

# 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](https://github.com/JusSolo/Mineria_Proyecto2.git)

## Análisis exploratorio

### descripción de las variables:

- **SalePrice:** cuantitativa, el precio de venta de la propiedad
- **MSSubClass:** cuantitativa, la clase de construcción
- **MSZoning:** cualitativa, la clasificación general de zona
- **LotFrontage:** cuantitativa, pies lineales de calle conectada a la propiedad
- **LotArea:** cuantitativa, tamaño del lote en pies cuadrados
- **Street:** cualitativa, tipo de acceso vial
- **Alley:** tipo de acceso por callejón
- **LotShape:** cualitativa, forma general de la propiedad
- **LandContour:** cualitativa, relieve de la propiedad
- **Utilities:** cualitativa, tipo de servicios públicos disponibles
- **LotConfig:** cualitativa, configuración del lote
- **LandSlope:** cualitativa, pendiente de la propiedad
- **Neighborhood:** cualitativa, ubicaciones físicas dentro de los límites de la ciudad de Ames
- **Condition1:** cualitativa, proximidad a la carretera principal o ferrocarril
- **Condition2:** cualitativa, proximidad a la carretera principal o ferrocarril (si hay un segundo presente)
- **BldgType:** cualitativa, tipo de vivienda
- **HouseStyle:** cualitativa, estilo de vivienda
- **OverallQual:** cuantitativa, calidad general de materiales y acabados
- **OverallCond:** cuantitativa, calificación general de la condición
- **YearBuilt:** cuantitativa, año de construcción original
- **YearRemodAdd:** cuantitativa, año de remodelación
- **RoofStyle:** cualitativa, tipo de techo
- **RoofMatl:** cualitativa, material del techo
- **Exterior1st:** cualitativa, revestimiento exterior de la casa
- **Exterior2nd:** cualitativa, revestimiento exterior de la casa (si se utiliza más de un material)
- **MasVnrType:** cualitativa, tipo de revestimiento de mampostería
- **MasVnrArea:** cuantitativa, área del revestimiento de mampostería en pies cuadrados
- **ExterQual:** cualitativa, calidad del material exterior
- **ExterCond:** cualitativa, condición actual del material en el exterior
- **Foundation:** cualitativa, tipo de cimentación
- **BsmtQual:** cualitativa, altura del sótano
- **BsmtCond:** cualitativa, condición general del sótano
- **BsmtExposure:** cualitativa, paredes del sótano de salida directa o a nivel de jardín
- **BsmtFinType1:** cualitativa, calidad del área terminada del sótano (tipo 1)
- **BsmtFinSF1:** cuantitativa, pies cuadrados terminados, tipo 1
- **BsmtFinType2:** cualitativa, calidad del área terminada del sótano (tipo 2, si está presente)
- **BsmtFinSF2:** cuantitativa, pies cuadrados terminados, tipo 2

- **BsmtUnfSF**: cuantitativa, pies cuadrados sin terminar del área del sótano
- **TotalBsmtSF**: cuantitativa, total de pies cuadrados del área del sótano
- **Heating**: cualitativa, tipo de calefacción
- **HeatingQC**: cualitativa, calidad y condición de la calefacción
- **CentralAir**: cualitativa, aire acondicionado central
- **Electrical**: cualitativa, sistema eléctrico
- **X1stFlrSF**: cuantitativa, pies cuadrados del primer piso
- **X2ndFlrSF**: cuantitativa, pies cuadrados del segundo piso
- **LowQualFinSF**: cuantitativa, pies cuadrados terminados de baja calidad (todas las plantas)
- **GrLivArea**: cuantitativa, pies cuadrados de área habitable sobre el nivel del suelo
- **BsmtFullBath**: cuantitativa, baños completos en el sótano
- **BsmtHalfBath**: cuantitativa, medios baños en el sótano
- **FullBath**: cuantitativa, baños completos sobre el nivel del suelo
- **HalfBath**: cuantitativa, medios baños sobre el nivel del suelo
- **BedroomAbvGr**: cuantitativa, número de dormitorios sobre el nivel del sótano
- **KitchenAbvGr**: cuantitativa, número de cocinas
- **KitchenQual**: cualitativa, calidad de la cocina
- **TotRmsAbvGrd**: cuantitativa, total de habitaciones sobre el nivel del suelo (no incluye baños)
- **Functional**: cualitativa, calificación de funcionalidad de la vivienda
- **Fireplaces**: cuantitativa, número de chimeneas
- **FireplaceQu**: cualitativa, calidad de la chimenea
- **GarageType**: cualitativa, ubicación del garaje
- **GarageYrBlt**: cuantitativa, año en que se construyó el garaje
- **GarageFinish**: cualitativa, acabado interior del garaje
- **GarageCars**: cuantitativa, capacidad del garaje en número de autos
- **GarageArea**: cuantitativa, tamaño del garaje en pies cuadrados
- **GarageQual**: cualitativa, calidad del garaje
- **GarageCond**: cualitativa, condición del garaje
- **PavedDrive**: cualitativa, acceso pavimentado
- **WoodDeckSF**: cuantitativa, área de la terraza de madera en pies cuadrados
- **OpenPorchSF**: cuantitativa, área del porche abierto en pies cuadrados
- **EnclosedPorch**: cuantitativa, área del porche cerrado en pies cuadrados
- **X3SsnPorch**: cuantitativa, área del porche de tres estaciones en pies cuadrados
- **ScreenPorch**: cuantitativa, área del porche con malla en pies cuadrados
- **PoolArea**: cuantitativa, área de la piscina en pies cuadrados
- **PoolQC**: cualitativa, calidad de la piscina
- **Fence**: cualitativa, calidad de la cerca
- **MiscFeature**: cualitativa, característica miscelánea no cubierta en otras categorías
- **MiscVal**: cuantitativa, valor en dólares de la característica miscelánea
- **MoSold**: cuantitativa, mes en que se vendió
- **YrSold**: cuantitativa, año en que se vendió
- **SaleType**: cualitativa, tipo de venta
- **SaleCondition**: cuantitativa, condición de la venta

## Estadísticas descriptivas

##	Id		MSSubClass	MSZoning	LotFrontage
##	Min.	:	1.0	Min.	: 20.0
##	1st Qu.	:	365.8	1st Qu.	: 20.0
##	Median	:	730.5	Median	: 50.0
##	Mean	:	730.5	Mean	: 56.9
##	3rd Qu.	:	1095.2	3rd Qu.	: 70.0
##	Max.	:	1460.0	Max.	: 190.0
##				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
##

```

## Variables numéricas

### Estadísticas descriptivas:

```
##      SalePrice      LotFrontage      LotArea      OverallQual
##  Min.   : 34900    Min.   : 21.00    Min.   : 1300    Min.   : 1.000
## 1st Qu.:129975    1st Qu.: 59.00    1st Qu.: 7554    1st Qu.: 5.000
## Median :163000    Median : 69.00    Median : 9478    Median : 6.000
## Mean   :180921    Mean   : 70.05    Mean   : 10517    Mean   : 6.099
## 3rd Qu.:214000    3rd Qu.: 80.00    3rd Qu.: 11602    3rd Qu.: 7.000
## Max.   :755000    Max.   :313.00    Max.   :215245    Max.   :10.000
##
##      NA's :259
##      OverallCond      YearBuilt      YearRemodAdd      MasVnrArea
##  Min.   :1.000    Min.   :1872    Min.   :1950    Min.   : 0.0
## 1st Qu.:5.000    1st Qu.:1954    1st Qu.:1967    1st Qu.: 0.0
## Median :5.000    Median :1973    Median :1994    Median : 0.0
## Mean   :5.575    Mean   :1971    Mean   :1985    Mean   : 103.7
## 3rd Qu.:6.000    3rd Qu.:2000    3rd Qu.:2004    3rd Qu.: 166.0
## Max.   :9.000    Max.   :2010    Max.   :2010    Max.   :1600.0
##
##      NA's :8
##      BsmtFinSF1      BsmtFinSF2      BsmtUnfSF      TotalBsmtSF
##  Min.   : 0.0    Min.   : 0.00    Min.   : 0.0    Min.   : 0.0
## 1st Qu.: 0.0    1st Qu.: 0.00    1st Qu.: 223.0    1st Qu.: 795.8
## Median : 383.5    Median : 0.00    Median : 477.5    Median : 991.5
## Mean   : 443.6    Mean   : 46.55    Mean   : 567.2    Mean   :1057.4
## 3rd Qu.: 712.2    3rd Qu.: 0.00    3rd Qu.: 808.0    3rd Qu.:1298.2
## Max.   :5644.0    Max.   :1474.00    Max.   :2336.0    Max.   :6110.0
##
##      X1stFlrSF      X2ndFlrSF      LowQualFinSF      GrLivArea
##  Min.   : 334    Min.   : 0    Min.   : 0.000    Min.   : 334
## 1st Qu.: 882    1st Qu.: 0    1st Qu.: 0.000    1st Qu.:1130
## Median :1087    Median : 0    Median : 0.000    Median :1464
## Mean   :1163    Mean   : 347    Mean   : 5.845    Mean   :1515
## 3rd Qu.:1391    3rd Qu.: 728    3rd Qu.: 0.000    3rd Qu.:1777
## Max.   :4692    Max.   :2065    Max.   :572.000    Max.   :5642
##
##      BsmtFullBath      BsmtHalfBath      FullBath      HalfBath
##  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   :0.4253    Mean   :0.05753    Mean   :1.565    Mean   :0.3829
## 3rd Qu.:1.0000    3rd Qu.:0.00000    3rd Qu.:2.000    3rd Qu.:1.0000
## Max.   :3.0000    Max.   :2.00000    Max.   :3.000    Max.   :2.0000
##
##      BedroomAbvGr      KitchenAbvGr      TotRmsAbvGrd      Fireplaces
##  Min.   :0.000    Min.   :0.000    Min.   : 2.000    Min.   :0.000
## 1st Qu.:2.000    1st Qu.:1.000    1st Qu.: 5.000    1st Qu.:0.000
## Median :3.000    Median :1.000    Median : 6.000    Median :1.000
## Mean   :2.866    Mean   :1.047    Mean   : 6.518    Mean   :0.613
## 3rd Qu.:3.000    3rd Qu.:1.000    3rd Qu.: 7.000    3rd Qu.:1.000
## Max.   :8.000    Max.   :3.000    Max.   :14.000    Max.   :3.000
##
##      GarageYrBlt      GarageCars      GarageArea      WoodDeckSF
##  Min.   :1900    Min.   :0.000    Min.   : 0.0    Min.   : 0.00
## 1st Qu.:1961    1st Qu.:1.000    1st Qu.: 334.5    1st Qu.: 0.00
```

```
## Median :1980    Median :2.000    Median : 480.0    Median : 0.00
## Mean   :1979    Mean   :1.767    Mean   : 473.0    Mean   : 94.24
## 3rd Qu.:2002    3rd Qu.:2.000    3rd Qu.: 576.0    3rd Qu.:168.00
## Max.   :2010    Max.   :4.000    Max.   :1418.0    Max.   :857.00
## NA's   :81
##   OpenPorchSF   EnclosedPorch   X3SsnPorch   ScreenPorch
## Min.    : 0.00   Min.    : 0.00   Min.    : 0.00   Min.    : 0.00
## 1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.: 0.00
## Median : 25.00   Median : 0.00   Median : 0.00   Median : 0.00
## Mean    : 46.66   Mean    : 21.95   Mean    : 3.41   Mean    : 15.06
## 3rd Qu.: 68.00   3rd Qu.: 0.00   3rd Qu.: 0.00   3rd Qu.: 0.00
## Max.    :547.00   Max.    :552.00   Max.    :508.00   Max.    :480.00
##
##   PoolArea   MiscVal   MoSold   YrSold
## Min.    : 0.000   Min.    : 0.00   Min.    : 1.000   Min.    :2006
## 1st Qu.: 0.000   1st Qu.: 0.00   1st Qu.: 5.000   1st Qu.:2007
## Median : 0.000   Median : 0.00   Median : 6.000   Median :2008
## Mean    : 2.759   Mean    : 43.49   Mean    : 6.322   Mean    :2008
## 3rd Qu.: 0.000   3rd Qu.: 0.00   3rd Qu.: 8.000   3rd Qu.:2009
## Max.    :738.000   Max.    :15500.00   Max.    :12.000   Max.    :2010
##
```

Podemos notar que la gran mayoría de variables tienen datos en todas las filas. Excepto para algunos casos que a continuación se explica como se llenaran los datos faltantes:

- LotFrontage: pies lineales de calle conectada a la propiedad, si hay un Na se va a considerar que la propiedad no tiene acceso directo a la calle (por ejemplo tiene derecho de paso con otra propiedad que si tiene acceso a la calle). por lo que tiene sentido remplazar los Na's por 0.
- MasVnrArea : por razon similar tiene sentido remplazar los Na's por 0
- GarageYrBlt : el año de construccion del garaje, en este caso el Na's tiene sentido porque si una casa no tiene garaje, este no tiene año de construccion. Pero para no eliminar las casas sin garaje se va a remplazar los Na's por la mediana de la columna por ser un estadístico mas robusto que la media, es decir que no tiende a sesgar los datos.

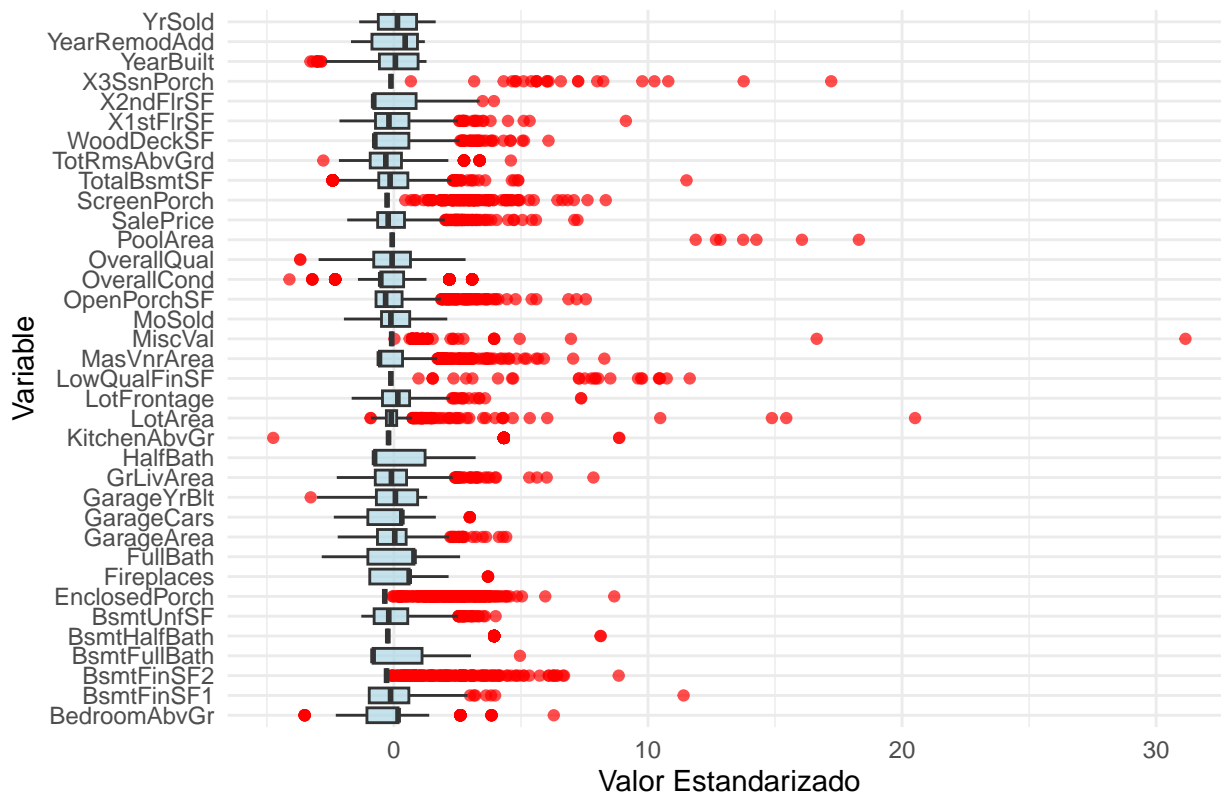
#### variables completadas

```
##   LotFrontage   MasVnrArea   GarageYrBlt
## Min.    : 0.00   Min.    : 0.0   Min.    :1900
## 1st Qu.: 42.00   1st Qu.: 0.0   1st Qu.:1962
## Median : 63.00   Median : 0.0   Median :1980
## Mean    : 57.62   Mean    :103.1   Mean    :1979
## 3rd Qu.: 79.00   3rd Qu.:164.2   3rd Qu.:2001
## Max.    :313.00   Max.    :1600.0   Max.    :2010
```

Otra cosa que se puede notar al leer los resultados del summary es que algunas variables como MiscVal, teniendo una media de 43 y un tercer cuartil de 0.00 tienen probablemente muchos datos atípicos de gran valor que sesgan la media. Vamos a normalizar los datos para poder hacer diagramas de caja y bigotes de todas variables a la vez. Para analizar visualmente el summary.

## Analisis de grafica de caja y bigotes

### Grafica de caja y bigotes



De esta esta grafica podemos notar que hay variables demasiado dispersas que podrían no ser buenas para elaborar una regresion. Otras que tienen casi todos los valores nulos. Tomando en cuenta que son variables de casas, puede ser que mansiones o casas muy grandes tengan otra escala de valores. Puede ser una buena idea crear clusters, alguna funcion que le asigne un cluster a una casa dada y crear regresiones lineales diferentes para cada cluster. Otra opcion sería no tomar en cuenta esas variables, quitar los datos atipicos y crear un modelo que sea bueno prediciendo el valor de una propiedad “usual”. Para calcular el valor de mansiones o similares puede que el modelo no vaya a ver bueno. Para indagar más en estas opciones se hara un analisis de correlacion y un clustering.

## Prueba de normalidad en las variables cuantitativas

Tabla de resultados de la prueba de Shapiro-Wilk

##	Variable	P_Value
## SalePrice	SalePrice	3.206142e-33
## LotFrontage	LotFrontage	9.099048e-30
## LotArea	LotArea	7.933654e-58
## OverallQual	OverallQual	2.686457e-22
## OverallCond	OverallCond	6.774229e-37
## YearBuilt	YearBuilt	2.770220e-26
## YearRemodAdd	YearRemodAdd	6.720280e-34
## MasVnrArea	MasVnrArea	4.382067e-48
## BsmtFinSF1	BsmtFinSF1	2.813854e-35
## BsmtFinSF2	BsmtFinSF2	1.850254e-58
## BsmtUnfSF	BsmtUnfSF	1.639911e-25
## TotalBsmtSF	TotalBsmtSF	1.611332e-27



```

## X1stFlrSF          X1stFlrSF 4.513223e-26
## X2ndFlrSF          X2ndFlrSF 2.514882e-41
## LowQualFinSF      LowQualFinSF 9.589248e-64
## GrLivArea          GrLivArea 6.597611e-26
## BsmtFullBath       BsmtFullBath 3.760666e-47
## BsmtHalfBath       BsmtHalfBath 1.466616e-60
## FullBath           FullBath 4.231488e-44
## HalfBath           HalfBath 4.581582e-48
## BedroomAbvGr       BedroomAbvGr 4.115551e-35
## KitchenAbvGr       KitchenAbvGr 4.221203e-61
## TotRmsAbvGrd       TotRmsAbvGrd 2.004964e-23
## Fireplaces         Fireplaces 4.830980e-42
## GarageYrBlt        GarageYrBlt 7.080806e-26
## GarageCars         GarageCars 2.301685e-36
## GarageArea         GarageArea 4.016963e-15
## WoodDeckSF         WoodDeckSF 3.227985e-41
## OpenPorchSF        OpenPorchSF 1.135905e-43
## EnclosedPorch      EnclosedPorch 4.849485e-56
## X3SsnPorch         X3SsnPorch 8.307268e-64
## ScreenPorch        ScreenPorch 3.305688e-59
## PoolArea           PoolArea 7.111538e-65
## MiscVal            MiscVal 1.529907e-64
## MoSold             MoSold 3.178973e-17
## YrSold             YrSold 3.420194e-30

```

Se puede constatar que todas las variables tienen  $p - valor < 0.001$ , por lo que se puede concluir que ninguna de las variables sigue una distribución normal. Pero para un modelo de regresión lineal no hay un supuesto de normalidad en la distribución de las variables, esto no debería de ser un problema. Dicho usar k-medias para elaborar un clustering puede no ser la mejor opción.

## Esto es Lo que te toca Mica:

Analisis de correlacion

Analisis de grupos (Clustering jerárquico y k-medias {si quieres})

Conjuntos de entrenamiento y prueba