

Informe Proyecto 2 entrega 1

Juan Luis Solórzano (carnet: 201598)

Micaela Yataz (carnet: 18960)

2025-01-20

git: https://github.com/JusSolo/Mineria_Proyecto2.git

Análisis exploratorio

descripción de las variables:

SalePrice: el precio de venta de la propiedad en dólares. Esta es la variable objetivo que se intenta predecir.
MSSubClass: la clase de construcción MSZoning: la clasificación general de zonificación LotFrontage: pies lineales de calle conectada a la propiedad LotArea: tamaño del lote en pies cuadrados Street: tipo de acceso vial Alley: tipo de acceso por callejón LotShape: forma general de la propiedad LandContour: planitud de la propiedad Utilities: tipo de servicios públicos disponibles LotConfig: configuración del lote LandSlope: pendiente de la propiedad Neighborhood: ubicaciones físicas dentro de los límites de la ciudad de Ames Condition1: proximidad a la carretera principal o ferrocarril Condition2: proximidad a la carretera principal o ferrocarril (si hay un segundo presente) BldgType: tipo de vivienda HouseStyle: estilo de vivienda OverallQual: calidad general de materiales y acabados OverallCond: calificación general de la condición YearBuilt: año de construcción original YearRemodAdd: año de remodelación RoofStyle: tipo de techo RoofMatl: material del techo Exterior1st: revestimiento exterior de la casa Exterior2nd: revestimiento exterior de la casa (si se utiliza más de un material) MasVnrType: tipo de revestimiento de mampostería MasVnrArea: área del revestimiento de mampostería en pies cuadrados ExterQual: calidad del material exterior ExterCond: condición actual del material en el exterior Foundation: tipo de cimentación BsmtQual: altura del sótano BsmtCond: condición general del sótano BsmtExposure: paredes del sótano de salida directa o a nivel de jardín BsmtFinType1: calidad del área terminada del sótano (tipo 1) BsmtFinSF1: pies cuadrados terminados, tipo 1 BsmtFinType2: calidad del área terminada del sótano (tipo 2, si está presente) BsmtFinSF2: pies cuadrados terminados, tipo 2 BsmtUnfSF: pies cuadrados sin terminar del área del sótano TotalBsmtSF: total de pies cuadrados del área del sótano Heating: tipo de calefacción HeatingQC: calidad y condición de la calefacción CentralAir: aire acondicionado central Electrical: sistema eléctrico 1stFlrSF: pies cuadrados del primer piso 2ndFlrSF: pies cuadrados del segundo piso LowQualFinSF: pies cuadrados terminados de baja calidad (todas las plantas) GrLivArea: pies cuadrados de área habitable sobre el nivel del suelo BsmtFullBath: baños completos en el sótano BsmtHalfBath: medios baños en el sótano FullBath: baños completos sobre el nivel del suelo HalfBath: medios baños sobre el nivel del suelo Bedroom: número de dormitorios sobre el nivel del sótano Kitchen: número de cocinas KitchenQual: calidad de la cocina TotRmsAbvGrd: total de habitaciones sobre el nivel del suelo (no incluye baños) Functional: calificación de funcionalidad de la vivienda Fireplaces: número de chimeneas FireplaceQu: calidad de la chimenea GarageType: ubicación del garaje GarageYrBlt: año en que se construyó el garaje GarageFinish: acabado interior del garaje GarageCars: capacidad del garaje en número de autos GarageArea: tamaño del garaje en pies cuadrados GarageQual: calidad del garaje GarageCond: condición del garaje PavedDrive: acceso pavimentado WoodDeckSF: área de la terraza de madera en pies cuadrados OpenPorchSF: área del porche abierto en pies cuadrados EnclosedPorch: área del porche cerrado en pies cuadrados 3SsnPorch: área del porche de tres estaciones en pies cuadrados ScreenPorch: área del porche con malla en pies cuadrados PoolArea: área de la piscina en pies cuadrados PoolQC: calidad de la piscina Fence: calidad de la cerca MiscFeature: característica miscelánea no cubierta en otras categorías MiscVal: valor en dólares de la característica miscelánea MoSold: mes en que se vendió YrSold: año en que se vendió SaleType: tipo de venta SaleCondition: condición de la venta

Estadísticas descriptivas de

`summary(datos)`

```
##           Id           MSSubClass           MSZoning           LotFrontage
## Min.      : 1.0      Min.      : 20.0      Length:1460      Min.      : 21.00
## 1st Qu.: 365.8      1st Qu.: 20.0      Class :character      1st Qu.: 59.00
## Median : 730.5      Median : 50.0      Mode  :character      Median : 69.00
## Mean    : 730.5      Mean    : 56.9                        Mean    : 70.05
## 3rd Qu.:1095.2      3rd Qu.: 70.0                        3rd Qu.: 80.00
## Max.    :1460.0      Max.    :190.0                        Max.    :313.00
##                                     NA's    :259
##           LotArea           Street           Alley           LotShape
## Min.      : 1300      Length:1460      Length:1460      Length:1460
## 1st Qu.: 7554      Class :character      Class :character      Class :character
## Median : 9478      Mode  :character      Mode  :character      Mode  :character
## Mean     : 10517
## 3rd Qu.: 11602
## Max.     :215245
##
##           LandContour           Utilities           LotConfig           LandSlope
## Length:1460      Length:1460      Length:1460      Length:1460
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##           Neighborhood           Condition1           Condition2           BldgType
## Length:1460      Length:1460      Length:1460      Length:1460
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##           HouseStyle           OverallQual           OverallCond           YearBuilt
## Length:1460      Min.      : 1.000      Min.      :1.000      Min.      :1872
## Class :character      1st Qu.: 5.000      1st Qu.:5.000      1st Qu.:1954
## Mode  :character      Median : 6.000      Median :5.000      Median :1973
##                                     Mean    : 6.099      Mean    :5.575      Mean    :1971
##                                     3rd Qu.: 7.000      3rd Qu.:6.000      3rd Qu.:2000
##                                     Max.     :10.000      Max.     :9.000      Max.     :2010
##
##           YearRemodAdd           RoofStyle           RoofMatl           Exterior1st
## Min.      :1950      Length:1460      Length:1460      Length:1460
## 1st Qu.:1967      Class :character      Class :character      Class :character
## Median :1994      Mode  :character      Mode  :character      Mode  :character
## Mean     :1985
## 3rd Qu.:2004
## Max.     :2010
##
##           Exterior2nd           MasVnrType           MasVnrArea           ExterQual
## Length:1460      Length:1460      Min.      : 0.0      Length:1460
```

```

## Class :character   Class :character   1st Qu.: 0.0   Class :character
## Mode :character   Mode :character   Median : 0.0   Mode :character
##                                     Mean : 103.7
##                                     3rd Qu.: 166.0
##                                     Max. :1600.0
##                                     NA's :8
## ExterCond          Foundation          BsmtQual          BsmtCond
## Length:1460        Length:1460        Length:1460        Length:1460
## Class :character   Class :character   Class :character   Class :character
## Mode :character   Mode :character   Mode :character   Mode :character
##
##
##
## BsmtExposure        BsmtFinType1        BsmtFinSF1        BsmtFinType2
## Length:1460        Length:1460        Min. : 0.0        Length:1460
## Class :character   Class :character   1st Qu.: 0.0        Class :character
## Mode :character   Mode :character   Median : 383.5      Mode :character
##                                     Mean : 443.6
##                                     3rd Qu.: 712.2
##                                     Max. :5644.0
##
## BsmtFinSF2          BsmtUnfSF          TotalBsmtSF        Heating
## Min. : 0.00         Min. : 0.0         Min. : 0.0         Length:1460
## 1st Qu.: 0.00       1st Qu.: 223.0     1st Qu.: 795.8     Class :character
## Median : 0.00       Median : 477.5     Median : 991.5     Mode :character
## Mean : 46.55        Mean : 567.2       Mean :1057.4
## 3rd Qu.: 0.00       3rd Qu.: 808.0     3rd Qu.:1298.2
## Max. :1474.00       Max. :2336.0       Max. :6110.0
##
## HeatingQC           CentralAir          Electrical          X1stFlrSF
## Length:1460         Length:1460         Length:1460         Min. : 334
## Class :character     Class :character     Class :character     1st Qu.: 882
## Mode :character     Mode :character     Mode :character     Median :1087
##                                     Mean :1163
##                                     3rd Qu.:1391
##                                     Max. :4692
##
## X2ndFlrSF           LowQualFinSF        GrLivArea          BsmtFullBath
## Min. : 0            Min. : 0.000        Min. : 334          Min. :0.0000
## 1st Qu.: 0          1st Qu.: 0.000      1st Qu.:1130        1st Qu.:0.0000
## Median : 0          Median : 0.000      Median :1464        Median :0.0000
## Mean : 347          Mean : 5.845        Mean :1515          Mean :0.4253
## 3rd Qu.: 728        3rd Qu.: 0.000      3rd Qu.:1777        3rd Qu.:1.0000
## Max. :2065          Max. :572.000       Max. :5642          Max. :3.0000
##
## BsmtHalfBath         FullBath            HalfBath           BedroomAbvGr
## Min. :0.00000        Min. :0.000         Min. :0.0000        Min. :0.000
## 1st Qu.:0.00000      1st Qu.:1.000       1st Qu.:0.0000      1st Qu.:2.000
## Median :0.00000      Median :2.000       Median :0.0000      Median :3.000
## Mean :0.05753        Mean :1.565         Mean :0.3829        Mean :2.866
## 3rd Qu.:0.00000      3rd Qu.:2.000       3rd Qu.:1.0000      3rd Qu.:3.000
## Max. :2.00000        Max. :3.000         Max. :2.0000        Max. :8.000
##

```

```

## KitchenAbvGr KitchenQual TotRmsAbvGrd Functional
## Min. :0.000 Length:1460 Min. : 2.000 Length:1460
## 1st Qu.:1.000 Class :character 1st Qu.: 5.000 Class :character
## Median :1.000 Mode :character Median : 6.000 Mode :character
## Mean :1.047 Mean : 6.518
## 3rd Qu.:1.000 3rd Qu.: 7.000
## Max. :3.000 Max. :14.000
##
## Fireplaces FireplaceQu GarageType GarageYrBlt
## Min. :0.000 Length:1460 Length:1460 Min. :1900
## 1st Qu.:0.000 Class :character Class :character 1st Qu.:1961
## Median :1.000 Mode :character Mode :character Median :1980
## Mean :0.613 Mean :1979
## 3rd Qu.:1.000 3rd Qu.:2002
## Max. :3.000 Max. :2010
## NA's :81
## GarageFinish GarageCars GarageArea GarageQual
## Length:1460 Min. :0.000 Min. : 0.0 Length:1460
## Class :character 1st Qu.:1.000 1st Qu.: 334.5 Class :character
## Mode :character Median :2.000 Median : 480.0 Mode :character
## Mean :1.767 Mean : 473.0
## 3rd Qu.:2.000 3rd Qu.: 576.0
## Max. :4.000 Max. :1418.0
##
## GarageCond PavedDrive WoodDeckSF OpenPorchSF
## Length:1460 Length:1460 Min. : 0.00 Min. : 0.00
## Class :character Class :character 1st Qu.: 0.00 1st Qu.: 0.00
## Mode :character Mode :character Median : 0.00 Median : 25.00
## Mean : 94.24 Mean : 46.66
## 3rd Qu.:168.00 3rd Qu.: 68.00
## Max. :857.00 Max. :547.00
##
## EnclosedPorch X3SsnPorch ScreenPorch PoolArea
## Min. : 0.00 Min. : 0.00 Min. : 0.00 Min. : 0.000
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.000
## Median : 0.00 Median : 0.00 Median : 0.00 Median : 0.000
## Mean : 21.95 Mean : 3.41 Mean : 15.06 Mean : 2.759
## 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.00 3rd Qu.: 0.000
## Max. :552.00 Max. :508.00 Max. :480.00 Max. :738.000
##
## PoolQC Fence MiscFeature MiscVal
## Length:1460 Length:1460 Length:1460 Min. : 0.00
## Class :character Class :character Class :character 1st Qu.: 0.00
## Mode :character Mode :character Mode :character Median : 0.00
## Mean : 43.49
## 3rd Qu.: 0.00
## Max. :15500.00
##
## MoSold YrSold SaleType SaleCondition
## Min. : 1.000 Min. :2006 Length:1460 Length:1460
## 1st Qu.: 5.000 1st Qu.:2007 Class :character Class :character
## Median : 6.000 Median :2008 Mode :character Mode :character
## Mean : 6.322 Mean :2008
## 3rd Qu.: 8.000 3rd Qu.:2009

```

```
## Max.    :12.000    Max.    :2010
##
##    SalePrice
## Min.    : 34900
## 1st Qu.:129975
## Median :163000
## Mean    :180921
## 3rd Qu.:214000
## Max.    :755000
##
```

Prueba de normalidad en las variables cuantitativas

Analisis de correlacion

Analisis de grupos (K-medias o clustering jerárquico)

Conjuntos de entrenamiento y prueba