**House Price Prediction Using Linear Regression**

**Introduction**

This project aims to develop a machine learning model to predict house prices using key features such as living area size, number of bedrooms, and bathrooms. The dataset used is sourced from the "House Prices: Advanced Regression Techniques" competition on Kaggle.

**Objective**

To build a predictive model that accurately estimates the selling price of a house based on features like:

* GrLivArea: Above-ground living area (sq ft)
* BedroomAbvGr: Number of bedrooms above ground
* FullBath: Number of full bathrooms

**Methodology**

* **Tool Used**: Python
* **Libraries**: pandas, scikit-learn
* **Model Used**: LinearRegression from sklearn.linear\_model
* **Steps**:
  1. Load training and test datasets.
  2. Select relevant features for the model.
  3. Train the linear regression model on the training data.
  4. Predict the house prices for the test data.
  5. Present the predicted results.

Code

import pandas as pd  
from sklearn.linear\_model import LinearRegression # For regression modeling  
  
train\_df=pd.read\_csv('train.csv')  
test\_df=pd.read\_csv('test.csv')  
# Select features and target  
features = ['GrLivArea', 'BedroomAbvGr', 'FullBath']  
X\_train = train\_df[features]  
y\_train = train\_df['SalePrice']  
  
#training the model  
model=LinearRegression()  
model.fit(X\_train,y\_train)  
  
#predicting  
X\_test=test\_df[features]  
predict =model.predict(X\_test)  
  
#creating dataframe with prediction  
result=pd.DataFrame({  
 'Id':test\_df['Id'],  
 'predicted\_saleprice' :predict  
})  
  
#Display the predicted values  
print(result)

output

Id predicted\_saleprice

0 1461 120100.812977

1 1462 139898.208279

2 1463 202611.414586

3 1464 199859.871426

4 1465 192059.204300

... ... ...

1454 2915 113813.579125

1455 2916 113813.579125

1456 2917 100482.394785

1457 2918 100386.048506

1458 2919 243444.315076

[1459 rows x 2 columns]

Observation

The model uses a linear relationship to estimate prices.

As expected, larger homes and homes with more bathrooms typically have higher predicted prices.

**Conclusion**

This project demonstrated how a **Linear Regression model** can be used to predict house prices based on basic numerical features. The model is easy to interpret and serves as a baseline for more complex approaches. Future work could include feature engineering, use of advanced models like Random Forest or XGBoost, and error analysis.