**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**MÔN HỌC: Data Mining**

**BÁO CÁO CUỐI KÌ**

**ĐỀ TÀI: PHÂN TÍCH BỘ DỮ LIỆU NHÀ Ở AMES**

**GVHD: ThS. Nguyễn Văn Thành**

|  |  |
| --- | --- |
| **SVTH:** | **MSSV** |
| Lê Lương Trường An | 21133001 |
| Nguyễn Anh Tuấn | 21133113 |
| Trần Sĩ Nguyên | 21133059 |
| Hoàng Mạnh Đức | 21133027 |

**Lớp: DAMI330484\_23\_2\_02**

**Tp.Hồ Chí Minh, tháng 5 năm 202**

**BẢNG PHÂN CÔNG**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Họ và Tên** | **Nhiệm Vụ** | **Mức độ**  **Hoàn thành** | **Ưu Điểm** | **Nhược Điểm** |
| Lê Lương Trường An | + EDA và tiền xử lí dữ liệu + Thực hiện xây dựng và đánh giá mô hình Ridge Regression và K- Nearest Neighbors + Phân tích so sánh | 100% | +Hoàn thành tốt được nhiệm vụ làm mô hình trên.  +Tham gia tích cực trong làm việc nhóm.  +Hỗ trợ nhau trong việc thực hiện các mô hình. | Trong quá trình làm vẫn còn nhiều thiếu sót,mô hình chưa được hoàn thiện tốt. |
| Nguyễn Anh Tuấn | + EDA và tiền xử lí dữ liệu + Thực hiện xây dựng và đánh giá mô hình Support Vector Regression và Neural etwork + Phân tích so sánh | 100% | +Hoàn thành tốt được nhiệm vụ làm mô hình trên.  +Tham gia tích cực trong làm việc nhóm.  + | Trong quá trình làm vẫn còn nhiều thiếu sót,mô hình chưa được hoàn thiện tốt. |
| Trần Sĩ Nguyên | + EDA và tiền xử lí dữ liệu + Thực hiện xây dựng và đánh giá mô hình Ridge Gradient Boosting và Elastic Net + Phân tích so sánh | 100% | +Hoàn thành tốt được nhiệm vụ làm mô hình trên.  +Tham gia tích cực trong làm việc nhóm.  +Hỗ trợ nhau trong việc thực hiện các mô hình. | Trong quá trình làm vẫn còn nhiều thiếu sót,mô hình chưa được hoàn thiện tốt. |
| Hoàng Mạnh Đức | + EDA và tiền xử lí dữ liệu + Thực hiện xây dựng và đánh giá mô hình Decision Tree và Random Forest + Phân tích so sánh | 100% | +Hoàn thành tốt được nhiệm vụ làm mô hình trên.  +Tham gia tích cực trong làm việc nhóm. | Trong quá trình làm vẫn còn nhiều thiếu sót,mô hình chưa được hoàn thiện tốt. |

**Nhận xét của giáo viên:**

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

......................................................................................................................

*Ngày ......... tháng ......... năm .........*

*Giáo viên chấm điểm*

**MỤC LỤC**

[1. Giới thiệu về tập dữ liệu 1](#_Toc167998715)

[1.1 Bối cảnh 1](#_Toc167998716)

[1.2 Tập dữ liệu 1](#_Toc167998717)

[2. Làm sạch dữ liệu 4](#_Toc167998718)

[2.1. Xử lý dữ liệu khuyết 4](#_Toc167998719)

[2.2. Loại bỏ các ngoại lệ 7](#_Toc167998720)

[2.3. Xóa một số cột không quan trọng 9](#_Toc167998721)

[3. Báo cáo Phân tích Dữ liệu Khám phá (EDA) 9](#_Toc167998722)

[3.1. Phân phối biến mục tiêu 9](#_Toc167998723)

[3.2. Mối tương quan giữa các biến 10](#_Toc167998724)

[3.3. Mối quan hệ giữa biến mục tiêu và các biến khác 10](#_Toc167998725)

[3.4. Mối quan hệ giữa các biến dự đoán 11](#_Toc167998726)

[3.5. Feature Engineering 11](#_Toc167998727)

[4. Phân cụm bằng phương pháp K-Mean 11](#_Toc167998728)

[4.1. Dùng phương pháp elbow để xác định số cụm nên chia 12](#_Toc167998729)

[4.2. Sử dụng KMEAN 13](#_Toc167998730)

[5. Dự đoán giá nhà bằng các mô hình học máy 15](#_Toc167998731)

[5.1. Các mô hình sử dụng 15](#_Toc167998732)

[5.2. Các bước thực hiện 15](#_Toc167998733)

[Feature Scaling 15](#_Toc167998734)

[Chia tập train test 15](#_Toc167998735)

[Dùng mô hình 15](#_Toc167998736)

[5.3. Minh họa một mô hình 16](#_Toc167998737)

[5.4. Tìm ra mô hình tốt nhất 17](#_Toc167998738)

[5.5. Tìm ra các thuộc tính có ảnh hưởng nhất đến mô hình 18](#_Toc167998739)

[6. Kết luận 18](#_Toc167998740)

# 1. Giới thiệu về tập dữ liệu

## ****1.1 Bối cảnh****

Hàng nghìn ngôi nhà được giao dịch mỗi ngày, và mỗi người mua đều đối mặt với những câu hỏi then chốt như: "Giá trị thực sự của ngôi nhà này là bao nhiêu?" và "Liệu tôi có đang trả một mức giá công bằng hay không?" Đây là những câu hỏi thiết yếu mà mỗi người mua đều tự hỏi khi đứng trước quyết định quan trọng về tài chính và cuộc sống của mình.

Trong bối cảnh này, Nhóm xin giới thiệu một mô hình học máy tiên tiến được thiết kế để dự đoán giá nhà một cách chính xác, dựa trên các thông tin liên quan đến ngôi nhà như kích thước, năm xây dựng, và nhiều yếu tố khác. Mô hình này không chỉ giúp người mua hiểu rõ hơn về giá trị thực sự của ngôi nhà mà họ quan tâm, mà còn giúp họ đưa ra quyết định thông minh hơn, đảm bảo rằng họ sẽ không trả quá cao cho ngôi nhà mà họ yêu thích.

Quá trình phát triển và đánh giá mô hình của chúng tôi sẽ được trình bày chi tiết, với mỗi bước được minh họa bằng mã code cụ thể cùng với kết quả đầu ra tương ứng. Điều này không chỉ giúp đảm bảo tính minh bạch trong nghiên cứu của chúng tôi mà còn tạo điều kiện thuận lợi cho việc tái tạo lại quy trình này bởi các nhà nghiên cứu và nhà phát triển khác.

Trong báo cáo này, nhóm chúng em sử dụng ngôn ngữ lập trình Python, một công cụ mạnh mẽ và linh hoạt trong lĩnh vực khoa học dữ liệu và học máy. Nhóm chúng em cũng tận dụng một số gói Python quan trọng để xây dựng và tối ưu hóa mô hình của mình, bao gồm Pandas, NumPy, Scikit-learn, và Matplotlib.

Bằng cách kết hợp kiến thức chuyên sâu về học máy với việc sử dụng các công cụ mạnh mẽ của Python, nhóm chúng em tin rằng mô hình dự đoán giá nhà của mình sẽ mang lại giá trị thực tiễn cao, giúp người mua nhà đưa ra những quyết định sáng suốt và tự tin hơn.

## 1.2 Tập dữ liệu

Trong báo cáo này, nhóm sẽ sử dụng bộ dữ liệu về nhà ở được trình bày bởi De Cock (2011). Bộ dữ liệu này mô tả các giao dịch bán nhà ở tại Ames, Iowa từ năm 2006 đến năm 2010. Bộ dữ liệu chứa một số lượng lớn các biến liên quan đến việc xác định giá nhà

Tập dữ liệu gồm: 2930 dòng và 82 cột.

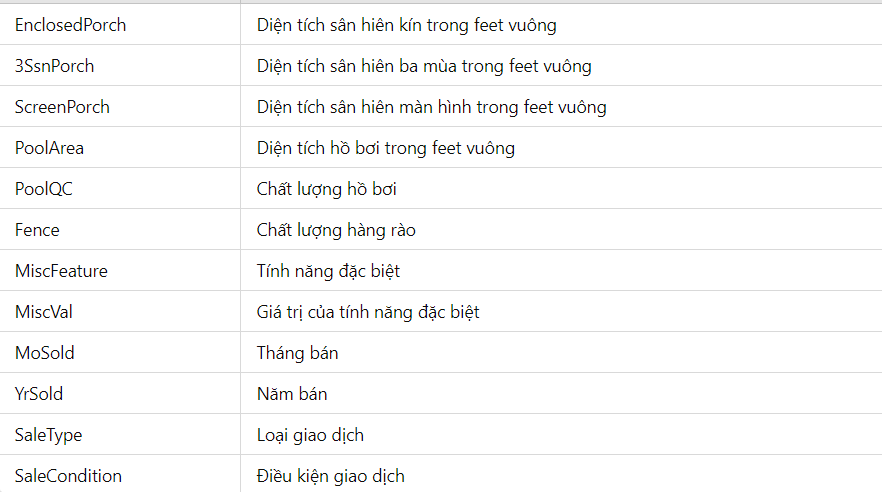








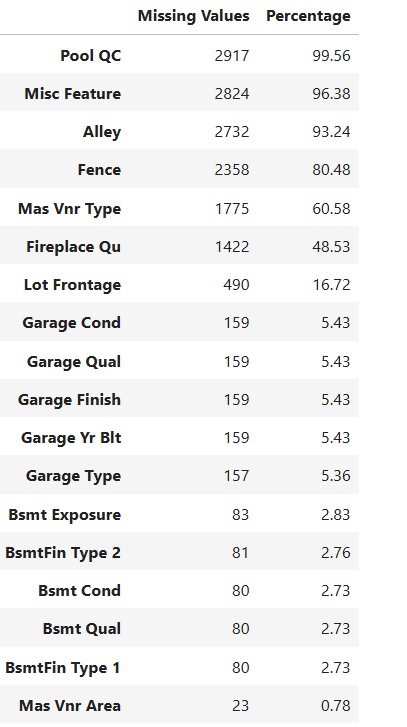


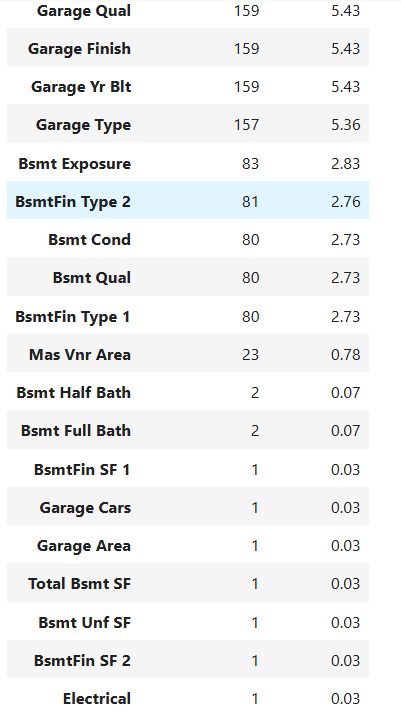


# Làm sạch dữ liệu

## 2.1. Xử lý dữ liệu khuyết

Chúng ta nên giải quyết vấn đề thiếu giá trị vì một số mô hình học máy không chấp nhận dữ liệu thiếu giá trị. Đầu tiên, hãy xem số lượng giá trị còn thiếu trong tập dữ liệu của chúng tôi. Ta sẽ xem gái trị phần tram của các giá trị bị thiếu cho mỗi cột thực sự chứa các giá trị bị thiếu (theo thứ tự giảm dần).

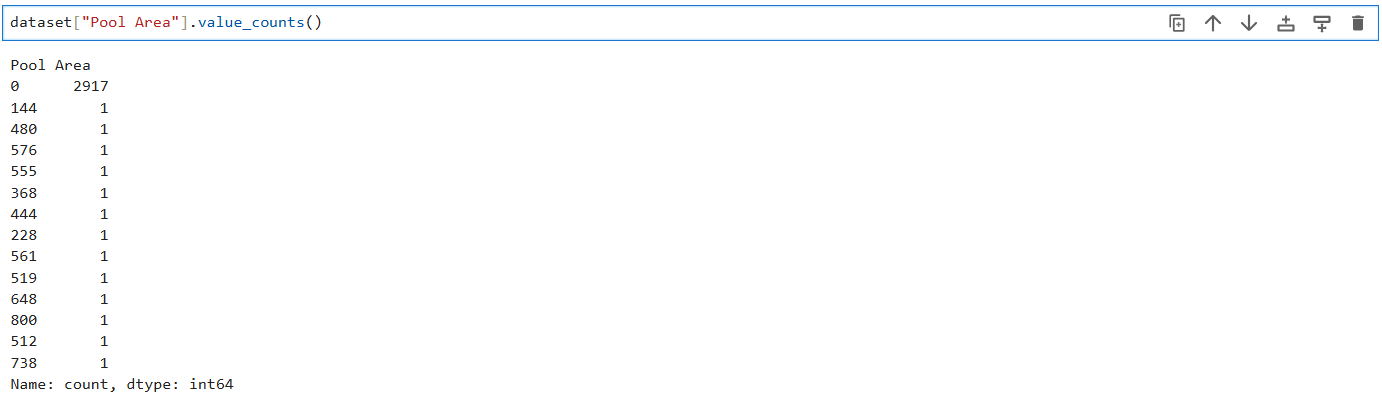




Tùy thuộc vào thuộc tính của dữ liệu bị khuyết thì chúng ta sẽ điền giá trị khuyết hợp lý.

Ví dụ một trường hơp như cột Pool QC

Tỷ lệ phần trăm giá trị bị thiếu trong cột Pool QC là 99,56%, tỷ lệ này rất cao. Giá trị còn thiếu trong cột này biểu thị rằng ngôi nhà tương ứng không có hồ bơi. Để xác minh điều này, chúng ta hãy xem các giá trị của cột Pool Area



Chúng ta có thể thấy rằng có 2917 mục trong cột Pool Area có giá trị 0. Điều này xác minh giả thuyết rằng mỗi ngôi nhà không có hồ bơi đều thiếu giá trị trong cột Pool QC và giá trị 0 trong Pool Area cột. Vì vậy, hãy điền các giá trị còn thiếu vào cột Pool QC bằng "No Pool".

Các trường hợp khác điền khuyết tương tự như ở cột Pool QC

## 2.2. Loại bỏ các ngoại lệ

Có nhiều lý do khiến việc loại bỏ ngoại lệ (outlier) trong một tập dữ liệu là cần thiết, tùy thuộc vào mục đích sử dụng dữ liệu và phương pháp phân tích ta sử dụng. Dưới đây là một số lý do chính:

1. Cải thiện độ chính xác của mô hình: Ngoại lệ có thể ảnh hưởng đáng kể đến độ chính xác của các mô hình học máy và thống kê. Do nằm ngoài xu hướng chung của dữ liệu, chúng có thể khiến mô hình học sai quy luật và đưa ra dự đoán sai lệch. Loại bỏ ngoại lệ có thể giúp giảm thiểu ảnh hưởng của chúng và cải thiện hiệu suất tổng thể của mô hình.

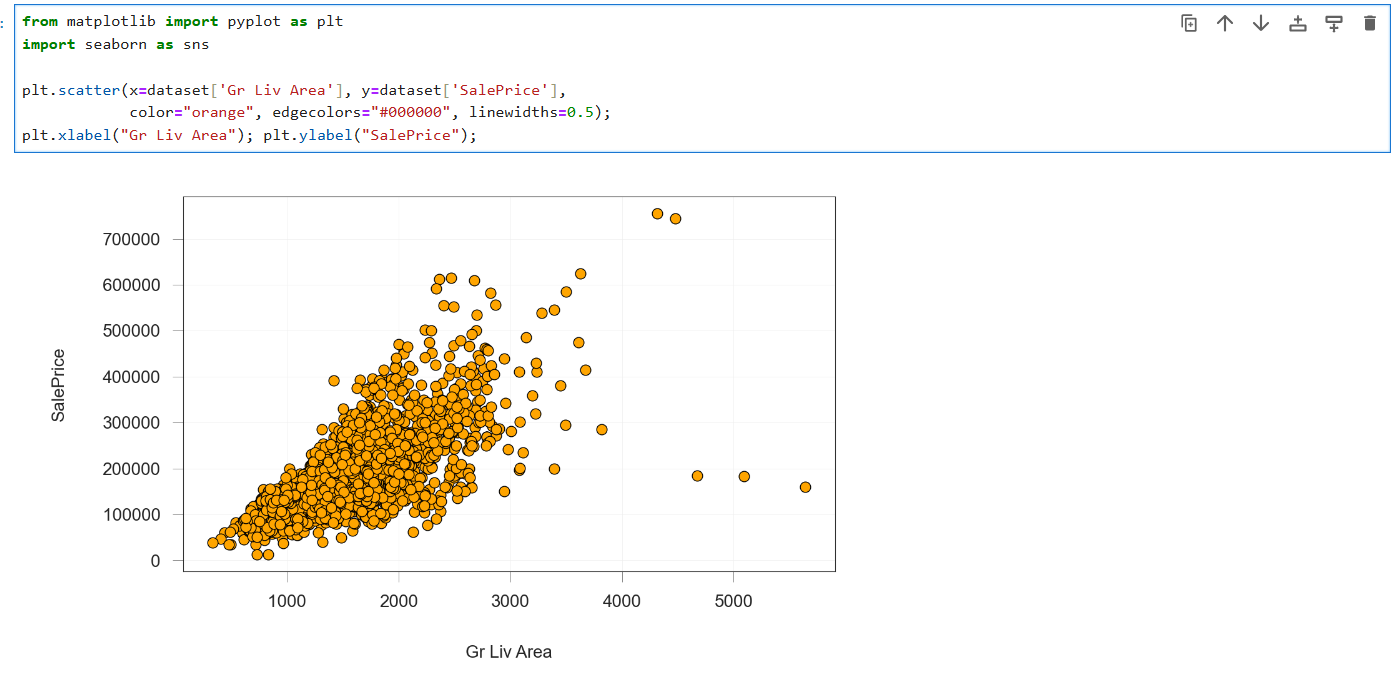
2. Tăng độ tin cậy của kết quả: Khi thực hiện phân tích dữ liệu, việc loại bỏ ngoại lệ có thể giúp đảm bảo độ tin cậy và chính xác của kết quả. Ngoại lệ có thể làm sai lệch các phép tính thống kê và che khuất các mẫu quan trọng trong dữ liệu. Loại bỏ chúng có thể giúp bạn có được bức tranh rõ ràng hơn về xu hướng thực sự trong dữ liệu và đưa ra kết luận chính xác hơn.

3. Loại bỏ nhiễu dữ liệu: Ngoại lệ có thể xuất phát từ nhiều nguồn khác nhau, chẳng hạn như lỗi đo lường, nhập dữ liệu sai sót hoặc các sự kiện bất thường. Loại bỏ chúng có thể giúp loại bỏ nhiễu dữ liệu và cải thiện chất lượng tổng thể của tập dữ liệu.

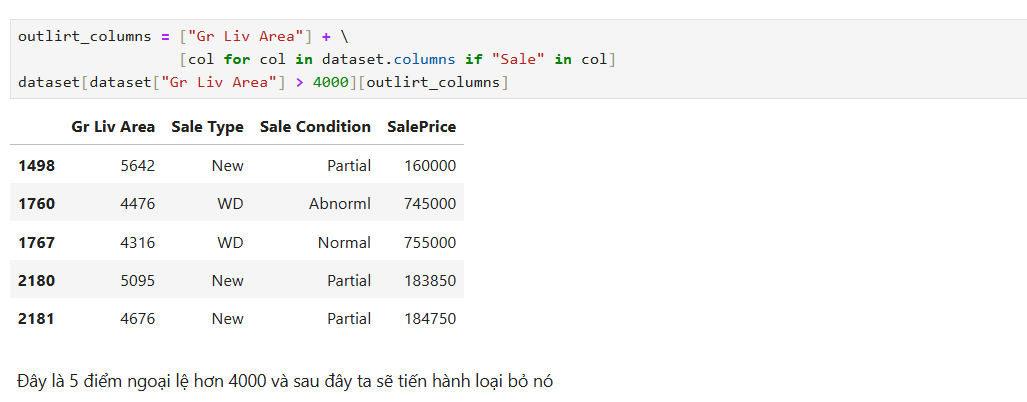
4. Chuẩn bị dữ liệu cho các thuật toán: Một số thuật toán học máy và thống kê nhạy cảm với ngoại lệ và có thể bị ảnh hưởng tiêu cực bởi chúng. Loại bỏ ngoại lệ trước khi áp dụng các thuật toán này có thể giúp đảm bảo rằng chúng hoạt động chính xác và hiệu quả.

5. Dễ dàng trực quan hóa dữ liệu: Khi trực quan hóa dữ liệu, ngoại lệ có thể che khuất các mẫu quan trọng và khiến việc giải thích dữ liệu trở nên khó khăn. Loại bỏ chúng có thể giúp tạo ra các biểu đồ và đồ thị rõ ràng hơn, dễ hiểu hơn.

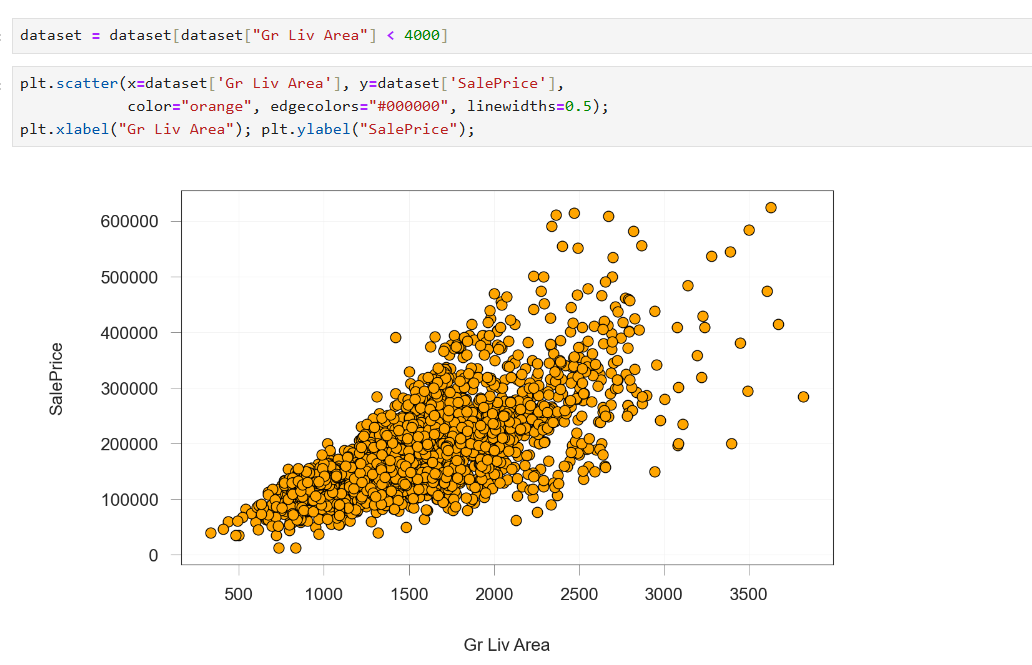
Chúng ta lập biểu đồ `SalePrice` dựa trên `Gr Liv Area` để phát hiện các điểm ngoại lệ.



Từ điểm dữ liệu ta thấy các điểm dữ liệu trên 4000 có các điểm ngoại lệ. Bây giờ, chúng tôi sẽ xóa chúng khỏi tập dữ liệu. Ta có thể làm như vậy bằng cách giữ các điểm dữ liệu có Gr Liv Area dưới 4.000. Nhưng trước tiên chúng ta hãy xem các hàng tập dữ liệu tương ứng với các giá trị bất thường này:



Kết quả sau khi xóa bỏ



## Xóa một số cột không quan trọng

Ở đây có 2 cột cần xóa là Order và PID bởi nó không tác động đến sự phân tích hay mô hình:



# 3. Báo cáo Phân tích Dữ liệu Khám phá (EDA)

## 3.1. Phân phối biến mục tiêu

Trong tập dữ liệu của chúng tôi, biến quan trọng nhất cần khám phá là biến mục tiêu (SalePrice). Để hiểu rõ hơn về phân bố của nó, chúng tôi bắt đầu bằng việc vẽ biểu đồ violin cho biến này. Chiều rộng của violin biểu thị tần số của các điểm dữ liệu trong từng khoảng giá. Ví dụ, nếu chiều rộng của biểu đồ lớn nhất ở khoảng từ 300.000 đến 400.000, điều đó có nghĩa là có nhiều điểm dữ liệu nằm trong khoảng giá này.

**Kết quả từ đồ thị violin:**

- Hầu hết giá nhà rơi vào khoảng từ 100.000 đến 250.000.

- Các đường đứt nét trên đồ thị biểu thị vị trí của ba phần tư: Q1, Q2 (trung vị) và Q3.

Sau đó, chúng tôi vẽ sơ đồ hình hộp (box plot) của SalePrice để hiểu rõ hơn về giá trị tối thiểu, tối đa và các phần tư của SalePrice.

## 3.2. Mối tương quan giữa các biến

Chúng tôi muốn xem các biến trong tập dữ liệu có mối tương quan với nhau như thế nào và đặc biệt là mối tương quan giữa các biến dự đoán và biến mục tiêu (SalePrice).

Ví dụ về mối tương quan giữa Lot Area và SalePrice:

- Mối tương quan dương: Khi Lot Area tăng, SalePrice cũng tăng.

- Mối tương quan âm: Khi Lot Area tăng, SalePrice giảm.

- Không tương quan: Không có mối quan hệ rõ ràng giữa Lot Area và SalePrice.

Mối tương quan được biểu thị bằng giá trị từ -1 đến +1, trong đó +1 biểu thị mối tương quan dương cao nhất, -1 biểu thị mối tương quan âm cao nhất và 0 biểu thị không có mối tương quan.

**Các mối tương quan đáng chú ý trong tập dữ liệu:**

* **Tương quan dương cao:**

- Garage Cars và Garage Area có mối tương quan dương cao vì khi diện tích gara tăng, sức chứa ô tô cũng tăng.

- Gr Liv Area và TotRms AbvGrd có mối tương quan dương cao vì khi diện tích sống trên mặt đất tăng, số phòng cũng tăng.

* **Tương quan nghịch:**

- Bsmt Unf SF có tương quan nghịch với BsmtFin SF 1: Khi diện tích tầng hầm chưa hoàn thiện tăng, diện tích tầng hầm đã hoàn thiện giảm.

- Bsmt Unf SF có tương quan nghịch với Bsmt Full Bath: Khi diện tích tầng hầm chưa hoàn thiện tăng, số lượng phòng tắm hoàn chỉnh giảm.

* **Tương quan với biến mục tiêu (SalePrice):**

- Overall Qual và Gr Liv Area có mối tương quan thuận chiều cao với SalePrice.

- Các biến khác như Year Built, Year Remod/Add, Mas Vnr Area, Total Bsmt SF, 1st Flr SF, Full Bath, Garage Cars và Garage Area cũng có mối tương quan thuận với SalePrice.

## 3.3. Mối quan hệ giữa biến mục tiêu và các biến khác

**Tương quan dương cao:**

- Mối quan hệ giữa SalePrice và Overall Qual: Tỷ lệ thuận, giá nhà tăng khi chất lượng tổng thể tăng.

- Mối quan hệ giữa SalePrice và Gr Liv Area: Tỷ lệ thuận, giá nhà tăng khi diện tích sống tăng.

**Tương quan dương vừa phải:**

- Các biến như Year Built, Year Remod/Add, Mas Vnr Area, Total Bsmt SF, 1st Flr SF, Full Bath, Garage Cars và Garage Area có mối tương quan dương vừa phải với SalePrice.

## 3.4. Mối quan hệ giữa các biến dự đoán

**Tương quan dương:**

- Garage Cars và Garage Area: Dữ liệu tập trung cao nhất khi Garage Cars là 2 và Garage Area nằm trong khoảng từ 450 đến 600 ft².

- Gr Liv Area và TotRms AbvGrd: Dữ liệu tập trung cao nhất khi Gr Liv Area nằm trong khoảng từ 800 đến 2000 ft² và TotRms AbvGrd là 6.

**Tương quan âm:**

- Bsmt Unf SF và BsmtFin SF 1: Khi diện tích tầng hầm chưa hoàn thiện tăng, diện tích tầng hầm đã hoàn thiện giảm.

- Bsmt Unf SF và Bsmt Full Bath: Không có mối quan hệ rõ ràng, nhưng có xu hướng diện tích tầng hầm chưa hoàn thiện tăng, số lượng phòng tắm hoàn chỉnh giảm.

## 3.5. Feature Engineering

**Tạo các tính năng phái sinh mới:**

- Tạo các đặc điểm đa thức từ Overall Qual và Gr Liv Area: Bình phương và lập phương của các giá trị ban đầu.

- Tạo một tính năng là tích của Overall Qual và Gr Liv Area.

**Loại bỏ các đặc điểm có mối tương quan cao để tránh đa cộng tuyến:**

- Loại bỏ Garage Cars vì nó có mối tương quan cao với Garage Area.

- Loại bỏ TotRms AbvGrd vì nó có mối tương quan cao với Gr Liv Area.

**Xử lý các biến thứ tự:**

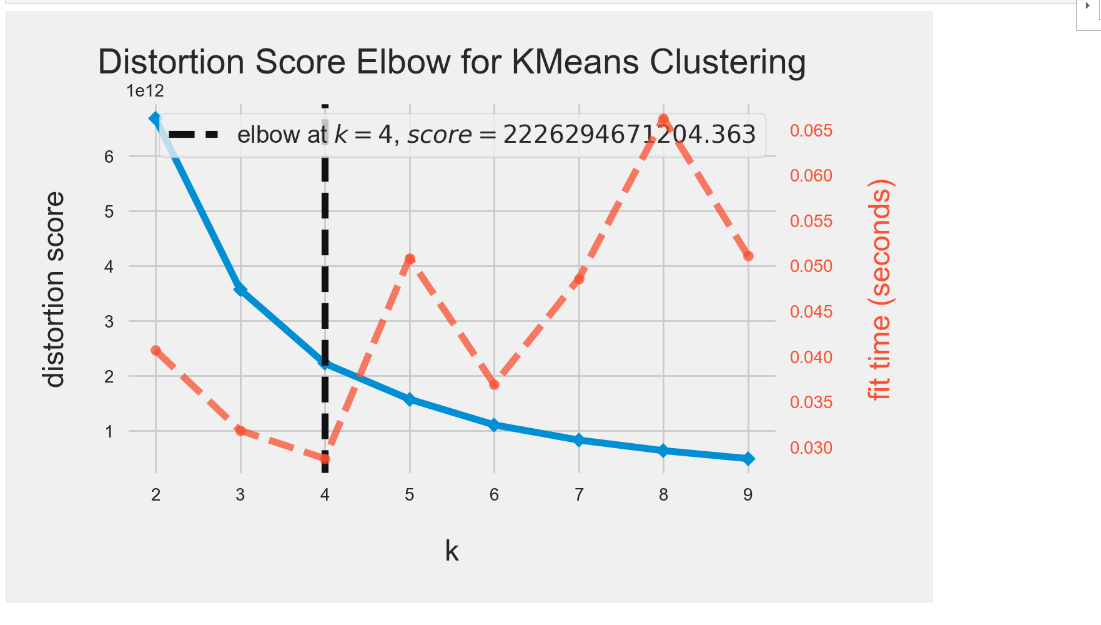
- Biến thứ tự như Bsmt Cond, Exter Qual, Exter Cond, Bsmt Qual, v.v., sẽ được ánh xạ thành các giá trị số để các mô hình học máy hiểu rằng chúng đại diện cho thứ hạng.

Việc chuyển đổi và ánh xạ này giúp mô hình học máy xử lý dữ liệu hiệu quả hơn và hiểu đúng về thứ tự các giá trị.

# 4. Phân cụm bằng phương pháp K-Mean

Để có cái nhìn tổng quát hơn về sự phân bố giá nahf qua các thuộc tính khác nhau trong tập dữ liệu.

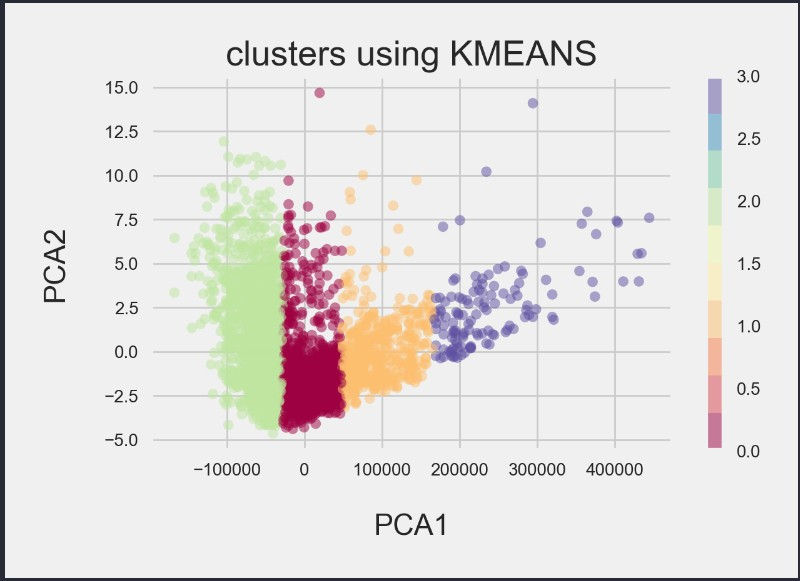
## 4.1. Dùng phương pháp elbow để xác định số cụm nên chia

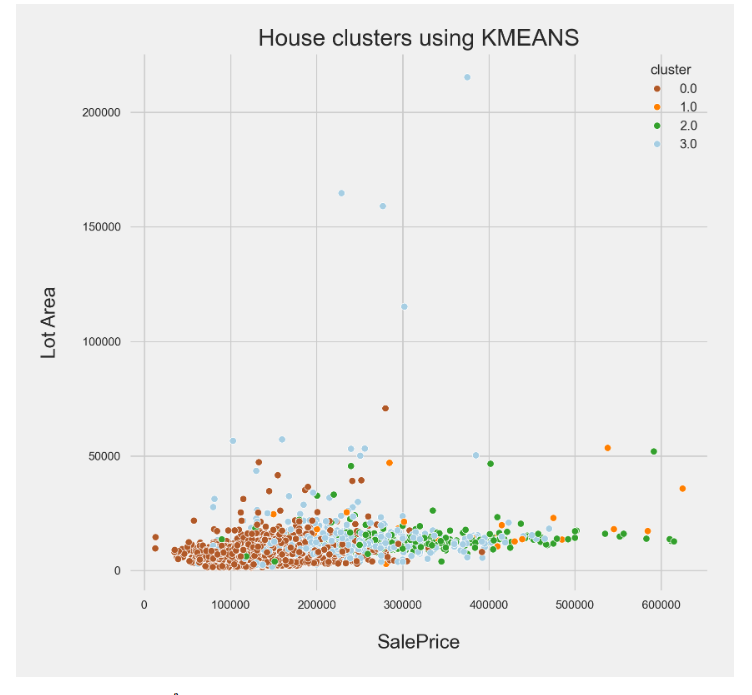


Ta có thể chia thành 4 cụm dựa vào elbow

## 4.2. Sử dụng KMEAN







# Dự đoán giá nhà bằng các mô hình học máy

## 5.1. Các mô hình sử dụng

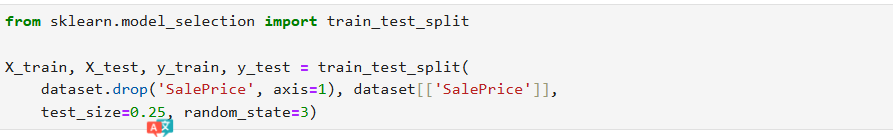
* Linear Regression(Ridge Regression, Elastic Net)
* Nearest Neighbors
* Support Vector Regression
* Decision Trees
* Neural Networks
* Random Forest
* Gradient Boosting

## 5.2. Các bước thực hiện

### Feature Scaling

Để làm cho tất cả các thuật toán hoạt động chính xác với dữ liệu của Nhóm, Nhóm cần mở rộng quy mô các tính năng trong tập dữ liệu của mình. Để làm được điều đó, chúng ta sẽ sử dụng một hàm hữu ích có tên StandardScaler() từ gói Python Scikit-Learn phổ biến. Hàm này chuẩn hóa các đặc điểm bằng cách trừ giá trị trung bình và chia tỷ lệ thành phương sai đơn vị. Nó hoạt động trên từng tính năng một cách độc lập

### Chia tập train test



### Dùng mô hình

Đối với mỗi kỹ thuật được đề cập trong phần trước (Hồi quy tuyến tính, Hàng xóm gần nhất, Máy vectơ hỗ trợ, v.v.), Nhóm sẽ làm theo các bước sau để xây dựng mô hình:

* Chọn thuật toán thực hiện kỹ thuật tương ứng
* Tìm kiếm tổ hợp tham số hiệu quả cho thuật toán đã chọn
* Tạo mô hình sử dụng các tham số tìm được
* Huấn luyện (fit) mô hình trên tập dữ liệu huấn luyện
* Kiểm thử mô hình trên tập dữ liệu thử nghiệm và nhận kết quả

**Tìm tham số phù hơp**

Sử dụng Scikit-Learn, Nhóm có thể xây dựng mô hình ví dụ như cây quyết định chẳng hạn như sau:

model = DecisionTreeRegressor(max\_depth=14, min\_samples\_split=5, max\_features=20)

Có thể làm điều này nhưng có thể đạt được hiệu suất tốt hơn nếu chọn các giá trị tốt hơn cho các tham số max\_deep, min\_samples\_split và max\_features. Để làm như vậy, Nhóm sẽ kiểm tra nhiều kết hợp tham số và chọn kết hợp mang lại điểm cao nhất. Scikit-Learn cung cấp một chức năng hữu ích cho mục đích đó: GridSearchCV(). Vì vậy, với ví dụ trên, chúng ta sẽ làm như sau:

parameter\_space = {

"max\_depth": [7, 15],

"min\_samples\_split": [5, 10],

"max\_features": [30, 45]

}

clf = GridSearchCV(DecisionTreeRegressor(), parameter\_space, cv=4,

scoring="neg\_mean\_absolute\_error")

clf.fit(X\_train, y\_train)

Đôi khi, khi số lượng kết hợp tham số lớn, GridSearchCV() có thể mất rất nhiều thời gian để chạy. Vì vậy, ngoài GridSearchCV(), đôi khi sẽ sử dụng RandomizedSearchCV() tương tự như GridSearchCV() nhưng thay vì sử dụng tất cả các kết hợp tham số, nó chọn một số kết hợp ngẫu nhiên được chỉ định bởi n\_iter. Đối với ví dụ trên, chúng ta có thể sử dụng RandomizedSearchCV() như sau:

clf = RandomizedSearchCV(DecisionTreeRegressor(), parameter\_space, cv=4,

scoring="neg\_mean\_absolute\_error", n\_iter=100)

Điều này sẽ làm cho RandomizedSearchCV() chọn ngẫu nhiên 100 kết hợp tham số.

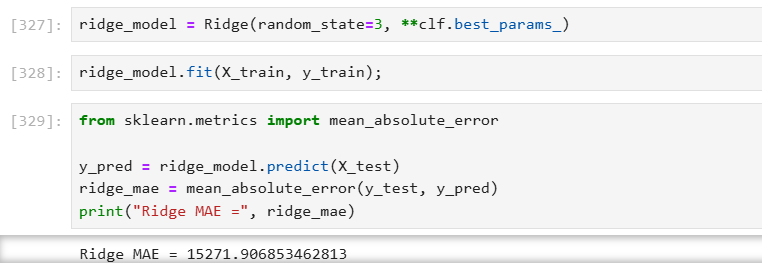
## 5.3. Minh họa một mô hình

**Các mô hình khác cũng tương tự**.

Nhóm sẽ tạo ra các bộ các tham số đầu vào cho mô hình trong biến parameter\_space. Sau đó GridSearchCV sẽ tìm ra bộ tham số tối ưu nhất.

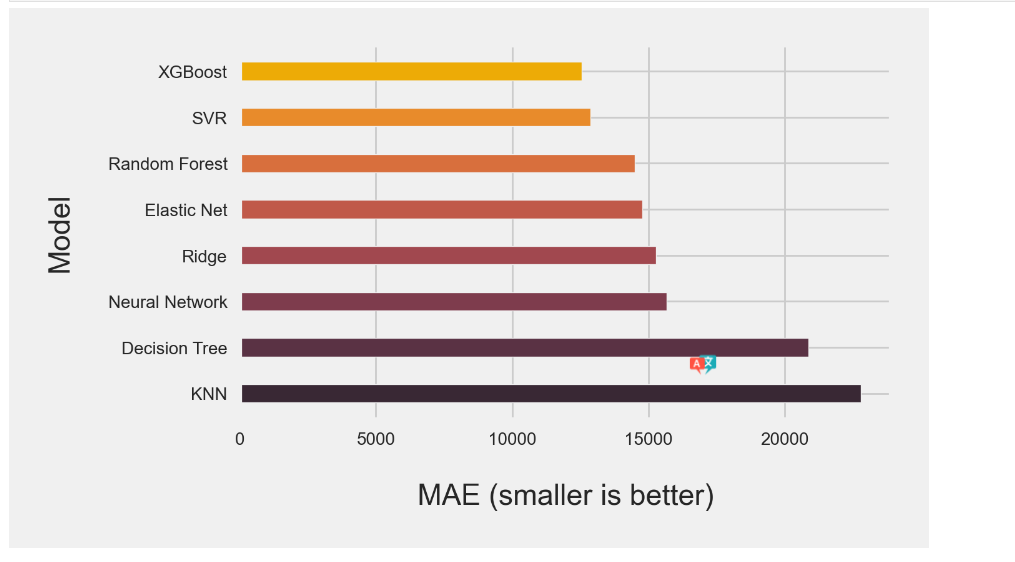


Sau đó dùng bộ tham số này để đưa vào mô hình sau đó tính MAE cho mô hình

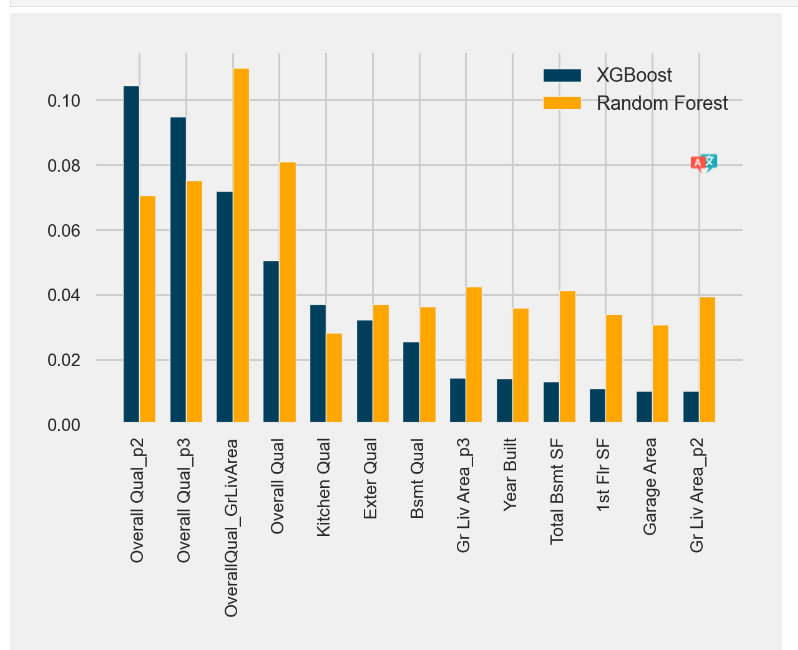


## Tìm ra mô hình tốt nhất

Mô hình tốt nhất là mô hình có MAE nhỏ nhất trong tất cả các mô hình:



## Tìm ra các thuộc tính có ảnh hưởng nhất đến mô hình



# 6. Kết luận

Trong báo cáo này, nhóm chúng em đã xây dựng một số mô hình hồi quy để dự đoán giá của một số căn nhà dựa trên một số đặc điểm của căn nhà đó. Nhóm chúng em đã đánh giá và so sánh từng mô hình để xác định mô hình có hiệu suất cao nhất. Nhóm chúng em cũng đã xem xét cách mà một số mô hình xếp hạng các đặc điểm theo mức độ quan trọng của chúng. Trong bài viết này, nhóm chúng em đã tuân theo quy trình khoa học dữ liệu bắt đầu từ việc thu thập dữ liệu, sau đó là làm sạch và tiền xử lý dữ liệu, tiếp theo là khám phá dữ liệu và xây dựng các mô hình, sau đó là đánh giá kết quả và truyền đạt chúng thông qua các biểu đồ.

Như một khuyến nghị, nhóm chúng em đề xuất sử dụng mô hình này (hoặc một phiên bản của nó được huấn luyện với dữ liệu mới hơn) bởi những người muốn mua nhà trong khu vực được bao phủ bởi tập dữ liệu để có một ý tưởng về giá thực tế. Mô hình cũng có thể được sử dụng với các tập dữ liệu mà bao gồm các thành phố và khu vực khác nhau miễn là chúng chứa các đặc điểm giống nhau. Nhóm chúng em cũng đề xuất rằng mọi người cân nhắc các đặc điểm đã được xem xét là quan trọng nhất như đã thấy trong phần trước; điều này có thể giúp họ ước lượng giá nhà tốt hơn.