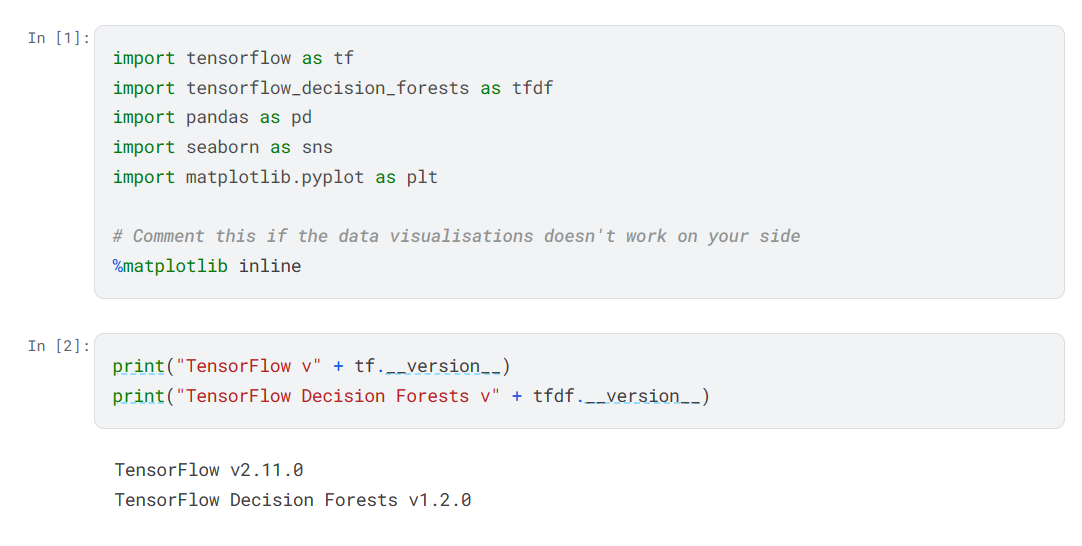
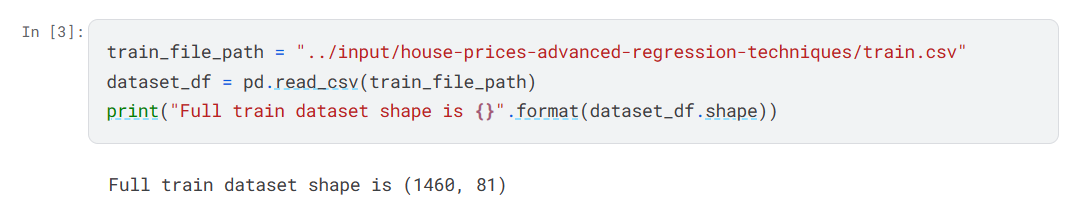
1. Miêu tả dữ liệu
   1. Tổng quan

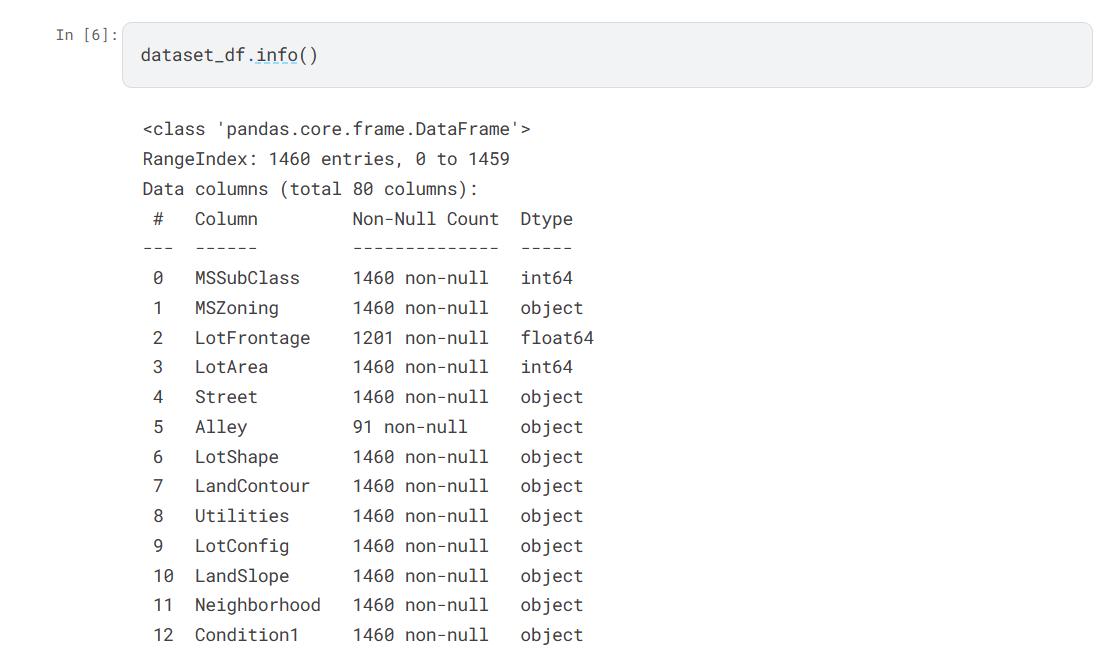
Nguồn dữ liệu: Dữ liệu được cung cấp bởi giảng viên bộ môn

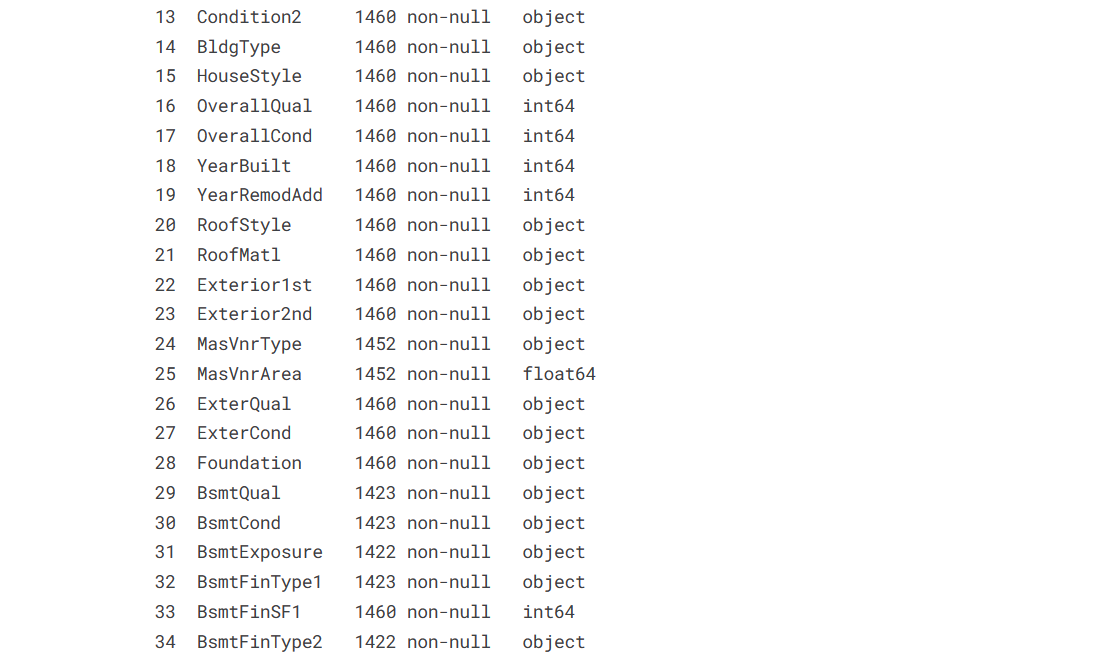
Import the library

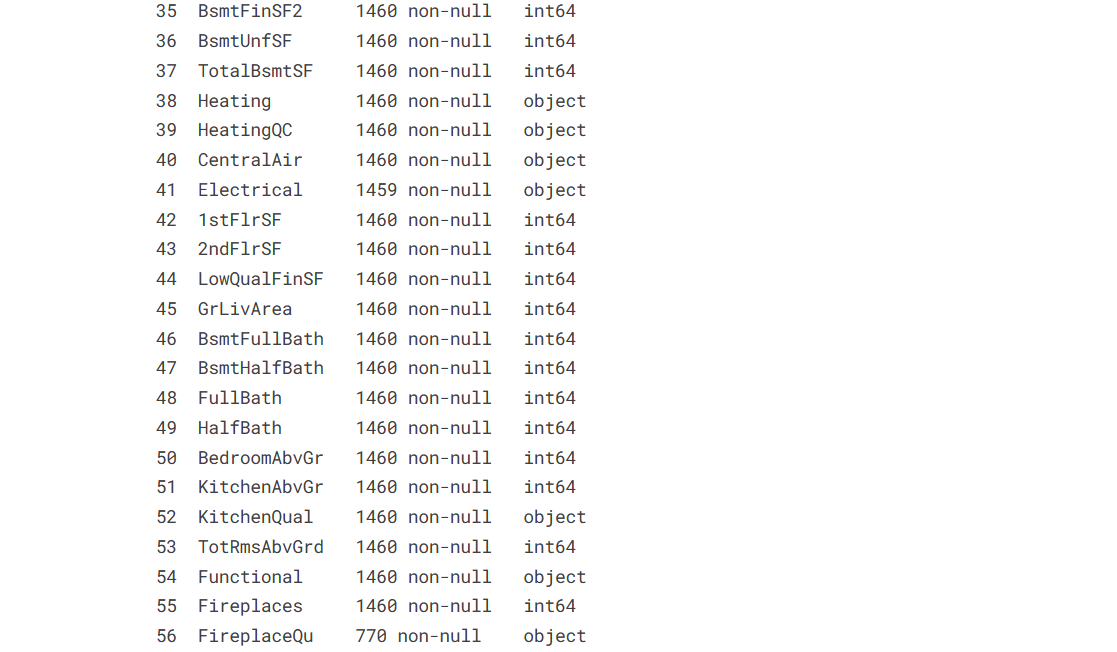


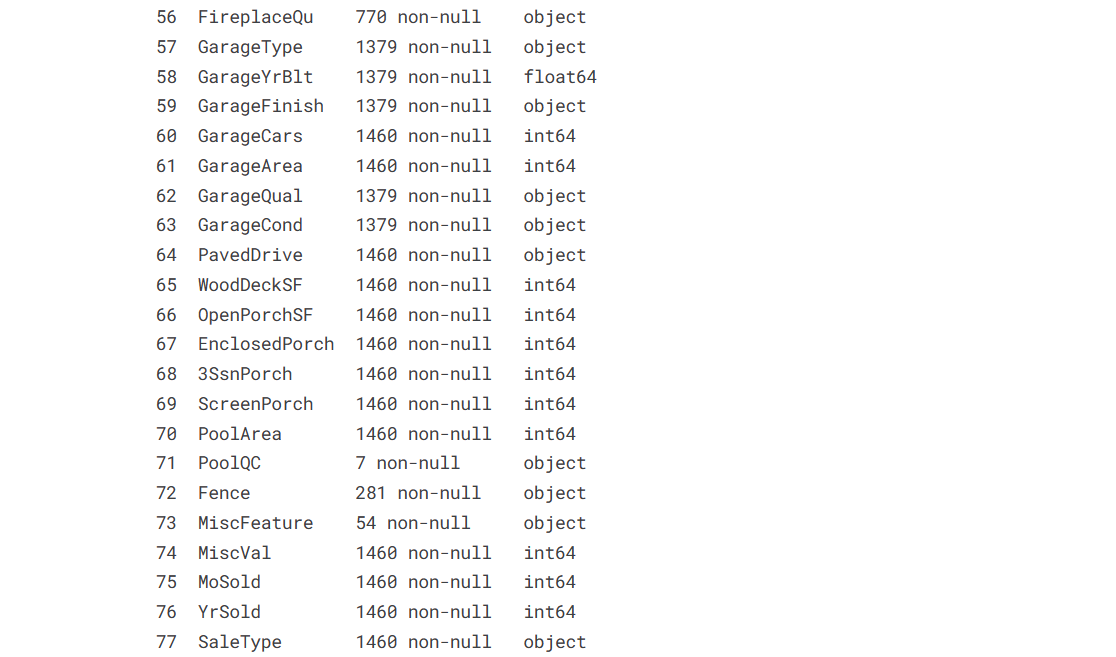
Load the dataset

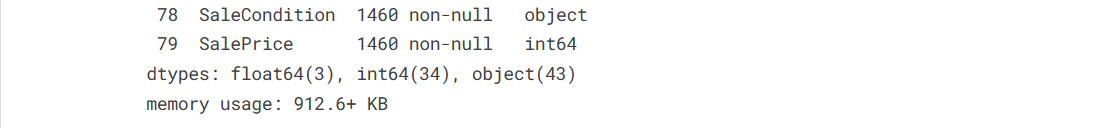












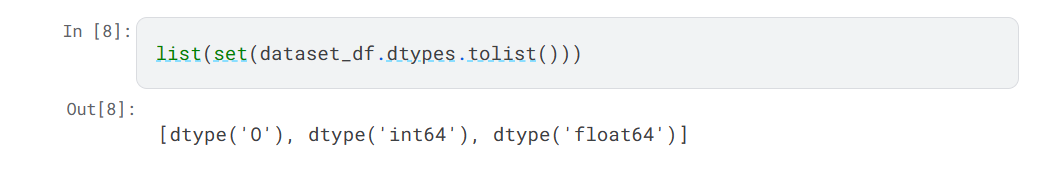
* Dữ liệu chứa 80 thuộc tính và 1459 bản ghi
* 80 thuộc tính gồm:
* MSSubClass: loại ngôi nhà
* MSZoning: khu vực phân loại sử dụng đất
* LotFrontage: rộng của mặt tiền lô đất (trước nhà).
* LotArea: diện tích của mỗi lô đất
* Street: phố của nhà
* Alley: con đường hẻm sau của ngôi nhà.
* LotShape: hình dạng tổng quát của lô đất
* LandContour: độ cao của đất so với mức nước biển xung quanh
* Utilities: các tiện ích công cộng có sẵn cho mỗi ngôi nhà
* LotConfig: cách lô đất được cấu hình trong khu vực
* LandSlope: độ dốc của lô đất.
* Neighborhood: khu vực xung quanh mỗi ngôi nhà
* Condition1: điều kiện gần ngôi nhà
* Condition2: điều kiện gần ngôi nhà thứ hai (nếu có).
* BldgType: loại kiểu xây dựng của ngôi nhà
* HouseStyle: phong cách kiến trúc của ngôi nhà
* OverallQual: tổng điểm chất lượng của ngôi nhà, đánh giá bởi người sử dụng.
* OverallCond: tổng điểm điều kiện tổng thể của ngôi nhà, đánh giá bởi người sử dụng.
* YearBuilt: năm xây dựng của ngôi nhà.
* YearRemodAdd: năm sửa chữa hoặc cải tạo gần đây nhất của ngôi nhà.
* RoofStyle: loại kiểu mái của ngôi nhà.
* RoofMatl: chất liệu mái của ngôi nhà.
* Exterior1st: vật liệu bề mặt bên ngoài chính của ngôi nhà.
* Exterior2nd: vật liệu bề mặt bên ngoài phụ của ngôi nhà (nếu có).
* MasVnrType: loại vật liệu vây cổ điển (nếu có).
* MasVnrArea: diện tích của vật liệu vây cổ điển trong square feet (nếu có).
* ExterQual: chất lượng vật liệu bề mặt bên ngoài của ngôi nhà.
* ExterCond: tình trạng vật liệu bề mặt bên ngoài của ngôi nhà.
* Foundation: loại nền móng của ngôi nhà.
* BsmtQual: chất lượng nền móng dưới lòng đất (nếu có).
* BsmtCond: tình trạng nền móng dưới lòng đất (nếu có).
* BsmtExposure: phần tiếp xúc với mặt ngoài của nền móng dưới lòng đất.
* BsmtFinType1: chất lượng hoàn thiện của khu vực dưới lòng đất.
* BsmtFinSF1: diện tích của khu vực hoàn thiện thứ nhất (type 1) trong square feet.
* BsmtFinType2: chất lượng hoàn thiện của khu vực dưới lòng đất (nếu có thêm).
* BsmtFinSF2: diện tích của khu vực hoàn thiện thứ hai (type 2) trong square feet.
* BsmtUnfSF: diện tích của khu vực chưa hoàn thiện trong square feet.
* TotalBsmtSF: tổng diện tích của tất cả các khu vực dưới lòng đất trong square feet.
* Heating: hệ thống sưởi.
* HeatingQC: chất lượng của hệ thống sưởi.
* CentralAir: có trung tâm điều hòa không khí hay không.
* Electrical: loại hệ thống điện của ngôi nhà.
* 1stFlrSF: diện tích của tầng trệt trong square feet.
* 2ndFlrSF: diện tích của tầng 2 trong square feet.
* LowQualFinSF: diện tích của các khu vực hoàn thiện chất lượng thấp trong square feet.
* GrLivArea: diện tích sống trên mặt đất trong square feet.
* BsmtFullBath: số lượng phòng tắm đầy đủ dưới lòng đất.
* BsmtHalfBath: số lượng phòng tắm bán dưới lòng đất.
* FullBath: số lượng phòng tắm đầy đủ trên mặt đất.
* HalfBath: số lượng phòng tắm bán trên mặt đất.
* BedroomAbvGr: số lượng phòng ngủ trên mặt đất.
* KitchenAbvGr: số lượng phòng bếp trên mặt đất.
* KitchenQual: chất lượng của phòng bếp.
* TotRmsAbvGrd: tổng số phòng trên mặt đất (không bao gồm phòng tắm).
* Functional: tình trạng chức năng của ngôi nhà sau xây dựng.
* Fireplaces: số lượng lò sưởi.
* FireplaceQu: chất lượng lò sưởi.
* GarageType: loại hình garađ.
* GarageYrBlt: năm xây dựng garađ.
* GarageFinish: hoàn thiện của garađ.
* GarageCars: số lượng ô tô có thể chứa trong garađ.
* GarageArea: diện tích garađ trong square feet.
* GarageQual: chất lượng garađ.
* GarageCond: tình trạng garađ.
* PavedDrive: lối đi lát đường.
* WoodDeckSF: diện tích sân vườn bằng gỗ trong square feet.
* OpenPorchSF: diện tích sân hiên mở ra ngoài trong square feet.
* EnclosedPorch: diện tích sân hiên kín trong square feet.
* 3SsnPorch: diện tích sân hiên ba mùa trong square feet.
* ScreenPorch: diện tích sân hiên được lọc trong square feet.
* PoolArea: diện tích bể bơi trong square feet.
* PoolQC: chất lượng của hồ bơi (nếu có)
* Fence: loại hàng rào xung quanh ngôi nhà.
* MiscFeature: các tính năng đặc biệt khác không được phân loại trong các đặc trưng trước đó.
* MiscVal: giá trị của các tính năng đặc biệt khác (nếu có).
* MoSold: tháng bán nhà.
* YrSold: năm bán nhà.
* SaleType: loại hình bán nhà.
* SaleCondition: điều kiện bán nhà.
* SalePrice: giá bán của nhà.

Miêu tả dữ liệu:



Numerical data distribution

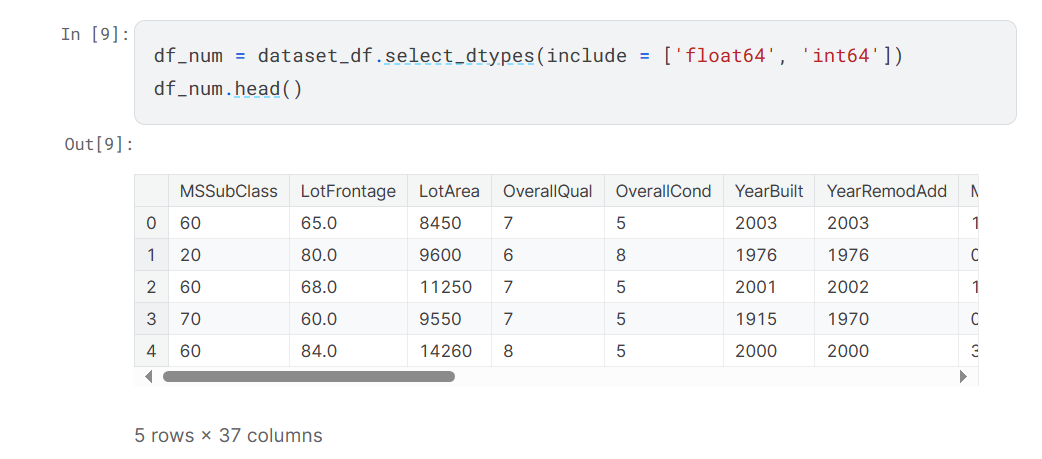
Bây giờ chúng ta sẽ xem xét cách các đặc trưng số học được phân phối. Để làm điều này, trước tiên hãy liệt kê tất cả các loại dữ liệu từ tập dữ liệu của chúng ta và chỉ chọn các loại dữ liệu số học.



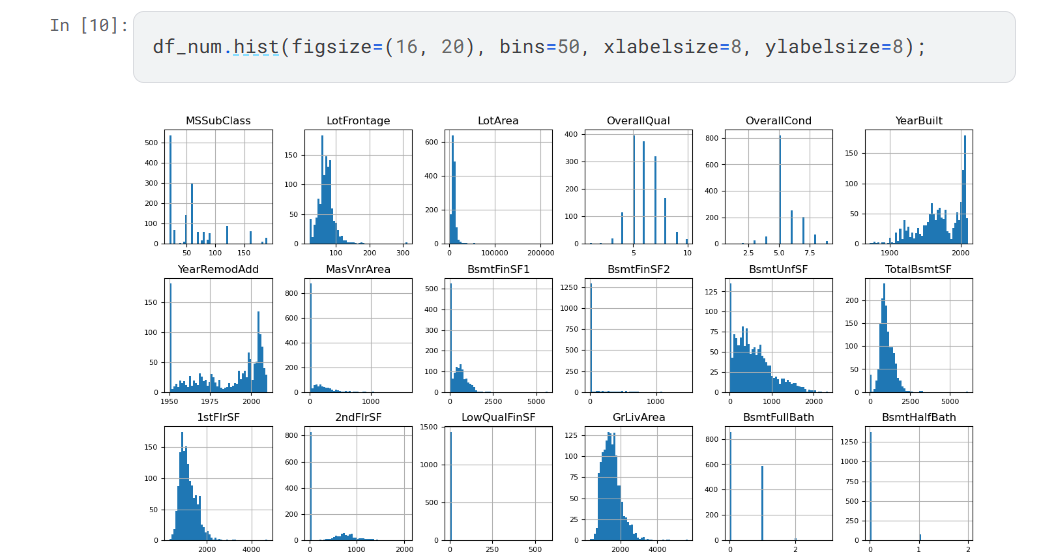
Có ba loại dữ liệu trong tập dữ liệu của chúng ta:

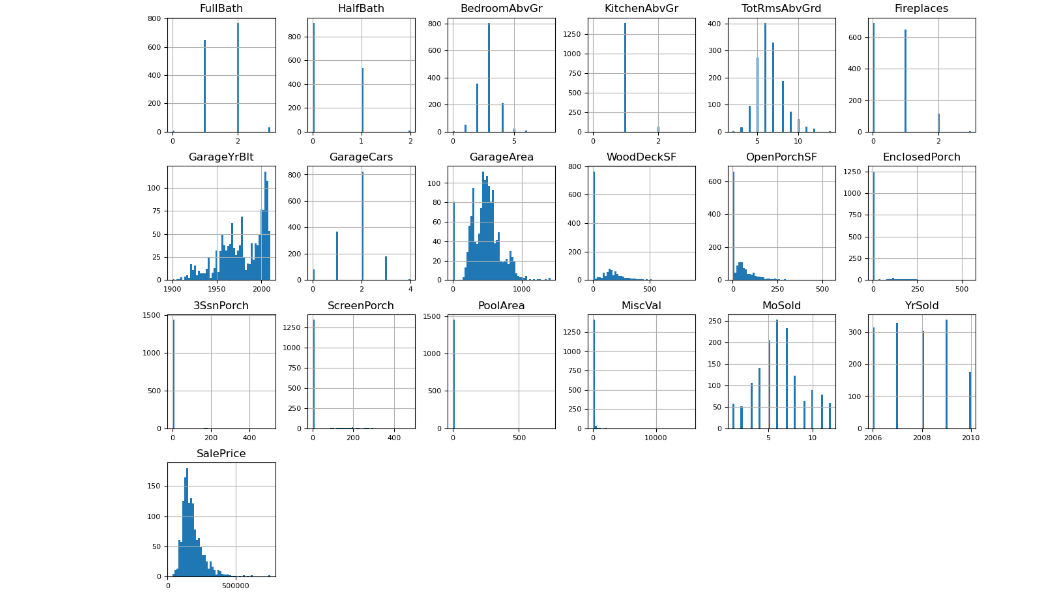
* dtype('O'): Dữ liệu đối tượng (object), thường là các chuỗi ký tự.
* dtype('int64'): Dữ liệu số nguyên 64-bit.
* dtype('float64'): Dữ liệu số thực 64-bit.

Chúng ta sẽ chỉ chọn các đặc trưng có kiểu dữ liệu số học (int64 và float64) để phân tích phân phối.

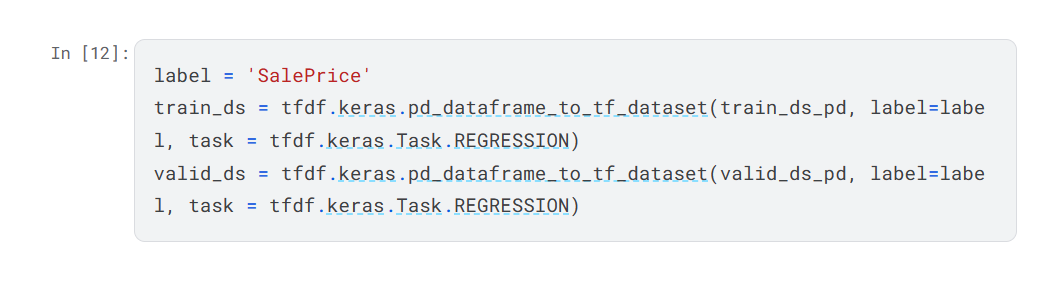


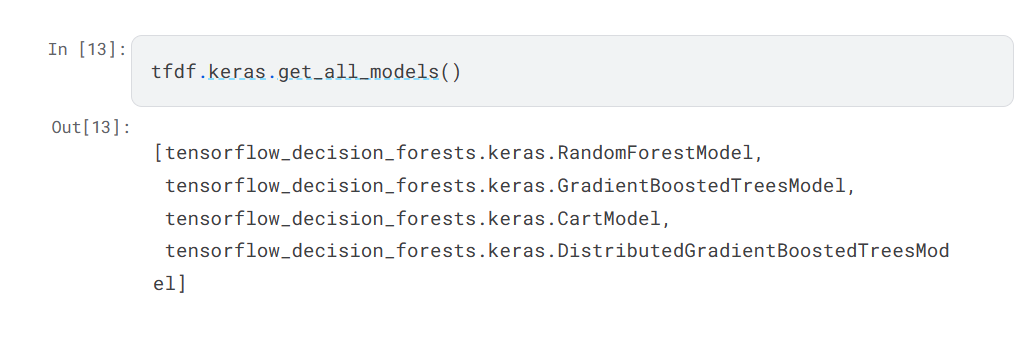
Bây giờ chúng ta hãy vẽ biểu đồ phân phối cho tất cả các đặc điểm số











Danh sách các mô hình TensorFlow Decision Forests:

* **tensorflow\_decision\_forests.keras.RandomForestModel**: Mô hình Rừng Ngẫu Nhiên.
* **tensorflow\_decision\_forests.keras.GradientBoostedTreesModel**: Mô hình Máy Tăng Cường Độ Gradient.
* **tensorflow\_decision\_forests.keras.CartModel**: Mô hình Cây Quyết Định.
* **tensorflow\_decision\_forests.keras.DistributedGradientBoostedTreesModel**: Mô hình Máy Tăng Cường Độ Gradient Phân Tán.