

**The Superior University**

***Session 2023-2025***

***Department of Software Engineering***

***Faculty of Computer Science & Information Technology***

***The Superior University, Lahore***

***Course: Programming For Artificial Intelligence***

***Course Instructor:Sir Rasikh ali***

***Semester 4 BSAI***

***Fall 2025***

***Name:Sara hanif***

***Roll no:082***

***LAB TASK 2***

***Data set download from kaggle:***

<https://www.kaggle.com/competitions/house>[-pricing-predicition](https://www.kaggle.com/code/umehabibaakbarali/house-pricing-predicition)

***Environment setup and dataloading:***

* The notebook starts by importing necessary libraries such as pandas, numpy, and os.
* It lists the files available in the input directory using os.walk.
* The dataset (train.csv) is loaded into a pandas DataFrame (df).

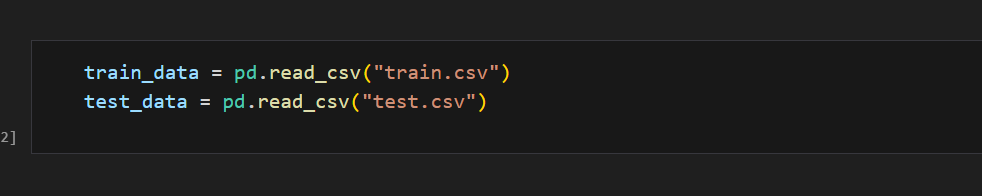
Data exploration:

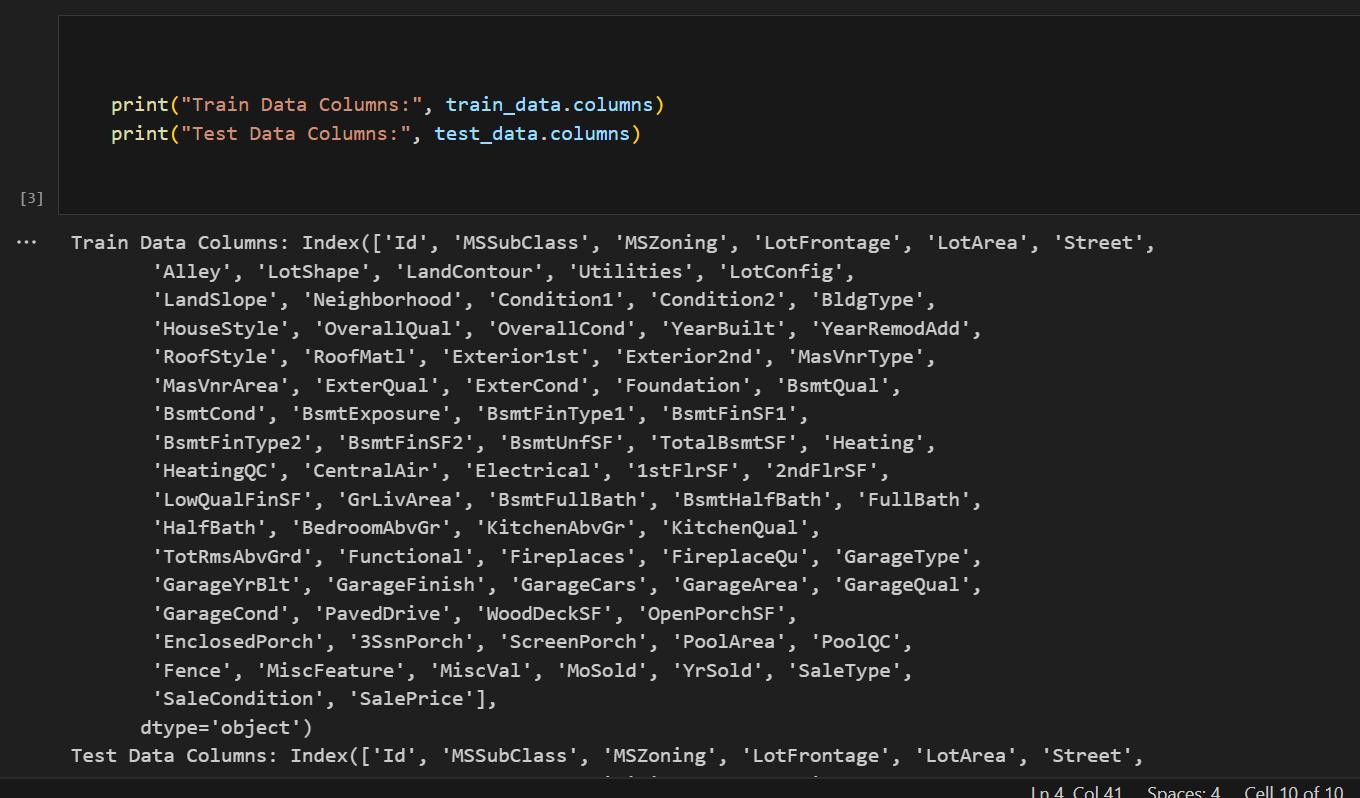
• The first few rows of the dataset are displayed using df.head(5) to get an initial look at the data.

**IMPORT LIBARARIES:**

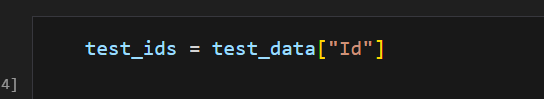
******

**LOAD DATASETS:**





**Use "Id" instead of "PassengerId" for house dataset:**

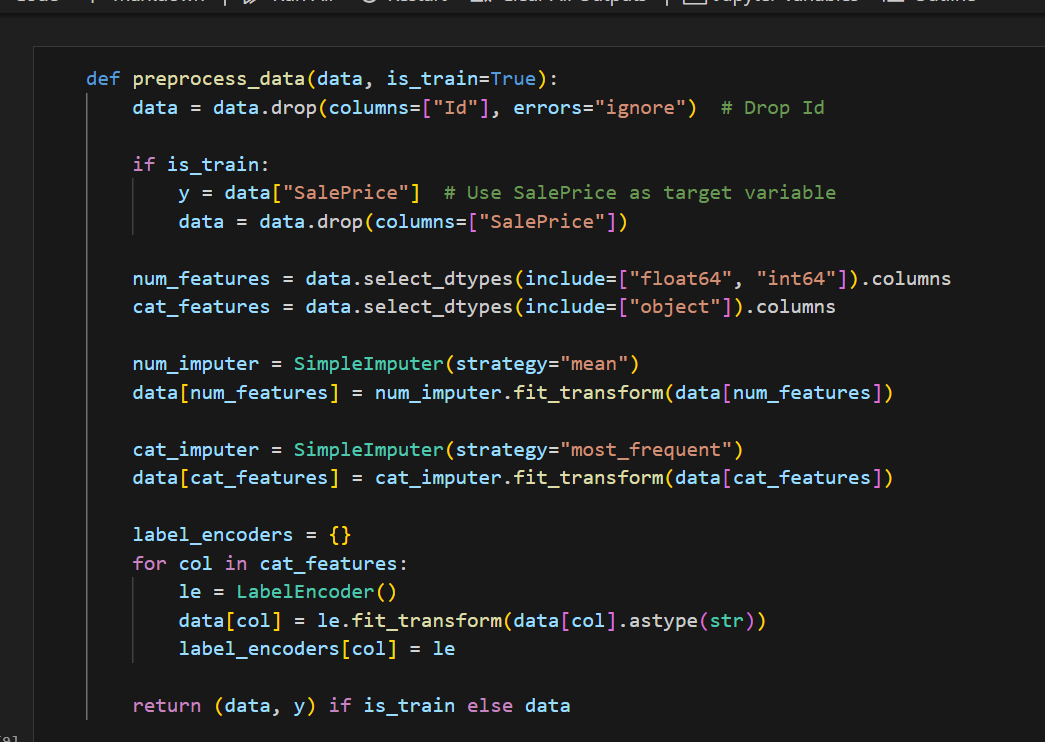
****

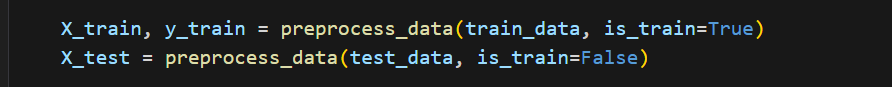
**Data preprocessing:**

* **Handling Missing Values:**
  + Columns with more than 50% missing values are dropped.

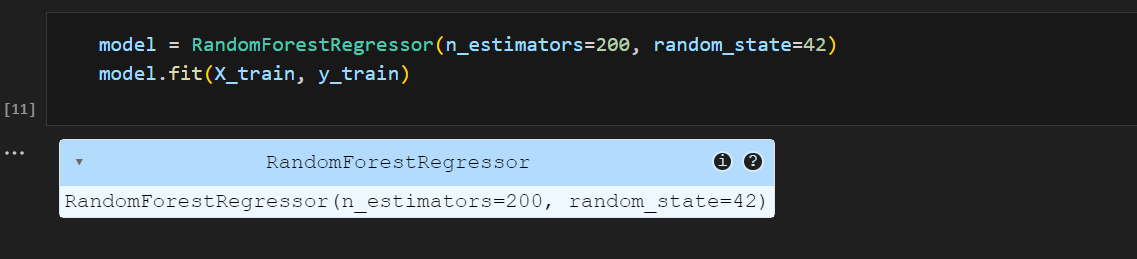
o Missing values in numerical columns are filled with the median, and missing values in categorical columns are filled with the string 'Unknown'.

* **Encoding Categorical Variables:**
  + Categorical columns are one-hot encoded using pd.get\_dummies.

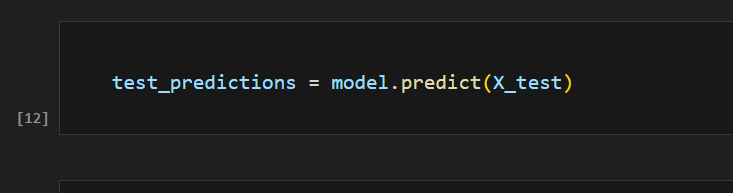
****

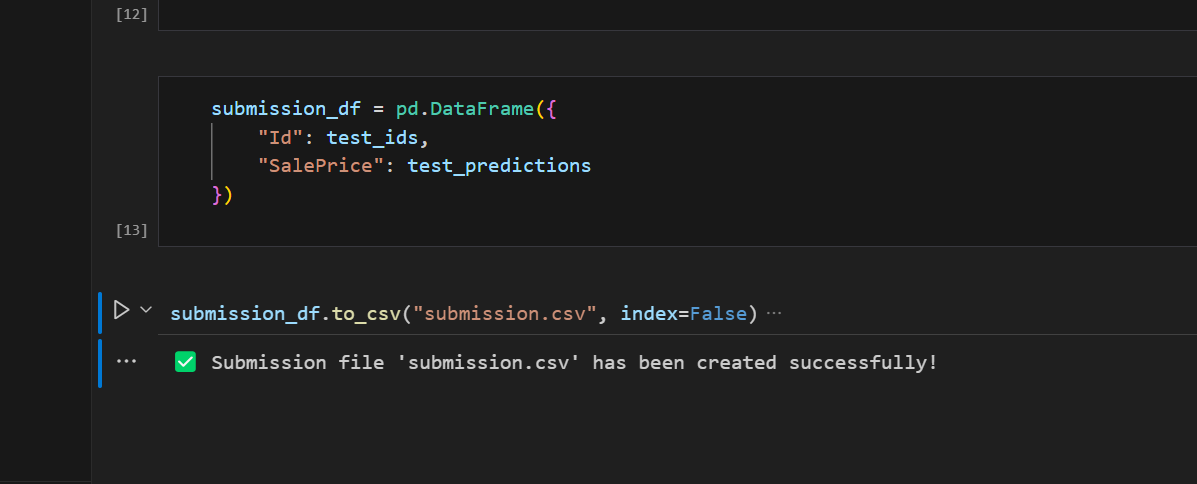
**Preprocess data:** ****

**Train Random Forest Regressor:**

****

**Predict house prices:**

****

**submission file:** ****