***California Housing Price Prediction***

DESCRIPTION:

The project aims at building a model of housing prices to predict median house values in California using the provided dataset. This model should learn from the data and be able to predict the median housing price in any district, given all the other metrics.

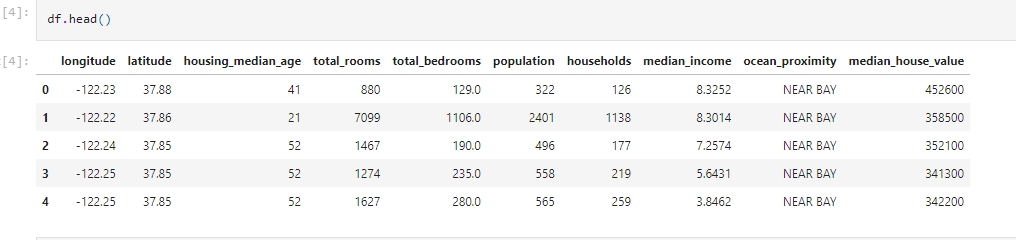
Districts or block groups are the smallest geographical units for which the US Census Bureau  
publishes sample data (a block group typically has a population of 600 to 3,000 people). There are 20,640 districts in the project dataset.

**Domain**: Finance and Housing

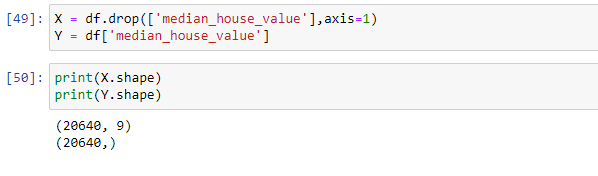
1. **Load the data**:



1. **Print first few rows of this data.**



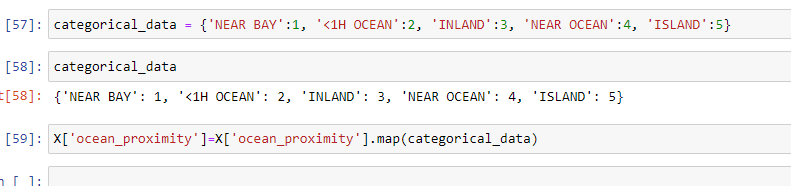
1. **Extract input (X) and output (Y) data from the dataset.**



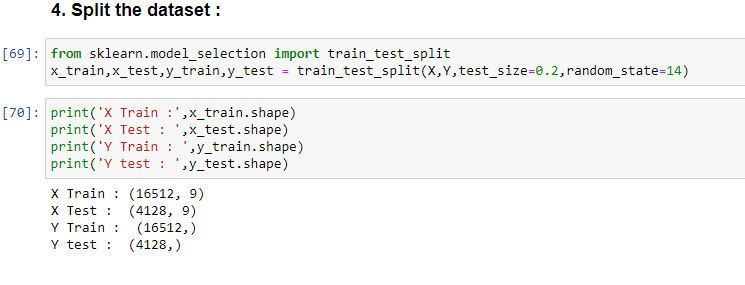
2. **Handle missing values** :



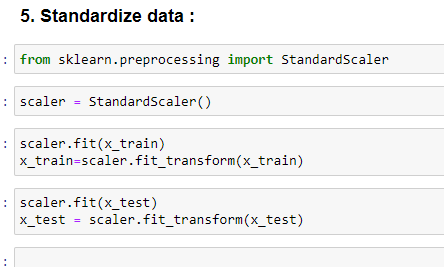
1. **Encode categorical data** :



1. **Split the dataset** :



1. **Standardize data** :

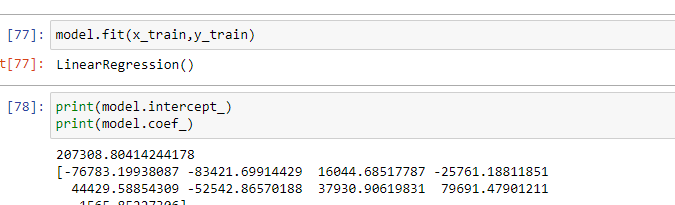


1. **Perform Linear Regression**:

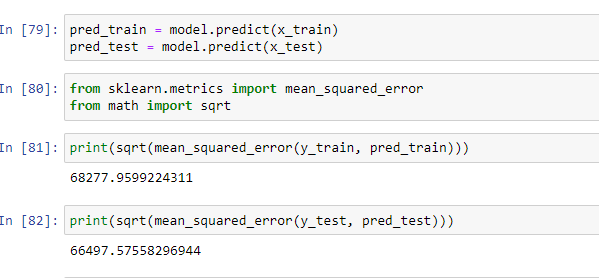
* Perform Linear Regression on training data.

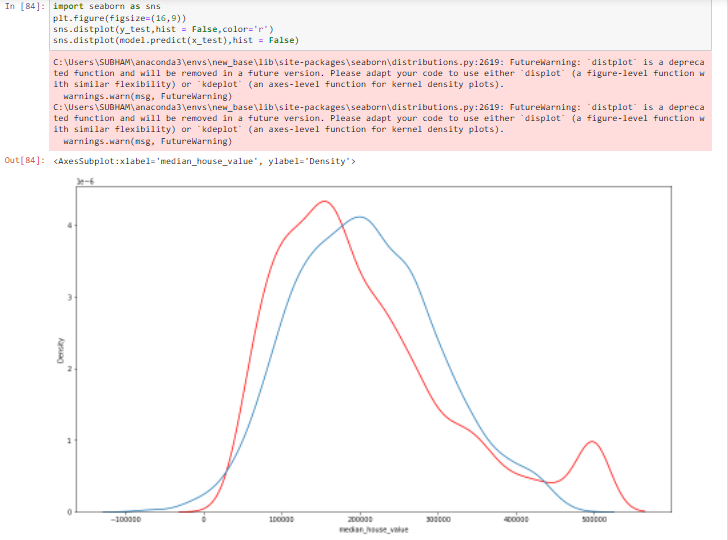


* Predict output for test dataset using the fitted model.



* Print root mean squared error (RMSE) from Linear Regression.





7. **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.

