**Housing Prices**

**Data Codebook**

# **Team** Members: Eilon Bar, Yotam Shavit, Orel Strassman, Ilan Kerpel TeamNumber: 9

# **Data Overview**

## **Credentials**

This data set can be downloaded from Kaggle datasets.

The direct link to the data is:

<https://www.kaggle.com/c/house-prices-advanced-regression-techniques>

## **Business Goals**

This data was collected to answer these questions:

1. Predict the price of a house according to several parameters of the house.
2. Can we scale the parameters according to their contribution to price?
3. Which subset of parameters predicts the price of a house without sacrificing the accuracy too much?
4. Which hidden parameters contribute to the price besides the size of the house, the lot and its condition?

**Reviewers**

1. Real Estate agents
2. Real Estate Companies & REITs - Managers
3. Data Science Experts

**Github Repository**

<https://github.com/silkerp/IDC-BDA-Exercises>

## **Data Description**

This data set is a data frame of 79 variables over 2919 rows. Each row represents a house sale in the city of Ames, Iowa, USA.

There are 19 variables with some missing values in the data.

# **Variables description**

Per each variable describe:

a. Variable label

b. Variable full name or description

c. Possible values and value

d. Summary statistics

e. Missing values

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Type** | **Possible Values** | **Description** |
| Id | int64 | 1, 2, …, 2919 | Identifier |
| SalePrice | int64 | {256000, 106500, 208900, 169990…} | Sale Price |
| GrLivArea | int64 | {2054, 2057, 2058, 2060, 2062… | Above grade (ground) living area square feet |
| TotalBsmtSF | int64 | {0, 2076, 2077, 2078, 2109, 2110, 2121… | Total square feet of basement area |
| LotArea | int64 | {10240, 8197, 8198, 81991.. | Lot size in square feet |
| MSSubClass | int64 | {160, 70, 40, 75, 45, 80, 50, 20, 85, 180, 30, 120, 90, 60, 190} | Identifies the type of dwelling involved in the sale. |
| MSZoning | object | {'RH', 'RM', 'FV', 'C (all)', 'RL'} | Identifies the general zoning classification of  the sale. |
| LotFrontage | float64 | {nan, 21.0, 24.0.. | Linear feet of street connected to property |
| Street | object | {'Grvl', 'Pave'} | Type of road access to property |
| Alley | object | {nan, 'Pave', 'Grvl'} | Type of alley access to property |
| LotShape | object | {'IR1', 'Reg', 'IR3', 'IR2'} | General shape of property |
| LandContour | object | {'Bnk', 'HLS', 'Lvl', 'Low'} | Flatness of the property |
| Utilities | object | {'AllPub', 'NoSeWa'} | Type of utilities available |
| LotConfig | object | {'FR3', 'Corner', 'CulDSac', 'Inside', 'FR2'} | Lot configuration |
| LandSlope | object | {'Gtl', 'Sev', 'Mod'} | Slope of property |
| Neighborhood | object | {'SWISU', 'OldTown', 'Veenker', 'Mitchel', 'NPkVill', 'SawyerW', 'StoneBr', 'Gilbert', 'CollgCr', 'Timber', 'ClearCr', 'Crawfor', 'BrDale', 'BrkSide', 'NWAmes', 'IDOTRR', 'Blueste', 'Edwards', 'Somerst', 'Sawyer', 'Blmngtn', 'NoRidge', 'NAmes', 'NridgHt', 'MeadowV'} | Physical locations within Ames city limits |
| Condition1 | object | {'RRNn', 'PosA', 'RRNe', 'Feedr', 'Artery', 'RRAe', 'Norm', 'RRAn', 'PosN'} | Proximity to various conditions |
| Condition2 | object | {'RRNn', 'PosA', 'Feedr', 'Artery', 'RRAe', 'Norm', 'RRAn', 'PosN'} | Proximity to various conditions (if more than  one is present) |
| BldgType | object | {'TwnhsE', '2fmCon', 'Duplex', '1Fam', 'Twnhs'} | Type of dwelling |
| HouseStyle | object | {'2.5Unf', '2.5Fin', 'SLvl', '1Story', '1.5Unf', 'SFoyer', '2Story', '1.5Fin'} | Style of dwelling |
| OverallQual | int64 | {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} | Rates the overall material and finish of the  house |
| OverallCond | int64 | {1, 2, 3, 4, 5, 6, 7, 8, 9} | Rates the overall condition of the house |
| YearBuilt | int64 | {1872, 1875, 1880, 1882, 1885… | Original construction date |
| YearRemodAdd | int64 | {1950, 1951, 1952, 1953, 1954… | Remodel date (same as construction date if no  remodeling or additions) |
| RoofStyle | object | {'Flat', 'Gambrel', 'Gable', 'Hip', 'Shed', 'Mansard'} | Type of roof |
| RoofMatl | object | {'CompShg', 'Membran', 'Metal', 'WdShngl', 'Roll', 'WdShake', 'ClyTile', 'Tar&Grv'} | Roof material |
| Exterior1st | object | {'Wd Sdng', 'Stucco', 'Stone', 'BrkFace', 'ImStucc', 'BrkComm', 'AsphShn', 'VinylSd', 'MetalSd', 'HdBoard', 'CBlock', 'WdShing', 'CemntBd', 'Plywood', 'AsbShng'} | Exterior covering on house |
| Exterior2nd | object | {'Wd Sdng', 'Stucco', 'Stone', 'Other', 'BrkFace', 'ImStucc', 'AsphShn', 'VinylSd', 'Wd Shng', 'CmentBd', 'MetalSd', 'HdBoard', 'CBlock', 'Brk Cmn', 'Plywood', 'AsbShng'} | Exterior covering on house (if more than one  material) |
| MasVnrType | object | {'None', 'Stone', nan, 'BrkFace', 'BrkCmn'} | Masonry veneer type |
| MasVnrArea | float64 | {nan, 0.0, 513.0… | Masonry veneer area in square feet |
| ExterQual | object | {'Fa', 'TA', 'Gd', 'Ex'} | Evaluates the quality of the material on the  exterior |
| ExterCond | object | {'Po', 'Gd', 'TA', 'Ex', 'Fa'} | Evaluates the present condition of the material  on the exterior |
| Foundation | object | {'Stone', 'CBlock', 'PConc', 'Slab', 'Wood', 'BrkTil'} | Type of foundation |
| BsmtQual | object | {nan, 'Gd', 'TA', 'Ex', 'Fa'} | Evaluates the height of the basement |
| BsmtCond | object | {nan, 'Po', 'Gd', 'TA', 'Fa'} | Evaluates the general condition of the basement |
| BsmtExposure | object | {nan, 'Gd', 'Av', 'Mn', 'No'} | Refers to walkout or garden level walls |
| BsmtFinType1 | object | {nan, 'ALQ', 'Unf', 'LwQ', 'Rec', 'GLQ', 'BLQ'} | Rating of basement finished area |
| BsmtFinSF1 | int64 | {0, 2, 16, 20… | Type 1 finished square feet |
| BsmtFinType2 | object | {nan, 'ALQ', 'Unf', 'LwQ', 'Rec', 'GLQ', 'BLQ'} | Rating of basement finished area (if multiple  types) |
| BsmtFinSF2 | int64 | {0, 1029, 1031, 531, 532… | Type 2 finished square feet |
| BsmtUnfSF | int64 | {0, 14, 15, 23, 26, 29, 30, 32, 35, 36… | Unfinished square feet of basement area |
| Heating | object | {'OthW', 'GasA', 'Wall', 'GasW', 'Floor', 'Grav'} | Type of heating |
| HeatingQC | object | {'Po', 'Gd', 'TA', 'Ex', 'Fa'} | Heating quality and condition |
| CentralAir | object | {'N', 'Y'} | Central air conditioning |
| Electrical | object | {nan, 'FuseF', 'SBrkr', 'FuseA', 'FuseP', 'Mix'} | Electrical system |
| 1stFlrSF | int64 | {2053, 2069, 2073, 2076, 2084… | First Floor square feet |
| 2ndFlrSF | int64 | {0, 2065, 110, 167, 192, 208, 213… | Second floor square feet |
| LowQualFinSF | int64 | {0, 513, 514, 515, 384, 390, 392, 397, 528… | Low quality finished square feet (all floors) |
| BsmtFullBath | int64 | {0, 1, 2, 3} | Basement full bathrooms |
| BsmtHalfBath | int64 | {0, 1, 2} | Basement half bathrooms |
| FullBath | int64 | {0, 1, 2, 3} | Full bathrooms above grade |
| HalfBath | int64 | {0, 1, 2} | Half baths above grade |
| BedroomAbvGr | int64 | {0, 1, 2, 3, 4, 5, 6, 8} | Bedrooms above grade (does NOT include basement  bedrooms) |
| KitchenAbvGr | int64 | {0, 1, 2, 3} | Kitchens above grade |
| KitchenQual | object | {'Fa', 'TA', 'Gd', 'Ex'} | Kitchen quality |
| TotRmsAbvGrd | int64 | {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14} | Total rooms above grade (does not include  bathrooms) |
| Functional | object | {'Mod', 'Maj2', 'Min1', 'Min2', 'Sev', 'Typ', 'Maj1'} | Home functionality (Assume typical unless  deductions are warranted) |
| Fireplaces | int64 | {0, 1, 2, 3} | Number of fireplaces |
| FireplaceQu | object | {nan, 'Po', 'Gd', 'TA', 'Ex', 'Fa'} | Fireplace quality |
| GarageType | object | {nan, 'Attchd', '2Types', 'Detchd', 'CarPort', 'Basment', 'BuiltIn'} | Garage location |
| GarageYrBlt | float64 | {nan, 1900.0, 1906.0,... | Year garage was built |
| GarageFinish | object | {nan, 'Unf', 'Fin', 'RFn'} | Interior finish of the garage |
| GarageCars | int64 | {0, 1, 2, 3, 4} | Size of garage in car capacity |
| GarageArea | int64 | {0, 160, 164, 180, 186, 189… | Size of garage in square feet |
| GarageQual | object | {nan, 'Po', 'Gd', 'TA', 'Ex', 'Fa'} | Garage quality |
| GarageCond | object | {nan, 'Po', 'Gd', 'TA', 'Ex', 'Fa'} | Garage condition |
| PavedDrive | object | {'N', 'Y', 'P'} | Paved driveway |
| WoodDeckSF | int64 | {0, 517, 519, 12, 24, 536, 26… | Wood deck area in square feet |
| OpenPorchSF | int64 | {0, 4, 8, 10, 11, 12, 523, 15… | Open porch area in square feet |
| EnclosedPorch | int64 | {0, 19, 20, 24, 30, 32, 34, 36, 37, 39… | Enclosed porch area in square feet |
| 3SsnPorch | int64 | {0, 320, 130, 162, 196, 96, 290, 168, 140, 238, 144, 304, 180, 245, 182, 407, 23, 153, 508, 216} | Three season porch area in square feet |
| ScreenPorch | int64 | {0, 128, 130, 259, 260, 385, 263, 265, 266… | Screen porch area in square feet |
| PoolArea | int64 | {0, 512, 576, 480, 738, 519, 648, 555} | Pool area in square feet |
| PoolQC | object | {nan, 'Gd', 'Fa', 'Ex'} | Pool quality |
| Fence | object | {nan, 'GdPrv', 'MnPrv', 'GdWo', 'MnWw'} | Fence quality |
| MiscFeature | object | {nan, 'Othr', 'Shed', 'Gar2', 'TenC'} | Miscellaneous feature not covered in other  categories |
| MiscVal | int64 | {0, 15500, 400, 1300, 800, 3500, 1200, 560, 54, 700, 450, 2500, 2000, 600, 350, 480, 620, 8300, 500, 1400, 1150} | $Value of miscellaneous feature |
| MoSold | int64 | {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} | Month Sold (MM) |
| YrSold | int64 | {2006, 2007, 2008, 2009, 2010} | Year Sold (YYYY) |
| SaleType | object | {'CWD', 'New', 'ConLD', 'ConLw', 'Con', 'ConLI', 'COD', 'WD', 'Oth'} | Type of sale |
| SaleCondition | object | {'Family', 'Partial', 'Normal', 'AdjLand', 'Abnorml', 'Alloca'} | Condition of sale |

**Related Academic Articles (patent or blog):**

Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project.

<http://jse.amstat.org/v19n3/decock.pdf>

# Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data.

<https://www.sciencedirect.com/science/article/pii/S0957417414007325>

Predicting House Price Using Regression Algorithm for Machine Learning

# <https://yalantis.com/blog/predictive-algorithm-for-house-price/>