**Applied Data Science With Python**

Course-End Project Problem Statement



**Course-End Project: Feature Engineering**

**Project Statement:**

While searching for the dream house, the buyer looks at various factors, not just at the height of the basement ceiling or the proximity to an east-west railroad.

Using the dataset, find the factors that influence price negotiations while buying a house.

There are 79 explanatory variables describing every aspect of residential homes in Ames, Iowa.

**Dataset Description:**

| **Variable** | **Description** |
| --- | --- |
| SalePrice | The property's sale price is 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 the 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 bathrooms 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 the garage in car capacity |
| GarageArea | Size of the 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 |

**Note:**

1. Download the “PEP1.csv” using the link given in the Feature Engineering project problem statement
2. For a detailed description of the dataset, you can download and refer to data\_description.txt using the link given in the Feature Engineering project problem statement

**Perform the following steps:**

1. Understand the dataset:
   1. Identify the shape of the dataset
   2. Identify variables with null values
   3. Identify variables with unique values
2. Generate a separate dataset for numerical and categorical variables
3. EDA of numerical variables:
   1. Missing value treatment
   2. Identify the skewness and distribution
   3. Identify significant variables using a correlation matrix
   4. Pair plot for distribution and density
4. EDA of categorical variables
   1. Missing value treatment
   2. Count plot and box plot for bivariate analysis
   3. Identify significant variables using p-values and Chi-Square values
5. Combine all the significant categorical and numerical variables
6. Plot box plot for the new dataset to find the variables with outliers

**Note:** The last two points are performed to make the new dataset ready for training and prediction.