**Experiment 02**

**Aim:** Simple Linear Regression in Python

**Tools:** Jupyter Notebook, Linear Regression Dataset (Kaggle)

**Theory:**

What is simple linear regression? Simple linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line. Both variables should be quantitative.

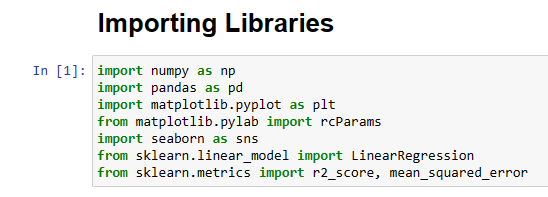
About Dataset used in this experiment:

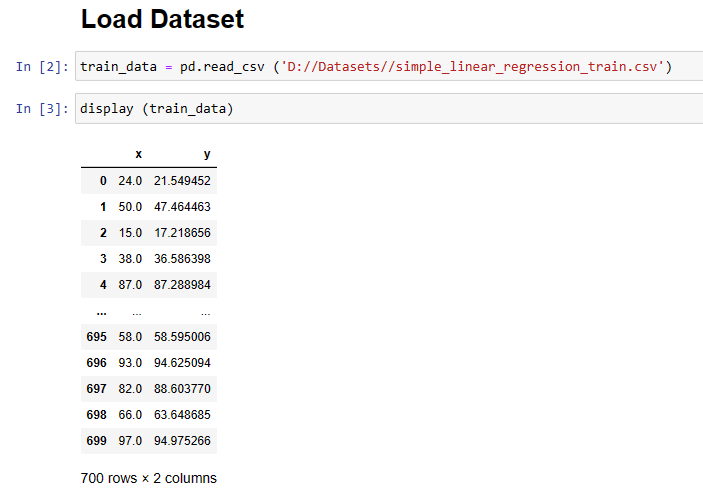
The training dataset is a CSV file with 700 data pairs (x,y). The x-values are numbers between 0 and 100. The corresponding y-values have been generated using the Excel function NORMINV(RAND(), x, 3). Consequently, the best estimate for y should be x.  
The test dataset is a CSV file with 300 data pairs.

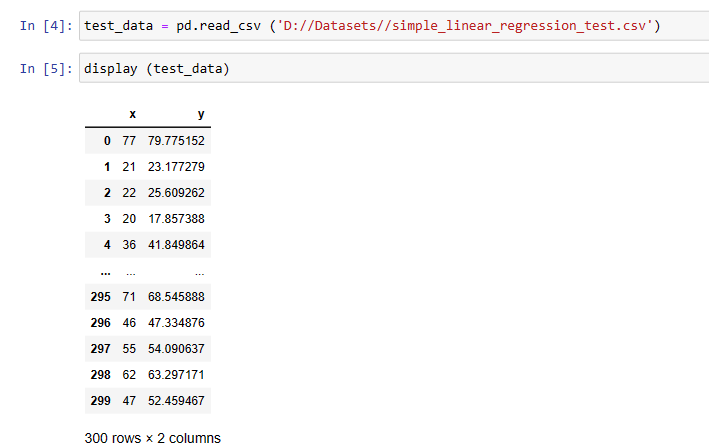
We will be using SciKit Learn for Linear Regression Algorithm.

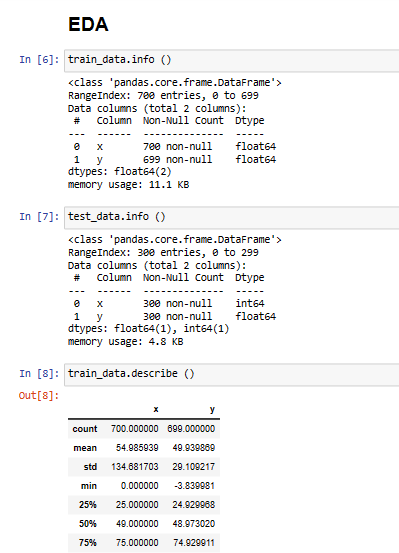
**scikit-learn** (formerly **scikits.learn** and also known as **sklearn**) is a [free software](https://en.wikipedia.org/wiki/Free_software) [machine learning](https://en.wikipedia.org/wiki/Machine_learning) [library](https://en.wikipedia.org/wiki/Library_(computing)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) [programming language](https://en.wikipedia.org/wiki/Programming_language). It features various [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support-vector machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [*k*-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN](https://en.wikipedia.org/wiki/DBSCAN), and is designed to interoperate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy" \o "NumPy) and [SciPy](https://en.wikipedia.org/wiki/SciPy" \o "SciPy). Scikit-learn is a [NumFOCUS](https://en.wikipedia.org/w/index.php?title=NumFOCUS&action=edit&redlink=1" \o "NumFOCUS (page does not exist)) fiscally sponsored project.

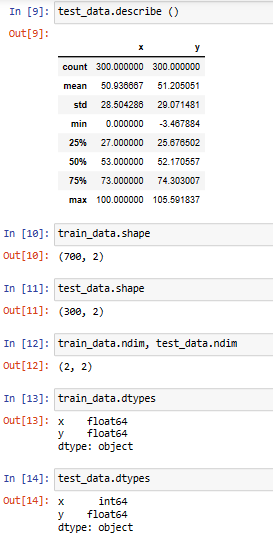
**Code & Output:**

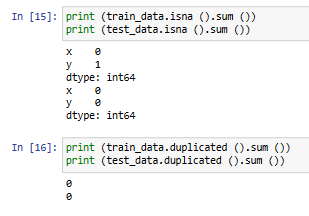
****

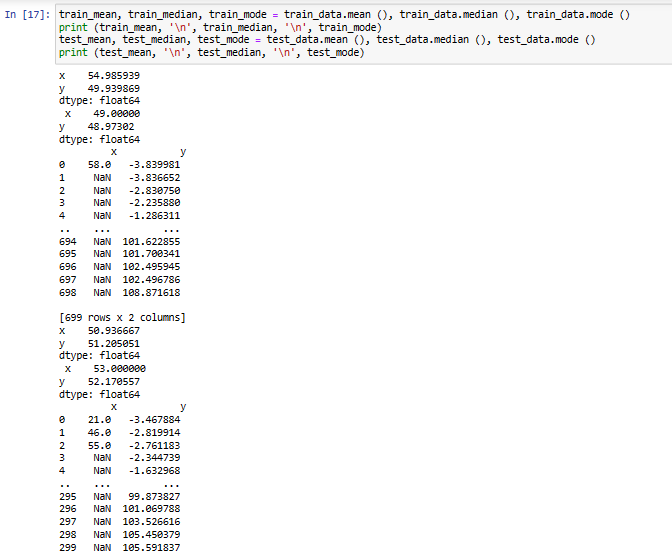
****

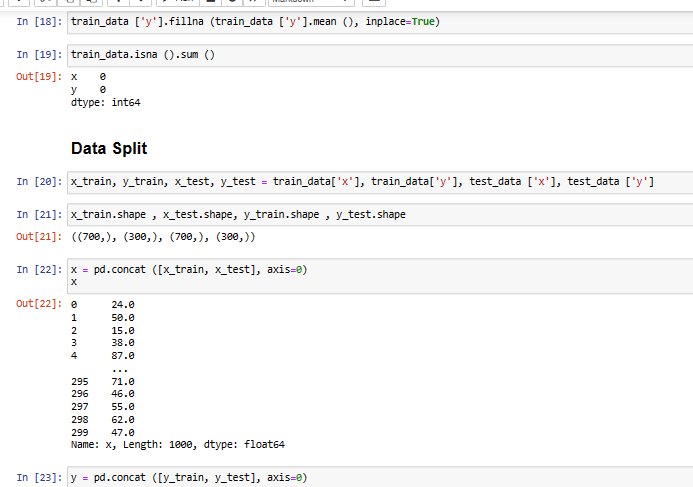
****

****

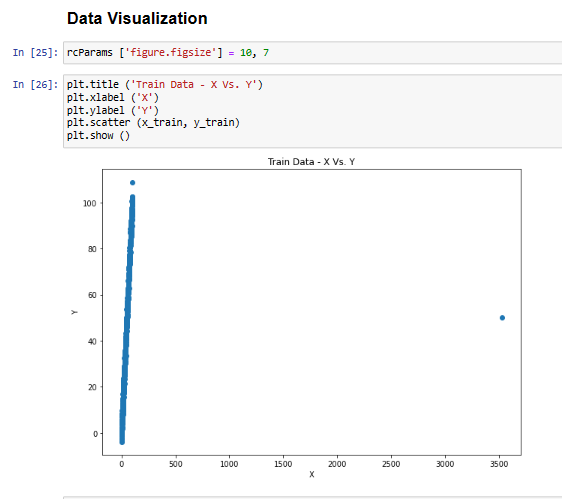
****

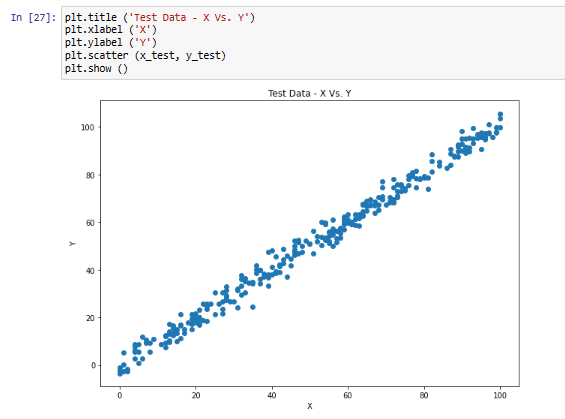
****

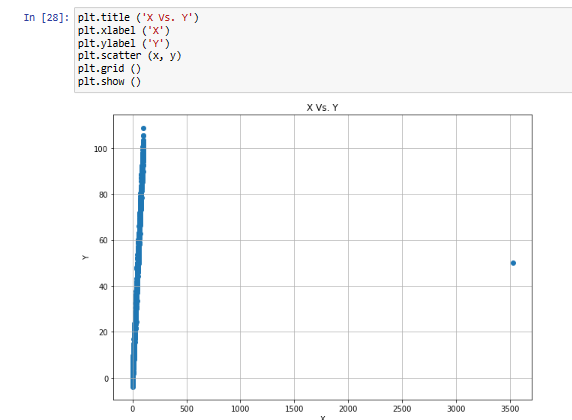
****

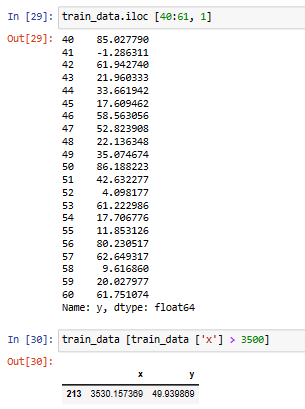
****

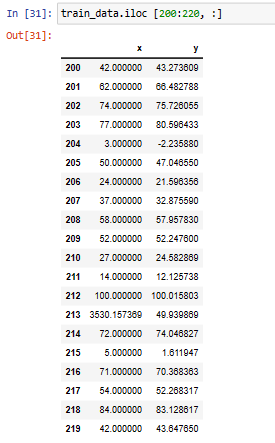
****

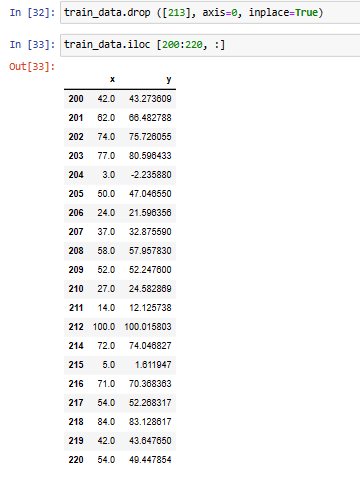
****

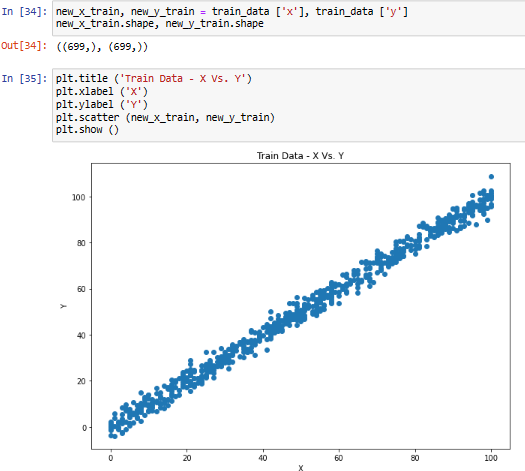
****

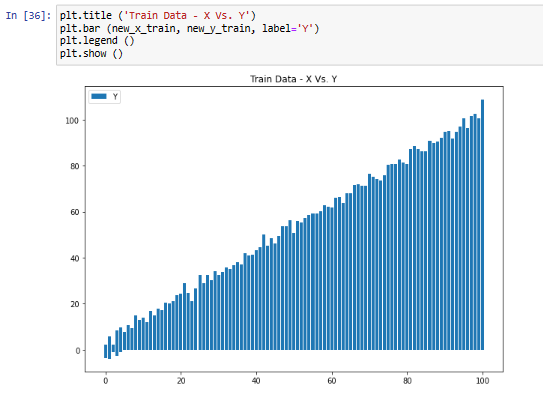
****

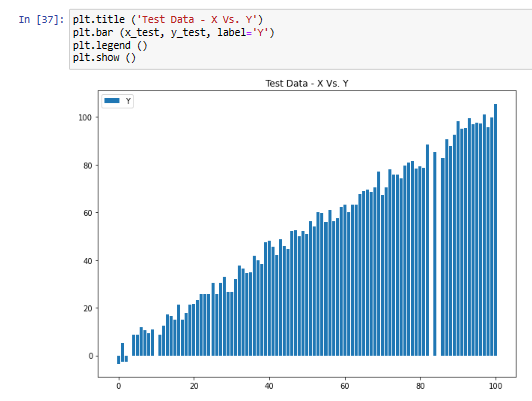
****

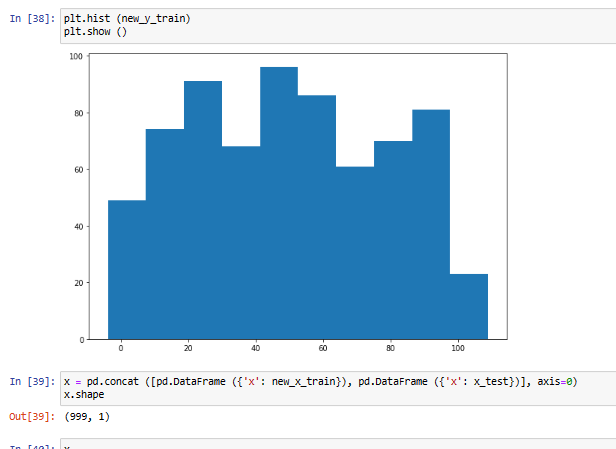
****

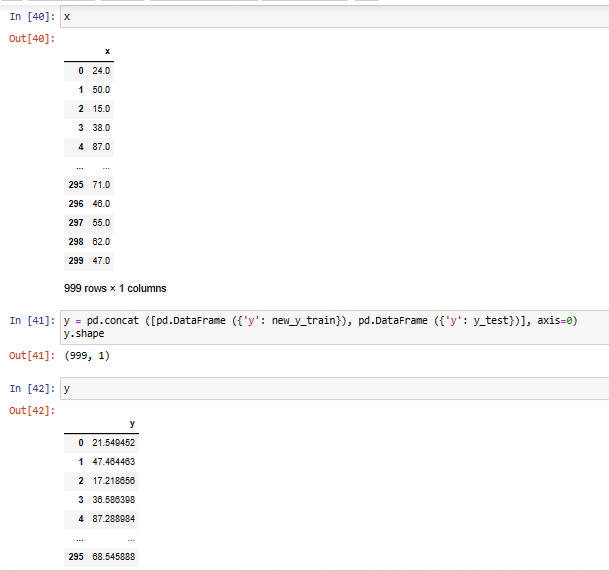
****

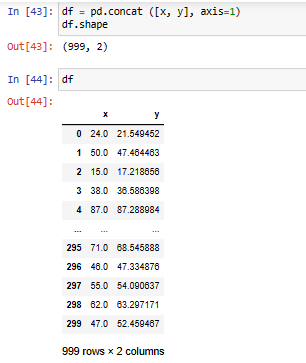
****

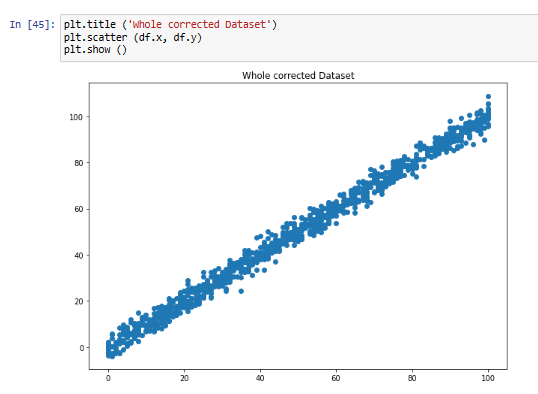
****

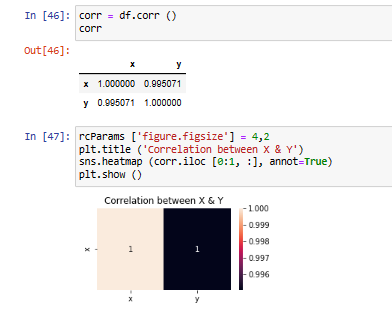
****

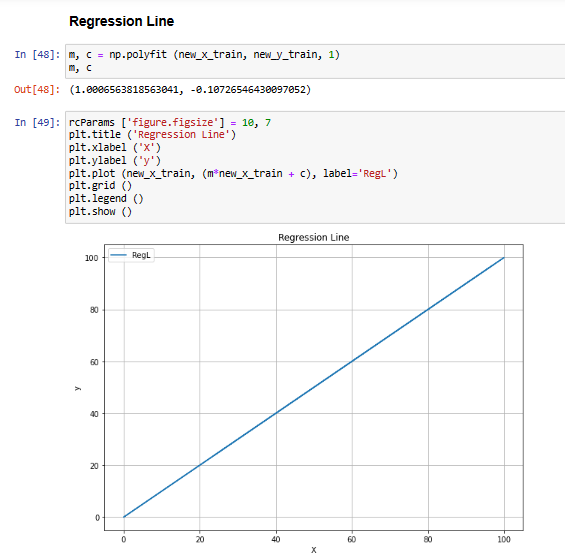
****

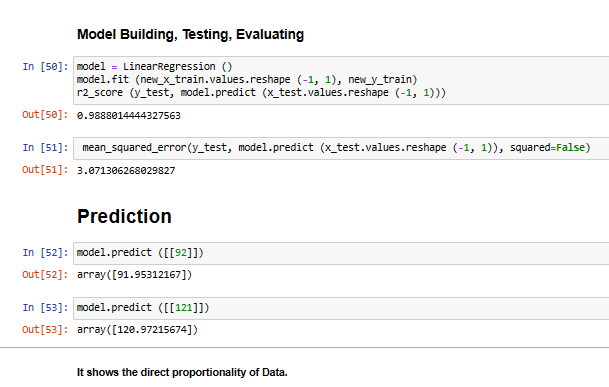
****

****

****

****

****

****

**Conclusion:** We have performed the given experiment successfully.