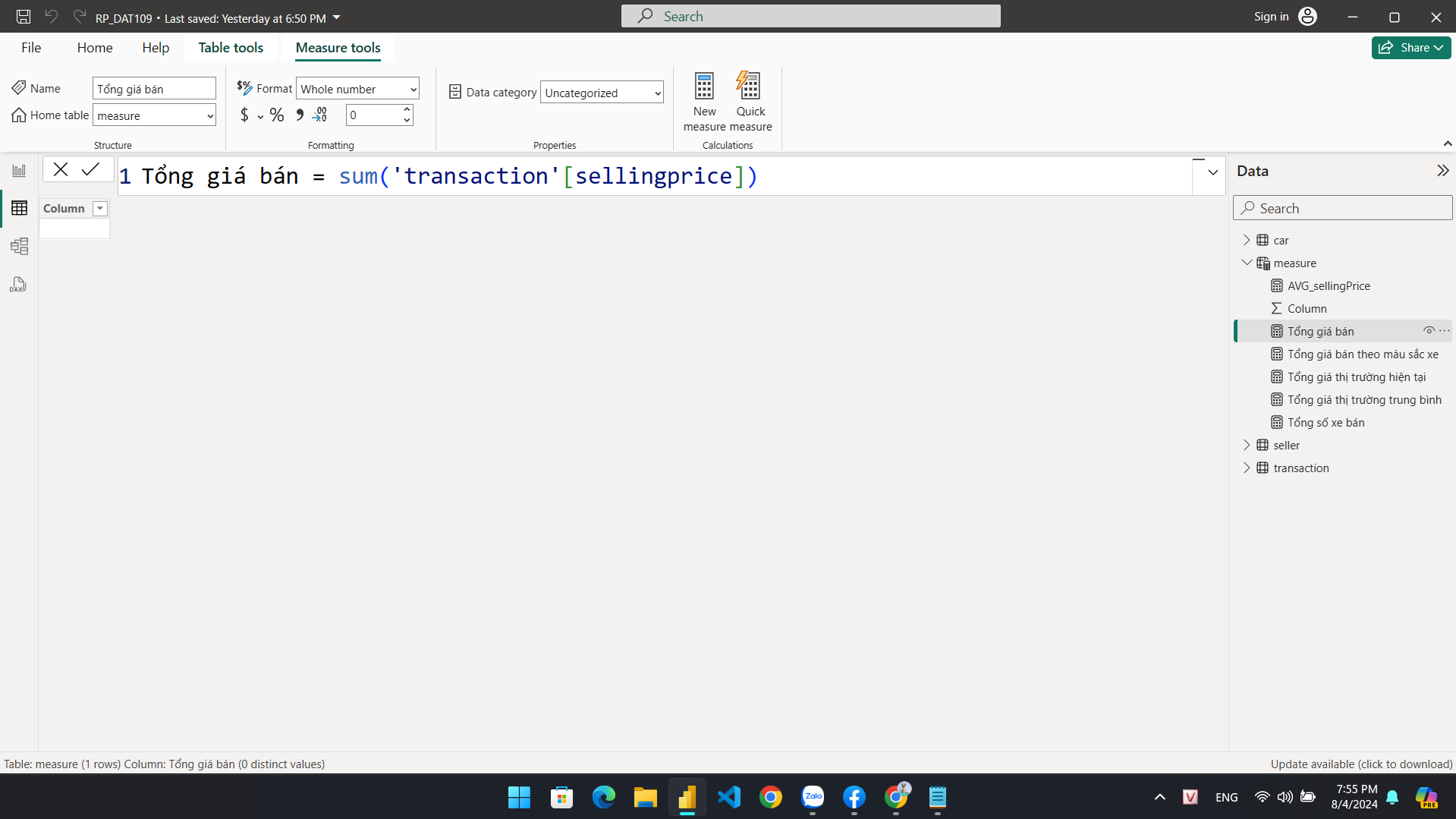
#### 4.3.1.4 Tạo measure cho tổng số xe bán ra



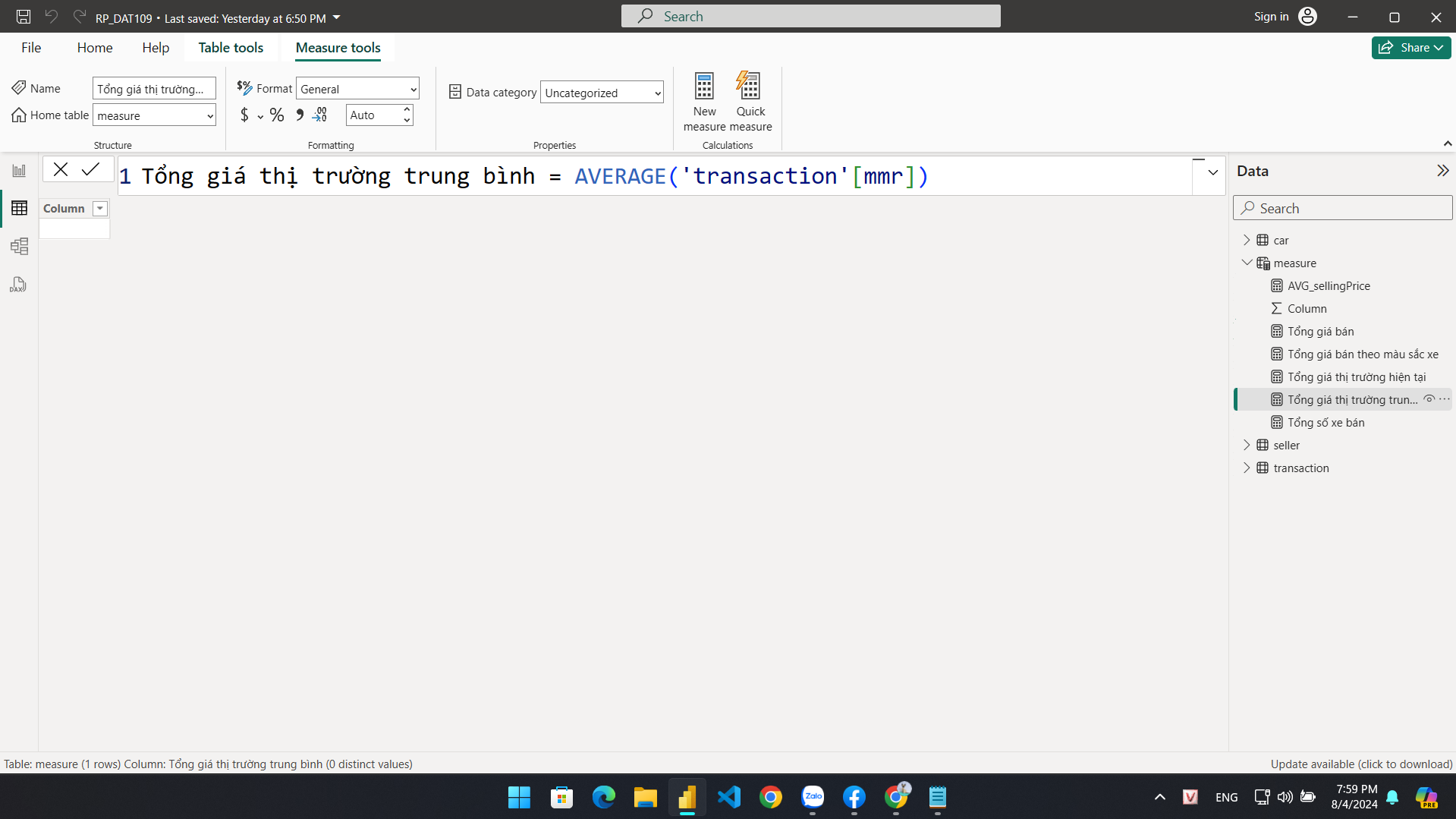
#### 4.3.1.5 Tạo measure cho tổng giá thị trường hiện tại



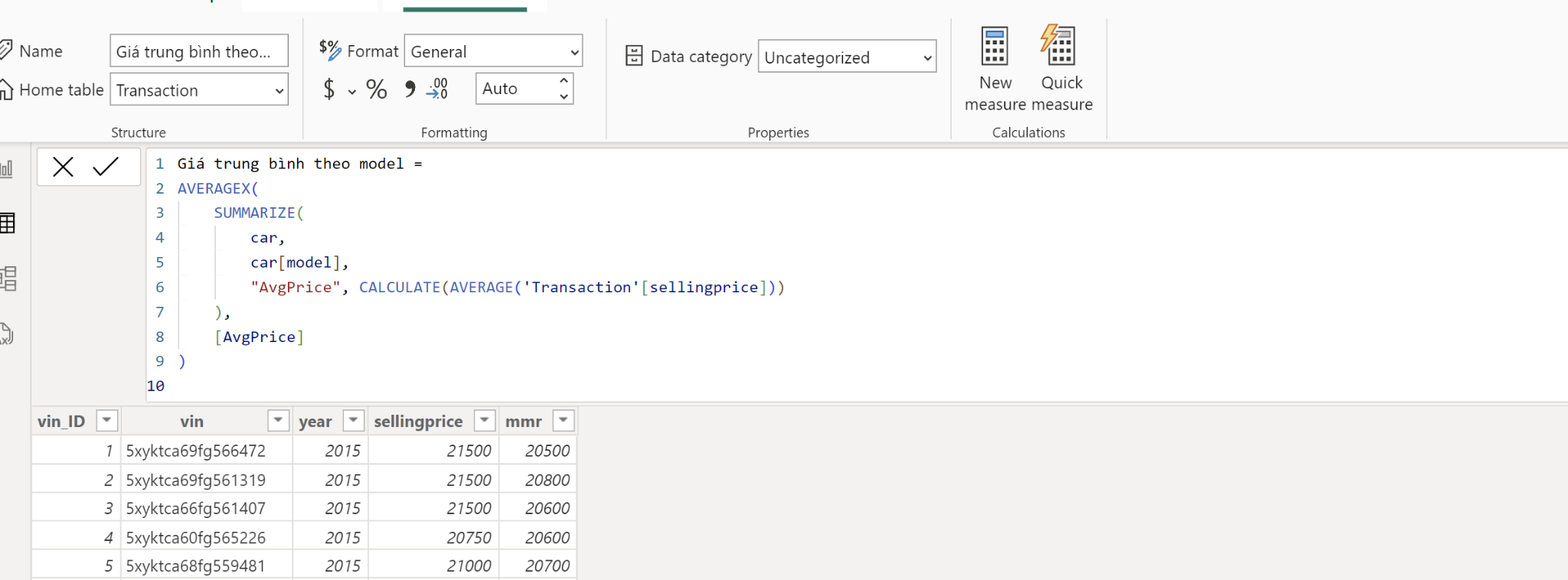
#### 4.3.1.6 Tạo measure cho tổng giá bán ra



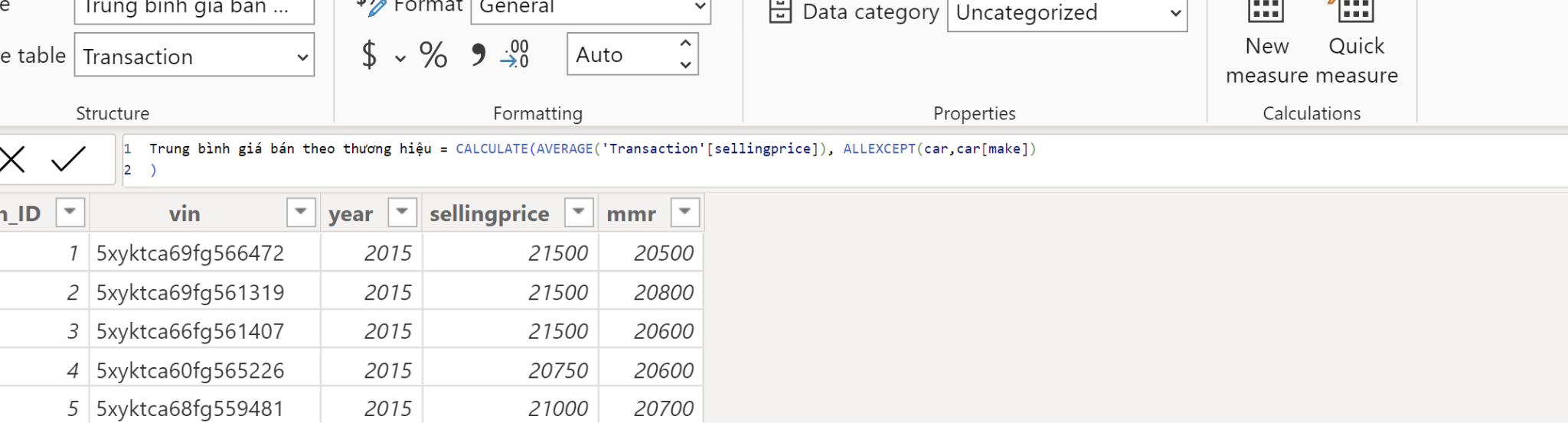
#### 4.3.1.7 Tạo measure cho tổng giá thị trường trung bình



#### 4.3.1.8 Tạo measure cho giá trung bình theo model



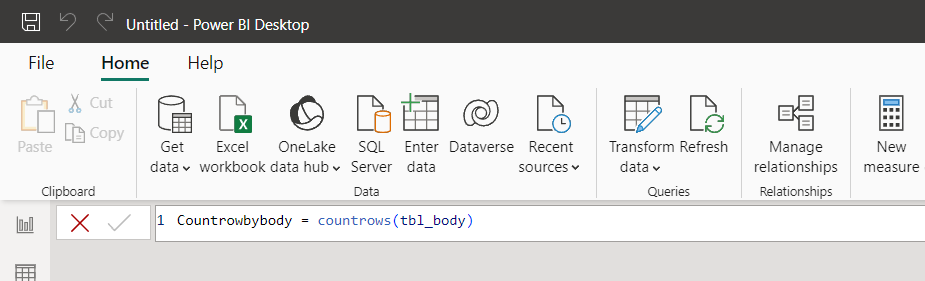
#### 4.3.1.9 Tạo measure cho trung bình giá bán theo thương hiệu



#### 4.3.1.10 Tạo measure cho đếm số lượt bán theo hãng xe.

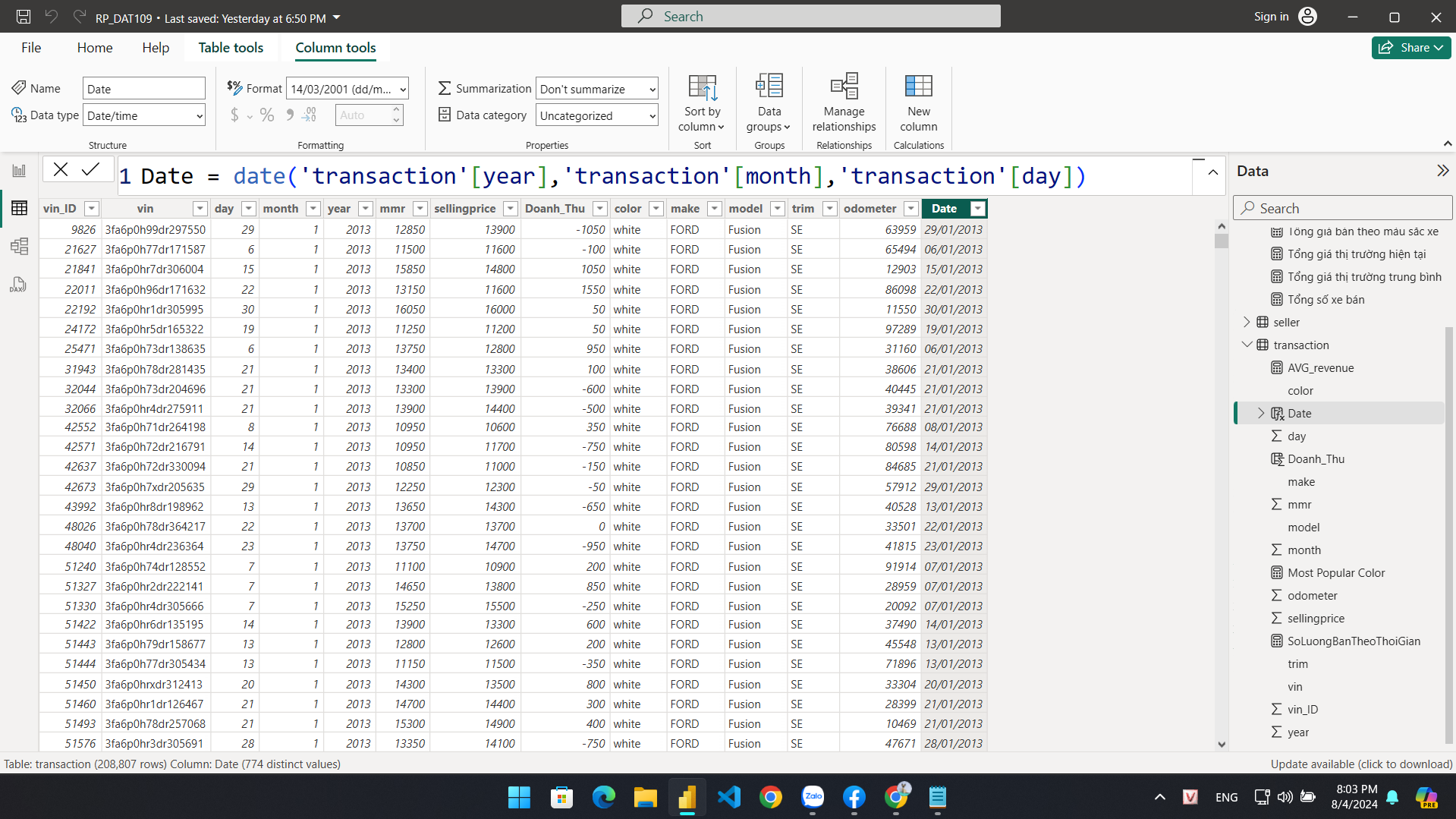


#### 4.3.1.10 Tạo measure cho đếm số lượt bán theo kiểu dáng xe.

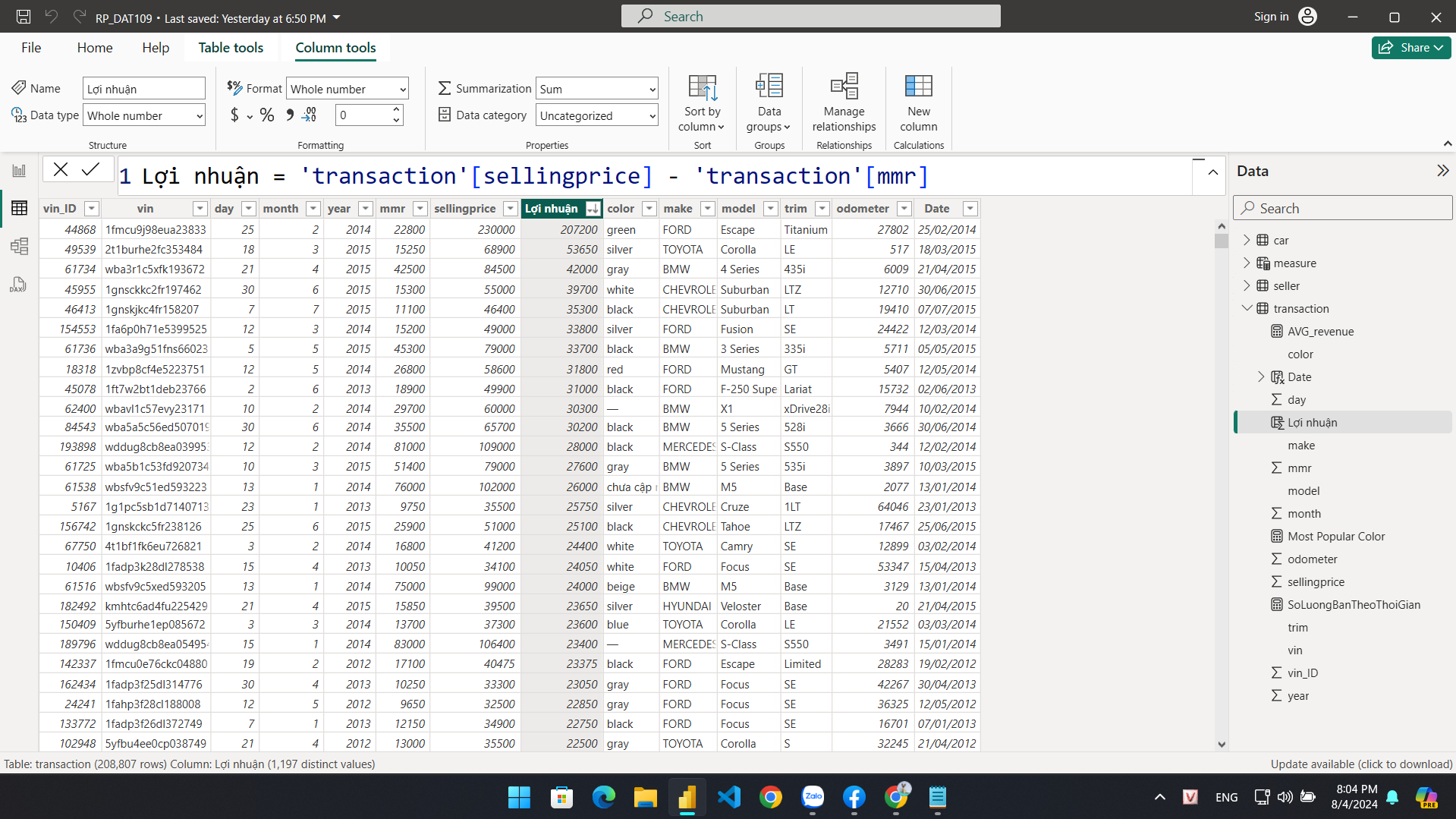


### 4.3.2 Calculated column

#### 4.3.2.1 Tạo cột Date trong bảng transaction

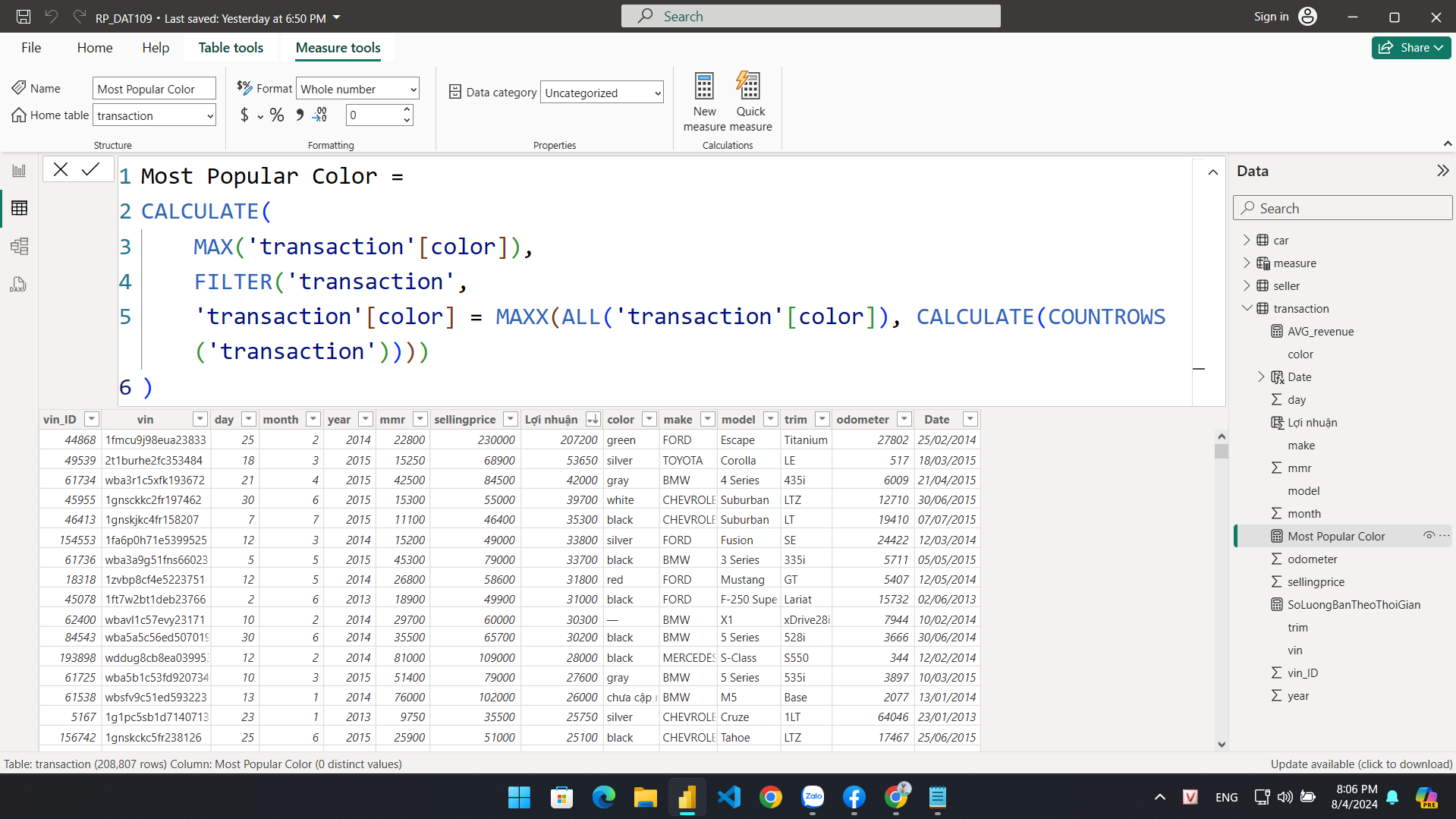


#### 4.3.2.2 Tạo cột lợi nhuận trong bảng transaction



### 4.3.3 Filter

#### 4.3.3.1 Tạo filter màu sắc xe phổ biến nhất



# 5 Trực quan hóa dữ liệu

## 5.1 Các kỹ thuật trực quan hóa

| *Loại trực quan hóa dữ liệu* | *Mục đích* | *Thực hiện trong dự án (Có/Không)* |
| --- | --- | --- |
| **Trực quan hóa dữ liệu chuỗi thời gian** (Temporal Data Visualization) | Trình bày các đối tượng một chiều tuyến tính như đồ thị đường, biểu đồ đường hoặc dòng thời gian | Có |
| **Trực quan hóa dữ liệu phân cấp** (Hierarchical Data Visualization) | Hiển thị các nhóm hoặc tập hợp các mục có liên kết chung với một mục cha | Có |
| **Trực quan hóa dữ liệu mạng** (Network Data Visualization) | Trình bày dữ liệu dưới dạng điểm trên một đồ thị hoặc sử dụng các biểu đồ phức tạp hơn | Không |
| **Trực quan hóa dữ liệu đa chiều** (Multidimensional Data Visualization) | So sánh hai hoặc nhiều yếu tố dữ liệu và biểu thị các phần của tổng thể thuộc từng danh mục | Có |
| **Trực quan hóa dữ liệu không gian địa lý** (Geospatial Data Visualization) | Trình bày dữ liệu liên quan đến các vị trí ngoài đời thực | Không |

## 5.2 Các nguyên tắc trực quan hóa

| *Nguyên tắc* | *Ứng dụng trong dự án* |
| --- | --- |
| Chọn đúng loại biểu đồ | * Sử dụng biểu đồ cột để hiển thị khối lượng giao dịch và tổng giá trị giao dịch. * Sử dụng biểu đồ đường để thể hiện xu hướng giá điều chỉnh theo ngày. * Sử dụng biểu đồ boxplot để biểu diễn biên độ giá điều chỉnh. |
| Không phải tất cả data đều quan trọng | * Chỉ hiển thị những mã chứng khoán tăng và giảm đáng chú ý. * Tập trung vào các thông tin quan trọng như tổng giá trị giao dịch, tổng khối lượng giao dịch, số lượng mã tăng và giảm. |
| Biểu đồ thể hiện đúng tương quan số liệu thực tế | * Các biểu đồ được thiết kế để phản ánh chính xác sự thay đổi của giá cả và khối lượng giao dịch. * Tránh làm sai lệch số liệu thông qua việc giữ tỷ lệ khung hình và tỷ lệ giá trị chính xác. |
| Sử dụng màu sắc hợp lý khi chuyển dữ liệu sang dạng biểu đồ | * Màu sắc được sử dụng để phân biệt rõ ràng các mã chứng khoán khác nhau và tình trạng (tăng, giảm, tham chiếu). * Sử dụng màu sắc để nhấn mạnh các yếu tố quan trọng mà không gây rối mắt cho người xem. |
| Luôn đảm bảo dữ liệu được trình bày một cách đơn giản và hiệu quả | - Trình bày dữ liệu một cách rõ ràng và trực quan để người xem dễ dàng nắm bắt thông tin quan trọng.  - Lược bỏ các yếu tố không cần thiết, giữ cho bảng điều khiển đơn giản và dễ hiểu. |

**Quan trọng nhất trong dự án**

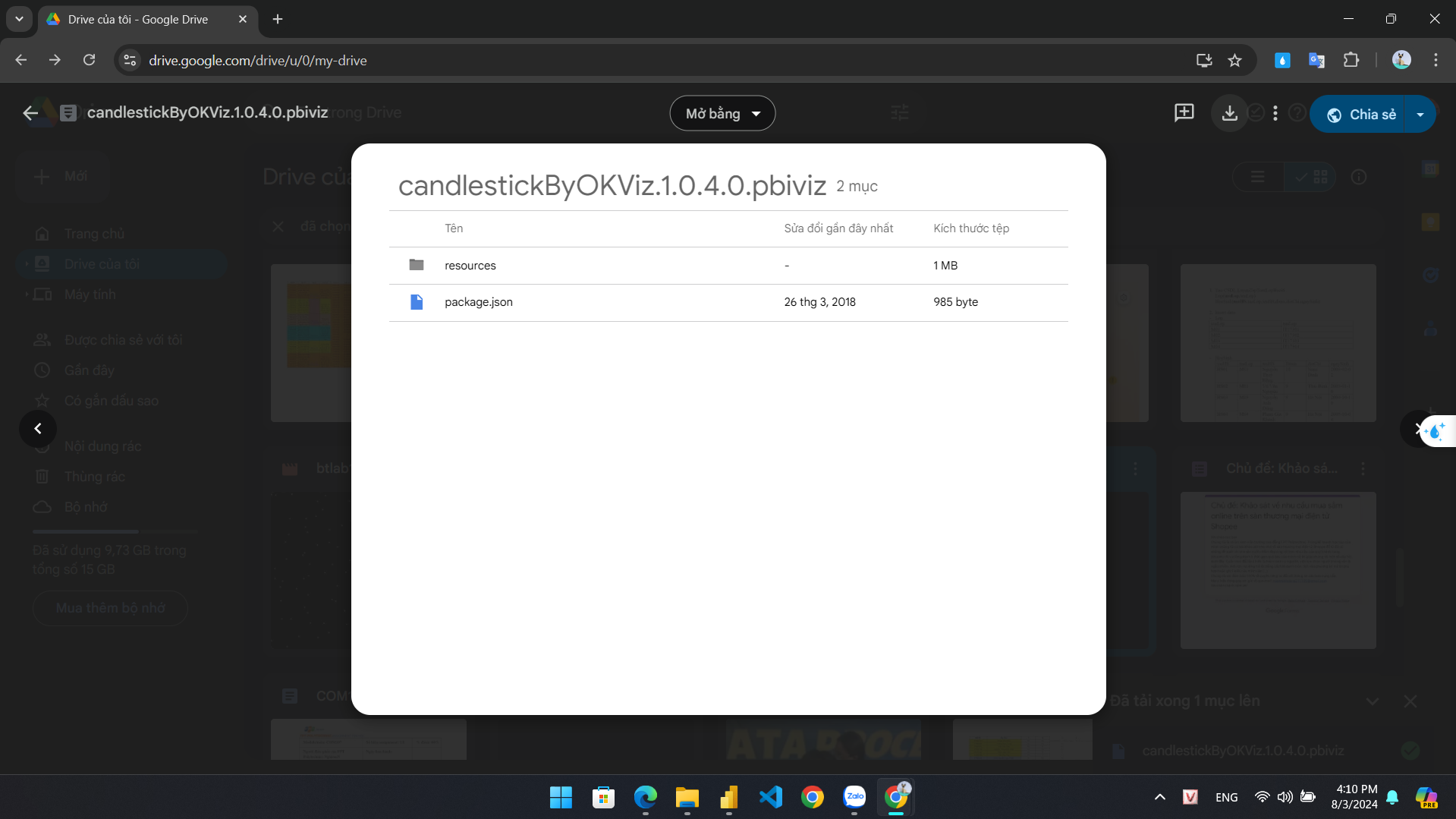
| *Nguyên tắc* | *Lí do quan trọng* | *Chi tiết* |
| --- | --- | --- |
| Chọn đúng loại biểu đồ | Tính rõ ràng và dễ hiểu | * Chọn đúng loại biểu đồ giúp người xem dễ dàng nắm bắt thông tin mà bạn muốn truyền đạt. * Mỗi loại biểu đồ có thế mạnh riêng trong việc biểu diễn một dạng dữ liệu cụ thể. * Ví dụ, biểu đồ cột phù hợp để so sánh các giá trị, biểu đồ đường tốt cho việc hiển thị xu hướng theo thời gian, và biểu đồ boxplot hiệu quả trong việc thể hiện phân bố và biến động của dữ liệu. |
| Tối ưu hóa việc truyền đạt thông tin | * Khi sử dụng loại biểu đồ phù hợp, thông tin sẽ được truyền đạt một cách trực quan và dễ dàng hơn. Điều này giúp người xem nhanh chóng nhận ra các điểm quan trọng và xu hướng trong dữ liệu. |
| Giảm thiểu sự nhầm lẫn | * Việc chọn sai loại biểu đồ có thể dẫn đến hiểu lầm và sai lệch trong việc diễn giải dữ liệu. Điều này có thể gây ra những quyết định không chính xác dựa trên thông tin bị hiểu sai. |
| Tăng hiệu quả phân tích | Khi dữ liệu được biểu diễn một cách rõ ràng và chính xác, người dùng có thể dễ dàng phân tích và đưa ra những kết luận đúng đắn.  Điều này rất quan trọng trong các báo cáo và bảng điều khiển dữ liệu, nơi mà quyết định thường được đưa ra dựa trên các phân tích trực quan. |

## 5.3 Trình bày cách thêm visual mới

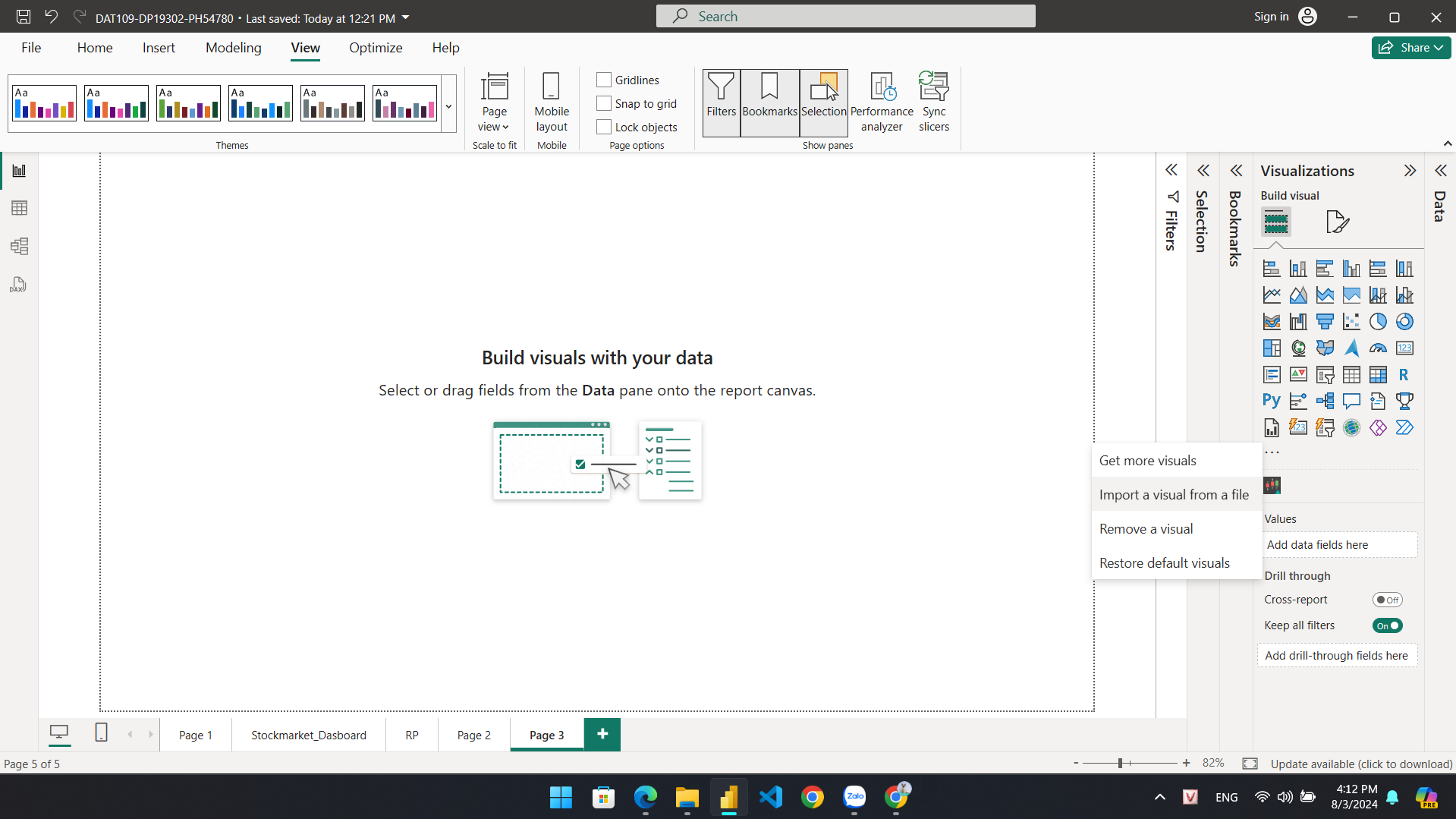
**Các bước thêm một visual mới (Candlestick chart)**

1. **Mở đường link và tải tài nguyên trong này về( click vào biểu tượng tải xuống phía trên góc phải)**

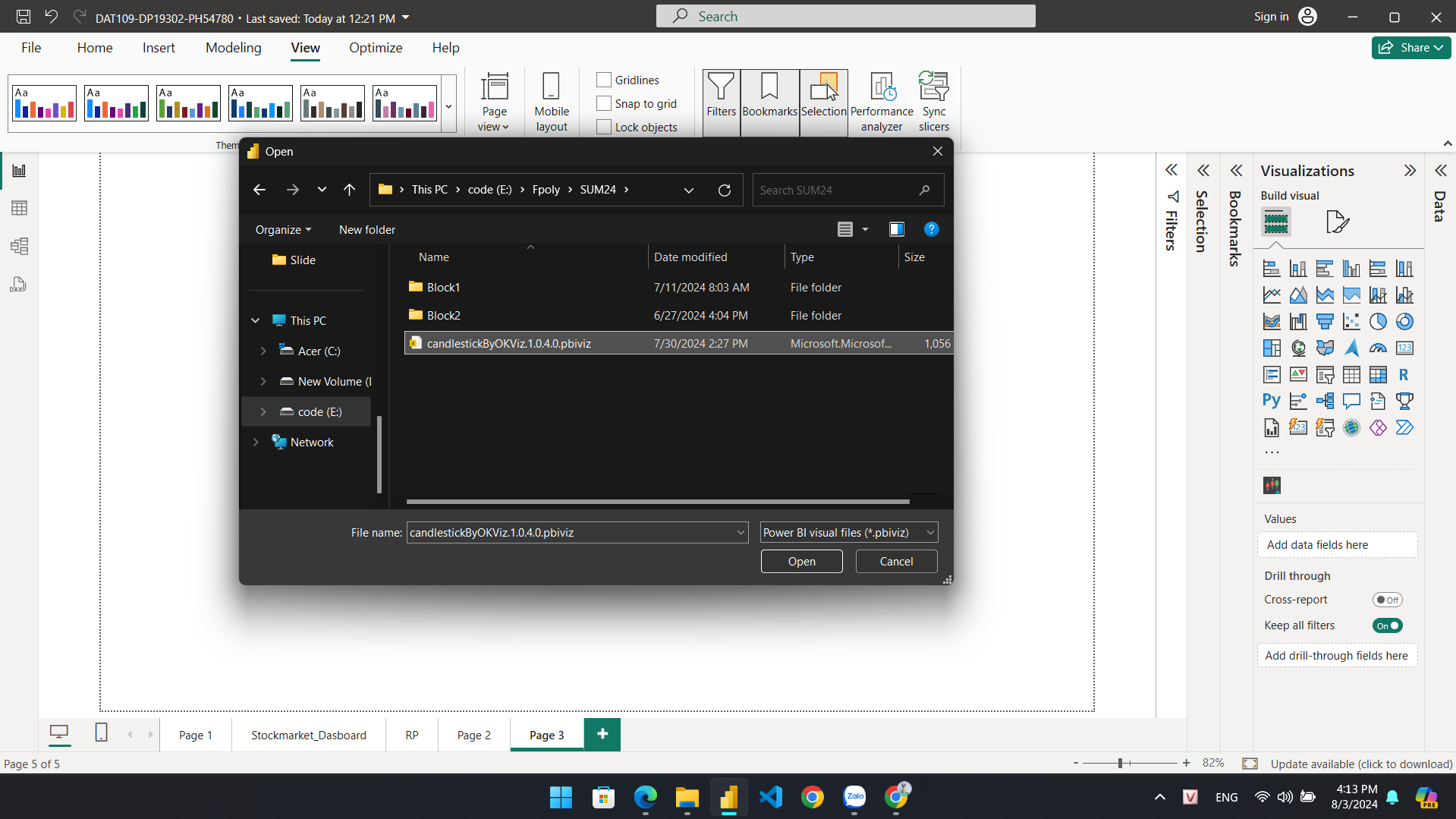
[**https://drive.google.com/file/d/1JBnrdBaNaZqGXd0kZRq3Pgx\_jtbfRLqh/view?usp=sharing**](https://drive.google.com/file/d/1JBnrdBaNaZqGXd0kZRq3Pgx_jtbfRLqh/view?usp=sharing)

****

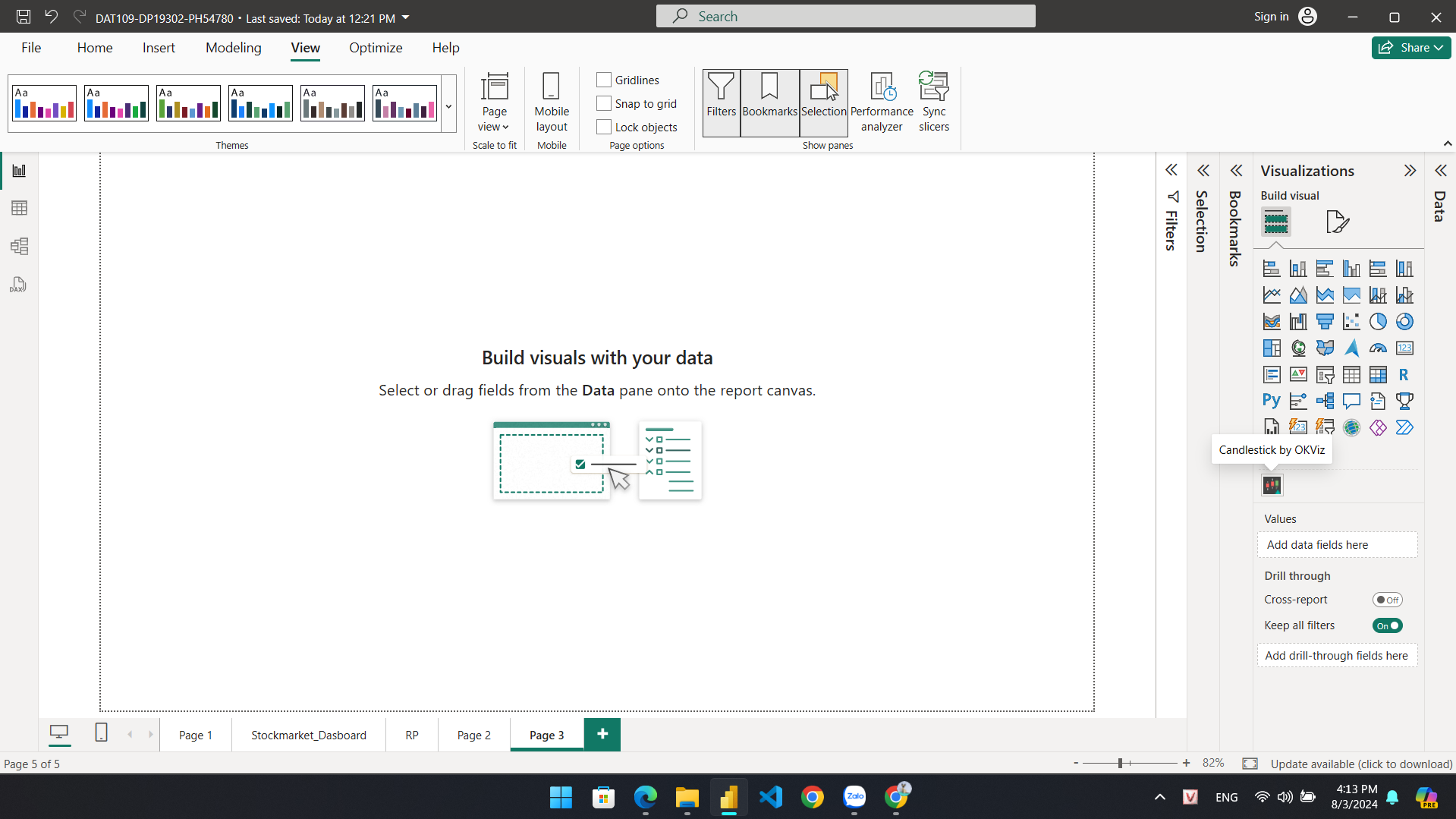
1. **Sau đó mở Power Bi, trong thẻ visualiaztion chọn mục ba chấm và import a visual from a file**

****

1. **Tiếp đó mở file vừa tải về và import vào**

****

1. **Xuất hiện biểu tượng này và đã cài đặt thành công**

****

## 5.4 Trình bày tạo các report cho dự án

| **Bộ lọc** | |
| --- | --- |
| Bộ lọc | Minh họa |
| Tạo bộ lọc theo ngày giao dịch xe |  |
| Tạo bộ lọc theo hãng xe |  |
| Tạo bộ lọc theo màu sắc xe |  |
| **Thống kê** | |
| Thống kê | Minh họa |
| Thống kê theo doanh thu bán xe các hãng |  |
| Thống kê tổng số lượng bán xe của các hãng |  |
| Thống kê giá bán trung bình |  |
| **Tạo biểu đồ** | |
| Tạo biểu đồ | Minh họa |
| Biểu đồ giá bán trung bình theo các năm |  |
| Biểu đồ màu sắc xe được ưa chuộng nhất |  |
| Biểu đồ thể hiện mối tương quan giữa số km đi được và giá bán xe |  |
| Tạo biểu đồ thể hiện lợi nhuận trung bình của các hãng xe |  |

# 6 Xây dựng báo cáo

## 6.1 Dashboard và report

| *Lĩnh vực* | *Cách tối ưu hóa* |
| --- | --- |
| Áp dụng các bộ lọc để giới hạn số lượng dữ liệu | * **Ít dữ liệu hơn giúp cải thiện tốc độ tải hình ảnh:** Giảm khối lượng dữ liệu để dashboard và báo cáo tải nhanh hơn. * **Đặt câu hỏi về nhu cầu dữ liệu:** Xác định liệu người xem có cần thấy tất cả dữ liệu không, từ đó áp dụng bộ lọc để chỉ hiển thị thông tin cần thiết. * **Giảm thiểu số lượng dữ liệu hiển thị:** Xem xét các cách để giảm số lượng dữ liệu hiển thị mà không làm giảm trải nghiệm của người dùng. |
| Kiểm soát số lượng hình ảnh trên trang báo cáo | * **Giới hạn số lượng hình ảnh trên một trang báo cáo:** Đặt mức giới hạn hợp lý cho số lượng hình ảnh để đảm bảo trang báo cáo không quá tải. * **Dễ hiểu hơn và ổn định hơn:** Giúp người xem dễ hiểu báo cáo hơn và đảm bảo báo cáo vận hành ổn định hơn. |
| Đảm bảo hiệu suất cao của các hình ảnh tùy chỉnh | * **Tác động lớn của hình ảnh trực quan:** Nhận thức rằng hình ảnh có ảnh hưởng mạnh đến toàn bộ báo cáo. * **Đánh giá chất lượng và số lượng hình ảnh:** Đánh giá một cách chính xác chất lượng và số lượng của các hình ảnh để tối ưu hóa hiệu suất. |
| Hiểu rõ ý nghĩa về hiệu suất của dashboard | * **Sử dụng bộ nhớ cache:** Khi truy xuất dữ liệu từ bộ nhớ cache, dashboard sẽ hoạt động tốt hơn và nhất quán hơn so với việc chỉ dựa vào nguồn dữ liệu gốc. |
| Tối ưu hóa môi trường vận hành | * **Cấu hình cài đặt dung lượng:** Tối ưu hóa cài đặt dung lượng của môi trường như Power BI. * **Định cỡ cổng dữ liệu và giảm độ trễ mạng:** Tinh chỉnh cổng dữ liệu và giảm độ trễ mạng để cải thiện hiệu suất tổng thể. |

## 6.2 Xây dựng báo cáo

### 6.2.1 Dashboard vs Report

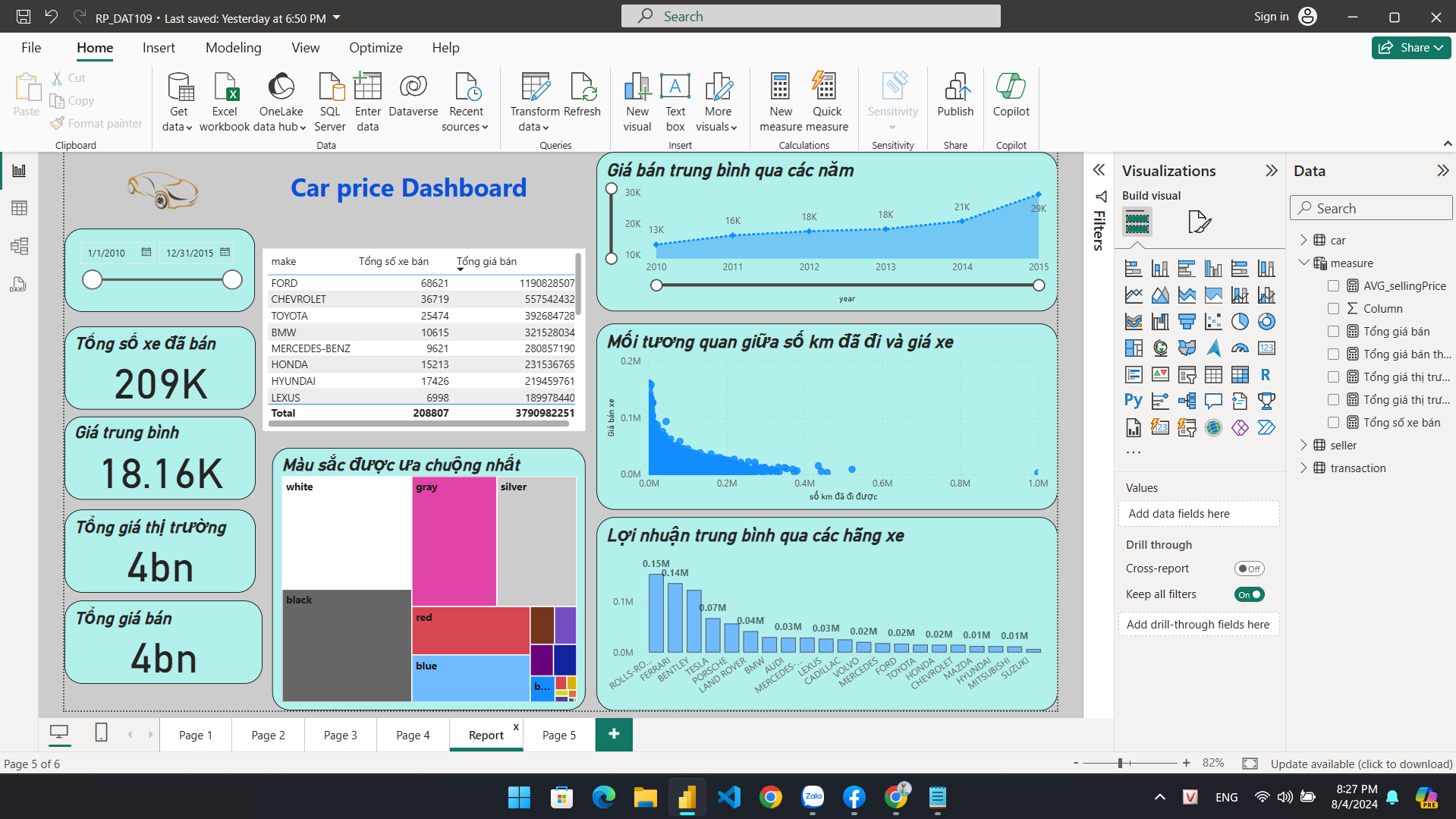
| *Item* | *Dashboard* | *Report* |
| --- | --- | --- |
| Pages | Chỉ một trang. | Có thể tạo một hoặc nhiều trang. |
| Data sources | Tạo từ nhiều dataset hoặc reports. | Tạo từ một dataset. |
| Visualization | Xây dựng thông tin chi tiết về dữ liệu bằng cách sử dụng biểu đồ, hình ảnh sinh động, lưu đồ. | Xây dựng trang tổng hợp thông tin, cung cấp góc nhìn tổng quan. |
| Available in Power BI Desktop | Không thể tạo. | Có thể tạo và xem. |
| Filters and Slicers | Không thể dung, bị giới hạn trên trang đơn. | Có thể dùng các loại filter, highlight và slide khác nhau. |
| User Interactivity | Cho phép pin visuals từ report và datasets trên canvas đơn, làm cho nó đơn giản để nhóm những gì cần thiết cho người dùng. | Tập trung hơn vào khả năng trực quan hóa, áp dụng các phép biến đổi trên một tập dữ liệu. |
| Favorite | Có thể gán nhiều dashboards | Có thể gán nhiều dashboards |
| Q&A feature | Có | Có thể phân quyền chỉnh sửa report. |
| Alerts | Gửi email cảnh báo khi đáp ứng điều kiện, tiêu chí cụ thể hoặc vượt qua giới hạn. | Không thể |
| Subscribe | Có | Có |
| See underlying dataset tables and fields | Không thể xem dataset nhưng có thể xuất dữ liệu này. | Có thể xem dữ liệu trong tab dữ liệu. |
| Purpose | Được sử dụng để giám sát cấp cao, thường theo thời gian thực hoặc gần thời gian thực, cung cấp cái nhìn tổng hợp về hiệu quả kinh doanh. | Được sử dụng để phân tích chuyên sâu và khám phá dữ liệu nhằm trả lời các câu hỏi kinh doanh phức tạp. |

**Giải thích**

| *Mô hình* | *Lí do* |
| --- | --- |
| Dashboard | * Theo dõi thời gian thực: Cung cấp cái nhìn tổng quan và cập nhật về giá cổ phiếu, khối lượng giao dịch, và biến động giá. * Tóm tắt thông tin quan trọng: Hiển thị các chỉ số chính, giúp theo dõi nhanh chóng. * Khả năng tương tác: Cho phép lọc và chọn dữ liệu để phân tích các cổ phiếu hoặc nhóm cổ phiếu cụ thể. * Ra quyết định nhanh chóng: Cung cấp thông tin ngay lập tức để hỗ trợ quyết định đầu tư. |
| Report | - Phân tích chi tiết: Cung cấp phân tích sâu về các yếu tố ảnh hưởng đến giá chứng khoán như báo cáo tài chính và tin tức thị trường.  Trình bày kết quả nghiên cứu: Sử dụng bảng, đồ thị để trình bày phân tích và xu hướng.  Khuyến nghị đầu tư: Đưa ra khuyến nghị dựa trên phân tích dữ liệu.  Lưu trữ và chia sẻ thông tin: Lưu trữ kết quả phân tích và chia sẻ với các bên liên quan. |

### 6.2.2 Dashboard

**Dashboard sử dụng trong dự án**



*Giải thích:*

**Sử dụng Dashboard đó trong dự án này là vì:**

### Tập Trung và Dễ Dàng Theo Dõi Thông Tin

### Tương Tác và Linh Hoạt

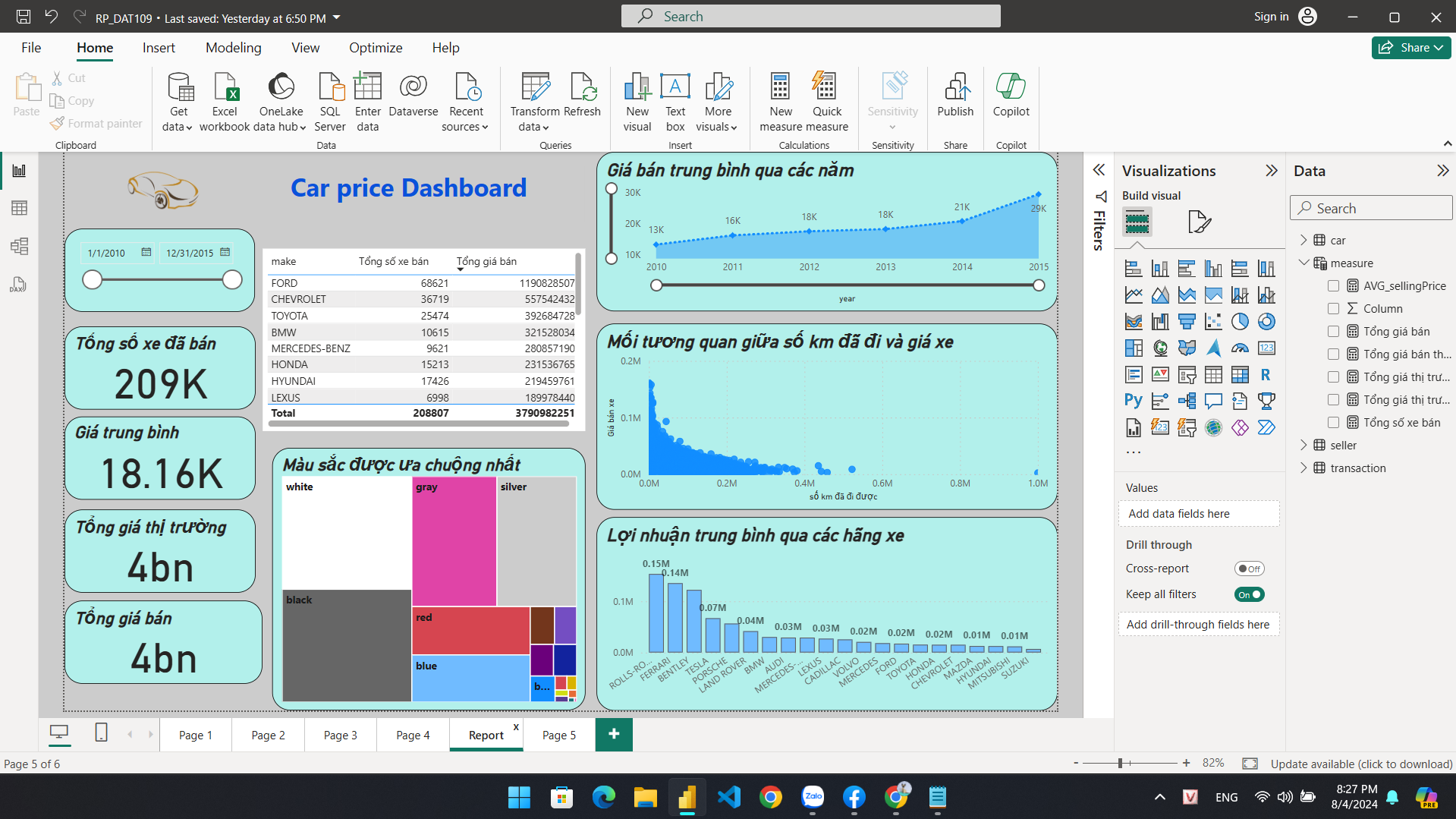
### Cải Thiện Hiệu Quả

### Hỗ Trợ Quyết Định Dựa Trên Dữ Liệu

### Báo Cáo và Truyền Thông

### 6.2.3 Report

**Report sử dụng trong dự án**



*Giải thích:*

**Bộ lọc khoảng thời gian**: Ở góc trên bên trái, có bộ lọc khoảng thời gian cho phép người dùng lọc dữ liệu giữa các ngày cụ thể.

**Tổng số xe đã bán**: Thẻ bên trái hiển thị tổng số xe đã bán

**Giá trung bình**: Thẻ tiếp theo hiển thị giá bán trung bình của các xe

**Tổng giá thị trường**: Thẻ này hiển thị tổng giá trị thị trường của các xe

**Tổng giá bán**: Tổng giá bán của tất cả các xe cũng được hiển thị

**Các hãng xe hàng đầu theo doanh số**: Bảng ở giữa liệt kê các hãng xe hàng đầu cùng với số lượng xe đã bán và tổng doanh thu cho mỗi hãng.

**Giá trung bình qua các năm**: Biểu đồ đường ở góc trên bên phải minh họa xu hướng giá xe trung bình qua các năm

**Số km đã đi và giá xe**: Biểu đồ phân tán ở giữa bên phải hiển thị mối quan hệ giữa số km đã đi và giá xe. Có vẻ như khi số km tăng, giá xe thường giảm.

**Màu sắc được ưa chuộng nhất**: Biểu đồ cây ở phần giữa dưới hiển thị các màu xe được ưa chuộng nhất, với trắng, xám và bạc là ba màu hàng đầu.

**Lợi nhuận trung bình qua các hãng xe**: Biểu đồ cột ở góc dưới bên phải chỉ ra lợi nhuận trung bình cho mỗi hãng xe. Rolls-Royce, Ferrari và Bentley có lợi nhuận trung bình cao nhất.

### 6.2.4 Bookmark

Sử dụng các Bookmark trong Report có tác dụng như:

**Tạo báo cáo tương tác**: Có thể tạo các kịch bản khác nhau trong cùng một báo cáo bằng cách sử dụng các bookmark để lưu trạng thái của bộ lọc, hiển thị các biểu đồ hoặc bảng khác nhau, và thậm chí thay đổi cách bố trí của báo cáo.

**Điều hướng dễ dàng**: Có thể tạo các nút hoặc liên kết để chuyển đổi giữa các bookmark, giúp điều hướng giữa các phần khác nhau của báo cáo một cách dễ dàng và trực quan.

**Thuyết trình**: Bookmark rất hữu ích khi bạn thuyết trình, cho phép chuẩn bị các bước cụ thể và chuyển qua lại giữa các trạng thái báo cáo mà không cần thay đổi thủ công trong lúc trình bày.

**Lưu lại các phân tích quan trọng**: Có thể lưu lại các trạng thái phân tích quan trọng mà bạn đã thực hiện, giúp bạn dễ dàng quay lại và xem lại hoặc chia sẻ với đồng nghiệp.

**Tạo các tính năng tương tác**: Bookmark kết hợp với Selection Pane và Visual Level Filters có thể tạo ra các tính năng tương tác như bật/tắt một số biểu đồ, bảng, hoặc hình ảnh.

#### 6.2.4.1 Tạo bookmark slicer chọn tất cả các ngày

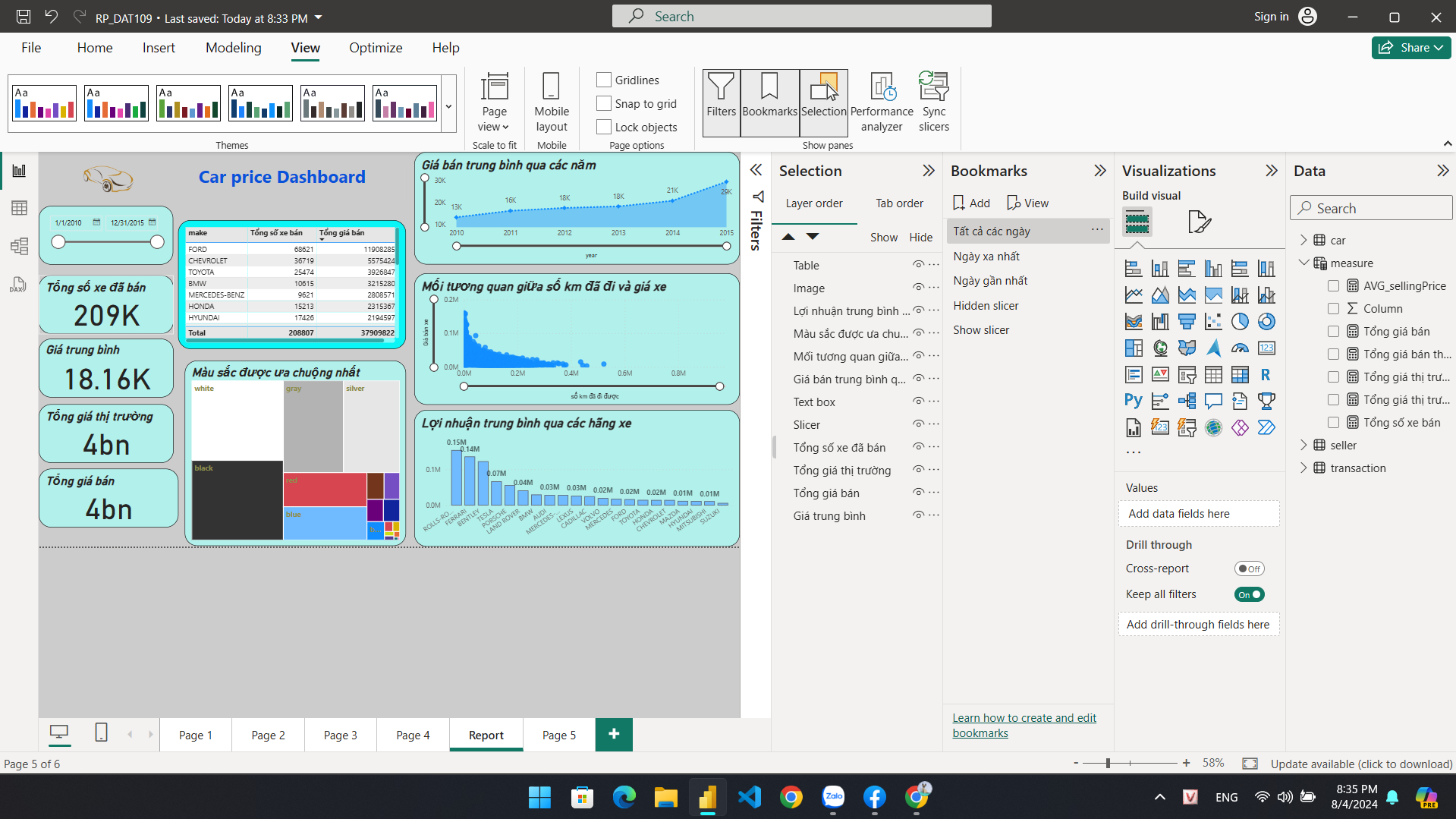
B1: chọn thẻ View

B2: chọn Bookmark

B3: Chọn add ở thanh bên phải mục Bookmarks

B4: Đổi tên Bookmark 1 thành tất cả các ngày

B5: Kéo Slicer ngày giao dịch sao cho hiển thị đủ tất cả các ngày



*Giải thích:* Giúp người xem có thể quay về trạng thái hiển thị thông tin tất cả các ngày (trạng thái ban đầu) một cách nhanh nhất.

#### 6.2.4.2 Tạo bookmark slicer chọn ngày gần nhất

B1: chọn thẻ View

B2: chọn Bookmark

B3: Chọn add ở thanh bên phải mục Bookmarks

B4: Đổi tên Bookmark 2 thành ngày gần nhất

B5: Kéo Slicer ngày giao dịch sao cho thanh ngày chỉ nhận giá trị gần nhất bên phải



*Giải thích:* Giúp người xem có thể quay về trạng thái hiển thị thông tin ngày gần nhất một cách nhanh chóng

#### 6.2.4.3 Tạo bookmark slicer chọn ngày xa nhất

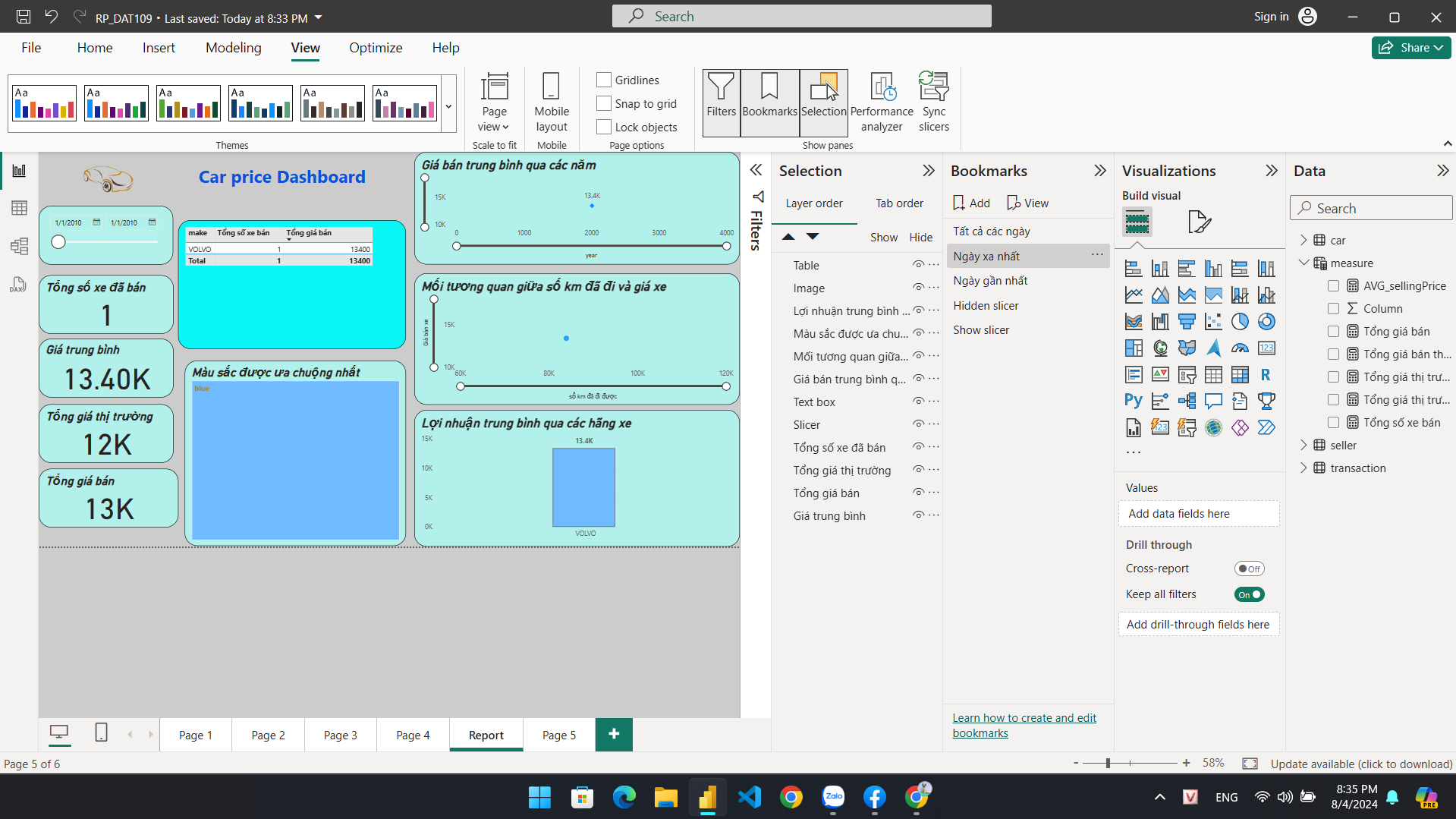
B1: chọn thẻ View

B2: chọn Bookmark

B3: Chọn add ở thanh bên phải mục Bookmarks

B4: Đổi tên Bookmark 3 thành ngày xa nhất

B5: Kéo Slicer ngày giao dịch sao cho thanh ngày chỉ nhận giá trị xa nhất bên trái



*Giải thích:* Giúp người xem có thể quay về trạng thái hiển thị thông ngày xa nhất một cách nhanh chóng.

#### 6.2.4.4 Tạo bookmark ẩn slicer

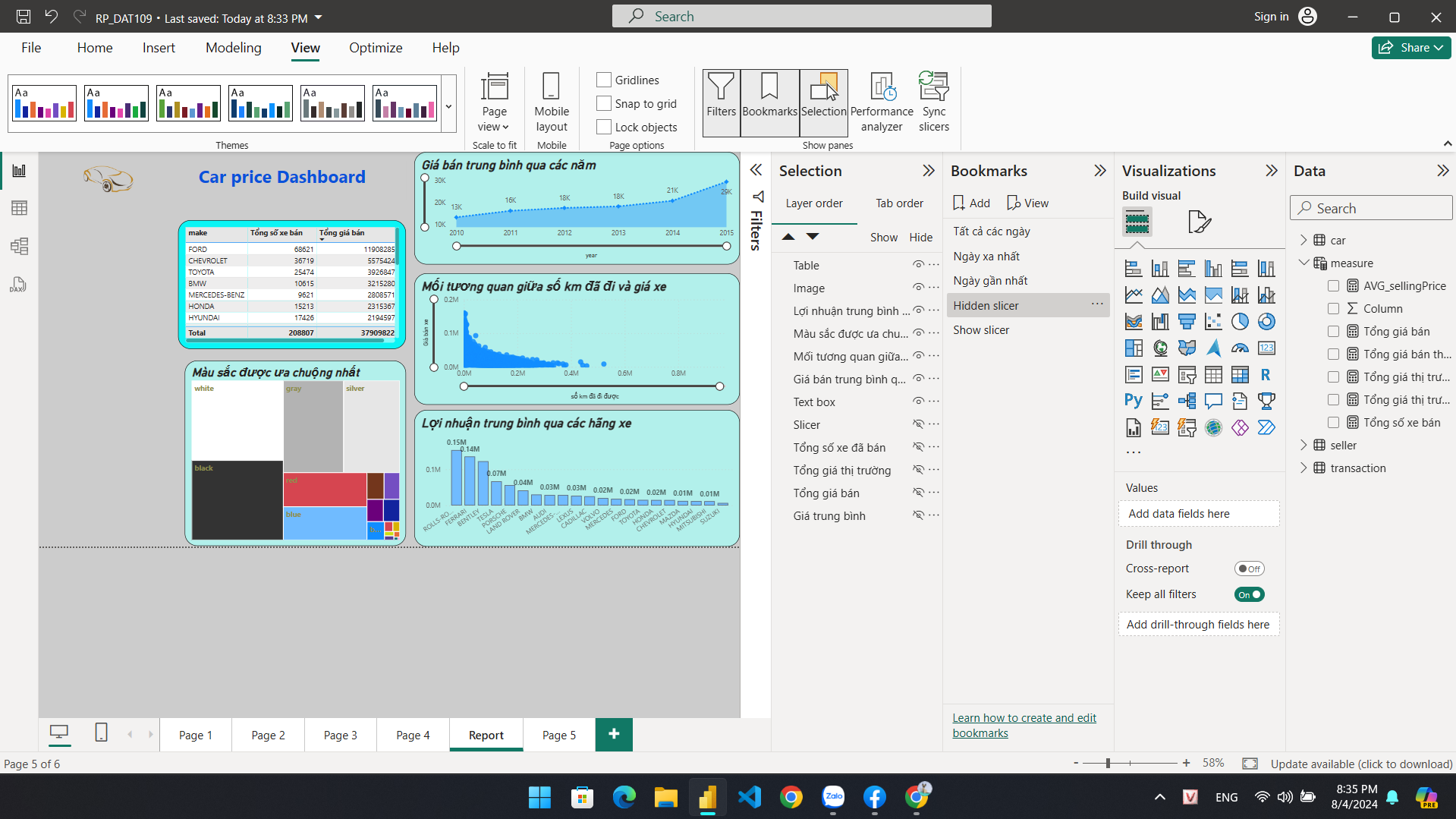
B1: chọn thẻ View

B2: chọn Bookmark & Selection

B3: Chọn add ở thanh bên phải mục Bookmarks

B4: Đổi tên Bookmark 4 thành Hidden Slicer

B5: Mở thẻ Selection sau đó tích ẩn các slicer



*Giải thích:* Giúp người dùng có thể ẩn đi các slicer thể hiện các card.

#### 6.2.4.5 Tạo bookmark Show Slicer

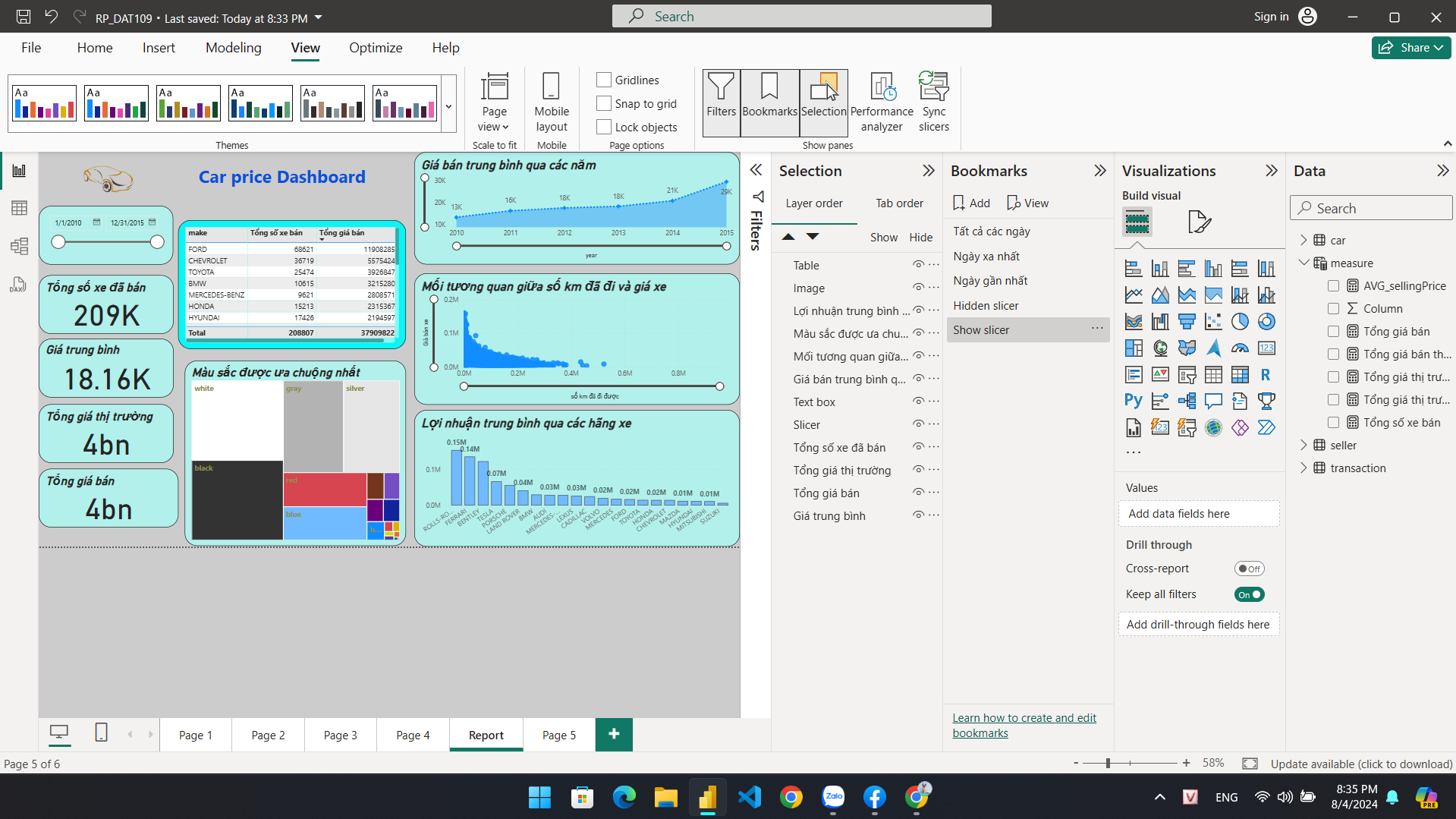
B1: chọn thẻ View

B2: chọn Bookmark & Selection

B3: Chọn add ở thanh bên phải mục Bookmarks

B4: Đổi tên Bookmark 5 thành Show Slicer

B5: Mở thẻ Selection sau đó bỏ tích ẩn các slicer



*Giải thích:* Giúp người dùng có thể hiện các slicer đã ẩn khi muốn xem lại …

# 7 KẾT LUẬN

## 7.1 Báo cáo

### 7.7.1 Các bước viết báo cáo

B1: Xem đối tượng

B2: Soạn thảo nội dung

B3: Phát triển nội dung

B4: Viết phần giới thiệu

B5: Thêm 1 kết luận ngắn

B6: Thêm các đối tượng trực quan hóa dữ liệu

B7: Chỉnh sửa trước khi gửi

### 7.1.2 Tổng hợp

**Chọn công nghệ:**

* Nền tảng báo cáo Power Bi
* Hệ thống quản lý cơ sở dữ liệu: SQL server
* Nền tảng xuất báo cáo: Xuất ra file PDF

**Chọn kiến trúc**

* Kiến trúc phân tầng: Bao gồm các tầng thu thập dữ liệu, lưu trữ dữ liệu, xử lý dữ liệu, phân tích dữ liệu và trình bày báo cáo.
* Tầng lưu trữ dữ liệu: Sử dụng SQL Server để lưu trữ dữ liệu.
* Tầng xử lý và phân tích dữ liệu: Sử dụng các công cụ phân tích của SQL Server và khả năng trực quan hóa của Power BI.
* Tầng trình bày: Sử dụng Power BI để tạo báo cáo và bảng điều khiển, xuất báo cáo qua file PDF

**Cách thu thập dữ liệu, nguồn**

* Bộ dữ liệu được thu thập từ các nguồn trang web khác nhau (có trả phí và tính phí)
* Phạm vi của bộ dữ liệu: Dữ liệu được thu thập hàng ngày qua các phiên giao dịch, bao gồm các chỉ số chính như giá mở cửa, giá đóng cửa, khối lượng giao dịch, …
* Nhân viên có thể xem báo cáo thông qua máy tính phân tích hoặc xuất qua PDF

**Báo cáo cung cấp thông tin**

* **Tổng số xe đã bán:** Tổng số xe ô tô đã được bán ra trong khoảng thời gian từ ngày 01/01/2010 đến ngày 31/12/2015
* **Giá trung bình:** Giá bán trung bình của các xe đã được bán trong khoảng thời gian.
* **Tổng giá trị thị trường:** Tổng giá trị thị trường của tất cả các xe đã bán.
* **Tổng giá bán:** Tổng số tiền thu được từ việc bán các xe.
* **Thông tin chi tiết về các hãng xe:** Bảng này liệt kê các hãng xe, tổng số xe bán được và tổng giá trị bán của từng hãng.
* **Biến động giá bán trung bình qua các năm:** Biểu đồ đường thể hiện sự biến động giá bán trung bình của các xe qua các năm từ 2010 đến 2015.
* **Mối tương quan giữa số km đã đi và giá xe:** Biểu đồ tán xạ cho thấy mối quan hệ giữa số km đã đi và giá bán của xe.
* **Sự phổ biến của màu sắc xe:** Biểu đồ dạng cây thể hiện tỷ lệ các màu sắc của xe được bán ra.
* **Lợi nhuận trung bình của các hãng xe:** Biểu đồ cột thể hiện lợi nhuận trung bình của các hãng xe.

**Các lưu ý cần thiết**

* **Đơn vị đo lường**: Kiểm tra đơn vị đo lường cho từng số liệu. Ví dụ, số xe bán được là "K" (nghìn), giá trung bình là "K" (nghìn đô la), tổng giá thị trường và tổng giá bán là "bn" (tỷ đô la).
* **Khoảng thời gian**: Chú ý đến khoảng thời gian được chọn trong bộ lọc ngày ở góc trên bên trái. Điều này xác định phạm vi dữ liệu được hiển thị trong báo cáo.
* **Ngữ cảnh của dữ liệu**: Hiểu rõ rằng dữ liệu trong báo cáo này chỉ bao gồm các xe bán từ năm 2010 đến 2015. Điều này ảnh hưởng đến các xu hướng và số liệu mà bạn thấy.
* **Mối tương quan giữa các số liệu**: Khi nhìn vào biểu đồ mối tương quan giữa số km đã đi và giá xe, cần lưu ý rằng có sự giảm giá rõ rệt khi số km đã đi tăng lên. Điều này có thể giúp bạn hiểu rõ hơn về giá trị còn lại của xe theo thời gian sử dụng.
* **Xu hướng qua các năm**: Biểu đồ giá bán trung bình qua các năm cho thấy xu hướng giá xe qua các năm. Điều này có thể giúp bạn nhận ra các biến động và lý do tiềm ẩn đằng sau sự thay đổi giá.
* **Lợi nhuận trung bình của các hãng xe**: Khi nhìn vào biểu đồ lợi nhuận trung bình của các hãng xe, hãy lưu ý rằng những hãng có lợi nhuận trung bình cao nhất có thể không bán nhiều xe nhất, nhưng giá trị mỗi xe có thể rất cao.
* **Sự phổ biến của màu sắc xe**: Biểu đồ màu sắc ưa chuộng nhất cho thấy màu sắc nào phổ biến nhất. Điều này có thể hữu ích trong việc đưa ra quyết định mua xe dựa trên xu hướng thị trường.
* **Đọc kỹ các chú thích và tiêu đề**: Đảm bảo đọc kỹ các chú thích và tiêu đề của các biểu đồ và bảng để hiểu rõ ngữ cảnh và ý nghĩa của dữ liệu được trình bày.

## 7.2 Khó khăn

**Chất lượng dữ liệu không đồng nhất**: Dữ liệu có thể bị thiếu, không chính xác hoặc không đồng nhất, gây khó khăn trong quá trình phân tích. Ví dụ, thông tin về số km đã đi hoặc điều kiện xe có thể bị nhập sai hoặc không đầy đủ.

**Độ phức tạp của dữ liệu**: Dữ liệu mua bán xe cũ thường rất phức tạp và đa dạng, có thể bao gồm nhiều loại xe, nhiều hãng xe, và các yếu tố khác nhau ảnh hưởng đến giá trị xe, làm cho việc phân tích trở nên phức tạp hơn.

**Tính đa chiều của dữ liệu**: Dữ liệu về xe cũ có nhiều chiều và yếu tố ảnh hưởng khác nhau (năm sản xuất, số km đã đi, điều kiện xe, v.v.), đòi hỏi phải sử dụng các kỹ thuật phân tích đa chiều và mô hình hóa phức tạp.

**Yếu tố thị trường không ổn định**: Thị trường xe cũ có thể thay đổi nhanh chóng do các yếu tố ngoại cảnh như thay đổi về kinh tế, luật pháp, và xu hướng tiêu dùng. Điều này làm cho việc dự báo và phân tích trở nên khó khăn hơn.

## 7.3 Thuận lợi

**Dữ liệu phong phú**: Bộ dữ liệu mua bán xe cũ thường rất phong phú, bao gồm nhiều thuộc tính như năm sản xuất, hãng xe, mẫu xe, điều kiện xe, số km đã đi, màu sắc, giá bán, và ngày bán. Điều này cho phép phân tích chi tiết và sâu sắc.

**Dễ dàng truy cập dữ liệu**: Các dữ liệu về mua bán xe cũ thường được các đại lý, trang web bán xe, và các sàn giao dịch ô tô thu thập và lưu trữ một cách có hệ thống, dễ dàng truy cập và sử dụng.

**Nhu cầu thị trường rõ ràng**: Dữ liệu về xe cũ có thể giúp xác định rõ nhu cầu thị trường, xu hướng về mẫu xe, màu sắc được ưa chuộng, và những yếu tố ảnh hưởng đến giá cả. Điều này hữu ích cho việc dự báo và đưa ra quyết định kinh doanh.

**Phân tích dễ dàng với công cụ hiện đại**: Có nhiều công cụ phân tích dữ liệu mạnh mẽ Power BI giúp dễ dàng trực quan hóa và phân tích dữ liệu.

## 7.4 Hướng phát triển

**Nâng cao chất lượng và độ chính xác của phân tích**: Sử dụng các thuật toán và công nghệ tiên tiến như Machine Learning và AI để cải thiện độ chính xác và hiệu quả của các phân tích dữ liệu chứng khoán.

**Tích hợp thêm nguồn dữ liệu**: Kết hợp dữ liệu chứng khoán với các nguồn dữ liệu khác như dữ liệu kinh tế, dữ liệu xã hội để có cái nhìn toàn diện và chính xác hơn.

**Phát triển giao diện người dùng**: Tạo ra các giao diện báo cáo dễ sử dụng và tương tác, giúp người dùng dễ dàng truy cập và hiểu thông tin từ báo cáo.

**Đảm bảo bảo mật và tuân thủ quy định**: Tiếp tục đầu tư vào các biện pháp bảo mật để bảo vệ dữ liệu và tuân thủ các quy định pháp luật về bảo mật dữ liệu.

**Đào tạo và nâng cao kỹ năng cho nhân viên**: Cung cấp đào tạo và hướng dẫn cho nhân viên về cách sử dụng các công cụ phân tích dữ liệu và đọc hiểu báo cáo chứng khoán để nâng cao hiệu quả công việc.

## 