

Multiple Linear Regression Analysis

Bett Kipkemoi

2026-01-10

Multiple Linear Regression Analysis

Task: Using publicly available data from the Kaggle Housing Prices Dataset (<https://www.kaggle.com/datasets/prevek18/ames-housing-dataset>), perform a multiple linear regression analysis to predict the sale price of houses based on features such as Lot Area, Overall Qual, Year Built, Total Bsmt SF, and Gr Liv Area.

To begin, we load libraries that will be necessary for performing data exploration, processing, visualization, model performance and evaluation, then read the data. The glimpse of the data is printed in the last section with the code.

R Code

```
knitr::opts_chunk$set(comment = NA)
library(conflicted)
library(tidyverse)
library(readr)
library(skimr)
library(car)
library(performance)
library(see)
ames <- read_csv("AmesHousing.csv")
#glimpse of the data
glimpse(ames)
summary(ames)
```

Ouput

Rows: 2,930

Columns: 82

```
$ Order      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1~
$ PID        <chr> "0526301100", "0526350040", "0526351010", "052635303~
$ `MS SubClass` <chr> "020", "020", "020", "020", "060", "060", "120", "12~
$ `MS Zoning`  <chr> "RL", "RH", "RL", "RL", "RL", "RL", "RL", "RL", "RL"~
$ `Lot Frontage` <dbl> 141, 80, 81, 93, 74, 78, 41, 43, 39, 60, 75, NA, 63,~
$ `Lot Area`   <dbl> 31770, 11622, 14267, 11160, 13830, 9978, 4920, 5005,~
$ Street      <chr> "Pave", "Pave", "Pave", "Pave", "Pave", "Pave", "Pav~
$ Alley       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ `Lot Shape`  <chr> "IR1", "Reg", "IR1", "Reg", "IR1", "IR1", "Reg", "IR~
$ `Land Contour` <chr> "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "Lvl", "HL~
$ Utilities    <chr> "AllPub", "AllPub", "AllPub", "AllPub", "AllPub", "A~
$ `Lot Config` <chr> "Corner", "Inside", "Corner", "Corner", "Inside", "I~
$ `Land Slope` <chr> "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gtl", "Gt~
```

\$ Neighborhood	<chr> "Names", "Names", "Names", "Names", "Gilbert", "Gilb~
\$ `Condition 1`	<chr> "Norm", "Feedr", "Norm", "Norm", "Norm", "Norm", "No~
\$ `Condition 2`	<chr> "Norm", "Norm", "Norm", "Norm", "Norm", "Norm", "Nor~
\$ `Bldg Type`	<chr> "1Fam", "1Fam", "1Fam", "1Fam", "1Fam", "1Fam", "Twn~
\$ `House Style`	<chr> "1Story", "1Story", "1Story", "1Story", "2Story", "2~
\$ `Overall Qual`	<dbl> 6, 5, 6, 7, 5, 6, 8, 8, 8, 7, 6, 6, 6, 7, 8, 8, 8, 9~
\$ `Overall Cond`	<dbl> 5, 6, 6, 5, 5, 6, 5, 5, 5, 5, 5, 7, 5, 5, 5, 5, 7, 2~
\$ `Year Built`	<dbl> 1960, 1961, 1958, 1968, 1997, 1998, 2001, 1992, 1995~
\$ `Year Remod/Add`	<dbl> 1960, 1961, 1958, 1968, 1998, 1998, 2001, 1992, 1996~
\$ `Roof Style`	<chr> "Hip", "Gable", "Hip", "Hip", "Gable", "Gable", "Gab~
\$ `Roof Matl`	<chr> "CompShg", "CompShg", "CompShg", "CompShg", "CompShg~
\$ `Exterior 1st`	<chr> "BrkFace", "VinylSd", "Wd Sdng", "BrkFace", "VinylSd~
\$ `Exterior 2nd`	<chr> "Plywood", "VinylSd", "Wd Sdng", "BrkFace", "VinylSd~
\$ `Mas Vnr Type`	<chr> "Stone", "None", "BrkFace", "None", "None", "BrkFace~
\$ `Mas Vnr Area`	<dbl> 112, 0, 108, 0, 0, 20, 0, 0, 0, 0, 0, 0, 0, 0, 0, 60~
\$ `Exter Qual`	<chr> "TA", "TA", "TA", "Gd", "TA", "TA", "Gd", "Gd", "Gd"~
\$ `Exter Cond`	<chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA"~
\$ Foundation	<chr> "CBlock", "CBlock", "CBlock", "CBlock", "PConc", "PC~
\$ `Bsmt Qual`	<chr> "TA", "TA", "TA", "TA", "Gd", "TA", "Gd", "Gd", "Gd"~
\$ `Bsmt Cond`	<chr> "Gd", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA"~
\$ `Bsmt Exposure`	<chr> "Gd", "No", "No", "No", "No", "No", "Mn", "No", "No"~
\$ `BsmtFin Type 1`	<chr> "BLQ", "Rec", "ALQ", "ALQ", "GLQ", "GLQ", "GLQ", "AL~
\$ `BsmtFin SF 1`	<dbl> 639, 468, 923, 1065, 791, 602, 616, 263, 1180, 0, 0, ~
\$ `BsmtFin Type 2`	<chr> "Unf", "LwQ", "Unf", "Unf", "Unf", "Unf", "Unf", "Un~
\$ `BsmtFin SF 2`	<dbl> 0, 144, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1120, 0, ~
\$ `Bsmt Unf SF`	<dbl> 441, 270, 406, 1045, 137, 324, 722, 1017, 415, 994, ~
\$ `Total Bsmt SF`	<dbl> 1080, 882, 1329, 2110, 928, 926, 1338, 1280, 1595, 9~
\$ Heating	<chr> "GasA", "GasA", "GasA", "GasA", "GasA", "GasA", "Gas~
\$ `Heating QC`	<chr> "Fa", "TA", "TA", "Ex", "Gd", "Ex", "Ex", "Ex", "Ex"~
\$ `Central Air`	<chr> "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y"~
\$ Electrical	<chr> "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr", "SBrkr"~
\$ `1st Flr SF`	<dbl> 1656, 896, 1329, 2110, 928, 926, 1338, 1280, 1616, 1~
\$ `2nd Flr SF`	<dbl> 0, 0, 0, 0, 701, 678, 0, 0, 0, 776, 892, 0, 676, 0, ~
\$ `Low Qual Fin SF`	<dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ `Gr Liv Area`	<dbl> 1656, 896, 1329, 2110, 1629, 1604, 1338, 1280, 1616, ~
\$ `Bsmt Full Bath`	<dbl> 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1~
\$ `Bsmt Half Bath`	<dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ `Full Bath`	<dbl> 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 3, 2, 1~
\$ `Half Bath`	<dbl> 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1~
\$ `Bedroom AbvGr`	<dbl> 3, 2, 3, 3, 3, 3, 2, 2, 2, 3, 3, 3, 3, 2, 1, 4, 4, 1~
\$ `Kitchen AbvGr`	<dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
\$ `Kitchen Qual`	<chr> "TA", "TA", "Gd", "Ex", "TA", "Gd", "Gd", "Gd", "Gd"~
\$ `TotRms AbvGrd`	<dbl> 7, 5, 6, 8, 6, 7, 6, 5, 5, 7, 7, 6, 7, 5, 4, 12, 8, ~
\$ Functional	<chr> "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Typ", "Ty~
\$ Fireplaces	<dbl> 2, 0, 0, 2, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1~
\$ `Fireplace Qu`	<chr> "Gd", NA, NA, "TA", "TA", "Gd", NA, NA, "TA", "TA", ~
\$ `Garage Type`	<chr> "Attchd", "Attchd", "Attchd", "Attchd", "Attchd", "A~
\$ `Garage Yr Blt`	<dbl> 1960, 1961, 1958, 1968, 1997, 1998, 2001, 1992, 1995~
\$ `Garage Finish`	<chr> "Fin", "Unf", "Unf", "Fin", "Fin", "Fin", "Fin", "RF~
\$ `Garage Cars`	<dbl> 2, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 3~
\$ `Garage Area`	<dbl> 528, 730, 312, 522, 482, 470, 582, 506, 608, 442, 44~
\$ `Garage Qual`	<chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA"~
\$ `Garage Cond`	<chr> "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA", "TA"~
\$ `Paved Drive`	<chr> "P", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y", "Y"~

```

$ `Wood Deck SF`      <dbl> 210, 140, 393, 0, 212, 360, 0, 0, 237, 140, 157, 483~
$ `Open Porch SF`    <dbl> 62, 0, 36, 0, 34, 36, 0, 82, 152, 60, 84, 21, 75, 0,~
$ `Enclosed Porch`   <dbl> 0, 0, 0, 0, 0, 0, 170, 0, 0, 0, 0, 0, 0, 0, 0,~
$ `3Ssn Porch`       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ `Screen Porch`     <dbl> 0, 120, 0, 0, 0, 0, 0, 144, 0, 0, 0, 0, 0, 140, 2~
$ `Pool Area`        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
$ `Pool QC`          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ Fence              <chr> NA, "MnPrv", NA, NA, "MnPrv", NA, NA, NA, NA, NA, NA~
$ `Misc Feature`     <chr> NA, NA, "Gar2", NA, NA, NA, NA, NA, NA, NA, NA, "She~
$ `Misc Val`         <dbl> 0, 0, 12500, 0, 0, 0, 0, 0, 0, 0, 0, 500, 0, 0, 0,~
$ `Mo Sold`          <dbl> 5, 6, 6, 4, 3, 6, 4, 1, 3, 6, 4, 3, 5, 2, 6, 6, 6, 6~
$ `Yr Sold`          <dbl> 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010, 2010~
$ `Sale Type`        <chr> "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD", "WD"~
$ `Sale Condition`   <chr> "Normal", "Normal", "Normal", "Normal", "Normal", "N~
$ SalePrice          <dbl> 215000, 105000, 172000, 244000, 189900, 195500, 2135~

```

Order	PID	MS SubClass	MS Zoning
Min. : 1.0	Length:2930	Length:2930	Length:2930
1st Qu.: 733.2	Class :character	Class :character	Class :character
Median :1465.5	Mode :character	Mode :character	Mode :character
Mean :1465.5			
3rd Qu.:2197.8			
Max. :2930.0			

Lot Frontage	Lot Area	Street	Alley
Min. : 21.00	Min. : 1300	Length:2930	Length:2930
1st Qu.: 58.00	1st Qu.: 7440	Class :character	Class :character
Median : 68.00	Median : 9436	Mode :character	Mode :character
Mean : 69.22	Mean : 10148		
3rd Qu.: 80.00	3rd Qu.: 11555		
Max. :313.00	Max. :215245		
NA's :490			

Lot Shape	Land Contour	Utilities	Lot Config
Length:2930	Length:2930	Length:2930	Length:2930
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Land Slope	Neighborhood	Condition 1	Condition 2
Length:2930	Length:2930	Length:2930	Length:2930
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Bldg Type	House Style	Overall Qual	Overall Cond
Length:2930	Length:2930	Min. : 1.000	Min. :1.000
Class :character	Class :character	1st Qu.: 5.000	1st Qu.:5.000
Mode :character	Mode :character	Median : 6.000	Median :5.000
		Mean : 6.095	Mean :5.563
		3rd Qu.: 7.000	3rd Qu.:6.000

Max. :10.000 Max. :9.000

Year Built	Year Remod/Add	Roof Style	Roof Matl
Min. :1872	Min. :1950	Length:2930	Length:2930
1st Qu.:1954	1st Qu.:1965	Class :character	Class :character
Median :1973	Median :1993	Mode :character	Mode :character
Mean :1971	Mean :1984		
3rd Qu.:2001	3rd Qu.:2004		
Max. :2010	Max. :2010		

Exterior 1st	Exterior 2nd	Mas Vnr Type	Mas Vnr Area
Length:2930	Length:2930	Length:2930	Min. : 0.0
Class :character	Class :character	Class :character	1st Qu.: 0.0
Mode :character	Mode :character	Mode :character	Median : 0.0
			Mean : 101.9
			3rd Qu.: 164.0
			Max. :1600.0
			NA's :23
Exter Qual	Exter Cond	Foundation	Bsmt Qual
Length:2930	Length:2930	Length:2930	Length:2930
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Bsmt Cond	Bsmt Exposure	BsmtFin Type 1	BsmtFin SF 1
Length:2930	Length:2930	Length:2930	Min. : 0.0
Class :character	Class :character	Class :character	1st Qu.: 0.0
Mode :character	Mode :character	Mode :character	Median : 370.0
			Mean : 442.6
			3rd Qu.: 734.0
			Max. :5644.0
			NA's :1

BsmtFin Type 2	BsmtFin SF 2	Bsmt Unf SF	Total Bsmt SF
Length:2930	Min. : 0.00	Min. : 0.0	Min. : 0
Class :character	1st Qu.: 0.00	1st Qu.: 219.0	1st Qu.: 793
Mode :character	Median : 0.00	Median : 466.0	Median : 990
	Mean : 49.72	Mean : 559.3	Mean :1052
	3rd Qu.: 0.00	3rd Qu.: 802.0	3rd Qu.:1302
	Max. :1526.00	Max. :2336.0	Max. :6110
	NA's :1	NA's :1	NA's :1
Heating	Heating QC	Central Air	Electrical
Length:2930	Length:2930	Length:2930	Length:2930
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

1st Flr SF	2nd Flr SF	Low Qual Fin SF	Gr Liv Area
Min. : 334.0	Min. : 0.0	Min. : 0.000	Min. : 334
1st Qu.: 876.2	1st Qu.: 0.0	1st Qu.: 0.000	1st Qu.:1126
Median :1084.0	Median : 0.0	Median : 0.000	Median :1442

Mean :1159.6	Mean : 335.5	Mean : 4.677	Mean :1500
3rd Qu.:1384.0	3rd Qu.: 703.8	3rd Qu.: 0.000	3rd Qu.:1743
Max. :5095.0	Max. :2065.0	Max. :1064.000	Max. :5642

Bsmt Full Bath	Bsmt Half Bath	Full Bath	Half Bath
Min. :0.0000	Min. :0.00000	Min. :0.000	Min. :0.0000
1st Qu.:0.0000	1st Qu.:0.00000	1st Qu.:1.000	1st Qu.:0.0000
Median :0.0000	Median :0.00000	Median :2.000	Median :0.0000
Mean :0.4314	Mean :0.06113	Mean :1.567	Mean :0.3795
3rd Qu.:1.0000	3rd Qu.:0.00000	3rd Qu.:2.000	3rd Qu.:1.0000
Max. :3.0000	Max. :2.00000	Max. :4.000	Max. :2.0000
NA's :2	NA's :2		

Bedroom AbvGr	Kitchen AbvGr	Kitchen Qual	TotRms AbvGrd
Min. :0.000	Min. :0.000	Length:2930	Min. : 2.000
1st Qu.:2.000	1st Qu.:1.000	Class :character	1st Qu.: 5.000
Median :3.000	Median :1.000	Mode :character	Median : 6.000
Mean :2.854	Mean :1.044		Mean : 6.443
3rd Qu.:3.000	3rd Qu.:1.000		3rd Qu.: 7.000
Max. :8.000	Max. :3.000		Max. :15.000

Functional	Fireplaces	Fireplace Qu	Garage Type
Length:2930	Min. :0.0000	Length:2930	Length:2930
Class :character	1st Qu.:0.0000	Class :character	Class :character
Mode :character	Median :1.0000	Mode :character	Mode :character
	Mean :0.5993		
	3rd Qu.:1.0000		
	Max. :4.0000		

Garage Yr Blt	Garage Finish	Garage Cars	Garage Area
Min. :1895	Length:2930	Min. :0.000	Min. : 0.0
1st Qu.:1960	Class :character	1st Qu.:1.000	1st Qu.: 320.0
Median :1979	Mode :character	Median :2.000	Median : 480.0
Mean :1978		Mean :1.767	Mean : 472.8
3rd Qu.:2002		3rd Qu.:2.000	3rd Qu.: 576.0
Max. :2207		Max. :5.000	Max. :1488.0
NA's :159		NA's :1	NA's :1
Garage Qual	Garage Cond	Paved Drive	Wood Deck SF
Length:2930	Length:2930	Length:2930	Min. : 0.00
Class :character	Class :character	Class :character	1st Qu.: 0.00
Mode :character	Mode :character	Mode :character	Median : 0.00
			Mean : 93.75
			3rd Qu.: 168.00
			Max. :1424.00

Open Porch SF	Enclosed Porch	3Ssn Porch	Screen Porch
Min. : 0.00	Min. : 0.00	Min. : 0.000	Min. : 0
1st Qu.: 0.00	1st Qu.: 0.00	1st Qu.: 0.000	1st Qu.: 0
Median : 27.00	Median : 0.00	Median : 0.000	Median : 0
Mean : 47.53	Mean : 23.01	Mean : 2.592	Mean : 16
3rd Qu.: 70.00	3rd Qu.: 0.00	3rd Qu.: 0.000	3rd Qu.: 0
Max. :742.00	Max. :1012.00	Max. :508.000	Max. :576

Pool Area	Pool QC	Fence	Misc Feature
Min. : 0.000	Length:2930	Length:2930	Length:2930

1st Qu.:	0.000	Class :	character	Class :	character	Class :	character
Median :	0.000	Mode :	character	Mode :	character	Mode :	character
Mean :	2.243						
3rd Qu.:	0.000						
Max. :	800.000						

Misc Val	Mo Sold	Yr Sold	Sale Type
Min. : 0.00	Min. : 1.000	Min. : 2006	Length:2930
1st Qu.: 0.00	1st Qu.: 4.000	1st Qu.: 2007	Class :character
Median : 0.00	Median : 6.000	Median : 2008	Mode :character
Mean : 50.64	Mean : 6.216	Mean : 2008	
3rd Qu.: 0.00	3rd Qu.: 8.000	3rd Qu.: 2009	
Max. : 17000.00	Max. : 12.000	Max. : 2010	

Sale Condition	SalePrice
Length:2930	Min. : 12789
Class :character	1st Qu.:129500
Mode :character	Median :160000
	Mean :180796
	3rd Qu.:213500
	Max. :755000