Documentation de CSV/XLSX

house_price.csv

File_properties

{'file_name': 'house_price.csv', 'date': '11-07-2023', 'Nombre de lignes': 1459, 'Missing values': 7000}

Variables

Id: {'type': 'int64', 'extremum': {'earliest_year': 1461, 'latest_year': 2919}, 'missing values': 0}

MSSubClass: {'type': 'int64', 'statistics': {'mean': 57.37834132967786, 'std': 42.73222773152917, 'min': 20, 'max': 190, 'q1': 20.0, 'q3': 70.0, 'shapiro_statistic': 0.8106269836425781, 'shapiro_pvalue': 2.6139244218547836e-38, 'outliers': 104}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

MSZoning: {'type': 'object', 'missing_values': 4, '>10% appearance': {'RL': 76.56, 'RM': 16.63}}

LotFrontage: {'type': 'float64', 'statistics': {'mean': 68.58035714285714, 'std': 22.36775792692783, 'min': 21.0, 'max': 200.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 227, 'missing values percent column': 15.558601782042494, 'missing values percent total': 0}

LotArea: {'type': 'int64', 'statistics': {'mean': 9819.161069225496, 'std': 4953.818777664235, 'min':

1470, 'max': 56600, 'q1': 7391.0, 'q3': 11517.5, 'shapiro_statistic': 0.7905782461166382, 'shapiro_pvalue': 9.530020661103432e-40, 'outliers': 60}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

Street: {'type': 'binary', 'missing_values': 0, 'appearance': {'Pave': 99.59, 'Grvl': 0.41}}

Alley: {'type': 'binary', 'missing_values': 1352, 'appearance': {'Grvl': 65.42, 'Pave': 34.58}}

LotShape: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 934, 'nbr of highest app': 1, '>10% appearance': {'Reg': 64.02, 'IR1': 33.17}}

LandContour: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 1311, 'nbr of highest app': 1, '>10% appearance': {'Lvl': 89.86}}

Utilities: {'type': 'Identifiant', 'missing_values': 2, 'unique value': 1, 'highest appearance': 1457, 'nbr of highest app': 1, '>10% appearance': {'AllPub': 100.0}}

LotConfig: {'type': 'object', 'missing_values': 0, '>10% appearance': {'Inside': 74.09, 'Corner': 17.0}}

LandSlope: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 3, 'highest appearance': 1396, 'nbr of highest app': 1, '>10% appearance': {'Gtl': 95.68}}

Documentation de CSV/XLSX

Neighborhood: {'type': 'object', 'missing_values': 0, '>10% appearance': {'NAmes': 14.94}}

Condition1: {'type': 'object', 'missing_values': 0, '>10% appearance': {'Norm': 85.74}}

Condition2: {'type': 'object', 'missing_values': 0, '>10% appearance': {'Norm': 98.97}}

BldgType: {'type': 'object', 'missing_values': 0, '>10% appearance': {'1Fam': 82.59}}

HouseStyle: {'type': 'object', 'missing_values': 0, '>10% appearance': {'1Story': 51.06, '2Story': 29.27, '1.5Fin': 10.97}}

OverallQual: {'type': 'int64', 'statistics': {'mean': 6.078821110349555, 'std': 1.4363191600333747, 'min': 1, 'max': 10, 'q1': 5.0, 'q3': 7.0, 'shapiro_statistic': 0.9469878077507019, 'shapiro_pvalue': 1.6973608339810768e-22, 'outliers': 2}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

OverallCond: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 9, 'highest appearance': 824, 'nbr of highest app': 1, '>10% appearance': {5: 56.48, 6: 19.12, 7: 12.68}}

YearBuilt: {'type': 'int64', 'extremum': {'earliest_year': 1879, 'latest_year': 2010}, 'missing values': 0}

Documentation de CSV/XLSX

YearRemodAdd: {'type': 'int64', 'extremum': {'earliest_year': 1950, 'latest_year': 2010}, 'missing values': 0}

RoofStyle: {'type': 'object', 'missing_values': 0, '>10% appearance': {'Gable': 80.12, 'Hip': 18.16}}

RoofMatl: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 1442, 'nbr of highest app': 1, '>10% appearance': {'CompShg': 98.83}}

Exterior1st: {'type': 'object', 'missing_values': 1, '>10% appearance': {'VinylSd': 34.98, 'MetalSd': 15.78, 'HdBoard': 15.09, 'Wd Sdng': 14.06}}

Exterior2nd: {'type': 'object', 'missing_values': 1, '>10% appearance': {'VinylSd': 34.98, 'MetalSd': 15.98, 'HdBoard': 13.65, 'Wd Sdng': 13.31}}

MasVnrType: {'type': 'object', 'missing_values': 16, '>10% appearance': {'None': 60.85, 'BrkFace': 30.08}}

MasVnrArea: {'type': 'float64', 'statistics': {'mean': 100.70914127423822, 'std': 177.5643846385316, 'min': 0.0, 'max': 1290.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 15, 'missing values percent column': 1.0281014393420151, 'missing values percent total': 0}

Documentation de CSV/XLSX

ExterQual: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 892, 'nbr of highest app': 1, '>10% appearance': {'TA': 61.14, 'Gd': 33.65}}

ExterCond: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 5, 'highest appearance': 1256, 'nbr of highest app': 1, '>10% appearance': {'TA': 86.09, 'Gd': 10.49}}

Foundation: {'type': 'object', 'missing_values': 0, '>10% appearance': {'PConc': 45.31, 'CBlock': 41.19, 'BrkTil': 11.31}}

BsmtQual: {'type': 'Identifiant', 'missing_values': 44, 'unique value': 4, 'highest appearance': 634, 'nbr of highest app': 1, '>10% appearance': {'TA': 44.81, 'Gd': 41.77}}

BsmtCond: {'type': 'Identifiant', 'missing_values': 45, 'unique value': 4, 'highest appearance': 1295, 'nbr of highest app': 1, '>10% appearance': {'TA': 91.58}}

BsmtExposure: {'type': 'Identifiant', 'missing_values': 44, 'unique value': 4, 'highest appearance': 951, 'nbr of highest app': 1, '>10% appearance': {'No': 67.21, 'Av': 13.92, 'Gd': 10.04}}

BsmtFinType1: {'type': 'Identifiant', 'missing_values': 42, 'unique value': 6, 'highest appearance': 431, 'nbr of highest app': 1, '>10% appearance': {'GLQ': 30.42, 'Unf': 29.71, 'ALQ': 14.75, 'Rec': 10.94}}

Documentation de CSV/XLSX

BsmtFinSF1: {'type': 'float64', 'statistics': {'mean': 439.2037037037037, 'std': 455.11188761326923, 'min': 0.0, 'max': 4010.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 1, 'missing values percent column': 0.06854009595613435, 'missing values percent total': 0}

BsmtFinType2: {'type': 'Identifiant', 'missing_values': 42, 'unique value': 6, 'highest appearance': 1237, 'nbr of highest app': 1, '>10% appearance': {'Unf': 87.3}}

BsmtFinSF2: {'type': 'float64', 'statistics': {'mean': 52.61934156378601, 'std': 176.69330052860468, 'min': 0.0, 'max': 1526.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 1, 'missing values percent column': 0.06854009595613435, 'missing values percent total': 0}

BsmtUnfSF: {'type': 'float64', 'statistics': {'mean': 554.2949245541838, 'std': 437.11050793486936, 'min': 0.0, 'max': 2140.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 1, 'missing values percent column': 0.06854009595613435, 'missing values percent total': 0}

TotalBsmtSF: {'type': 'float64', 'statistics': {'mean': 1046.1179698216736, 'std': 442.74671244182736, 'min': 0.0, 'max': 5095.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 1, 'missing values percent column': 0.06854009595613435, 'missing values percent total': 0}

Documentation de CSV/XLSX

Heating: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 1446, 'nbr of highest app': 1, '>10% appearance': {'GasA': 99.11}}

HeatingQC: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 5, 'highest appearance': 752, 'nbr of highest app': 1, '>10% appearance': {'Ex': 51.54, 'TA': 29.4, 'Gd': 15.97}}

CentralAir: {'type': 'binary', 'missing_values': 0, 'appearance': {'Y': 93.08, 'N': 6.92}}

Electrical: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 4, 'highest appearance': 1337, 'nbr of highest app': 1, '>10% appearance': {'SBrkr': 91.64}}

1stFlrSF: {'type': 'int64', 'extremum': {'earliest_year': 407, 'latest_year': 5095}, 'missing values': 0}

2ndFlrSF: {'type': 'int64', 'statistics': {'mean': 325.96778615490064, 'std': 420.4660584352281, 'min': 0, 'max': 1862, 'q1': 0.0, 'q3': 676.0, 'shapiro_statistic': 0.7628400921821594, 'shapiro_pvalue': 1.45622936412635e-41, 'outliers': 5}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

LowQualFinSF: {'type': 'int64', 'statistics': {'mean': 3.5435229609321452, 'std': 44.0281546339866, 'min': 0, 'max': 1064, 'q1': 0.0, 'q3': 0.0, 'shapiro_statistic': 0.05454540252685547, 'shapiro_pvalue': 0.0, 'outliers': 14}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

Documentation de CSV/XLSX

GrLivArea: {'type': 'int64', 'extremum': {'earliest_year': 407, 'latest_year': 5095}, 'missing values': 0}

BsmtFullBath: {'type': 'Identifiant', 'missing_values': 2, 'unique value': 4, 'highest appearance': 849, 'nbr of highest app': 1, '>10% appearance': {0.0: 58.27, 1.0: 40.08}}

BsmtHalfBath: {'type': 'Identifiant', 'missing_values': 2, 'unique value': 3, 'highest appearance': 1364, 'nbr of highest app': 1, '>10% appearance': {0.0: 93.62}}

FullBath: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 5, 'highest appearance': 762, 'nbr of highest app': 1, '>10% appearance': {2: 52.23, 1: 45.17}}

HalfBath: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 3, 'highest appearance': 921, 'nbr of highest app': 1, '>10% appearance': {0: 63.13, 1: 35.98}}

BedroomAbvGr: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 7, 'highest appearance': 792, 'nbr of highest app': 1, '>10% appearance': {3: 54.28, 2: 26.32, 4: 12.82}}

KitchenAbvGr: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 3, 'highest appearance': 1393, 'nbr of highest app': 1, '>10% appearance': {1: 95.48}}

KitchenQual: {'type': 'Identifiant', 'missing_values': 1, 'unique value': 4, 'highest appearance': 757,

'nbr of highest app': 1, '>10% appearance': {'TA': 51.92, 'Gd': 38.75}}

TotRmsAbvGrd: {'type': 'int64', 'statistics': {'mean': 6.385195339273475, 'std': 1.5083773876622801, 'min': 3, 'max': 15, 'q1': 5.0, 'q3': 7.0, 'shapiro_statistic': 0.9267451167106628, 'shapiro_pvalue': 4.277491022781633e-26, 'outliers': 21}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

Functional: {'type': 'object', 'missing_values': 2, '>10% appearance': {'Typ': 93.14}}

Fireplaces: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 5, 'highest appearance': 730, 'nbr of highest app': 1, '>10% appearance': {0: 50.03, 1: 42.36}}

FireplaceQu: {'type': 'Identifiant', 'missing_values': 730, 'unique value': 5, 'highest appearance': 364, 'nbr of highest app': 1, '>10% appearance': {'Gd': 49.93, 'TA': 38.27}}

GarageType: {'type': 'object', 'missing_values': 76, '>10% appearance': {'Attchd': 61.68, 'Detchd': 28.34}}

GarageYrBlt: {'type': 'Identifiant', 'missing_values': 78, 'unique value': 97, 'highest appearance': 77, 'nbr of highest app': 1, '>10% appearance': {}}

Documentation de CSV/XLSX

GarageFinish: {'type': 'Identifiant', 'missing_values': 78, 'unique value': 3, 'highest appearance': 625, 'nbr of highest app': 1, '>10% appearance': {'Unf': 45.26, 'RFn': 28.17, 'Fin': 26.57}}

GarageCars: {'type': 'Identifiant', 'missing_values': 1, 'unique value': 6, 'highest appearance': 770, 'nbr of highest app': 1, '>10% appearance': {2.0: 52.81, 1.0: 27.91, 3.0: 13.24}}

GarageArea: {'type': 'float64', 'statistics': {'mean': 472.76886145404666, 'std': 216.97416468040274, 'min': 0.0, 'max': 1488.0, 'q1': nan, 'q3': nan, 'shapiro_statistic': nan, 'shapiro_pvalue': 1.0, 'outliers': 0}, 'missing values': 1, 'missing values percent column': 0.06854009595613435, 'missing values percent total': 0}

GarageQual: {'type': 'Identifiant', 'missing_values': 78, 'unique value': 4, 'highest appearance': 1293, 'nbr of highest app': 1, '>10% appearance': {'TA': 93.63}}

GarageCond: {'type': 'Identifiant', 'missing_values': 78, 'unique value': 5, 'highest appearance': 1328, 'nbr of highest app': 1, '>10% appearance': {'TA': 96.16}}

PavedDrive: {'type': 'Identifiant', 'missing_values': 0, 'unique value': 3, 'highest appearance': 1301, 'nbr of highest app': 1, '>10% appearance': {'Y': 89.17}}

WoodDeckSF: {'type': 'int64', 'statistics': {'mean': 93.17477724468814, 'std': 127.70109578290862, 'min': 0, 'max': 1424, 'q1': 0.0, 'q3': 168.0, 'shapiro_statistic': 0.7424671649932861, 'shapiro_pvalue':

8.646011524884121e-43, 'outliers': 35}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

OpenPorchSF: {'type': 'int64', 'statistics': {'mean': 48.3139136394791, 'std': 68.85975370488065, 'min': 0, 'max': 742, 'q1': 0.0, 'q3': 72.0, 'shapiro_statistic': 0.7170445919036865, 'shapiro_pvalue': 3.2229864679470793e-44, 'outliers': 79}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

EnclosedPorch: {'type': 'int64', 'statistics': {'mean': 24.24331734064428, 'std': 67.20472248301603, 'min': 0, 'max': 1012, 'q1': 0.0, 'q3': 0.0, 'shapiro_statistic': 0.4137563705444336, 'shapiro_pvalue': 0.0, 'outliers': 251}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

3SsnPorch: {'type': 'int64', 'statistics': {'mean': 1.794379712131597, 'std': 20.200915327384696, 'min': 0, 'max': 360, 'q1': 0.0, 'q3': 0.0, 'shapiro_statistic': 0.06270909309387207, 'shapiro_pvalue': 0.0, 'outliers': 13}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

ScreenPorch: {'type': 'int64', 'statistics': {'mean': 17.064427690198766, 'std': 56.59035938863508, 'min': 0, 'max': 576, 'q1': 0.0, 'q3': 0.0, 'shapiro_statistic': 0.33829742670059204, 'shapiro_pvalue': 0.0, 'outliers': 140}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

PoolArea: {'type': 'int64', 'statistics': {'mean': 1.7443454420836189, 'std': 30.48119501238561, 'min': 0, 'max': 800, 'q1': 0.0, 'q3': 0.0, 'shapiro_statistic': 0.03167390823364258, 'shapiro_pvalue': 0.0, 'outliers': 6}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

PoolQC: {'type': 'binary', 'missing_values': 1456, 'appearance': {'Ex': 66.67, 'Gd': 33.33}}

Fence: {'type': 'object', 'missing_values': 1169, '>10% appearance': {'MnPrv': 59.31, 'GdPrv': 20.34, 'GdWo': 20.0}}

MiscFeature: {'type': 'Identifiant', 'missing_values': 1408, 'unique value': 3, 'highest appearance': 46, 'nbr of highest app': 1, '>10% appearance': {'Shed': 90.2}}

MiscVal: {'type': 'int64', 'extremum': {'earliest_year': 0, 'latest_year': 17000}, 'missing values': 0}

MoSold: {'type': 'int64', 'statistics': {'mean': 6.104180945853324, 'std': 2.7214987626111355, 'min': 1, 'max': 12, 'q1': 4.0, 'q3': 8.0, 'shapiro_statistic': 0.9724769592285156, 'shapiro_pvalue': 4.48853505480572e-16, 'outliers': 0}, 'missing values': 0, 'missing values percent column': 0.0, 'missing values percent total': 0}

YrSold: {'type': 'int64', 'extremum': {'earliest_year': 2006, 'latest_year': 2010}, 'missing values': 0}

Documentation de CSV/XLSX

SaleType: {'type': 'object', 'missing_values': 1, '>10% appearance': {'WD': 86.28}}

SaleCondition: {'type': 'object', 'missing_values': 0, '>10% appearance': {'Normal': 82.52}}