

Assessing Home Price Determinants: A Regression Analysis of Ames Housing Data

Ryan Croce Keely Kinnane Sai Teja Yapuram Kavindu Wellalage

Narrative

There are many factors ranging from location, material quality, property size, and age, that influence the real estate market. The Ames data set provides a comprehensive selection of these property features that we can use to study their relationship and impact on home sales price. We aim to analyse the relationship of variables in the data set with the sale price as the response. We will explore which variables influence the most towards the sales price and we will further analyse how well each variable follows the LINE assumptions and how we can fix in case of an assumption is violated. Our results can be used to assist home sellers, policymakers, and real estate professionals better identify important factors in home prices.

Dataset Description

The Ames Housing Data set curated by Dean De Cock from Truman State University. The original data set describes the sale of individual residential property in Ames, Iowa from 2006 to 2010. The original data set contains 2930 observations and a large number of explanatory variables (23 nominal, 23 ordinal, 14 discrete, and 20 continuous) involved in assessing home values. The version used for this project is a subset of the original data set that contains 766 homes and the 28 variables. The subset is selected by following the steps given below.

- Removed large houses with an above ground living area of more than 1500 square feet.
- Filtered the data set to include houses that had a 'Normal' sales condition.
- Included the top five neighborhoods with the most houses for sale.
- Removed rows with NA values.
- Renamed a few variables to make their meaning clearer.

Data Dictionary

The full data set contains 766 homes and the following 28 variables:

- **LotArea**: Lot size in square feet. A numeric variable.
- **Neighborhood**: Physical location within Ames city limits. A categorical variable with the following levels:
 - “NAmes” : North Ames
 - “Sawyer” : Sawyer
 - “OldTown” : Old Town

- “Edwards” : Edwards
- “CollgCr” : College Creek
- OverallQual: Overall quality of the house’s material and finish. The scale ranges from 1 (Very Poor) to 9 (Very Excellent). A numeric variable.
- OverallCond: Overall condition rating. The scale ranges from 1 (Very Poor) to 9 (Very Excellent). A numeric variable.
- YearBuilt: Original construction date. A numeric variable.
- YearRemodAdd: Remodel date. A numeric variable.
- BsmtFinSf1: Type 1 finished square feet. A numeric variable.
- BsmtFinSf2: Type 2 finished square feet. A numeric variable.
- TotalBsmtSf: Total square feet of basement area. A numeric variable.
- FirstFlrSf: First floor square feet. A numeric variable.
- SecondFlrSf: Second floor square feet. A numeric variable.
- GrLivArea: Above grade (ground) living area square feet. A numeric variable.
- BsmtFullBath: Number of full bathrooms in the basement. A numeric variable.
- BsmtHalfBath: Number of half baths in the basement. A numeric variable.
- FullBath: Number of full bathrooms above ground. A numeric variable.
- HalfBath: Number of half baths above ground. A numeric variable.
- BedroomAbvGr: Number of Bedrooms above ground. A numeric variable.
- KitchenAbvGr: Number of Kitchens above ground. A numeric variable.
- TotRmsAbvGrd: Total rooms above ground (does not include bathrooms). A numeric variable.
- Fireplaces: Number of fireplaces. A numeric variable. A numeric variable.
- GarageCars: Size of garage in car capacity. A numeric variable.
- GarageArea: Size of garage in square feet. A numeric variable.
- WoodDeckSf: Wood deck area in square feet. A numeric variable.
- OpenPorchSf: Open porch area in square feet. A numeric variable.
- EnclosedPorch: Enclosed porch area in square feet. A numeric variable.
- ThreeSsnProch: Three season porch area in square feet. A numeric variable.
- ScreenPorch: Screen porch area in square feet. A numeric variable.
- SalePrice: The property’s sale price in dollars. A numeric response variable.

Evidence of Data

```
data = read.csv('ames_housing.csv')
head(data, nrow=10)
```

```
##   LotArea Neighborhood OverallQual OverallCond YearBuilt YearRemodAdd
## 1   11622         Names           5           6      1961         1961
## 2   14267         Names           6           6      1958         1958
## 3   11241         Names           6           7      1970         1970
## 4   12537         Names           5           6      1971         2008
## 5    8450         Names           5           6      1968         1968
## 6    8400         Names           4           5      1970         1970
##   BsmtFinSf1 BsmtFinSf2 TotalBsmtSf FirstFlrSf SecondFlrSf GrLivArea
## 1         468         144         882         896           0         896
## 2          923           0        1329        1329           0        1329
## 3          578           0        1004        1004           0        1004
## 4          734           0        1078        1078           0        1078
## 5          775           0        1056        1056           0        1056
## 6          804          78         882         882           0         882
##   BsmtFullBath BsmtHalfBath FullBath HalfBath BedroomAbvGr KitchenAbvGr
```

## 1	0	0	1	0	2	1
## 2	0	0	1	1	3	1
## 3	1	0	1	0	2	1
## 4	1	0	1	1	3	1
## 5	1	0	1	0	3	1
## 6	1	0	1	0	2	1
##	TotRmsAbvGrd	Fireplaces	GarageCars	GarageArea	WoodDeckSf	OpenPorchSf
## 1	5	0	1	730	140	0
## 2	6	0	1	312	393	36
## 3	5	1	2	480	0	0
## 4	6	1	2	500	0	0
## 5	6	1	1	304	0	85
## 6	4	0	2	525	240	0
##	EnclosedPorch	ThreeSsnPorch	ScreenPorch	SalePrice		
## 1	0	0	120	105000		
## 2	0	0	0	172000		
## 3	0	0	0	149000		
## 4	0	0	0	149900		
## 5	184	0	0	142000		
## 6	0	0	0	126000		