## **Predictive Models – Using Linear and Logistic Regression**

## Course Number: 7745

## Name : PRANATHI SUNDARI

EMAIL : psndari1@memphis.edu

## **Introduction**

This assignment is for the purpose of implementing and exploring linear and logistic regression techniques for developing predictive models.

By fitting a line to observed data, a regression model describes the relationship between variables. Linear regression models use straight lines, while logistic and nonlinear regression models use curved lines. Regression estimates the change in the dependent variable when the independent variable changes.

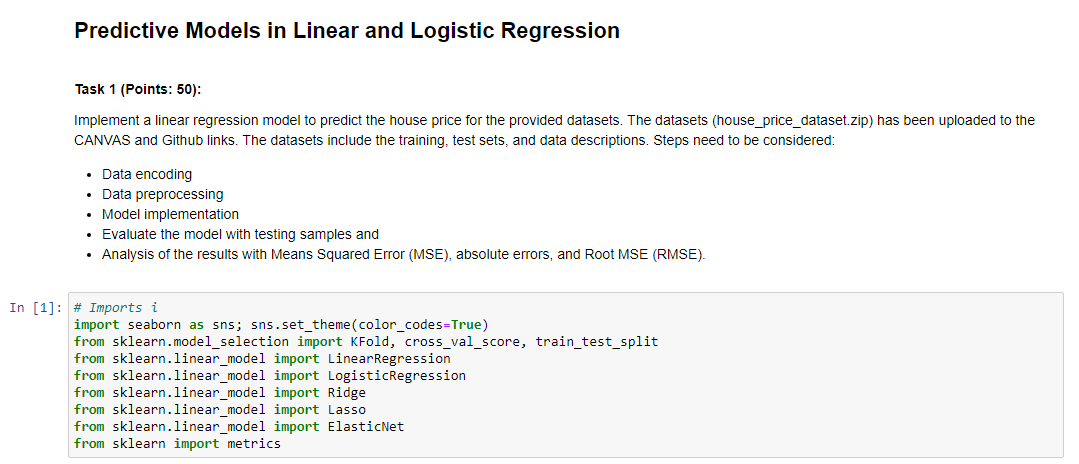
A linear regression technique defines a straight line between the independent and dependent variables. Use linear equations to determine the line of best fit (straight line) to your problem so that you can visualize and predict the output of your dependent variable. For example, linear regression can predict how an individual's obesity is related to work-life balance. Linear regression is further divided into simple and multiple linear regression, where one or more independent variables are used to predict the output.

Linear regression tries to version the connection among variables through fitting a equation to data. Only one variable is taken into account the explanatory variable and consequently the other variable is taken into account the variable quantity. for instance, a modeler can use a regression in the direction of the suggest version to suit somebody's weight to height. Before attempting to suit a linear version to the found data, the modeler ought to first determine whether there's an affiliation among the variables of interest. This doesn't usually imply that one variable drives the opposite (for example, better SAT scores don't always suggest higher university grades). Rather, it suggests that there's an intrinsic dating among the 2 variables. Scatterplots are beneficial for figuring out the power of affiliation among variables.  
Combining a linear regression version with the statistics will now no longer produce a beneficial version if there appears to be no courting among the proposed explanatory and established variables. The correlation coefficient (a range of from -1 to 1) suggests the diploma of courting among observations of variables. The system for linear regression is Y = a + bX where X is explanatory variable and Y is established variable. Slope of  road is b and y intersection (y with x = 0) is a.

Logistic regression is a data analysis approach that uses mathematics to determine the correlations between two data variables. The connection is then used to forecast the value of one of the parameters depending on the other. Predictions often have a limited number of outcomes, such as yes or no. Assume you want to predict whether or not a website visitor will hit the checkout button in their shopping basket. Logistic regression analysis examines previous visitor behavior, such as the amount of time spent on the website and the quantity of things in the shopping basket. It calculates that in the past if users stayed on the site for more than five minutes and added more than three things to the basket, they clicked the checkout button. The logistic regression function may then predict the behavior of a new website visitor based on this knowledge.

## **Methodology**

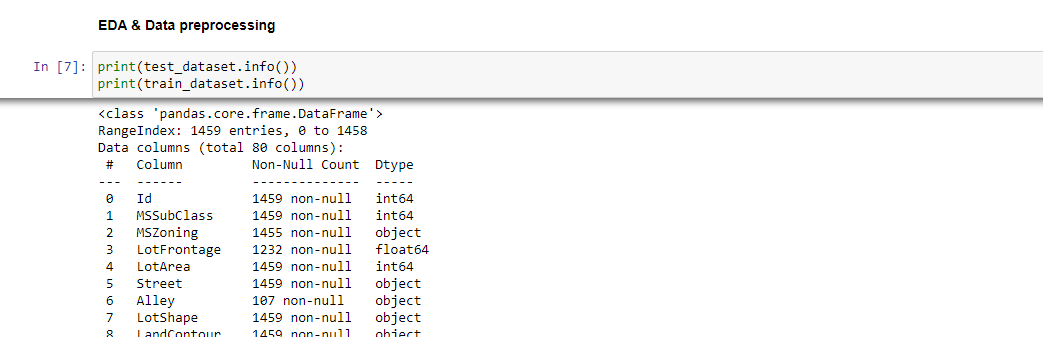
As demonstrated in the picture below, we began by importing all of the essential Python modules at the top of the notebook.



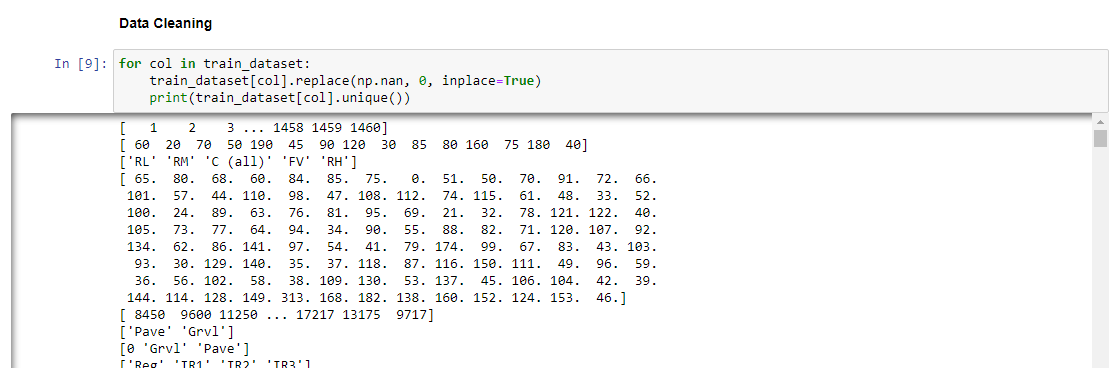
Importing the data, we imported our data using Pandas library in python



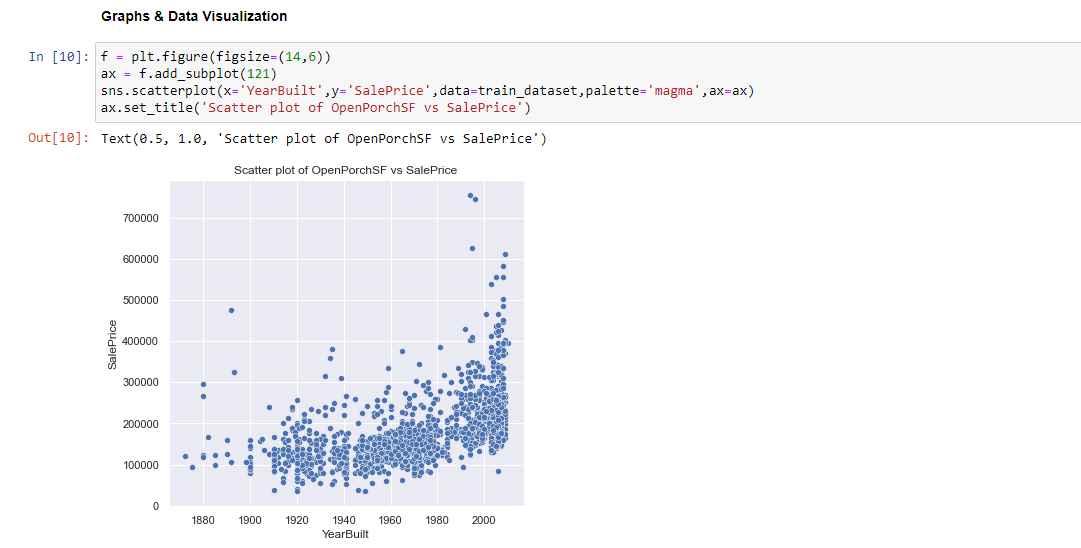
Then we cleaned and performed EDA and data visualization. Exploratory data analysis in statistics is a method of investigating data sets in order to summarize their primary aspects, which is typically done using statistical graphics and other data visualization tools. EDA is distinct from initial data analysis (IDA), which focuses on evaluating assumptions required for model fitting and hypothesis testing, as well as dealing with missing values and changing variables as needed. IDA is included in EDA. Data and information visualization is an interdisciplinary field that deals with data and information visualization. When the data or information is big, such as a time series, it is a highly efficient way of communication.



Cleaning data from a dataset, repairing or removing inaccurate, corrupted, malformed, duplicate, or incomplete data is known as data sanitization. When merging different data sources, there are several ways data can be duplicated or mislabeled.



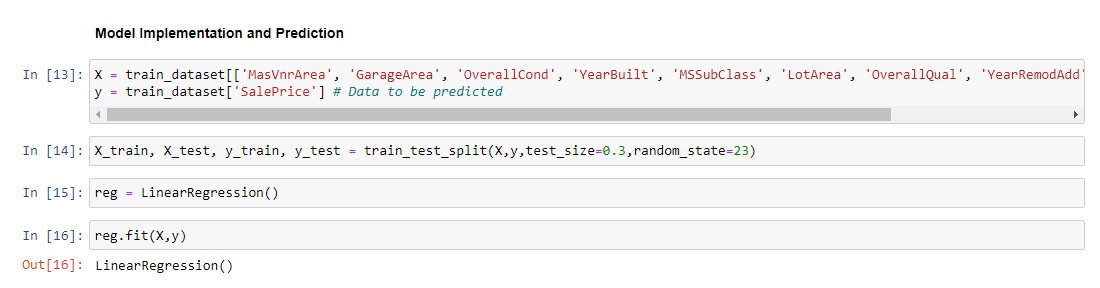
We moved to visualize the data in graphs and tables. These visualizations, which combine user-friendly and visually beautiful characteristics, speed up research and data processing while also serving as a potent communication tool.



## **Model Descriptions**

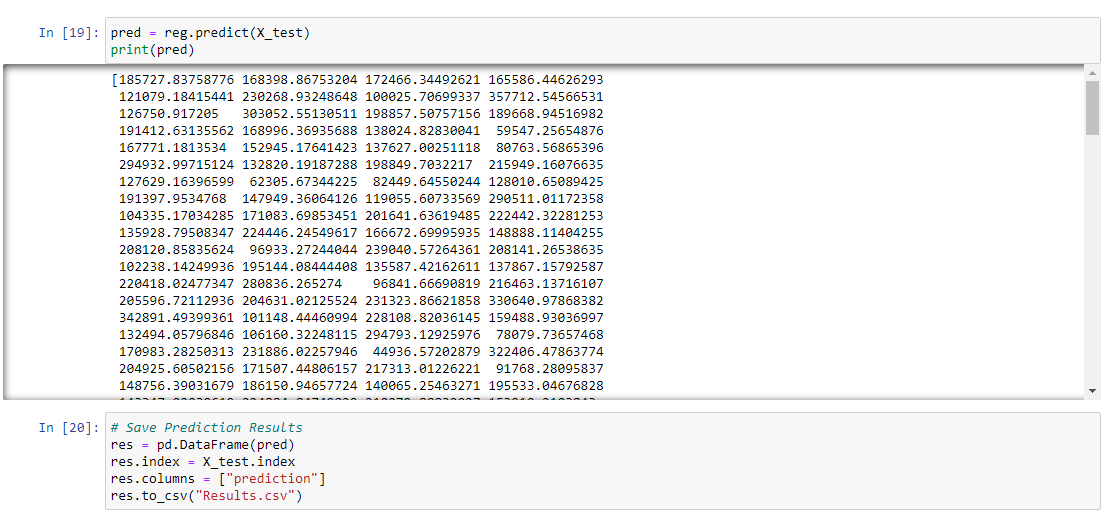
Linear models of effective connection presume that a brain region's many inputs are linearly separable. This assumption rules out activity-dependent connections that appear in one context but not another.

We made the train and split of our data for us to be able to make predictions



## Experiment and Results

The model is then implemented and predictions are made. As the name implies, linear regression essentially means fitting a line to the data that demonstrates a link between a target 'y' variable and the explanatory 'x' variables. We fitted our model using values and predicted the outcomes, which are given below. The results are saved in a CSV file, as specified in the instructions.



## **Conclusion**

Linear-regression fashions are quite trustworthy and provide a easy mathematical approach for generating predictions. Linear regression can be utilized in plenty of company and instructional settings. Linear regression is hired in plenty of fields, which includes biological, behavioral, environmental, and social sciences, as nicely as business. Linear regression fashions have proven to be a dependable and clinical approach of forecasting the future. Because linear regression is a nicely-installed statistical process, its traits are nicely understood and may be found out fast.

A positive coefficient indicates that when the independent variable's value increases, so does the mean of the dependent variable. As a result, we may deduce that the price of a property rises or falls dependent on factors such as the year the house was built, lot size, and other factors.

REFERENCES :

<https://machinelearningmastery.com/linear-regression-for-machine-learning/>

<https://towardsdatascience.com/linear-regression-model-machine-learning-9853450c8bce>

<https://medium.com/@minions.k/ridge-regression-l1-regularization-method-31b6bc03cbf>

<https://analyticsindiamag.com/hands-on-tutorial-on-elasticnet-regression/>

<https://www.jigsawacademy.com/blogs/ai-ml/lasso-regression>