House Price Prediction

Swati Gupta

17-08-2022

Outline

1. Business Context

2. Problem Statement

3. Solution Developed

4. Improvements to the Solution

5. Link to the working project

Business Context

This project aims to make it easier for individuals to sell their houses themselves. Core to this concept is an accurate prediction of a property its value. But rather than letting a realtor determine a good price, the model provides an indication of a properties value. This model should take into account all the relevant information of the house such as accessibility and location.

Problem Statement

Predict sales prices and practice feature engineering, RFs, and gradient boosting.

Price plays a big role in the decision to purchase a home. To predict the sale price for each house in hope of helping potential home buyers understand the Ames Iowa housing market and estimate a price for a home they desire..

For each Id in the test set, we must predict the value of the SalePrice variable. With 79 explanatory variables describing (almost) every aspect of residential homes in Ames, Iowa, we have to predict the final price of each home.

Solution Developed

We have to ask a home buyer to describe their dream house, and they probably won't begin with the height of the basement ceiling or the proximity to different modes of transportation. The dataset proves that much more influences price negotiations than the number of bedrooms or area of the property.

The dataset contained the following attributes, with 1460 entries. The data is remarkably clean, with no null values to strip out. We can focus on the data analysis and interpretation instead of the cleaning. Using the data type, I was able to identify the columns of numerical and categorical data. The last column, SalePrice, is what we are trying to predict. Before I built any models for this data, I wanted to visualize the data:

○ ***Summary of methodologies***

- Data collection

- Exploratory Data Analysis with Data Visualization

- Predictive analysis

- Split the data into training and testing

- Predictive analysis (LinearRegression)

Summary of all steps done in the project

Step 1 – Importing required libraries.

Step 2 – Reading our input data for House Price Prediction.

Step 3 – Describing our data.

Step 4 – Analyzing information from our data.

Step 5 – Plots to visualize data of House Price Prediction.

Step 6 – Scaling our data.

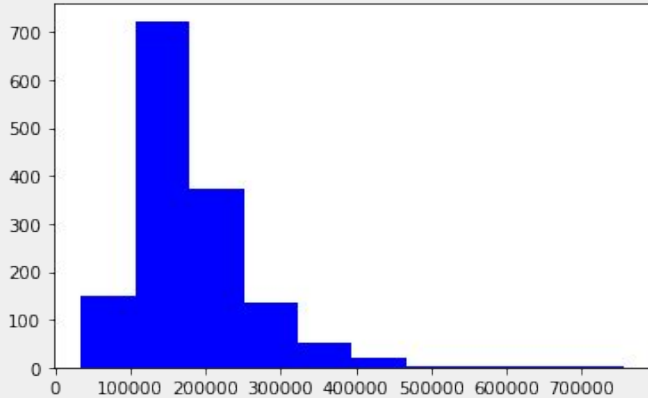Step 7 – Splitting our data for train and test purposes.

Step 8 – Training our Linear Regression model for House Price Prediction.

Step 9 – Lets visualize our predictions of House Price Prediction.

Step 10 – Plotting the residuals of our House Price Prediction model.

Step 11 – Observing the coefficients.

Data Visualisation

Histogram methods to represent the numerical data distribution it shows the data for SalePrice.


Displot provides access to several approaches for visualizing the univariate or bivariate distribution of data, including subsets of data defined by semantic mapping and faceting across multiple subplots.
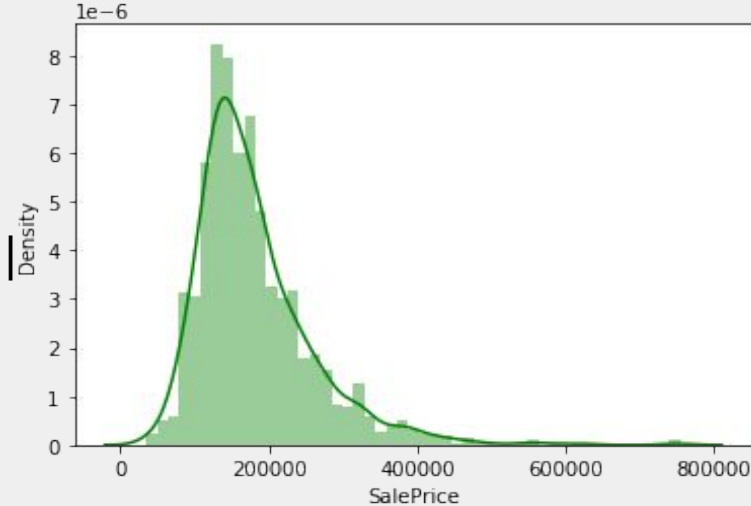
Scatter plots present the relationship between two variables in a data-set. It represents data points on a two-dimensional plane or on a **Cartesian system**. The independent variable or attribute is plotted on the X-axis, while the dependent variable is plotted on the Y-axis.

Here in the first graph the scatterplot is on GrLivArea and SalePrice

In the second graph the graph in between GarageArea and SalePrice
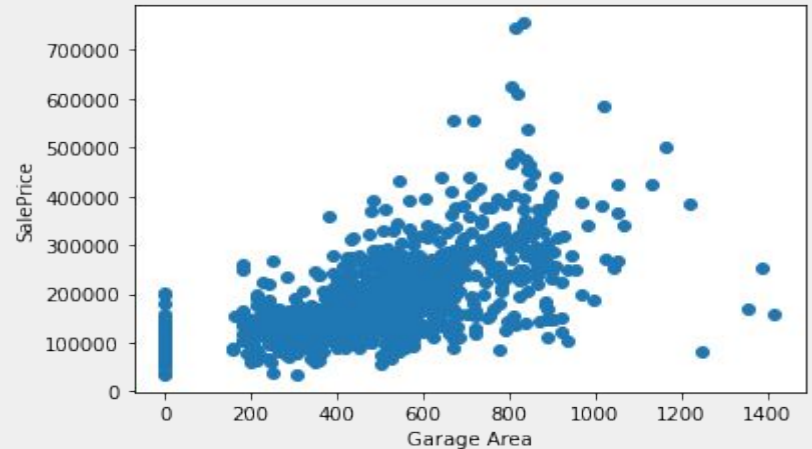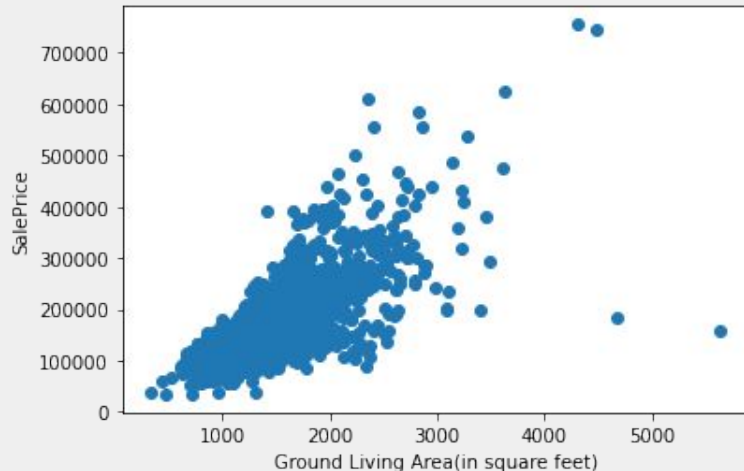
Improvements to the Solution

1 - Create an effective price prediction model

2 - Validate the model's prediction accuracy

3 - Identify the important home price attributes which feed the model's predictive power.

- Our model is giving a 0.79 out of 1 which is a very decent score.
- We can use Lasso and Ridge Regressions but they also performed nearly the same as Linear regression.
- It will help to improve the accuracy rate of our model.

Link to the Project - https://github.com/helloswati/kaggle_projects/blob/main/House_Prices/House_Prices.ipynb