

Proyecto 2

Rodrigo Mansilla, Javier Chen

2025-02-25

Introducción

Metodología

Exploración inicial de Datos

Dimensiones del Dataset

Table 1: Dimensiones del Dataset

Métrica	Valor
Número de filas	1460
Número de columnas	81

Primeras filas

Table 2: Primeras 6 filas (5 columnas)

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities
1	60	RL	65	8450	Pave	NA	Reg	Lvl	AllPub
2	20	RL	80	9600	Pave	NA	Reg	Lvl	AllPub
3	60	RL	68	11250	Pave	NA	IR1	Lvl	AllPub
4	70	RL	60	9550	Pave	NA	IR1	Lvl	AllPub
5	60	RL	84	14260	Pave	NA	IR1	Lvl	AllPub
6	50	RL	85	14115	Pave	NA	IR1	Lvl	AllPub

Últimas filas

Table 3: Últimas 6 filas (5 columnas)

	Id	MSSubClass	MSZoning	LotFrontage	LotArea
1455	1455	20	FV	62	7500
1456	1456	60	RL	62	7917
1457	1457	20	RL	85	13175

1458	1458	70	RL	66	9042
1459	1459	20	RL	68	9717
1460	1460	20	RL	75	9937

Observamos que el dataset contiene una complejidad adecuada y es necesaria la limpieza y transformación de datos para poder detectar relaciones, outliers y patrones en los datos.

Análisis descriptivo y Exploración de Variables

Estadísticas Descriptivas de variables numéricas

Table 4: Resumen Estadístico de Variables Numéricas

	count	mean	std	min	25%.25%	50%.50%	75%.75%	max	mediana
Id	1460	730.50	421.61	1	365.75	730.5	1095.25	1460	730.5
MSSubClass	1460	56.90	42.30	20	20.00	50.0	70.00	190	50.0
LotFrontage	1201	70.05	24.28	21	59.00	69.0	80.00	313	69.0
LotArea	1460	10516.83	9981.26	1300	7553.50	9478.5	11601.50	215245	9478.5
OverallQual	1460	6.10	1.38	1	5.00	6.0	7.00	10	6.0
OverallCond	1460	5.58	1.11	1	5.00	5.0	6.00	9	5.0
YearBuilt	1460	1971.27	30.20	1872	1954.00	1973.0	2000.00	2010	1973.0
YearRemodAdd	1460	1984.87	20.65	1950	1967.00	1994.0	2004.00	2010	1994.0
MasVnrArea	1452	103.69	181.07	0	0.00	0.0	166.00	1600	0.0
BsmtFinSF1	1460	443.64	456.10	0	0.00	383.5	712.25	5644	383.5
BsmtFinSF2	1460	46.55	161.32	0	0.00	0.0	0.00	1474	0.0
BsmtUnfSF	1460	567.24	441.87	0	223.00	477.5	808.00	2336	477.5
TotalBsmtSF	1460	1057.43	438.71	0	795.75	991.5	1298.25	6110	991.5
X1stFlrSF	1460	1162.63	386.59	334	882.00	1087.0	1391.25	4692	1087.0
X2ndFlrSF	1460	346.99	436.53	0	0.00	0.0	728.00	2065	0.0
LowQualFinSF	1460	5.84	48.62	0	0.00	0.0	0.00	572	0.0
GrLivArea	1460	1515.46	525.48	334	1129.50	1464.0	1776.75	5642	1464.0
BsmtFullBath	1460	0.43	0.52	0	0.00	0.0	1.00	3	0.0
BsmtHalfBath	1460	0.06	0.24	0	0.00	0.0	0.00	2	0.0
FullBath	1460	1.57	0.55	0	1.00	2.0	2.00	3	2.0
HalfBath	1460	0.38	0.50	0	0.00	0.0	1.00	2	0.0
BedroomAbvGr	1460	2.87	0.82	0	2.00	3.0	3.00	8	3.0
KitchenAbvGr	1460	1.05	0.22	0	1.00	1.0	1.00	3	1.0
TotRmsAbvGrd	1460	6.52	1.63	2	5.00	6.0	7.00	14	6.0
Fireplaces	1460	0.61	0.64	0	0.00	1.0	1.00	3	1.0
GarageYrBlt	1379	1978.51	24.69	1900	1961.00	1980.0	2002.00	2010	1980.0
GarageCars	1460	1.77	0.75	0	1.00	2.0	2.00	4	2.0
GarageArea	1460	472.98	213.80	0	334.50	480.0	576.00	1418	480.0
WoodDeckSF	1460	94.24	125.34	0	0.00	0.0	168.00	857	0.0
OpenPorchSF	1460	46.66	66.26	0	0.00	25.0	68.00	547	25.0
EnclosedPorch	1460	21.95	61.12	0	0.00	0.0	0.00	552	0.0
X3SsnPorch	1460	3.41	29.32	0	0.00	0.0	0.00	508	0.0
ScreenPorch	1460	15.06	55.76	0	0.00	0.0	0.00	480	0.0
PoolArea	1460	2.76	40.18	0	0.00	0.0	0.00	738	0.0
MiscVal	1460	43.49	496.12	0	0.00	0.0	0.00	15500	0.0
MoSold	1460	6.32	2.70	1	5.00	6.0	8.00	12	6.0
YrSold	1460	2007.82	1.33	2006	2007.00	2008.0	2009.00	2010	2008.0

	count	mean	std	min	25%.25%	50%.50%	75%.75%	max	mediana
SalePrice	1460	180921.20	79442.50	34900	129975.00	163000.0	214000.00	755000	163000.0

Ingeniería de Características y Transformación de Datos

Análisis de Correlaciones y Selección de Variables

Análisis de Grupos

Modelado

Discusión y Conclusiones