

Ames Housing price prediction Using Machine Learning techniques

Capstone 1 Project for Data science Career Track

The price of a house is dependent on various factors like size or area, how many bedrooms, location, the price of other houses, and many other factors. **Real estate investors** would like to find out the actual cost of the house in order to buy and sell real estate properties. They will lose money when they pay more than the current market cost of the house and when they sell for less than current market cost. **The banks** also want to find the current market price for the house, when they use someone's house as collateral for loans. Sometimes loan applicant overvalued their house to borrow the maximum loan from the bank. **Banks and financial institutions** also provide mortgage loan to home buyers. **Local home buyers** can also predict the price of the house to find out if a seller is asking for too much. **The local seller** can also predict their house price and find out how much is a fair market price.

Dataset :

- SalePrice - the property's sale price in dollars. This is the target variable that you're trying to predict.
- MSSubClass: The building class
- MSZoning: The general zoning classification
- LotFrontage: Linear feet of street connected to property
- LotArea: Lot size in square feet
- Street: Type of road access
- Alley: Type of alley access
- LotShape: General shape of property
- LandContour: Flatness of the property
- Utilities: Type of utilities available
- LotConfig: Lot configuration
- LandSlope: Slope of property
- Neighborhood: Physical locations within Ames city limits
- Condition1: Proximity to main road or railroad
- Condition2: Proximity to main road or railroad (if a second is present)
- BldgType: Type of dwelling
- HouseStyle: Style of dwelling
- OverallQual: Overall material and finish quality

- OverallCond: Overall condition rating
- YearBuilt: Original construction date
- YearRemodAdd: Remodel date
- RoofStyle: Type of roof
- RoofMatl: Roof material
- Exterior1st: Exterior covering on house
- Exterior2nd: Exterior covering on house (if more than one material)
- MasVnrType: Masonry veneer type
- MasVnrArea: Masonry veneer area in square feet
- ExterQual: Exterior material quality
- ExterCond: Present condition of the material on the exterior
- Foundation: Type of foundation
- BsmtQual: Height of the basement
- BsmtCond: General condition of the basement
- BsmtExposure: Walkout or garden level basement walls
- BsmtFinType1: Quality of basement finished area
- BsmtFinSF1: Type 1 finished square feet
- BsmtFinType2: Quality of second finished area (if present)
- BsmtFinSF2: Type 2 finished square feet
- BsmtUnfSF: Unfinished square feet of basement area
- TotalBsmtSF: Total square feet of basement area
- Heating: Type of heating
- HeatingQC: Heating quality and condition
- CentralAir: Central air conditioning
- Electrical: Electrical system
- 1stFlrSF: First Floor square feet
- 2ndFlrSF: Second floor square feet
- LowQualFinSF: Low quality finished square feet (all floors)
- GrLivArea: Above grade (ground) living area square feet
- BsmtFullBath: Basement full bathrooms
- BsmtHalfBath: Basement half bathrooms
- FullBath: Full bathrooms above grade
- HalfBath: Half baths above grade
- Bedroom: Number of bedrooms above basement level
- Kitchen: Number of kitchens
- KitchenQual: Kitchen quality
- TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)
- Functional: Home functionality rating
- Fireplaces: Number of fireplaces
- FireplaceQu: Fireplace quality
- GarageType: Garage location
- GarageYrBlt: Year garage was built
- GarageFinish: Interior finish of the garage
- GarageCars: Size of garage in car capacity
- GarageArea: Size of garage in square feet
- GarageQual: Garage quality
- GarageCond: Garage condition
- PavedDrive: Paved driveway

- WoodDeckSF: Wood deck area in square feet
- OpenPorchSF: Open porch area in square feet
- EnclosedPorch: Enclosed porch area in square feet
- 3SsnPorch: Three season porch area in square feet
- ScreenPorch: Screen porch area in square feet
- PoolArea: Pool area in square feet
- PoolQC: Pool quality
- Fence: Fence quality
- MiscFeature: Miscellaneous feature not covered in other categories
- MiscVal: \$Value of miscellaneous feature
- MoSold: Month Sold
- YrSold: Year Sold
- SaleType: Type of sale
- SaleCondition: Condition of sale

Approach

Data cleaning :

- > perform data type transformation of needed (time series data)
- > i will be separating numerical and categorical data as imputation idea as well as EDA approach will be different
- > outlier detection and understanding their implication i.e does removing those outlier does not make data partial towards certain decision

Initial Analysis:

- > Perform Exploratory data analysis based on variable time for e.g Scatter plot and histogram for continuous and categorical data .
- > Find correlation between variable independent variable to find redundant column
- > Also find relation between dependant and independent variable and significance of those independent variable with respect to dependant variable
- > will also be performing some feature engineering

Analysis

- > With respect to predicting price, regression will be go-to algorithm with use of backward elimination technique to select variables
- > will also be performing Advanced regression techniques like random forest and gradient boosting

> detecting important factors that affect the price of house

Writeup:

The deliverable will include a powerpoint presentation and python code. Powerpoint will include related EDA Graph, any correlation analysis , final model (regression) , confusion matrix , related statistics summary, recommendations