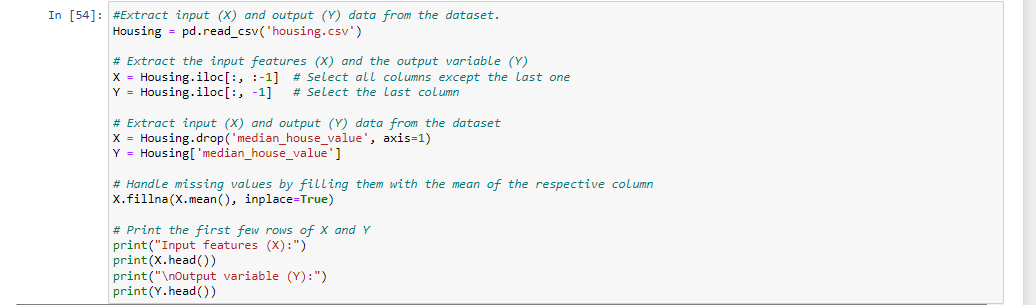
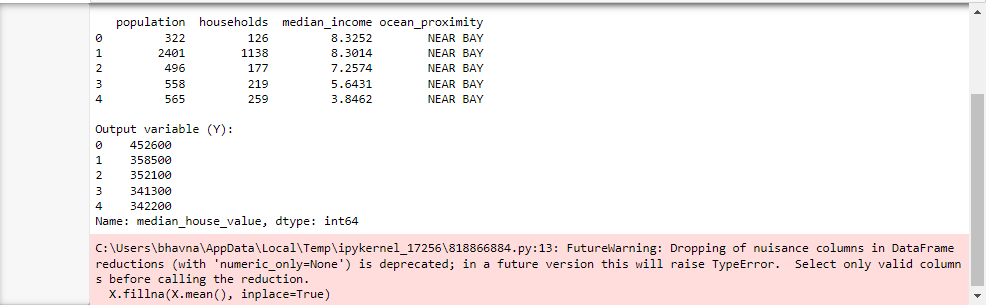
# **Load the data:**

* Read the “**housing.csv**” file from the folder into the program.
* Print first few rows of this data.



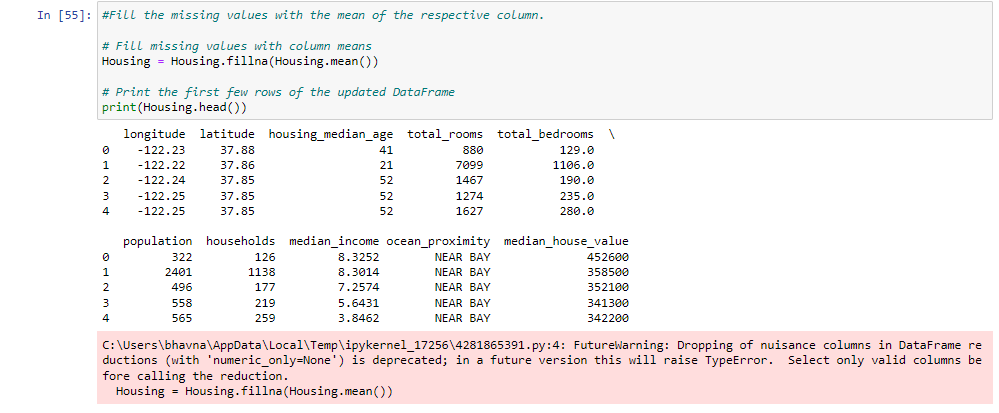
* Extract input (X) and output (Y) data from the dataset.





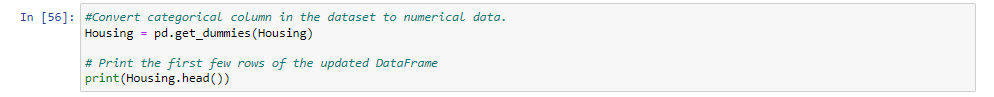
# **Handle missing values**:

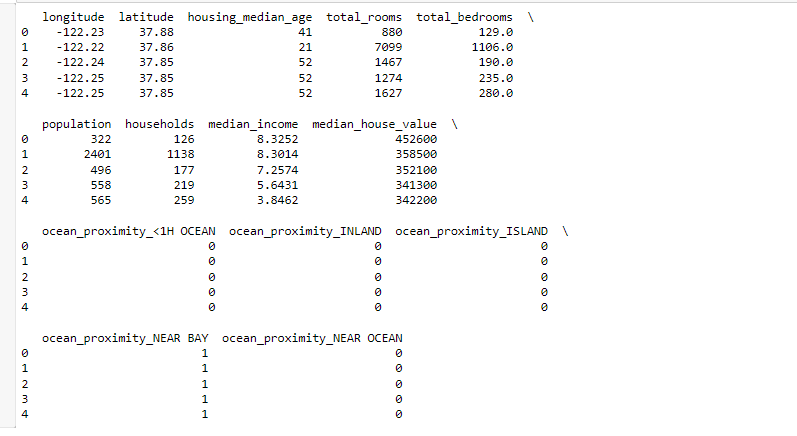
* Fill the missing values with the mean of the respective column.



# **Encode categorical data**:

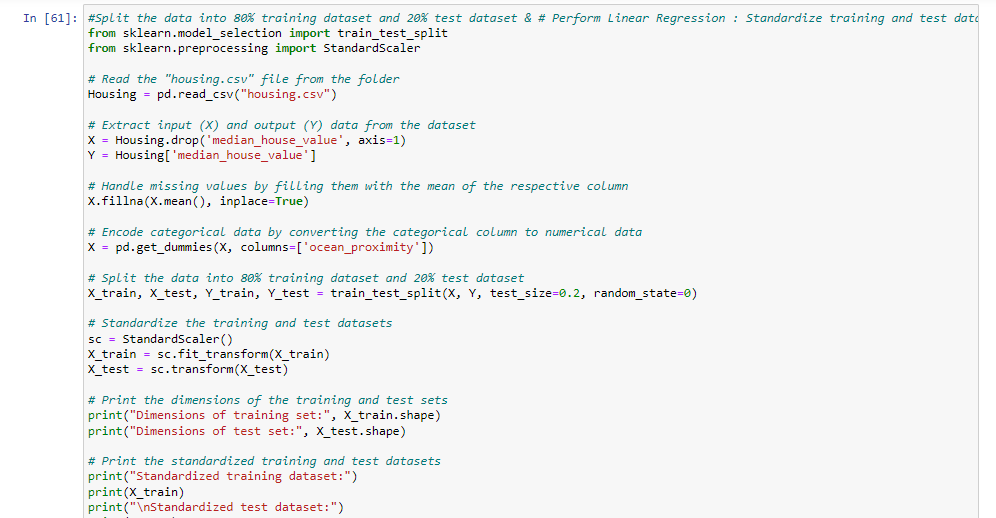
* Convert categorical column in the dataset to numerical data.

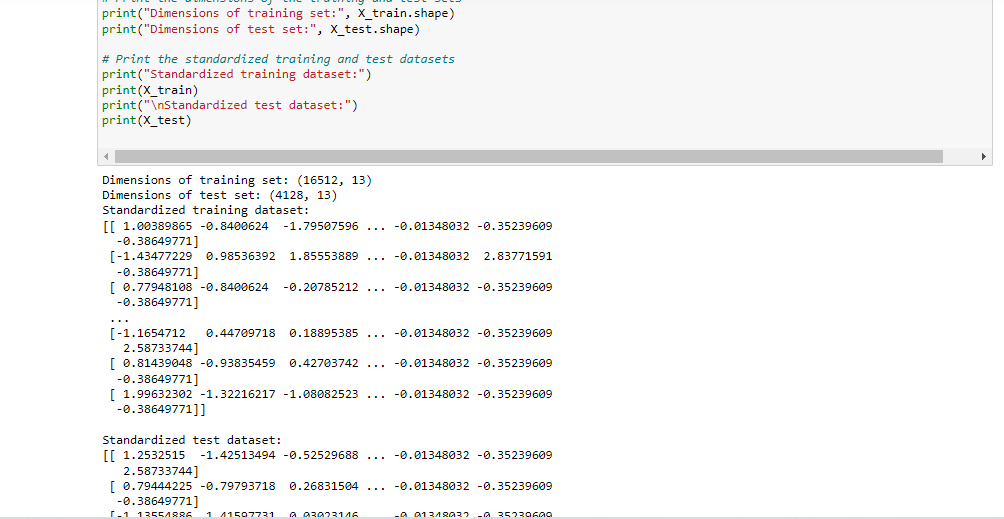


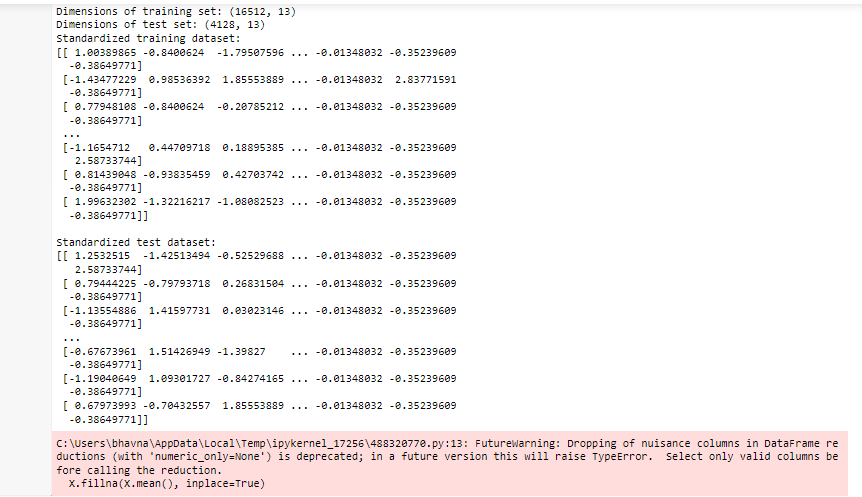


# **Split the dataset** & **Standardize data**

* Split the data into 80% training dataset and 20% test dataset.
* Standardize training and test datasets.

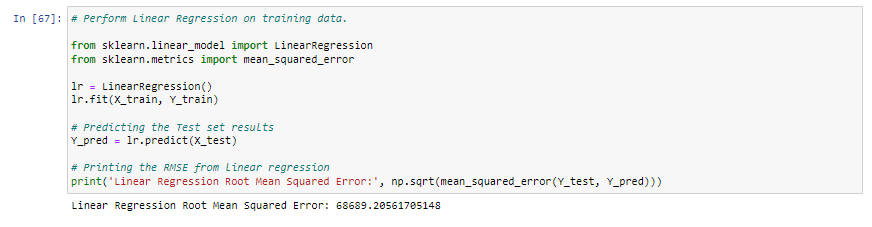






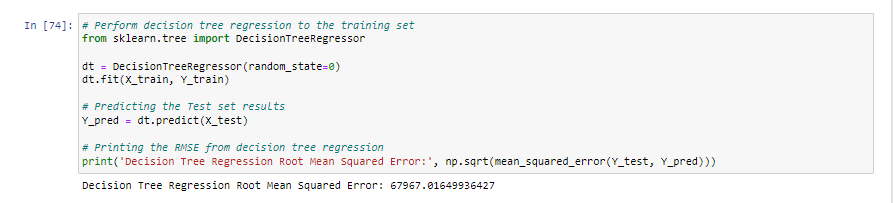
# **Perform Linear Regression**:

* Perform Linear Regression on training data.
* Predict output for test dataset using the fitted model.
* Print root mean squared error (RMSE) from Linear Regression.



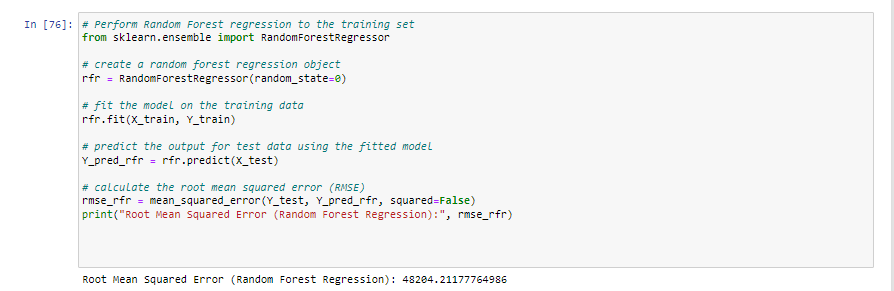
# **Perform Decision Tree Regression**:

* Perform Decision Tree Regression on training data.
* Predict output for test dataset using the fitted model.
* Print root mean squared error from Decision Tree Regression.



# **Perform Random Forest Regression**:

* Perform Random Forest Regression on training data.
* Predict output for test dataset using the fitted model.
* Print RMSE (root mean squared error) from Random Forest Regression.



# **Bonus exercise: Perform Linear Regression with one independent variable**:

* Extract just the median\_income column from the independent variables (from **X\_train** and **X\_test**).
* Perform Linear Regression to predict housing values based on **median\_income**.
* Predict output for test dataset using the fitted model.
* Plot the fitted model for training data as well as for test data to check if the fitted model satisfies the test data.



