****

**Superior University Lahore**

***Lab Task # 1***

**Name:** Muhammad Abdullah Kashif

**Roll No**: SU92-BSAIM-F23-070

**Instructor**: Sir Rasikh Ali

# Course: Programming for Artificial Intelligence (Lab)

House Price Prediction

**House Price Prediction Dataset Analysis**

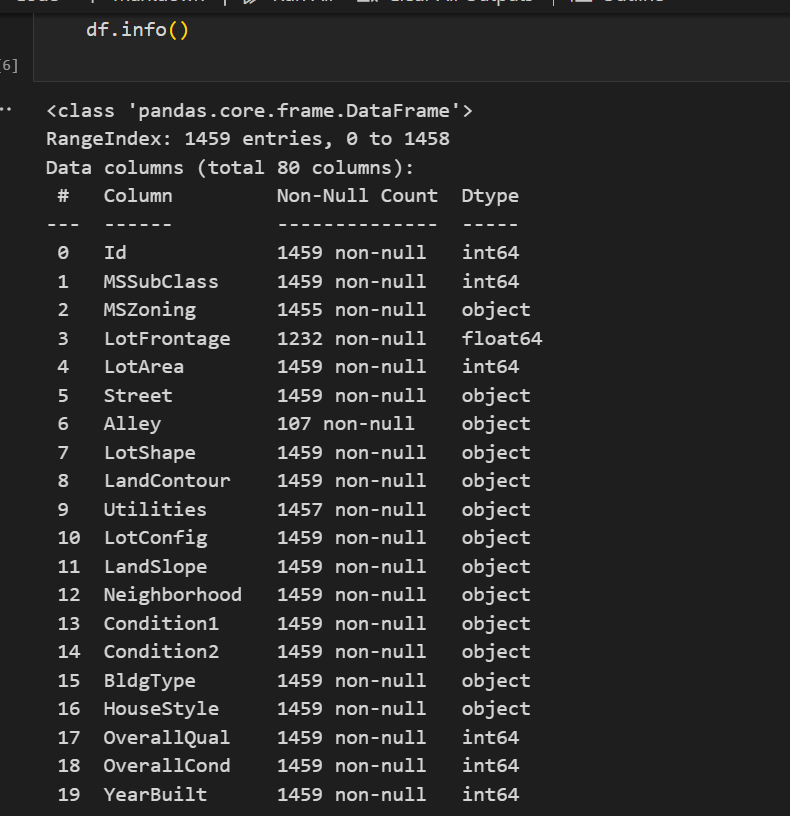
**1. Introduction**

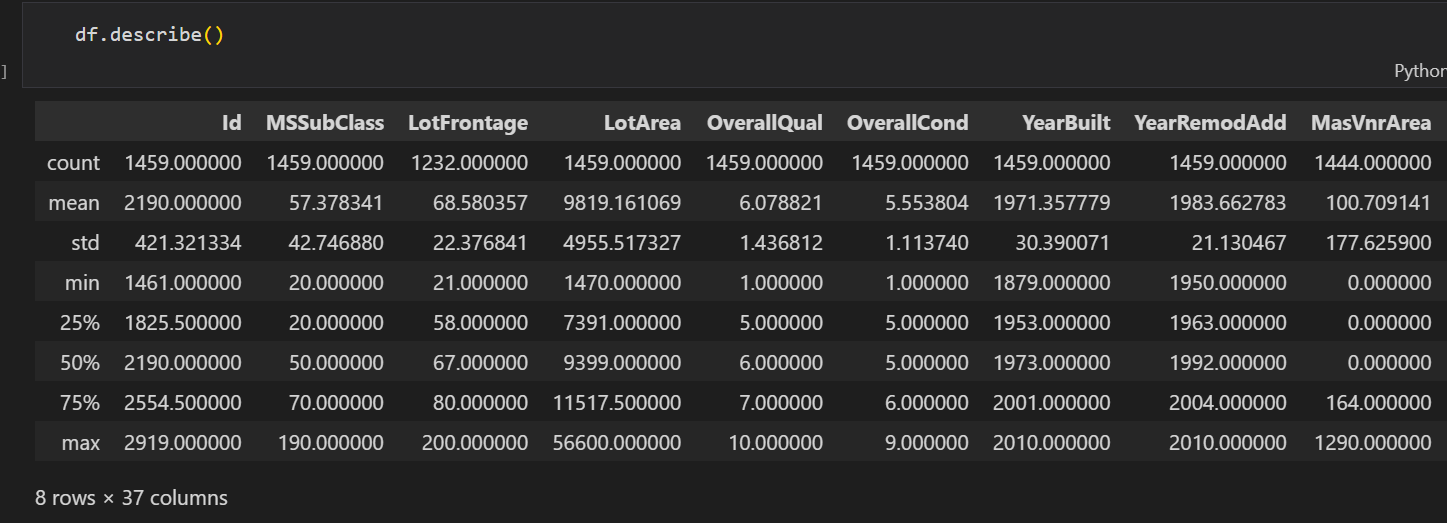
The house price prediction dataset is used for estimating the sale price of residential houses based on various features. The dataset consists of structural, locational, and economic attributes, which impact the valuation of houses. This analysis aims to explore, clean, and model the dataset to predict house prices effectively.

**2. Dataset Description**

The dataset comprises 81 features in the training set and 80 in the test set. It includes numerical and categorical variables that describe different aspects of a house, such as:

* **Structural Features**: Area, number of bedrooms, bathrooms, floors, basement size, etc.
* **Location Features**: Neighborhood, street type, proximity to roads, etc.
* **Condition & Quality Features**: Overall condition, material quality, year built/remodeled, etc.
* **Additional Features**: Garage, fireplace, pool, porch size, etc.



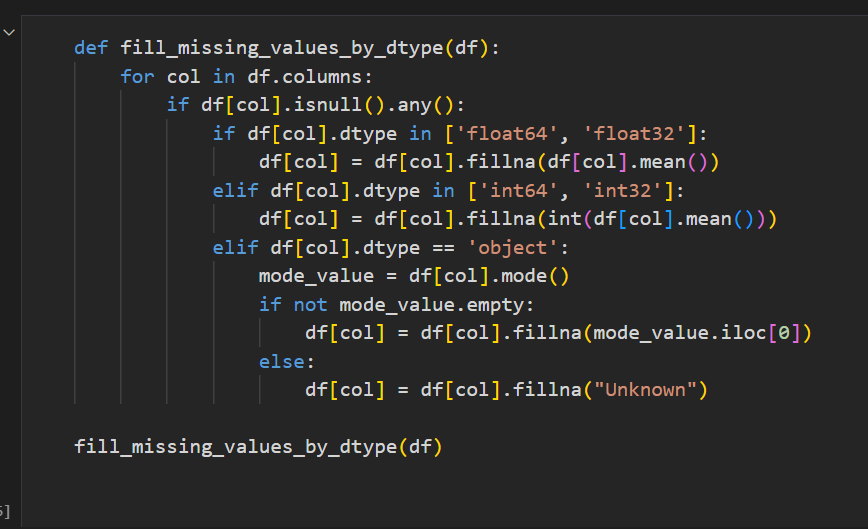


**3. Data Analysis & Preprocessing**

**3.1 Handling Missing Values**

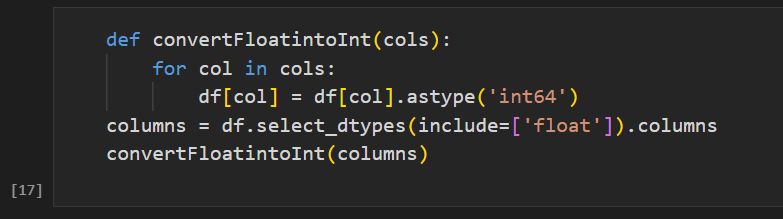
Several features contain missing values, including:

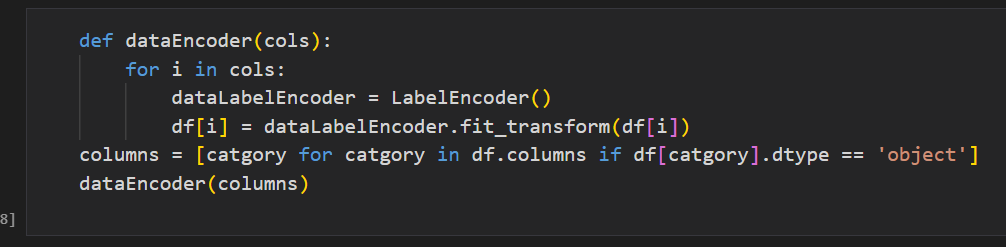
* LotFrontage: Missing values will be filled using the median based on neighborhood.
* Alley, PoolQC, Fence, MiscFeature: Many missing values indicate 'No Alley', 'No Pool', etc., so they will be replaced accordingly.
* GarageYrBlt, GarageType, FireplaceQu: Missing values will be imputed with appropriate methods based on related features.



**3.2 Feature Encoding & Transformation**

* Categorical variables will be encoded using label encoding or one-hot encoding.
* Skewed numerical features will be transformed using log transformation to improve model performance.

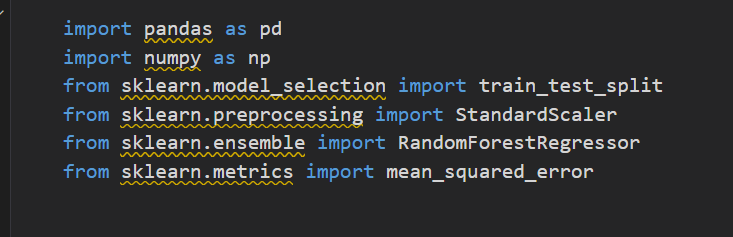


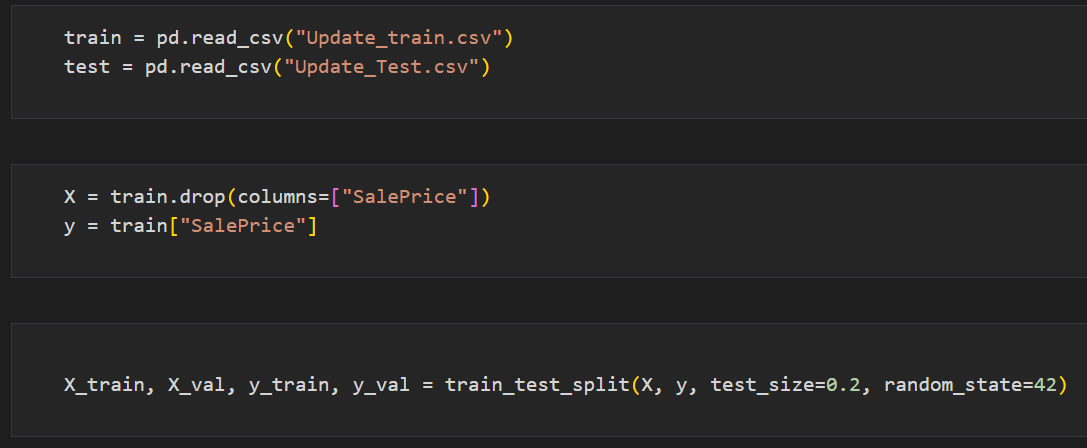


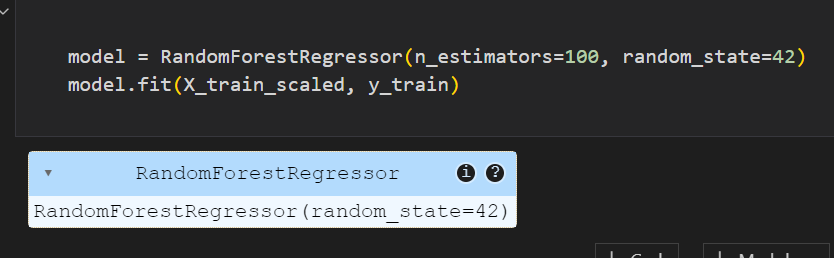
**4. Model Implementation**

Various regression models will be tested for house price prediction:

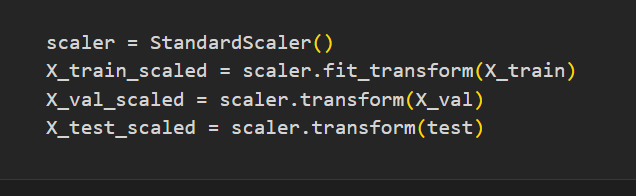
* **Random Forest** : To capture non-linear relationships.







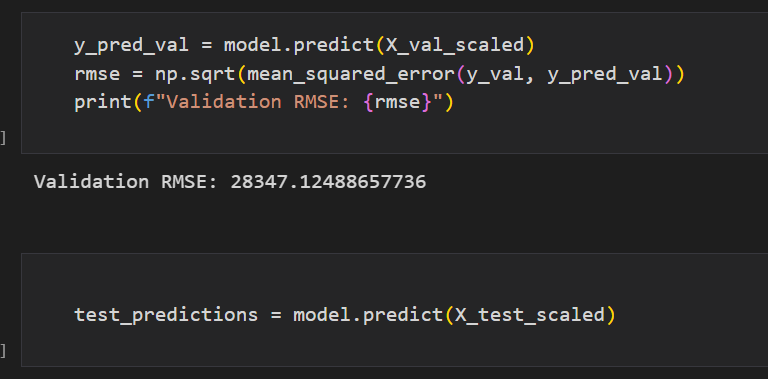
**5. Scaling using Standard Scaler:**

****

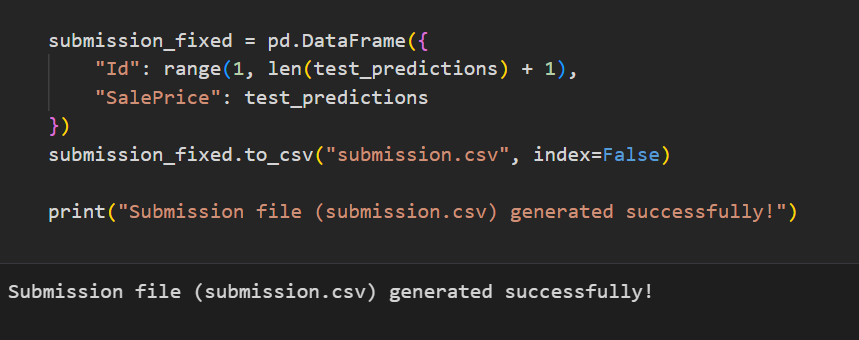
**6. Model Evaluation:**

The models will be evaluated using:

* **Root Mean Squared Error (RMSE)**: Measures prediction accuracy.
* **R-squared (R²)**: Indicates model fit quality.
* **Cross-validation**: To check model robustness.



**7. Prediction & Results:**



**8. Conclusion**

This analysis provides an in-depth exploration of the dataset, missing values handling, feature engineering, and modeling approaches. The best-performing model will be chosen based on predictive accuracy and stability.