

Yêu cầu 1: Chi tiết các bước Tiền xử lý và Khám phá dữ liệu được thực hiện bằng ngôn ngữ lập trình Python và được trình bày trong File: ETL\_EDA.pdf

#### 1.1.1 Cấu trúc dữ liệu:

Tập dữ liệu có 268255 dòng và 17 trường (columns), bao gồm thông tin liên quan đến:

- Thông tin cơ bản: Maker (Hãng xe), Genmodel, Genmodel\_ID, Adv\_ID
- Thời gian quảng cáo và đăng ký xe: Adv\_year, Adv\_month, Reg\_year
- Đặc điểm kỹ thuật và ngoại hình:
  - Màu xe: Color
  - Kiểu dáng: Bodytype
  - Số dặm đã chạy: Runned\_Miles
  - Dung tích động cơ: Engin\_size
  - Hộp số: Gearbox
  - Nhiên liệu: Fuel\_type
  - Số ghế (Seat\_num) và số cửa (Door\_num)

#### 1.1.2 Tiền xử lý dữ liệu & chuẩn hóa dữ liệu

Chuyển đổi kiểu dữ liệu

- Runned\_Miles: hiện đang lưu dưới dạng chuỗi, ví dụ “60,000”. Cần loại bỏ dấu phẩy và ép kiểu sang số (int hoặc float).
- Price: hiện là chuỗi, ví dụ “21,500”. Tương tự, bỏ dấu phẩy rồi chuyển sang số.

- Engin\_size: hiện là chuỗi, ví dụ “6.8L”. Loại bỏ ký tự “L”, sau đó ép phần số thành float (kết quả 6.8).

Xử lý dữ liệu thiếu

- Color (~8,15 % thiếu): điền giá trị phổ biến nhất (mode).
- Runned\_Miles (~0,39 % thiếu): số lượng rất nhỏ, xoá các bản ghi này.
- Engin\_size (~0,77 % thiếu): tách phần số, sau đó điền thiếu bằng trung vị (median).
- Seat\_num (~2,41 % thiếu): tính trung bình số ghế cho từng Genmodel, rồi điền vào.
- Door\_num (~1,70 % thiếu): lấy trung bình số cửa theo Genmodel hoặc Bodytype, rồi điền vào.

Loại bỏ giá trị bất thường

- Trường Adv\_month xuất hiện các giá trị không hợp lệ: 13, 17, 33.
- Giải pháp: lọc bỏ các bản ghi có Adv\_month nằm ngoài khoảng hợp lệ 1–12 trước khi phân tích.

1.1.3 Khám phá dữ liệu ( được trình bày chi tiết và nhận xét trong file ETL\_EDA.pdf)

1.1.4 Xây dựng mô hình Data Warehouse

- Em tiến hành thiết kế 7 bảng Dim và 1 bảng Fact
- Bảng Fact:

Bảng fact\_car\_listing\_full\_clean là bảng dữ liệu trung tâm trong mô hình dữ liệu sao (Star Schema), lưu trữ các thông tin chi tiết của từng bản ghi quảng cáo xe, dùng để phục vụ phân tích và tổng hợp theo các chiều (dimension) khác nhau.

Mô tả nội dung bảng:

Bảng fact\_car\_listing\_full\_clean bao gồm các thông tin sau:

- Thông tin thời gian:
  - Adv\_year, Adv\_month, YearMonth: thông tin thời điểm đăng quảng cáo, được ánh xạ đến bảng dim\_date.
- Thông tin xe và quảng cáo:
  - Adv\_ID: Mã định danh duy nhất cho mỗi quảng cáo xe.
  - Genmodel\_ID, Genmodel: Mã và tên dòng xe, được ánh xạ đến bảng dim\_genmodel.
  - Maker: Hãng xe, liên kết đến dim\_maker.
- Đặc điểm kỹ thuật:
  - Bodytype: Kiểu thân xe.
  - Fuel\_type: Loại nhiên liệu.
  - Gearbox: Loại hộp số.
  - Color: Màu sắc xe.
  - Engin\_size: Dung tích động cơ (đơn vị: lít), đã chuẩn hoá kiểu dữ liệu.
  - Seat\_num, Door\_num: Số ghế và số cửa của xe.
- Thông tin đăng ký và sử dụng:
  - Reg\_year: Năm đăng ký xe lần đầu.
  - Runned\_Miles: Số dặm đã sử dụng, đã được chuẩn hoá về kiểu số.

- Thông tin tài chính:
  - Price: Giá bán xe, đã được chuyển về định dạng số (int hoặc float) phục vụ phân tích.

Vai trò của bảng fact:

- Lưu trữ các biến số định lượng (số liệu thực) phục vụ phân tích, như Price, Runned\_Miles.
  - Làm trung tâm để liên kết tới các bảng dimension qua các khoá: Genmodel\_ID, Maker, Date\_ID, v.v.
  - Dùng làm nguồn dữ liệu chính cho biểu đồ
- 7 Bảng dim bao gồm:

Trong quá trình chuẩn hoá dữ liệu để phục vụ phân tích, các bảng dimension được tạo ra nhằm giảm thiểu trùng lặp, chuẩn hoá thông tin và tối ưu hiệu suất truy vấn. Dưới đây là mô tả chi tiết các bảng dimension đã xây dựng từ tập dữ liệu:

#### 1. dim\_date – Bảng thời gian quảng cáo

Bảng này phục vụ phân tích theo thời gian như xu hướng từng tháng, từng năm, hoặc theo quý.

#### 2. dim\_genmodel – Bảng dòng xe

Chứa thông tin chi tiết về từng dòng xe (Genmodel). Gồm các trường:

- Genmodel\_ID: Mã hoá dòng xe
- Lưu thông tin tên hãng xe (Maker), ví dụ: Toyota, Ford, Bentley,... giúp phân tích xu hướng thị trường theo từng hãng.
- Genmodel: Tên dòng xe (ví dụ: Camry, Accord, Arnage...)
- Genmodel\_Key: Khoá tự tăng để dùng làm khoá ngoại

Bảng này cho phép phân tích chuyên sâu theo từng dòng xe.

### 3. dim\_bodytype – Bảng kiểu thân xe

Chứa các kiểu dáng xe: Hatchback, SUV, Saloon,... Dùng để phân tích thói quen lựa chọn kiểu xe của người dùng.

### 4. dim\_fuel\_type – Bảng loại nhiên liệu

Chứa thông tin về loại nhiên liệu của xe: Petrol, Diesel, Electric,... Phục vụ phân tích xu hướng sử dụng năng lượng và tác động đến giá trị xe.

### 5. dim\_gearbox – Bảng loại hộp số

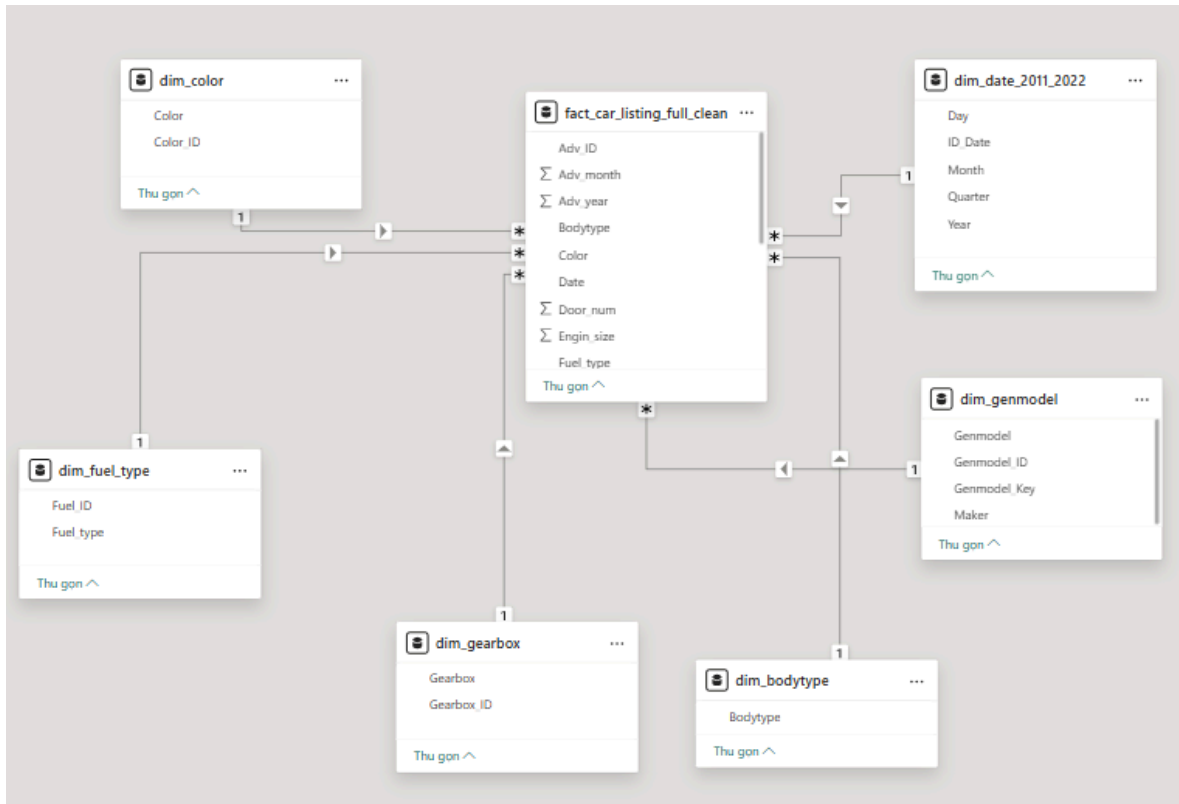
Lưu thông tin về loại hộp số: Manual, Automatic,...

### 6. dim\_color – Bảng màu sắc

Lưu màu sơn của xe: Black, Grey, White,... Dùng để thống kê màu xe phổ biến, hỗ trợ nhập hàng theo thị hiếu.

Tất cả các bảng trên đã được trích lọc từ tập dữ liệu chính, loại bỏ trùng lặp và được xuất thành các tệp .csv, sẵn sàng để sử dụng trong các bước xử lý và phân tích tiếp theo.

Dưới đây là Sơ đồ mô hình dữ liệu sao – Star Schema Diagram:



### 1.1.4 Xây dựng Dashboard

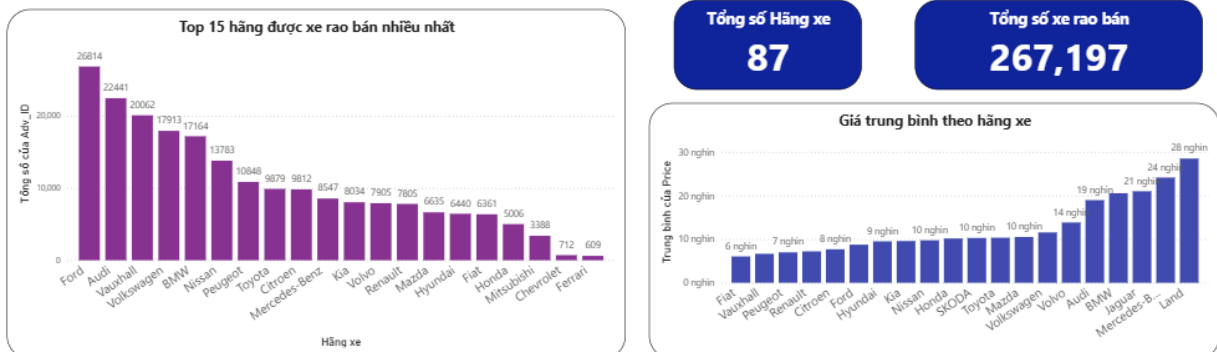
- Em đã tiến hành xây dựng 2 Dashboard được đính kèm cùng File tài liệu này, bao gồm: Báo cáo tổng quan và Báo cáo chi tiết

#### Yêu cầu 2: Phân tích cơ hội đầu tư

Câu 1: Nếu chị Linh muốn bắt đầu với 10 mẫu xe đầu tiên, bạn sẽ chọn loại nào? Tại sao?

Đối với chị Linh – người mới bước vào lĩnh vực kinh doanh xe cũ, việc lựa chọn các mẫu xe phù hợp để nhập hàng ban đầu là cực kỳ quan trọng nhằm đảm bảo an toàn vốn, dễ thanh khoản và học hỏi nhanh từ thị trường. Dưới đây là các tiêu chí lựa chọn xe nên ưu tiên:

1. Phổ biến, dễ bán: Ưu tiên các dòng xe có lượng rao bán cao, được nhiều người biết đến. Những mẫu xe này đã có chỗ đứng trên thị trường nên rất dễ tiêu thụ, giảm rủi ro tồn kho.
2. Giá hợp lý: Tập trung vào phân khúc giá tầm trung, phù hợp với đa số khách hàng cá nhân hoặc gia đình tại TP.HCM. Mức giá này cũng giúp chị Linh không cần đầu tư quá lớn mà vẫn đảm bảo vòng quay vốn nhanh.



Danh sách đề xuất 10 hãng xe phù hợp:

1. Ford
  - Số lượng rao bán: 26.814
  - Giá trung bình: khoảng 8.500 USD

## 2. Volkswagen

- Số lượng rao bán: 20.062
- Giá trung bình: khoảng 10.000 USD

## 3. Peugeot

- Số lượng rao bán: 10648
- Giá trung bình: khoảng 7.500 USD

## 4. Mazda

- Số lượng rao bán: 6635
- Giá trung bình: khoảng 10.000 USD

## 5. Hyundai

- Số lượng rao bán: 6440
- Giá trung bình: khoảng 9.000 USD

## 6. Renault

- Số lượng rao bán: 7805
- Giá trung bình: khoảng 8.000 USD
- Ghi chú: Hãng xe châu Âu phổ thông, giá mềm.

## 7. Nissan

- Số lượng rao bán: 13.763
- Giá trung bình: khoảng 9.000 USD
- Ghi chú: Phù hợp gia đình, dễ bảo dưỡng.

## 8. Citroen

- Số lượng rao bán: 9812

- Giá trung bình: khoảng 7.000 USD
- Ghi chú: Mẫu xe nhỏ, dễ sử dụng trong thành phố.

## 9. Kia

- Số lượng rao bán: 8034
- Giá trung bình: khoảng 9.000 USD
- Ghi chú: Xe phổ thông, linh hoạt, dễ tiếp cận.

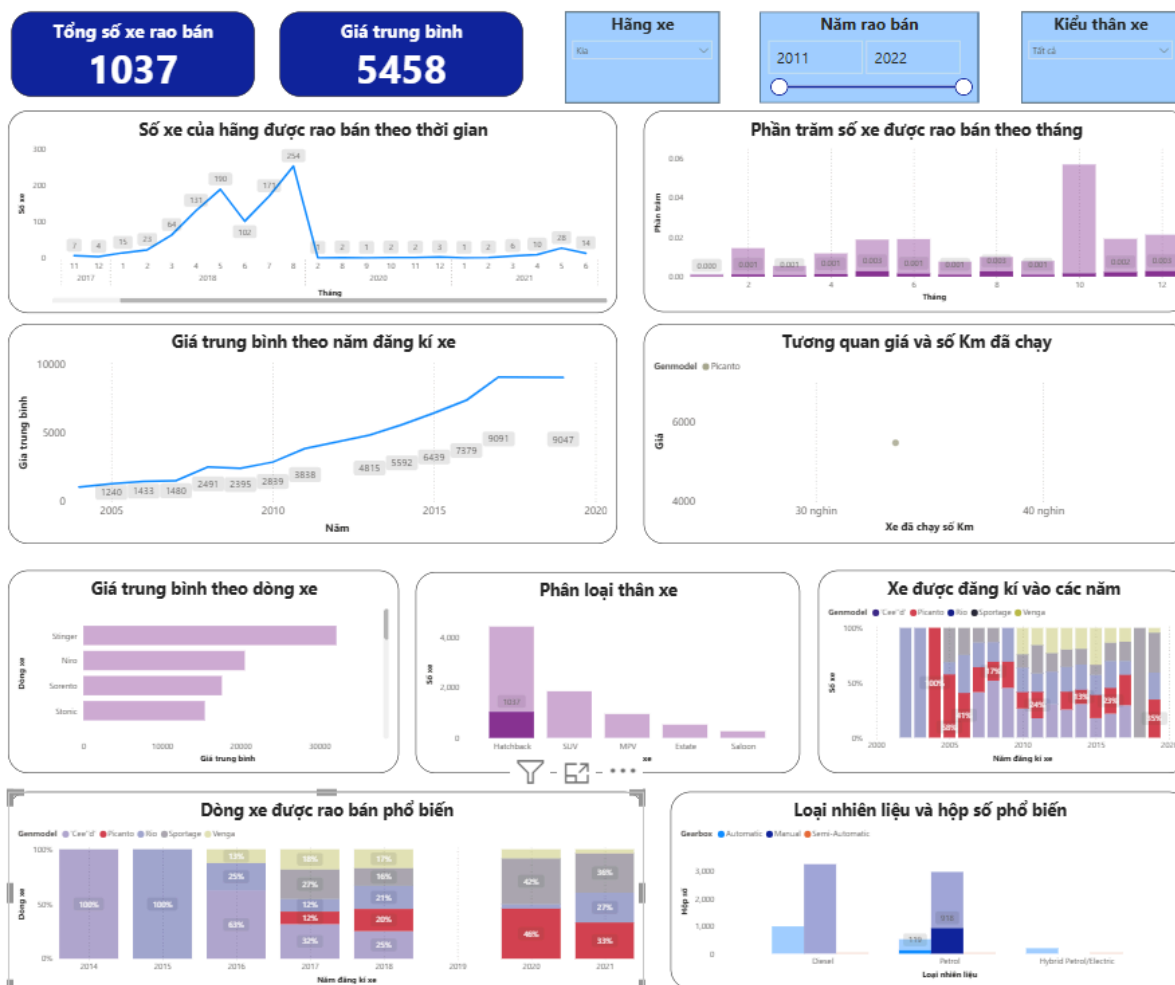
### 3. Tiêu chí tiếp theo để lựa chọn mẫu xe phù hợp

- Thương hiệu giữ giá: Độ giữ giá khi bán lại – đây là yếu tố then chốt nếu chị Linh muốn xoay vòng vốn mà không bị lỗ sâu, vì vậy ở đây em sử dụng biểu đồ liên quan đến thống kê là độ lệch chuẩn theo giá của các dòng xe của hãng để đánh giá mức độ biến động về giá:
- Tiết kiệm nhiên liệu: Chọn các mẫu xe sử dụng nhiên liệu xăng, hộp số tự động để phù hợp với nhu cầu đi lại thường ngày tại đô thị, vừa tiện lợi vừa tiết kiệm.
- Nguồn cung dồi dào: Ưu tiên nhập các dòng xe có nhiều lựa chọn trên thị trường, để tìm kiếm mẫu thay thế nếu cần mở rộng danh mục hoặc nhập thêm hàng.

Dưới đây là các mẫu xe em lựa chọn:

#### 1- KIA PICANTO

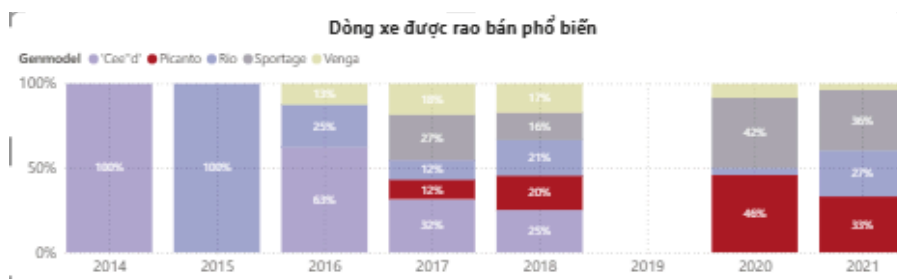




### Báo cáo chi tiết về dòng xe Picanto

Dựa vào báo cáo, với bộ lọc Hãng xe Kia, dòng xe KIA PICANTO thể hiện được lợi thế mạnh mẽ so với các dòng xe khác:

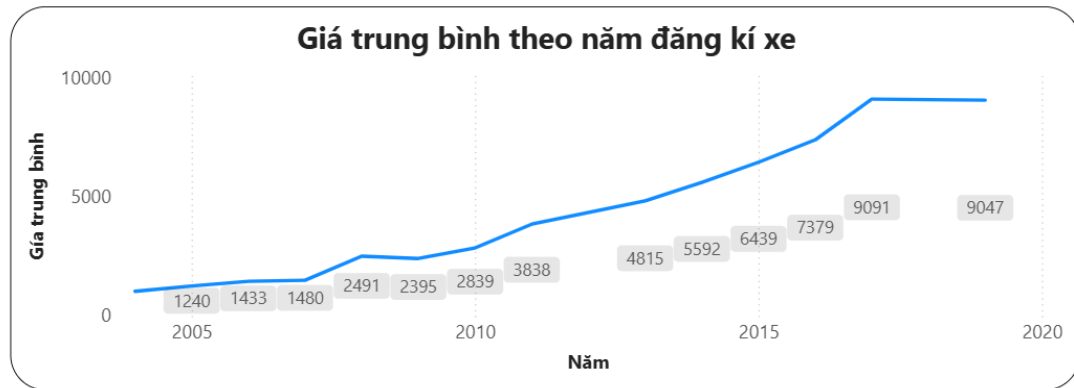
Nhận xét từ báo cáo về dòng xe Picanto:



Trong biểu đồ Dòng xe được rao bán phổ biến, Picanto chiếm tỷ lệ cao nhất từ 2019 đến 2021, có năm đạt tới 48%.

=>> Picanto là dòng có tính ổn định và phổ biến nhất – phù hợp chiến lược đầu tư dài hạn.

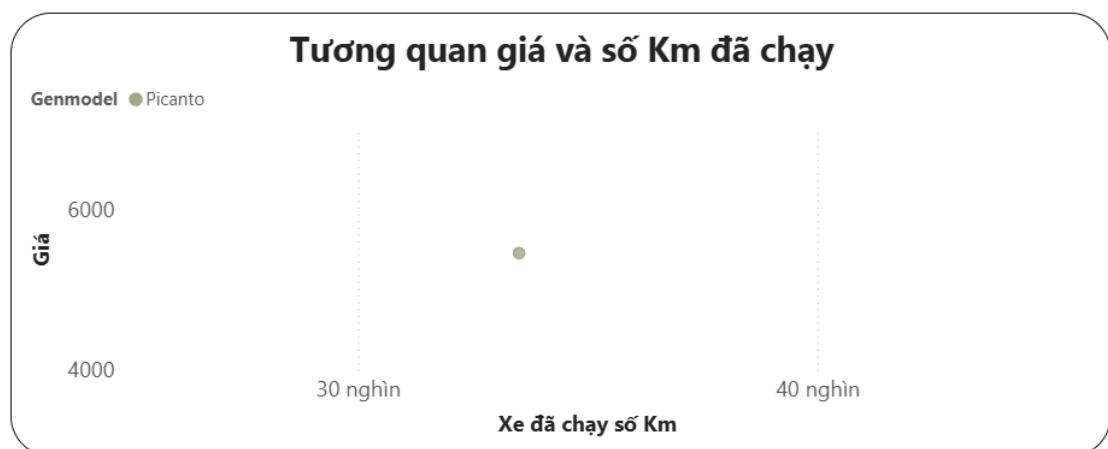
## 2. Giá trung bình theo năm đăng ký xe



- Tăng đều qua các năm, từ ~1,000 (2005) đến khoảng ~9,000 (sau 2018).
- Sau năm 2018, giá ổn định quanh mức cao (~9,000).

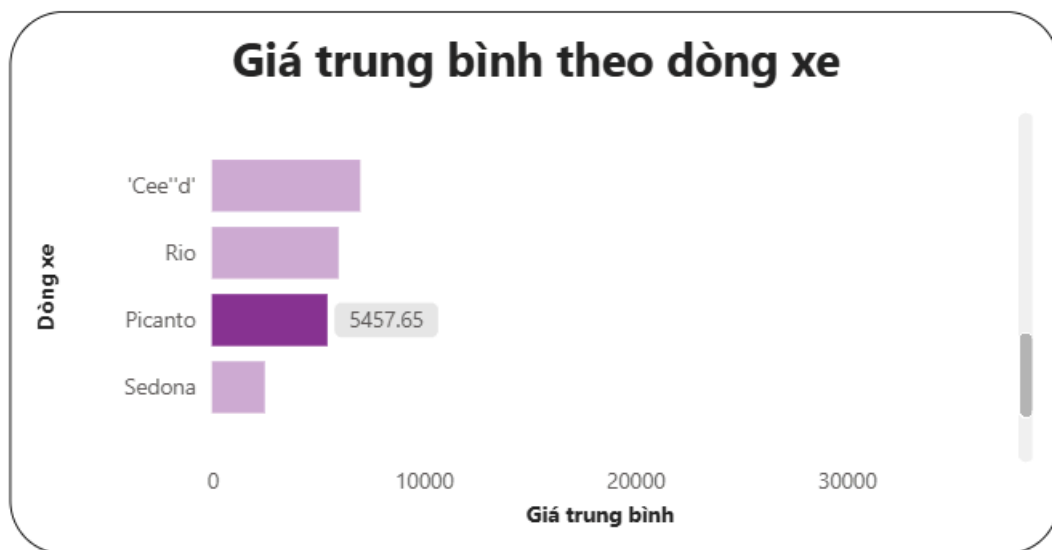
=> Xe Kia đăng ký gần đây giữ giá tốt, phù hợp đầu tư phân khúc xe mới 3–5 năm tuổi.

## 3. Tương quan giữa số Km đã chạy và Giá bán



- Với Picanto, điểm dữ liệu cho thấy: Km càng cao thì giá giảm (tương quan nghịch hợp lý).
- Biểu đồ còn đơn giản do ít điểm dữ liệu, cần mở rộng để có cái nhìn toàn diện hơn.
- => Cần chú ý số Km khi nhập Picanto – xe chạy dưới 30,000 km sẽ dễ bán và giá cao hơn.

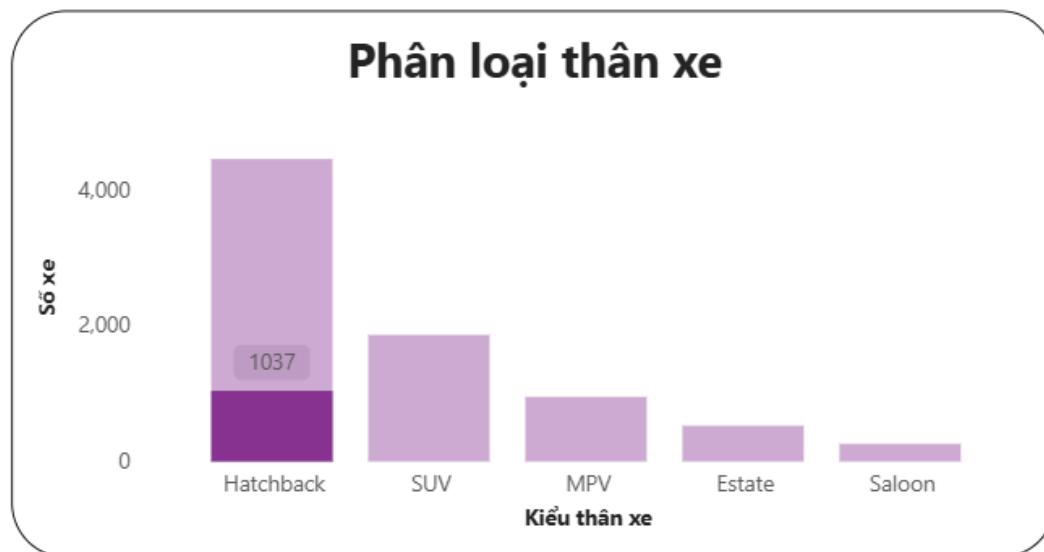
#### 4. Giá trung bình theo dòng xe



- Picanto có giá trung bình thấp, < 5,000.
- Các dòng đắt hơn: Pro\_Cee'd, Venga, Soul...

=> Picanto là lựa chọn tiết kiệm, dễ tiếp cận khách hàng phổ thông, phù hợp cho người mới bắt đầu kinh doanh xe.

#### 5. Phân loại thân xe

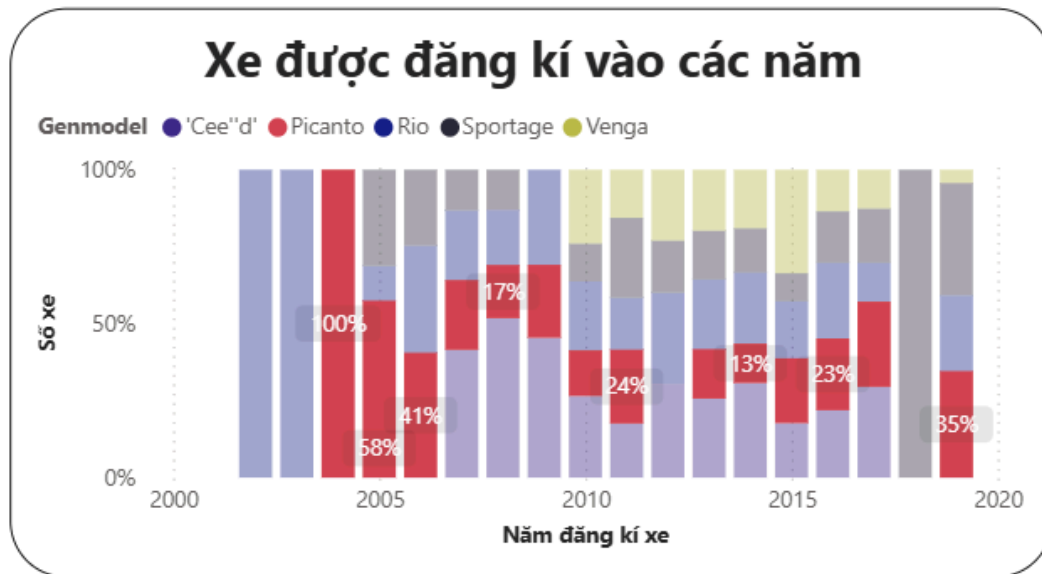


Picanto có thân xe chủ yếu là Hatchback là dòng xe đặc biệt phù hợp với môi trường đô thị nhờ thiết kế nhỏ gọn, dễ dàng luồn lách trong các con phố đông đúc – đặc biệt tại các thành phố lớn như TP.HCM hay Hà Nội. Với kích thước vừa phải, dòng xe này không chỉ giúp người dùng dễ dàng tìm chỗ đỗ xe mà còn mang lại khả năng tiết kiệm nhiên liệu tối ưu và dễ sử dụng cho người mới lái.

Bên cạnh đó, chi phí sở hữu Hatchback cũng khá thấp. Phần lớn các mẫu xe thuộc phân khúc này đều sử dụng động cơ dung tích nhỏ (dưới 1.8L), giúp tiết kiệm chi phí nhiên liệu trong quá trình sử dụng hàng ngày. Đồng thời, chi phí bảo trì và sửa chữa của Hatchback cũng ở mức hợp lý, rất phù hợp với nhóm khách hàng phổ thông hoặc các gia đình trẻ đang tìm kiếm một chiếc xe đô thị tiện dụng và kinh tế.

=> Kinh doanh dòng Hatchback như Picanto là hợp lý nhất cho Kia.

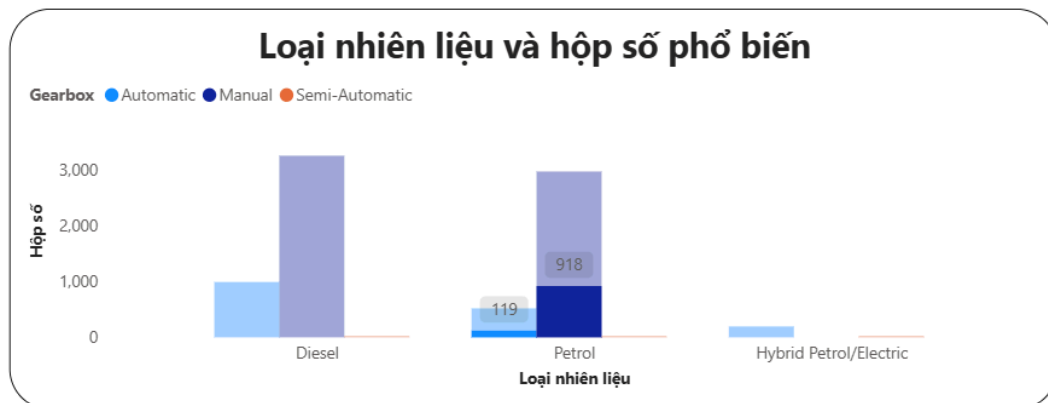
6. Xe được đăng kí vào các năm (theo dòng xe)



- Picanto xuất hiện ổn định, với tỷ lệ khá lớn.
- Các dòng khác như Cee'd, Sportage phân bố rải rác hơn.

=> Dòng Picanto có tính ổn định về nguồn cung – giúp dễ kiểm soát hàng nhập.

## 7. Loại nhiên liệu và hộp số phổ biến



- Xăng (Petrol) chiếm gần như toàn bộ → dễ tiếp nhiên liệu, chi phí thấp.

=> Cần ưu tiên chọn các xe Picanto xăng, số tự động để dễ bán và phù hợp thị trường.

Kết luận: Với sự phổ biến cao, chi phí đầu tư thấp, độ giữ giá ổn định, dễ sửa chữa và thân thiện với người sử dụng – Kia Picanto là lựa chọn lý

tưởng để khởi đầu kinh doanh xe cũ, đặc biệt tại các đô thị lớn như TP.HCM, nơi nhu cầu về xe nhỏ gọn, tiết kiệm và dễ sử dụng luôn cao.

Với logic tương tự, em lựa chọn các mẫu xe sau:

2. Hyundai – Santa Fe

3. Ford – Focus

4. Ford – Fiesta

5. Mazda – Mazda6

6. Nissan – Qashqai

7. Volkswagen – Passat

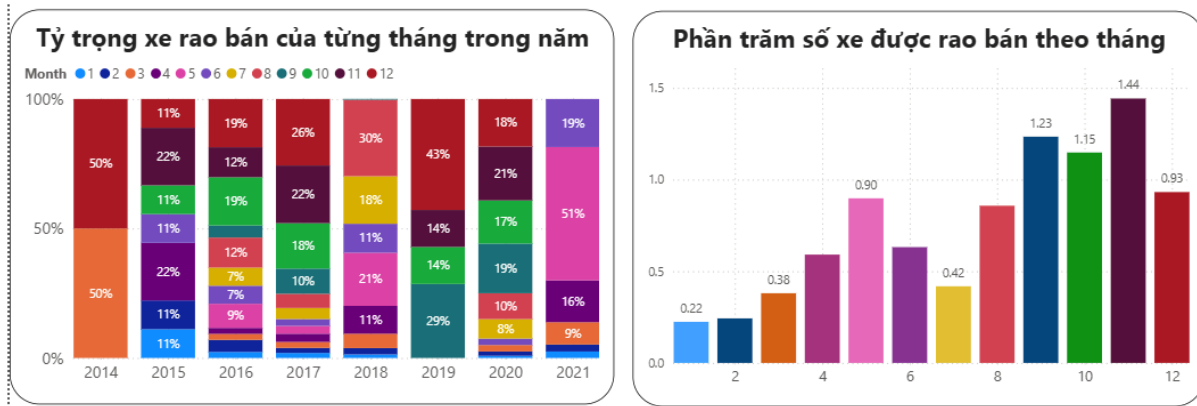
8. Peugeot – 308

9. Renault – Megane

10. Citroën – C3

**Câu 2. Có thời điểm nào trong năm mà lượng xe đăng bán tăng mạnh?  
Điều này ảnh hưởng gì đến chiến lược nhập – xuất xe?**

Dữ liệu ở đây cũng ảnh hưởng quyết định vì năm 2018 có số lượng rao bán gấp rất nhiều lần so với các năm khác => vì vậy ở đây em sẽ tiến hành tạo hàm DAX để Tính tổng % mà mỗi tháng chiếm trong từng năm, rồi cộng lại tất cả các năm → để biết tháng nào thường được rao bán nhiều nhất xét theo tỷ lệ % theo từng năm.



## 2 biểu đồ chính sử dụng để phân tích

Dựa trên biểu đồ “Tỷ trọng dòng xe được rao bán phổ biến theo các tháng” và “Phần trăm số xe được rao bán theo tháng”, ta có thể thấy rõ:

### 1. Thời điểm số lượng xe cũ được rao bán tăng mạnh

Theo biểu đồ bên phải:

- Tỷ lệ rao bán xe cũ cao nhất vào các tháng 10, 11 và 12:
  - Tháng 12: 1.44% (cao nhất)
  - Tháng 11: 1.35%
  - Tháng 10: 1.23%

Theo biểu đồ bên trái:

- Tỷ trọng rao bán ở các tháng cuối năm ngày càng chiếm ưu thế qua từng năm:
  - Năm 2020: tháng 10 chiếm đến 43%, tháng 11 chiếm 18%
  - Năm 2021: tháng 11 chiếm tới 51%

=> Kết luận: Cuối năm là mùa người dùng bán lại xe cũ nhiều nhất, do:

- Nhu cầu đổi xe đón Tết

- Cần tiền mặt cuối năm
- Doanh nghiệp thanh lý tài sản (xe công ty)

## **2. Chiến lược nhập xe**

Dựa trên số liệu, chị Linh nên: Tăng cường thu mua xe vào tháng 10–12:

- Đây là giai đoạn nguồn cung dồi dào, xe phong phú, dễ chọn lựa
- Cơ hội mặc cả tốt hơn, do cạnh tranh giữa người bán
- Có thể gom được nhiều xe chất lượng hơn với giá hợp lý

## **3. Chiến lược xuất xe**

Sau khi gom xe vào tháng 10–12, chị Linh có thể:

- Bán ngay trong mùa Tết Nguyên Đán: nhu cầu cao, biên lợi nhuận tốt
- Hoặc đầu năm sau (tháng 1–2): ít xe được rao bán hơn => ít cạnh tranh