**Experiment 5**

**Shashwat Tripathi**

**D15A Batch C  
RollNo: 64**

**AIM : Regression Analysis**

**THEORY :**

**a) Perform Linear and Logistic regression to find out relation between variables**

**b) Apply regression model technique to predict the data on above dataset.**

Understanding the Relationship Between Variables with Regression

Regression analysis is a statistical technique used to model the relationship between a dependent variable (what you're trying to predict) and one or more independent variables (what you believe influences the dependent variable). There are two main types of regression commonly used:

Linear Regression: This technique assumes a linear relationship between the independent and dependent variables. It estimates a straight line that best fits the data points, allowing you to predict the dependent variable based on the values of the independent variables.

Logistic Regression: Unlike linear regression, logistic regression is used for classification problems where the dependent variable is categorical (e.g., pass/fail, spam/not spam). It estimates the probability of an observation belonging to a specific category based on the independent variables.

Here's a breakdown of how these techniques can be applied to find the relationship between variables and predict data:

1. Data Preparation:

Assemble your dataset: This includes data points with values for both the dependent and independent variables.

Data Cleaning: Ensure your data is free from errors and missing values. Handle missing values through imputation or removal.

2. Model Selection:

Linear Regression: If your dependent variable is continuous, choose linear regression.

Logistic Regression: If your dependent variable is categorical, choose logistic regression.

3. Model Training:

Linear Regression: The model estimates a linear equation (y = mx + b) where 'y' is the dependent variable, 'x' is the independent variable, 'm' is the slope, and 'b' is the y-intercept.

Logistic Regression: The model estimates a sigmoid function that maps the input values (combination of independent variables) to a probability between 0 and 1 for each category of the dependent variable.

4. Model Evaluation:

Metrics: Evaluate the model's performance using metrics like Mean Squared Error (MSE) for linear regression or accuracy and precision/recall for logistic regression.

Visualization: Plot the predicted values against the actual values to visually assess the model's fit.

5. Prediction:

Once you have a trained and evaluated model, you can use it to predict the dependent variable for new data points based on the values of the independent variables.

Key Points:

Choosing the right regression technique depends on the nature of your dependent variable (continuous or categorical).

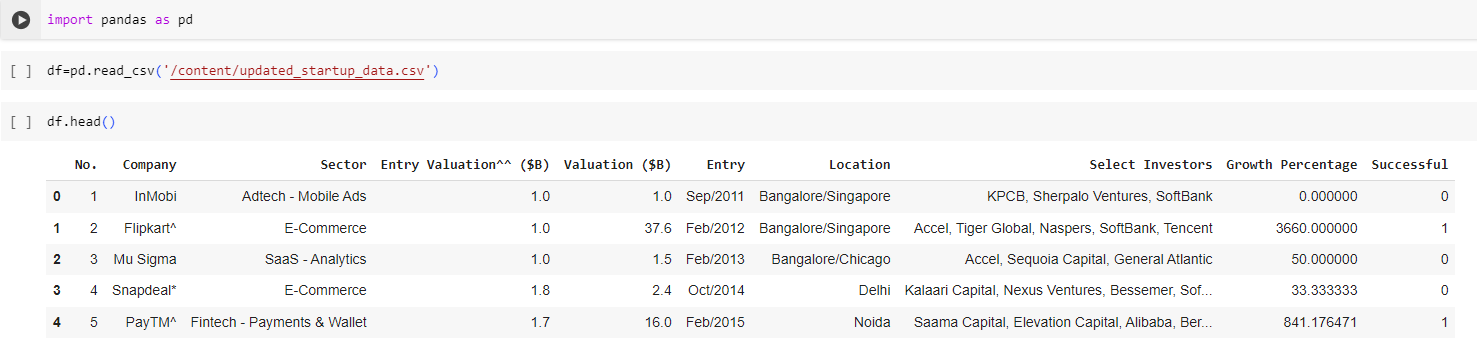
Both models involve training, evaluation, and then using the model for prediction.

The interpretation of the results differs between linear and logistic regression. In linear regression, the slope indicates the direction and strength of the relationship between the variables. In logistic regression, the coefficients help determine the impact of each independent variable on the probability of belonging to a particular category.

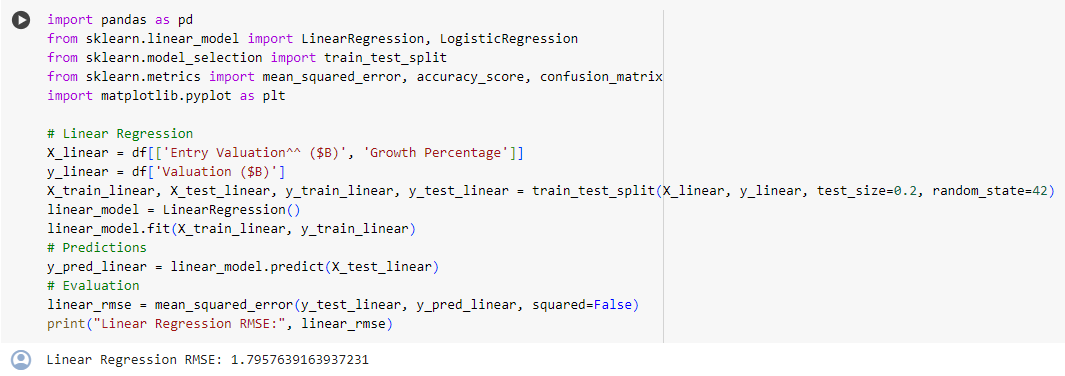
By applying these techniques, you can uncover the relationships between variables within your data and use those insights to make predictions on unseen data points

**OUTPUT :**

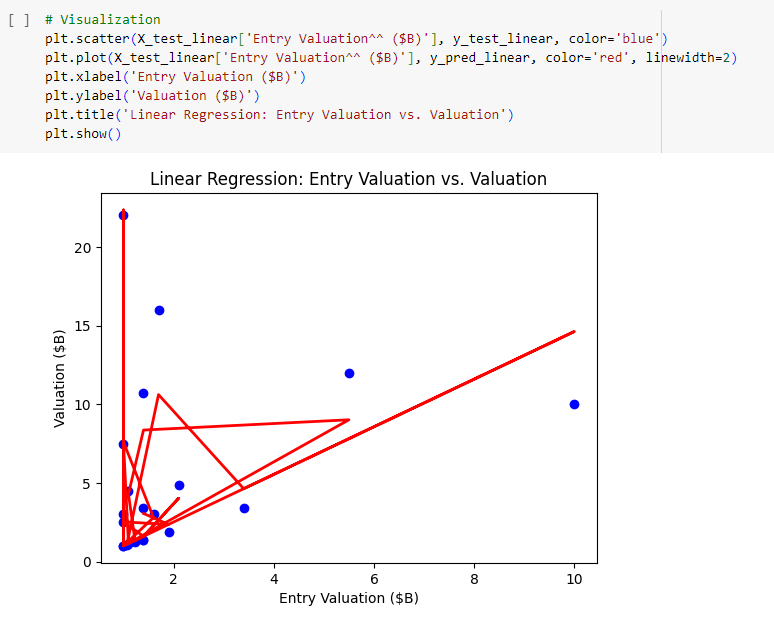
**Importing the Dataset** [https://www.kaggle.com/datasets/niekvanderzwaag/unicorn-startups](https://www.kaggle.com/datasets/niekvanderzwaag/unicorn-startups-cleaned)

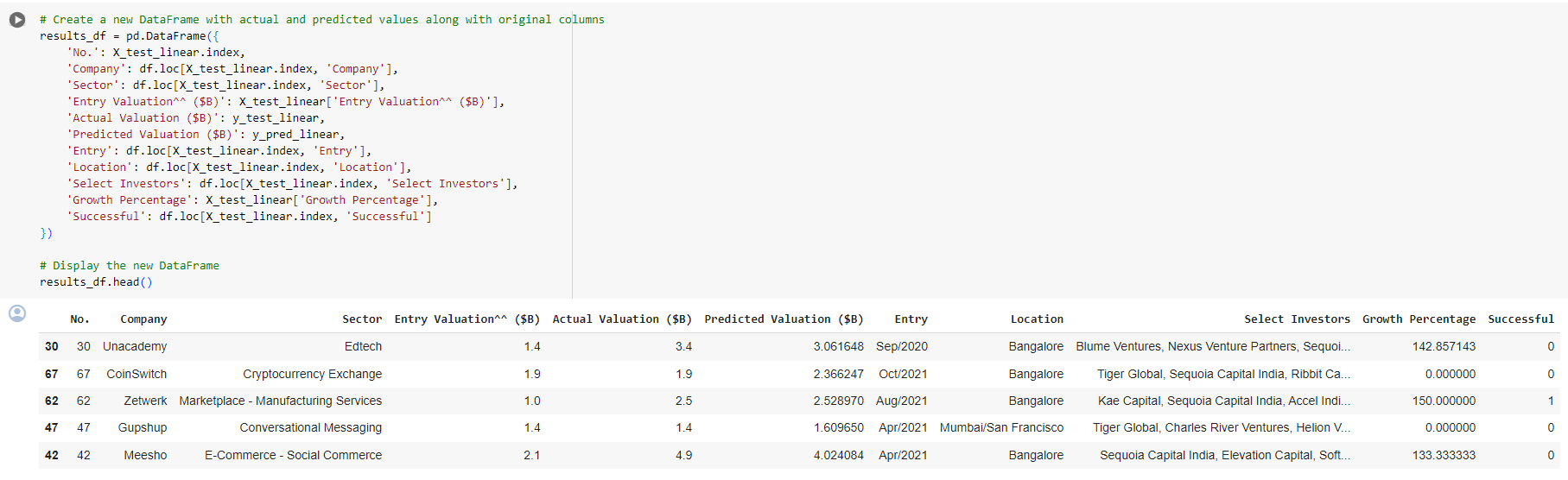


**Linear Regression**



**Visualization**

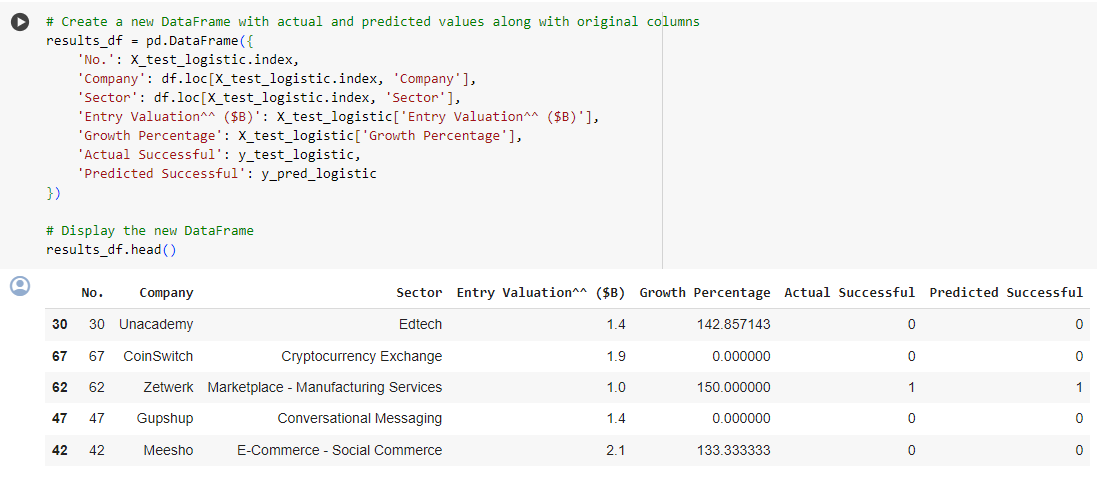




**Logistic Regression**



**Difference between prediction and actual value**



**CONCLUSION : Thus, we have performed various regression techniques and found useful insights.**