**Prediction of Survivals on Titanic Using Machine Learning Model**

Problem Statement - The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg during its maiden voyage from Southampton to New York City. There were an estimated 2,224 passengers and crew aboard the ship, and more than 1,500 died, making it one of the deadliest commercial peacetime maritime disasters in modern history. The RMS Titanic was the largest ship afloat at the time it entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. The Titanic was built by the Harland and Wolff shipyard in Belfast. Thomas Andrews, her architect, died in the disaster.

In this Project, we will analyze the Titanic data set and make two predictions. One prediction to see which passengers on board the ship would survive and then another prediction to see if we wouldn’t survive.

Steps to be taken in the Project is sub-divided into the following sections.



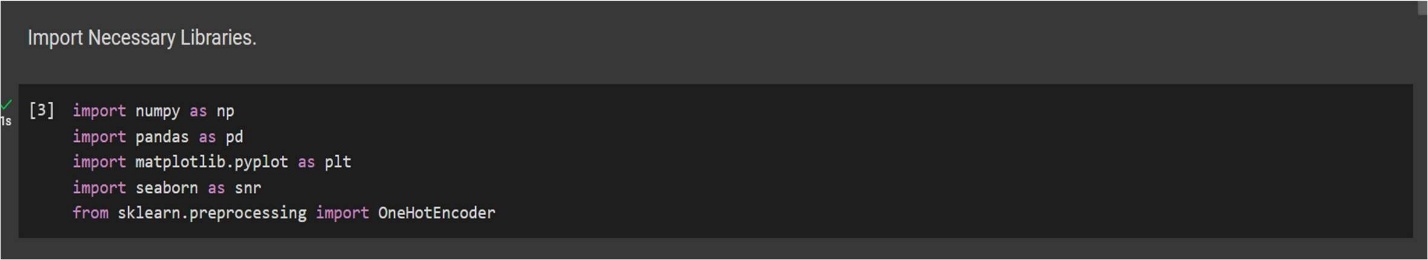
These are:

* Load the necessary libraries such as Numpy , Pandas , sklearn.model etc.
* Loading the dataset as csv file and showing first ten rows.
* Drop the unnecessary columns from the data.
* Calculate statistical values and round them up to 3 decimal places.
* Checking for null values and return their sum of numbers of true values in each column.
* Handle the null by mean of all values fill into them.
* Visualization of Passenger Survival data using Data Visualization with Python.
* Data preprocessing or (Data cleaning) performed by the one hot encoding in this process we change categorical data into numerical data and the technique is called feature Engineering.
* Splitting the cleaned data into dependent and independent variables.
* Splitting the data into train and test sets with train\_test\_split using sklearn library.
* Import different kind of Classification Models and Train that model with the help of .fit().
* Predicting the trained models and then checking their accuracy score and

confusion metrics of the model using confusion metrics & accuracy score.

* Then recall the train\_test\_split and split the data into training and testing set with different models.
* Then predicting the trained models and checking the accuracy of model and print the accuracy difference.
* And finally predict whether the Titanic Survivals classification generated or not.

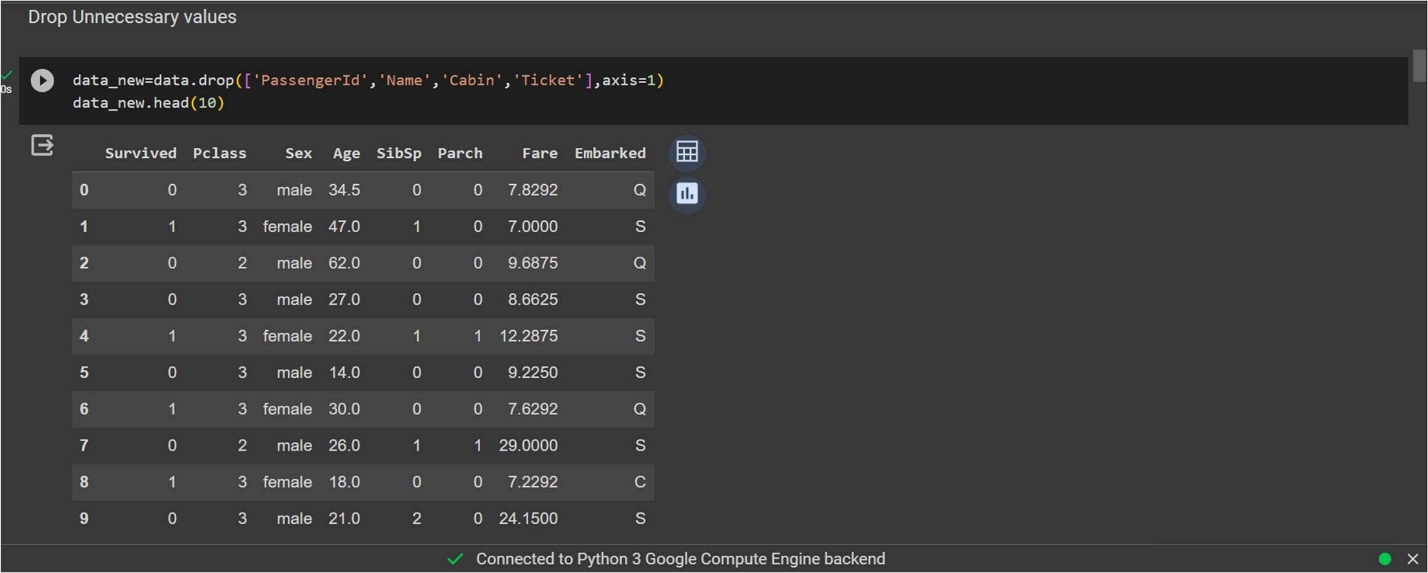
* Step-1 – Loading Necessary Libraries used in machine learning.



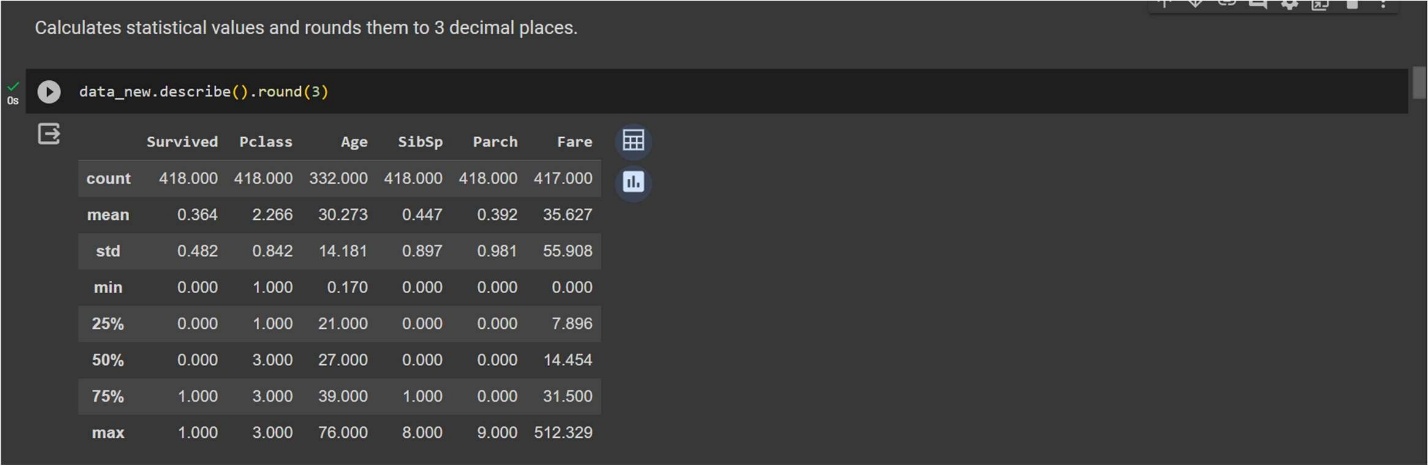
* Step-2 - Loading the dataset as csv file and showing first ten rows.



* Step-3 - Drop the unnecessary columns from the data.



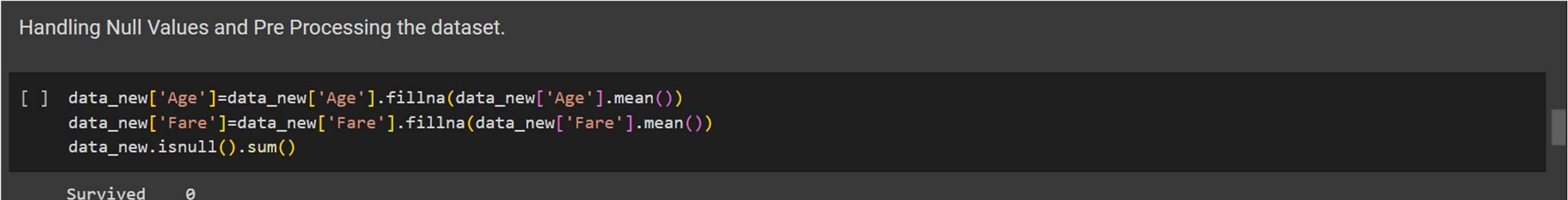
* Step-4 - Calculate statistical values and round them up to 3 decimal places.



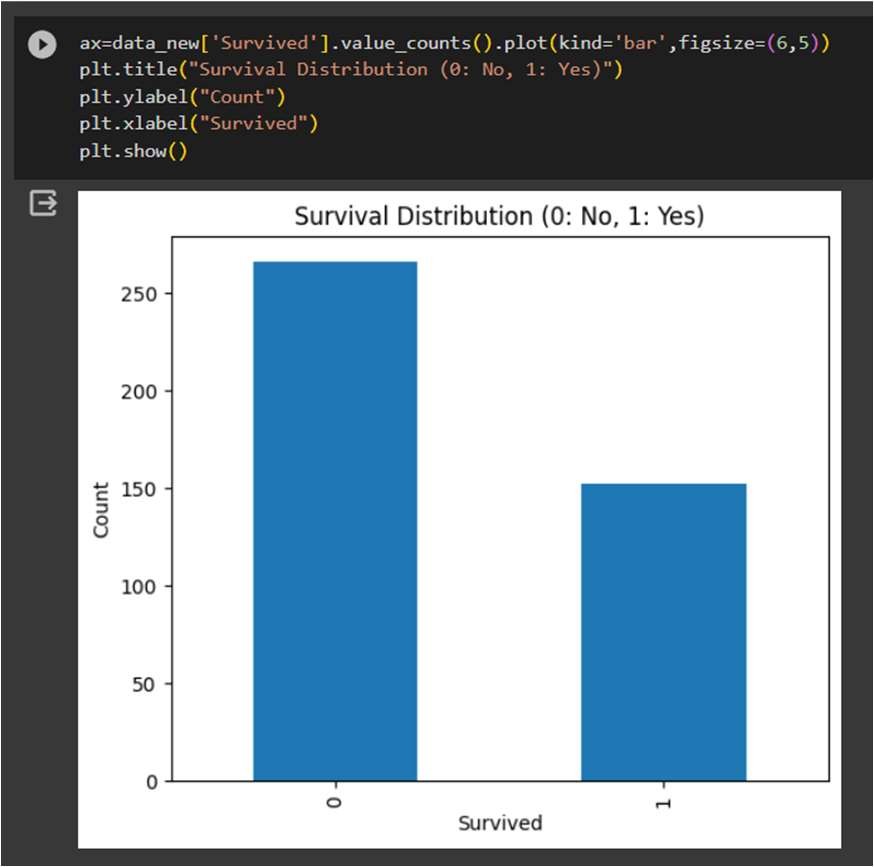
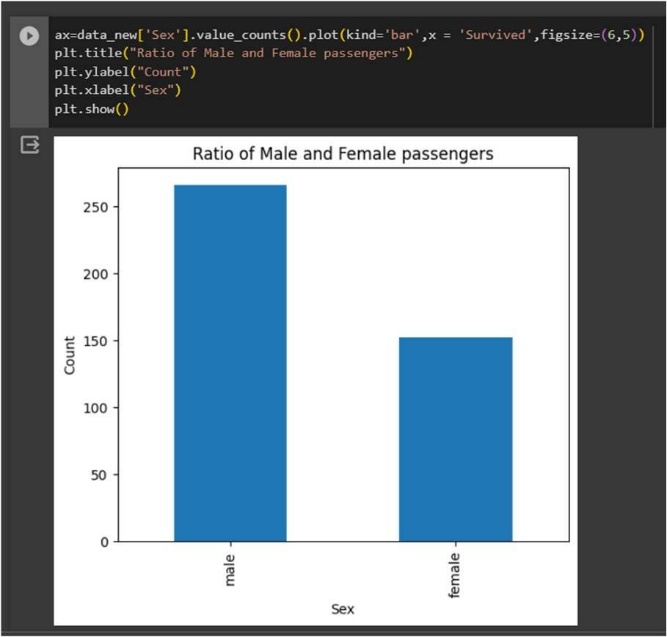
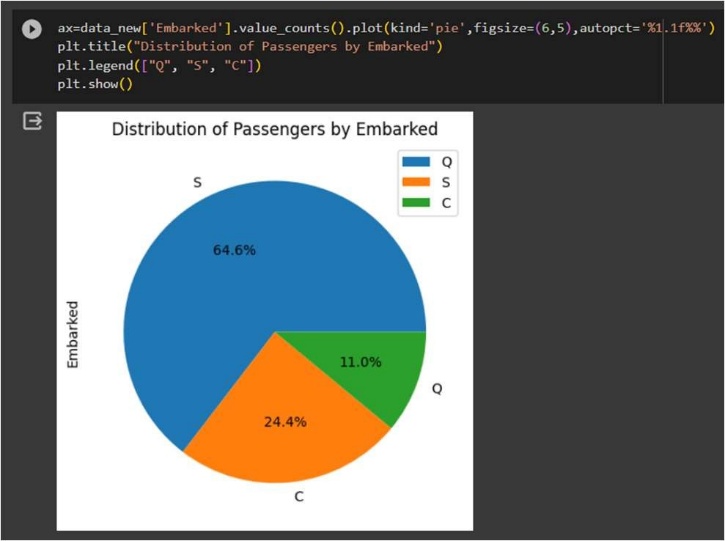
Step-5 - Checking for null values and return their sum of numbers of true values in each column.



* Step-6 - Handle the null by mean of all values fill into them.



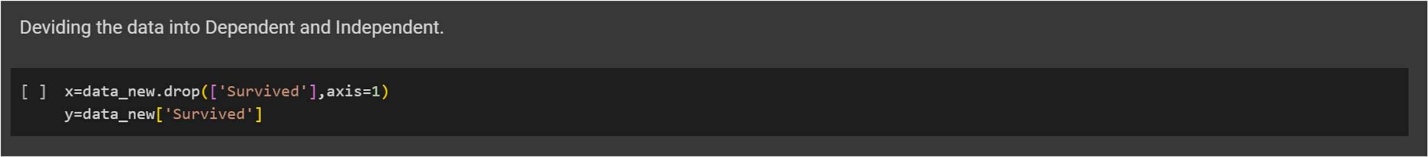
* Step-7 - Visualization of Passenger Survival using Data Visualization with Python.



* Step-8 - Data preprocessing or (Data cleaning) performed by the one hot encoding in this process we change categorical data into numerical data and the technique is called feature Engineering.



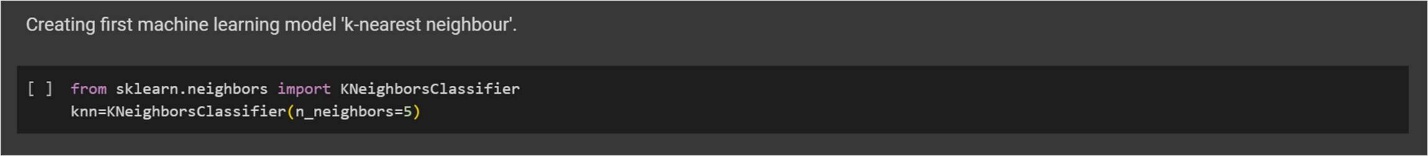
Step-9 - Splitting the cleaned data into dependent and independent variables.



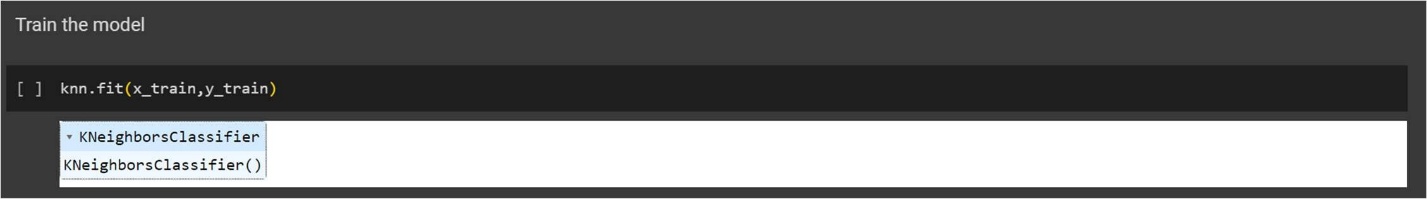
* Step-10 - Splitting the data into train and test sets with train\_test\_split using sklearn library.



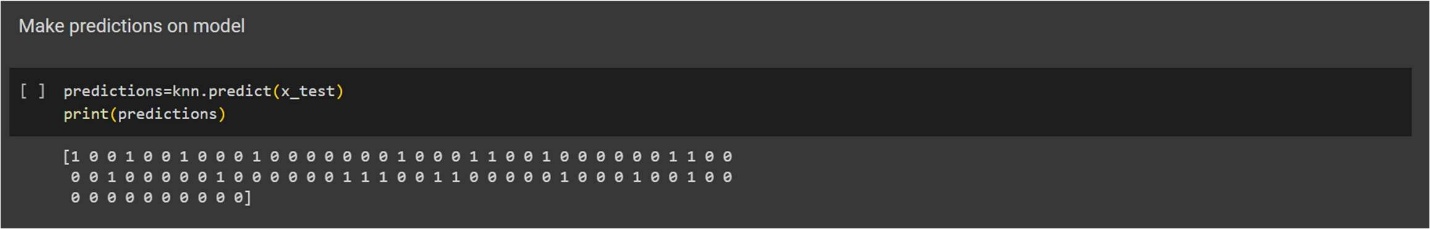
* Step-11 - Import first machine learning model K-Nearest neighbor taking n\_neighbor=5.



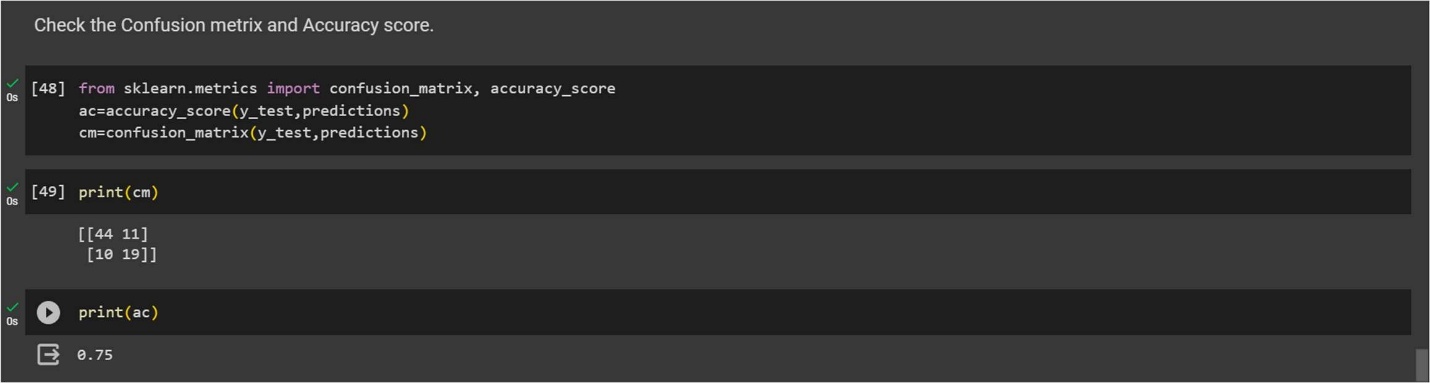
* Step-12 - Train the model using .fit() function.



* Step-13 - Predict the trained model using .predict() function.



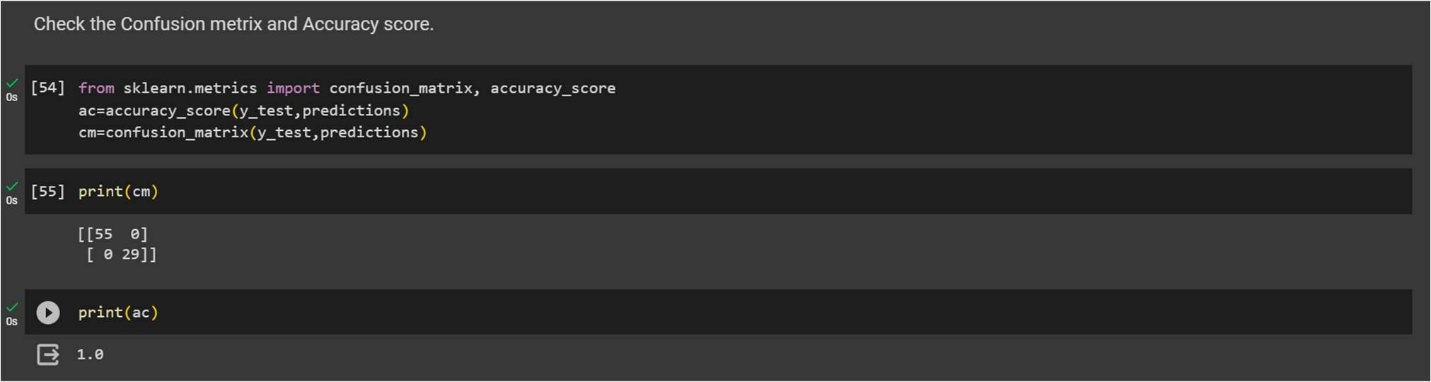
* Step-14 - Check the accuracy score and print a confusion metrics with confusion metrics & accuracy score.



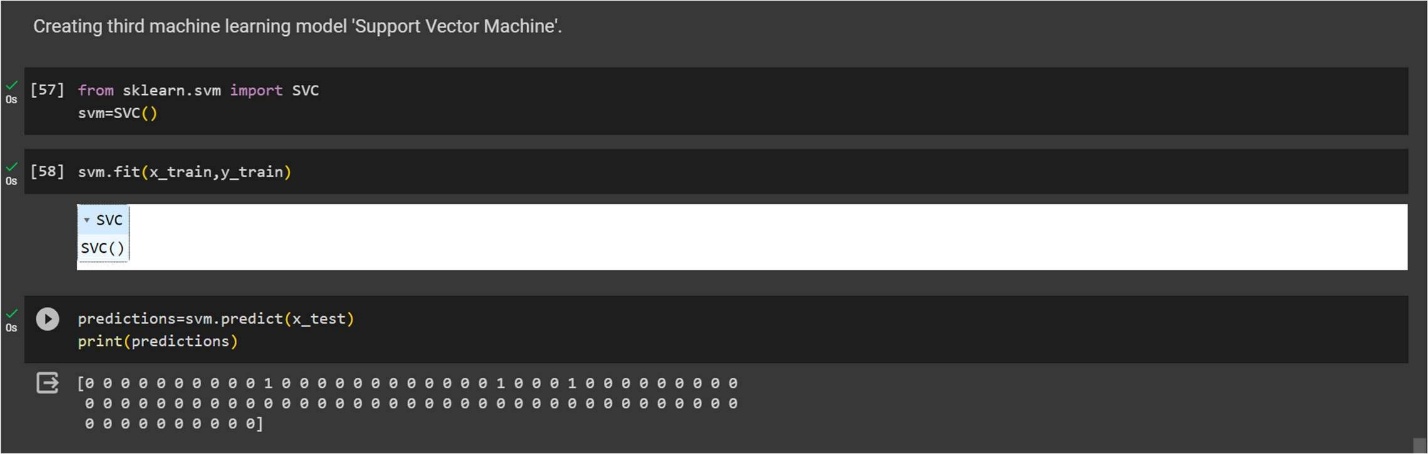
* Step-15 – Import the Second Machine Learning Model Decision Tree and train model and then make prediction.



* Step-16 - Print a confusion metrics and check accuracy score for Decision Tree Model.



Step-17 - Import the Third Machine Learning Model Support Vector Machine and train model and then make prediction.



* Step-18 - Print a confusion metrics and check accuracy score for Support Vector Machine Model.



Conclusion - The purpose of Project is to use the existing features of passengers onboard Titanic as predictors to predict their survival outcome, for 0 being dead and 1 being survived from the tragic ship crash. The K-Nearest neighbor is the is first classification model performed with k=5, and the I use Random Forest Classifier, and then I use the Support Vector Machine (SVM) analysis to improved performance. It is certain through the practice of model improvement, the SVM analysis is better performed than the KNearest Neighbor classification analysis and Random Forest analysis is also performed better than K-Nearest Neighbor for prediction accuracy.

However, even from the all three classification model, we can easily see that the Titanic survival outcome is highly depended on several predictors, such as sex, age and passenger class. In particular, Ratio of survived people are more while keeping other predictors conditions constant and lastly, people from a lower class are less likely to survived keeping other predictors conditions constant.

Prediction of IRIS flower using Machine Learning Model.

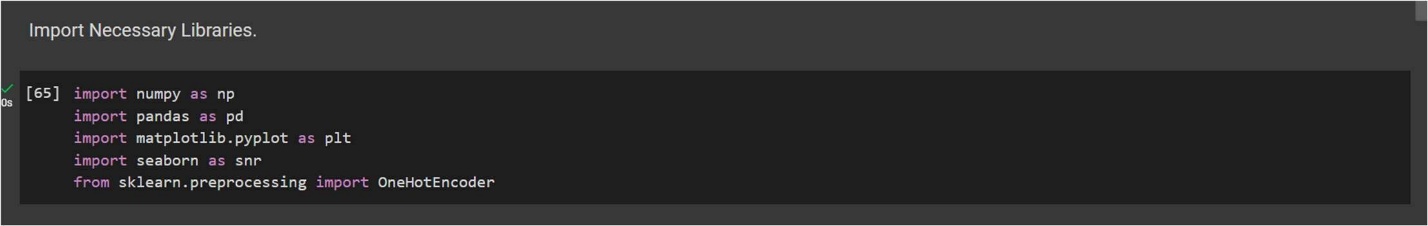
Problem Statement – In this project we make prediction of the physical parameters of three species of Iris flower — Iris-Versicolor, Iris-Setosa and Iris-Virginica. The numeric parameters which the dataset use contains are Sepal width, Sepal length, Petal width and Petal length. In this Project we will be predicting the classes of the flowers based on these parameters. The data consists of continuous numeric values which describe the dimensions of the respective features. We will be training the model based on these features.

Steps to be taken in the Project is sub-divided into the following sections. These are:



* Load the necessary libraries such as Numpy , Pandas , sklearn etc.
* Loading the dataset as csv file and showing first ten rows.
* Drop the unnecessary columns from the data.
* Calculate statistical values and round them up to 3 decimal places.
* Checking for null values and return their sum of numbers of true values in each column.
* Handle the null by mean of all values fill into them.
* Extracting all information about data.
* Checking shape of data and checking unique values in dependent variable.
* Visualization on different species of Iris flower using Python data visualization.
* Data preprocessing or (Data cleaning) performed by the one hot encoding in this process we change categorical data into numerical data and the technique is called feature Engineering.
* Splitting the cleaned data into dependent and independent variables.
* Splitting the data into train and test sets with train\_test\_split using sklearn library.
* Import different kind of Classification Models and Train that model with the help of .fit().
* Predicting the trained models and then checking their accuracy score and confusion metrics of the model using confusion metrics & accuracy score.
* Then recall the train\_test\_split and split the data into training and testing set with different models.
* Then predicting the trained models and checking the accuracy of model and check the accuracy difference.
* And finally predict whether the classification of different species of Iris is generated or not.

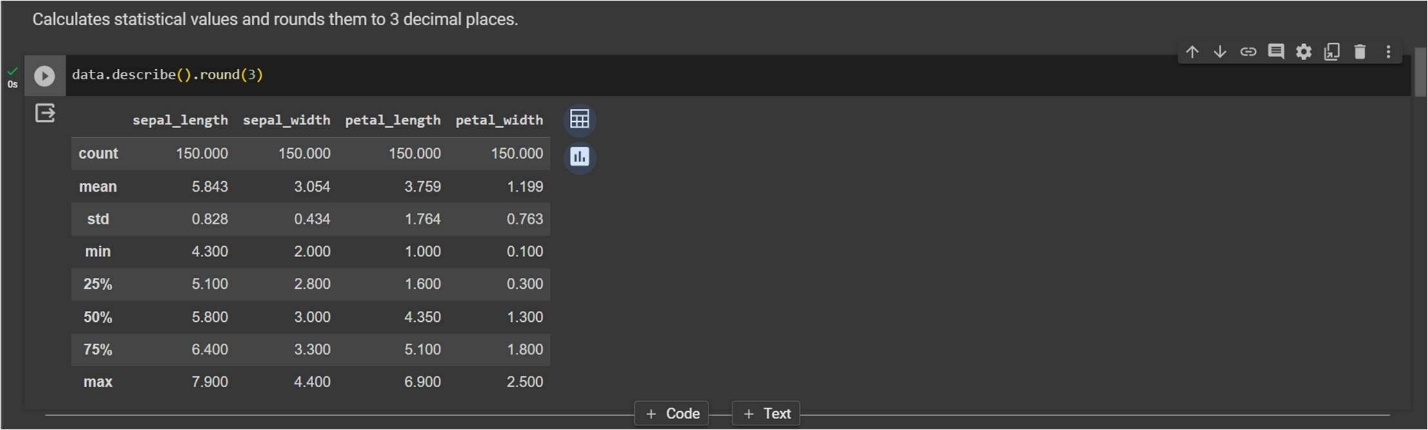
* Step-1 - Loading Necessary Libraries used in machine learning.



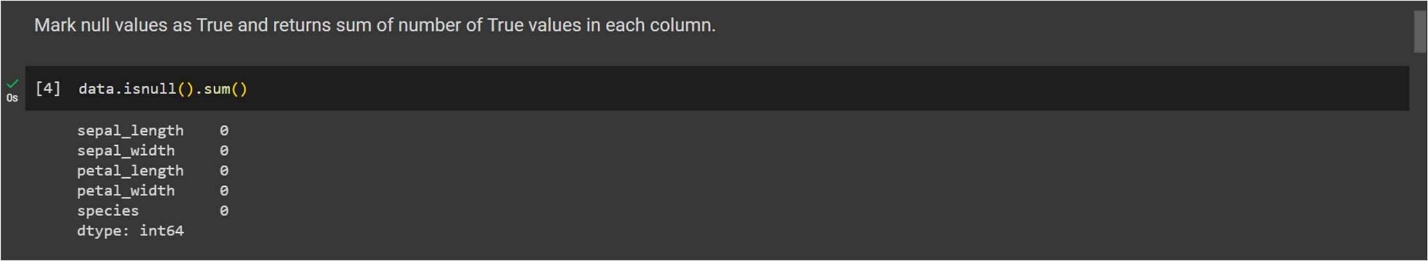
* Step-2 - Loading the dataset as csv file and showing first ten rows.



* Step-3 - Calculate statistical values and round them up to 3 decimal places.



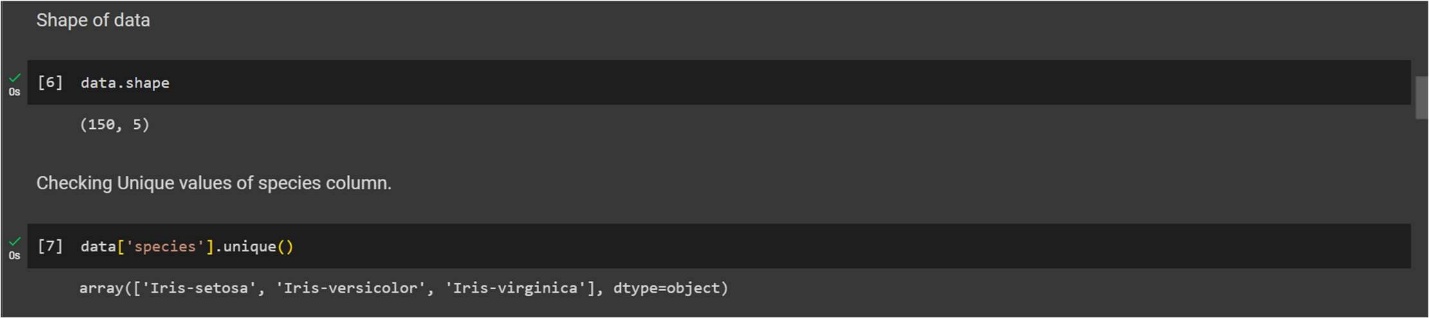
* Step-4 – Checking for null values and return their sum of numbers of true values in each column.



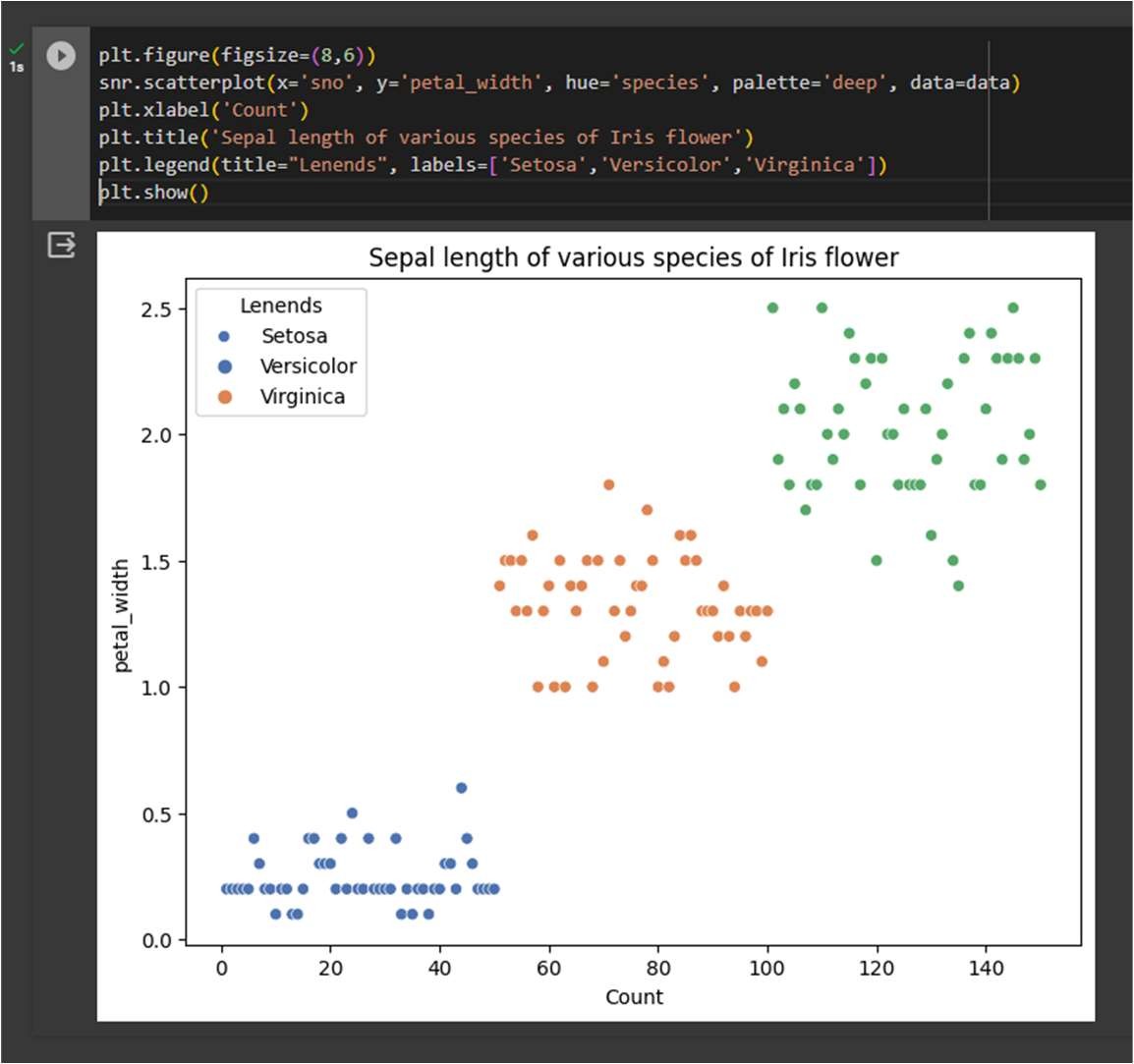
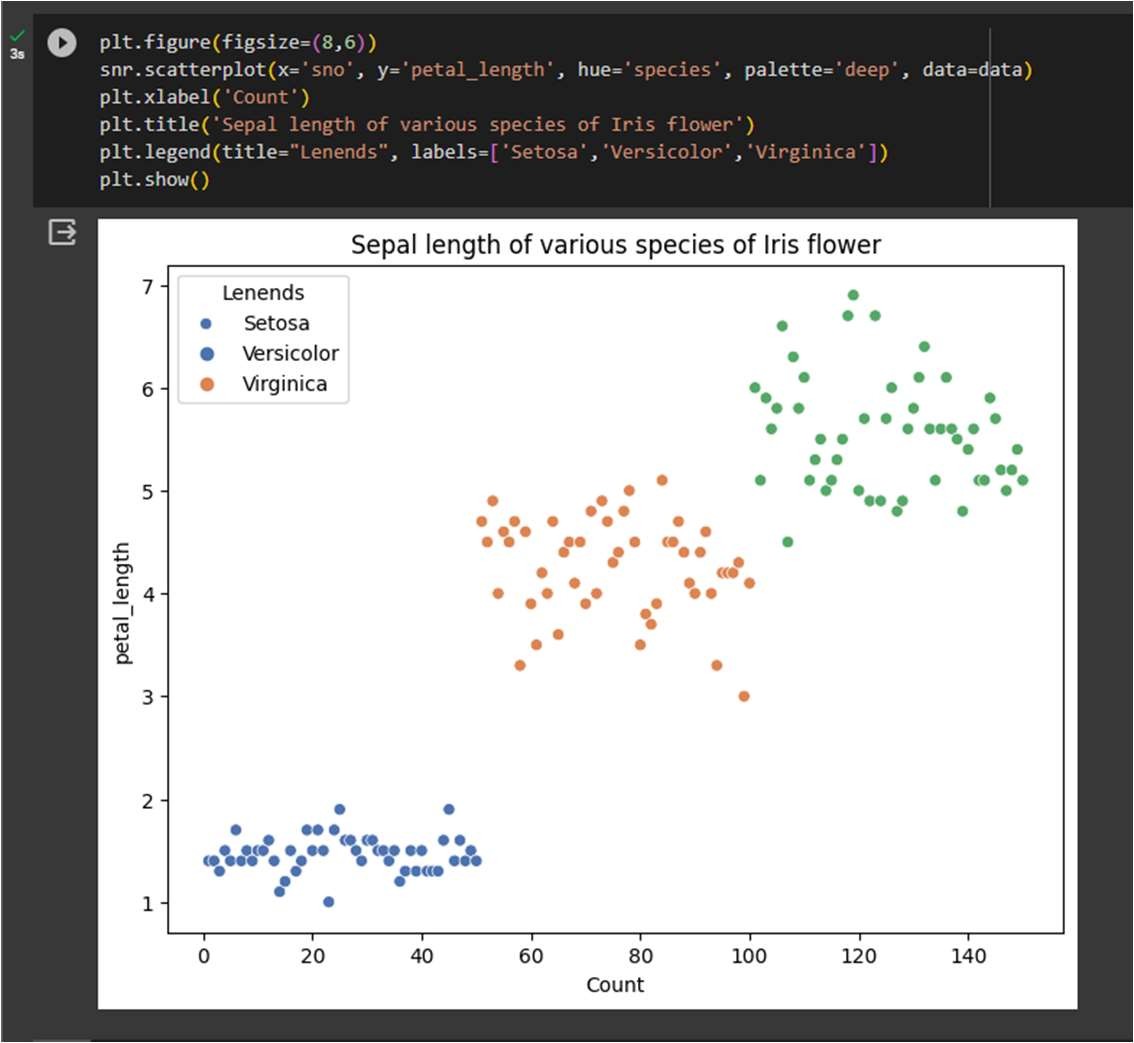
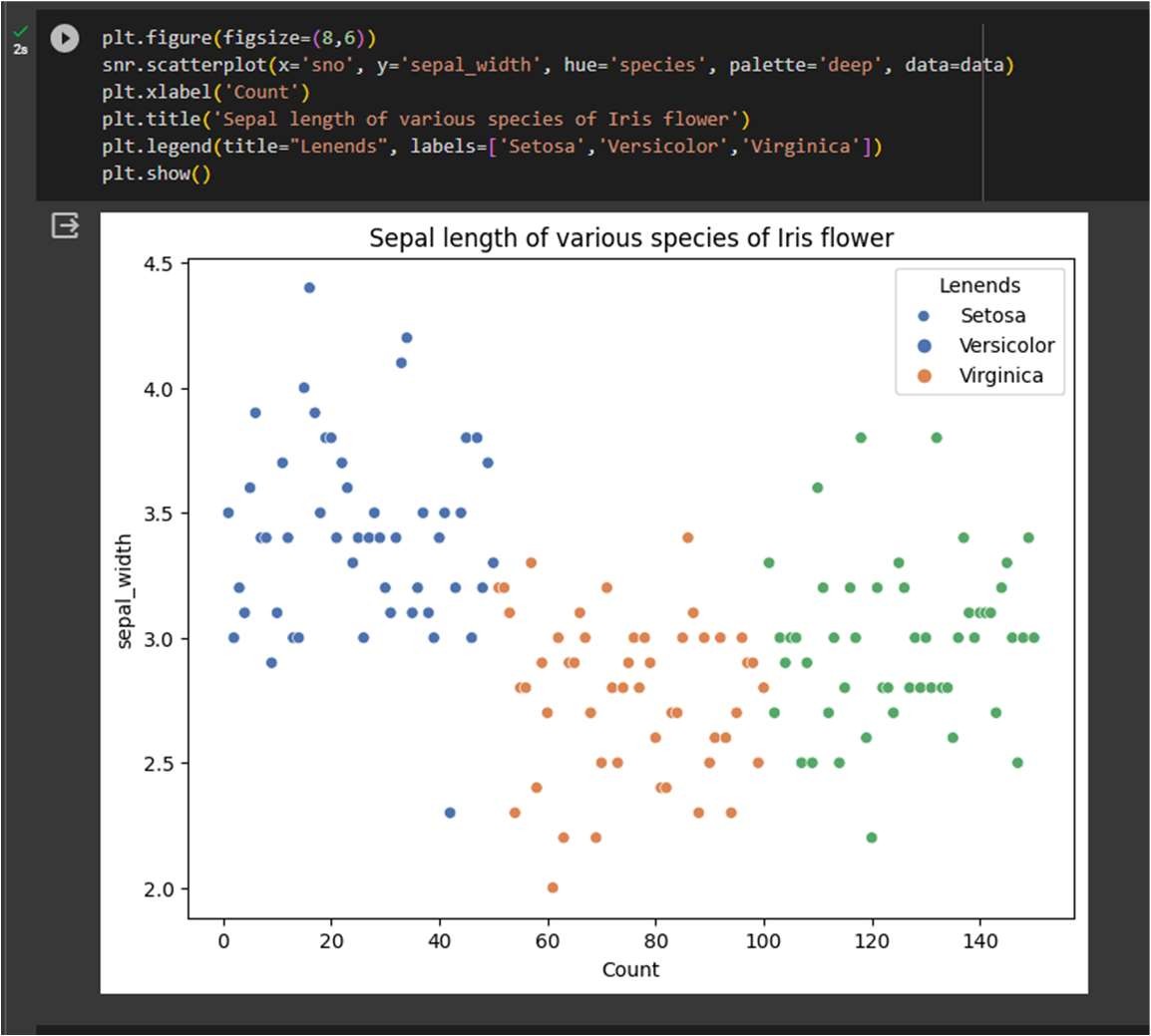
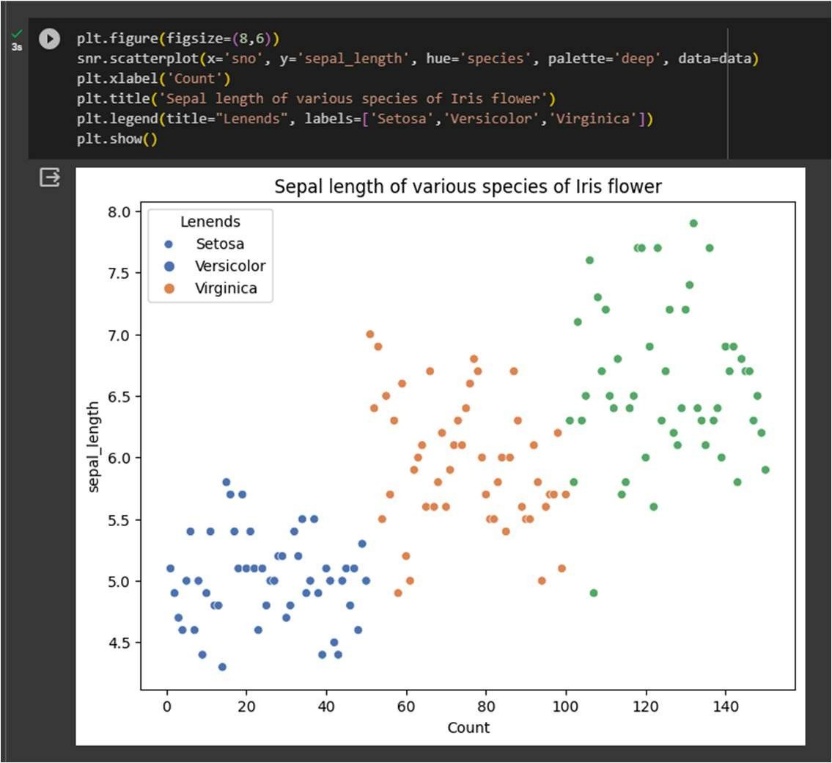
* Step-5 – Extracting all information about data.



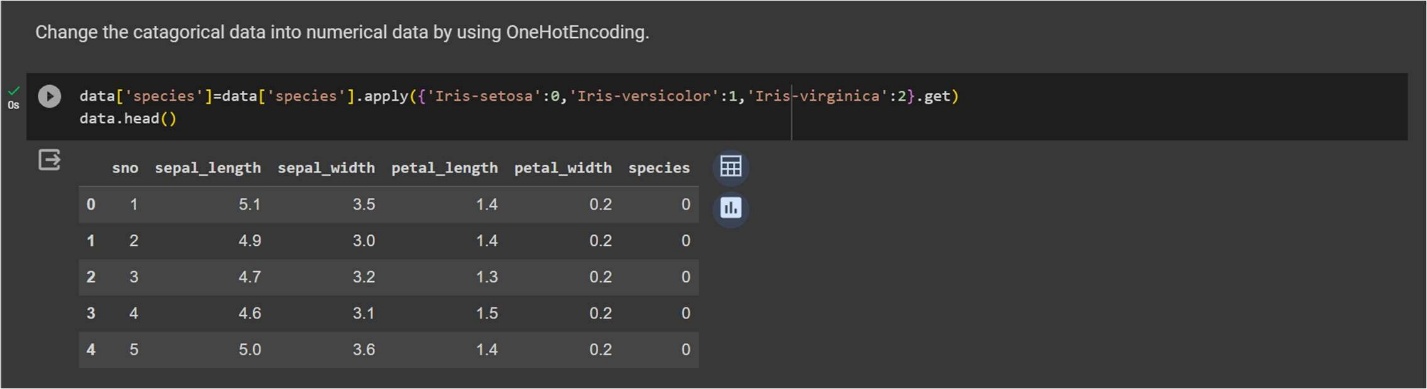
* Step-6 - Checking shape of data and checking unique values in dependent variable.



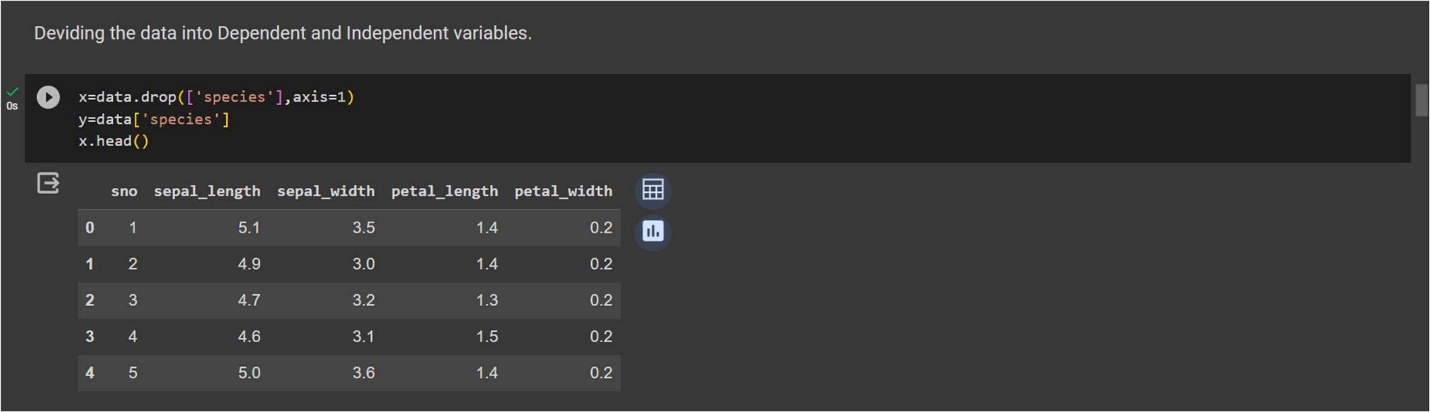
Step-7 - Visualization on different species of Iris flower using Python Data Visualization.



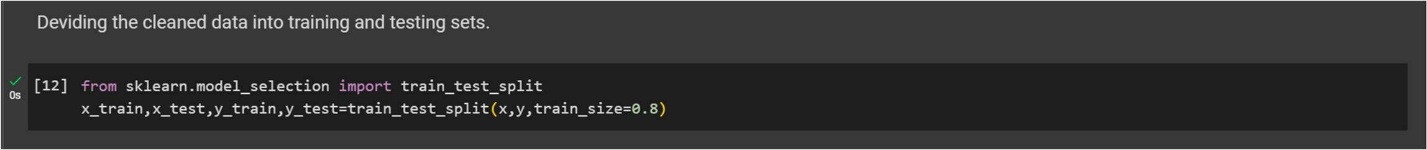
Step-8 – Data preprocessing or (Data cleaning) performed by the one hot encoding in this ocess we change categorical data into numerical data and the technique is called feature Engineering.



* Step-9 – Splitting the cleaned data into dependent and independent variables.



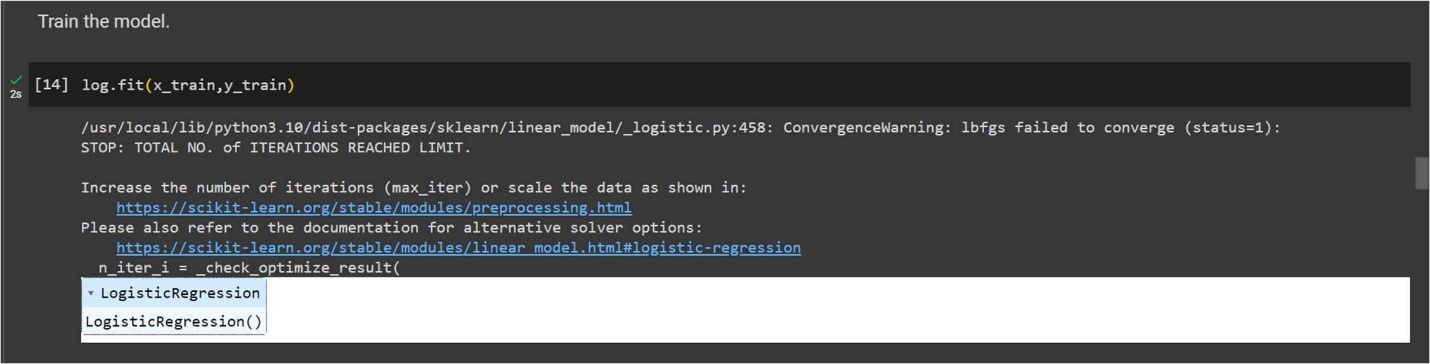
* Step-10 – Splitting the data into train and test sets with train\_test\_split using sklearn library.



* Step-11 – Import first machine learning model ‘Logistic regression’.



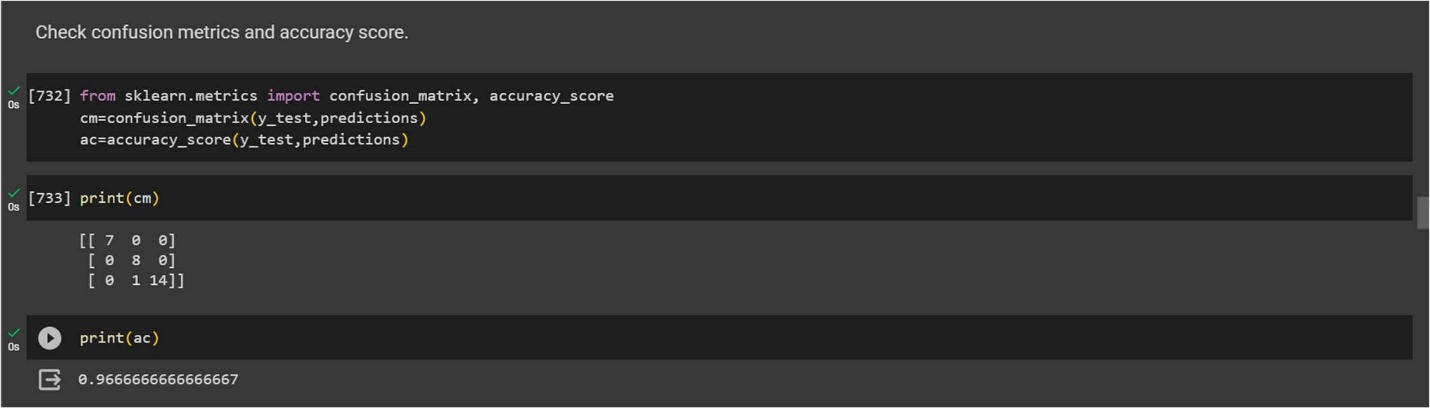
Step-12 – Train the model using .fit() function.



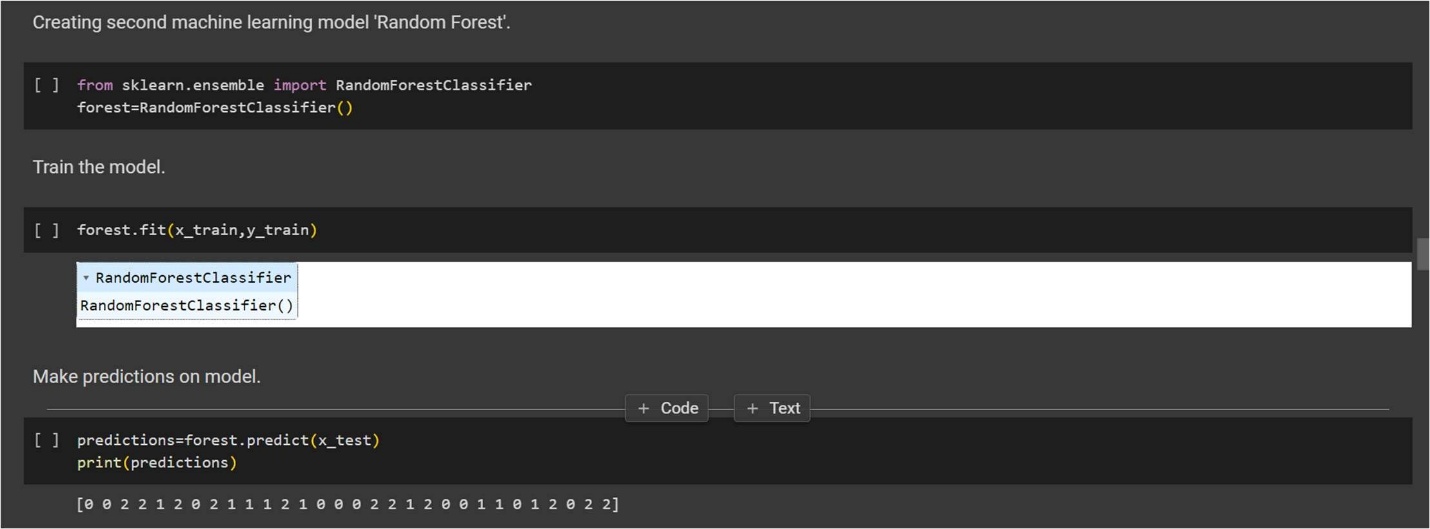
* Step-13 – Make predictions on model using .predict() function.



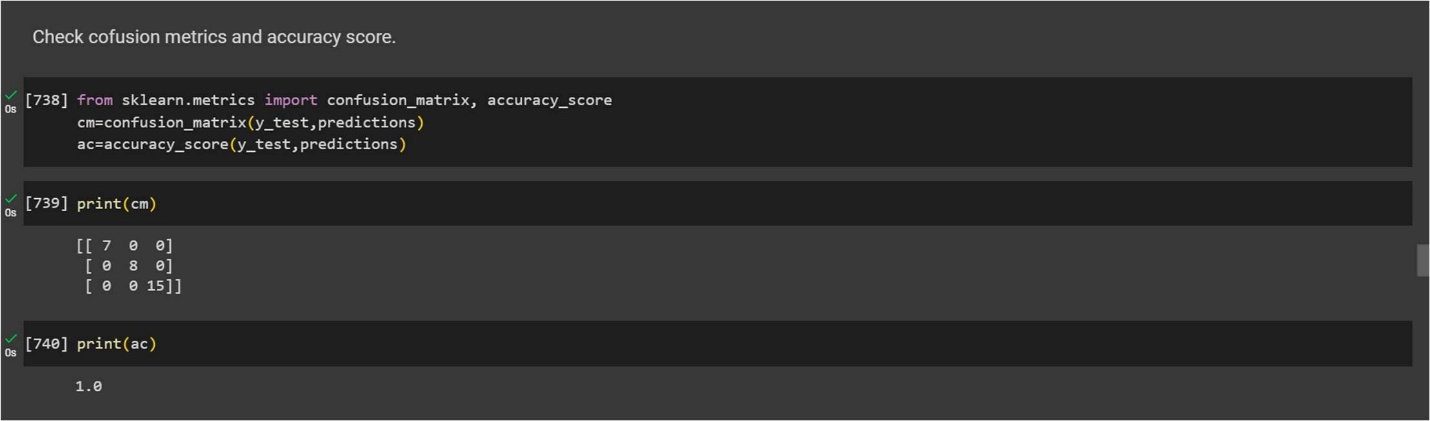
* Step-14 – Check the accuracy score and print a confusion metrics with confusion metrics & accuracy score.



Step-15 – Import the Second Machine Learning Model ‘Random Forest’ and train model and then make prediction.



* Step-16 – Print a confusion metrics and check accuracy score for Random forest Model.



Step-17 – Import the Third Machine Learning Model Support Vector Machine and train model and then make prediction.



Step-18 – Print a confusion metrics and check accuracy score for Support Vector Machine Model.



Conclusion – This project demonstrats the end-to-end process of building a machine learning model for Iris flower classification. It includes data preprocessing, exploratory data analysis, and model training. the first model is used in this project is Logistic Regression, second used is Random Forest, and third is Support vector Machine and model was trained with an accuracy of 97% approximate.

**Sales Predictions using Advertising Costs by using Machine Learning Model**

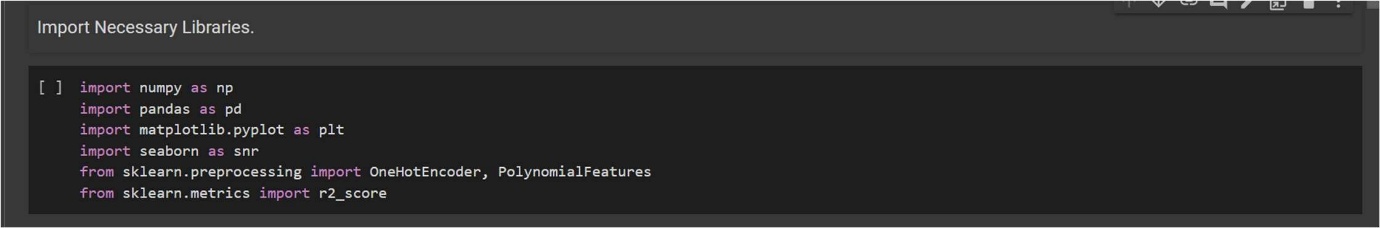
Problem Statement – In the business world, advertising is a crucial element for any company looking to promote its products or services. However, advertising costs can be substantial, and businesses need to determine the effectiveness of their advertising campaigns. This is where sales prediction comes in – it’s a critical aspect of advertising that helps companies understand how much revenue they can expect from their advertising campaigns. In this project I build a model which predicts sales based on the money spent on different advertising platforms for marketing.

Steps to be taken in the Project is sub-divided into the following sections. These are:

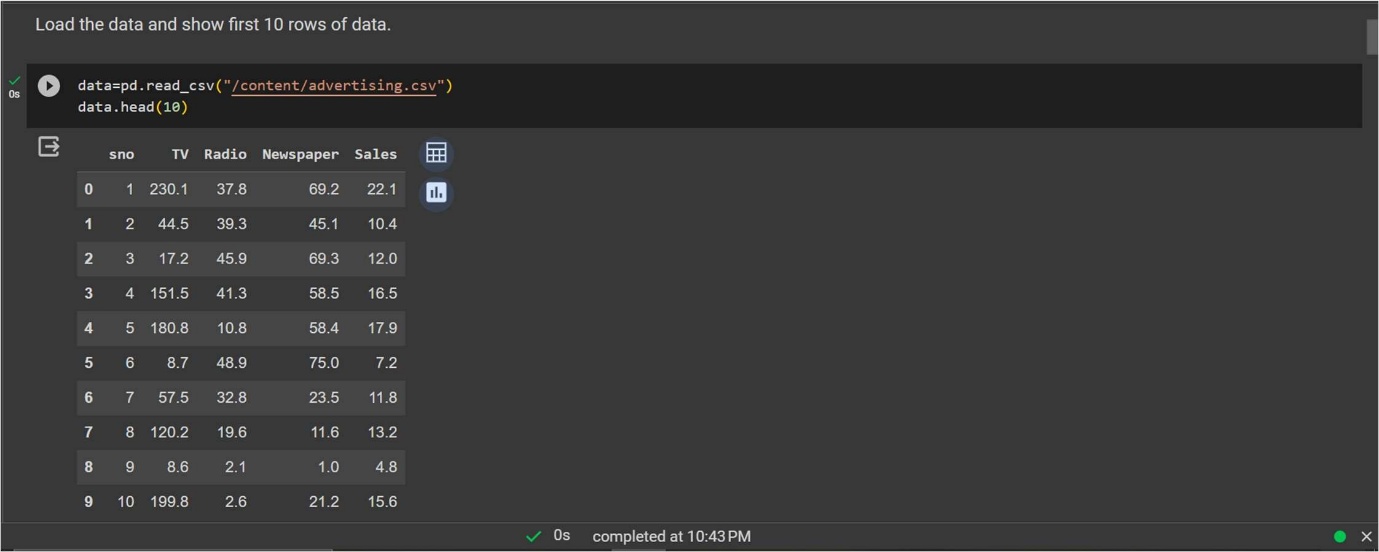


* Loading necessary libraries such as numpy , pandas , sklearn etc.
* Loading the dataset as CSV file and showing first 10 rows.
* Drop the unnecessary columns from dataset.
* Calculate statistical values and round them up to 3 decimal places. Checking for null values and return their sum of numbers of true values in each column.
* Handle the null by mean of all values fill into them.
* Extracting all information about data.
* Checking Shape of Data.
* Visualization of Sales by different source of Advertisement cost using Python data visualization.
* Data preprocessing or (Data cleaning) performed by the one hot encoding in this process we change categorical data into numerical data and the technique is called feature Engineering.
* Splitting the cleaned data into dependent and independent variables.
* Splitting the data into train and test sets with train\_test\_split using sklearn library.
* Import different kind of Regression Models and Train that model with the help of .fit().
* Predicting the trained models and then checking their accuracy of the model using accuracy score.
* Then recall the train\_test\_split and split the data into training and testing set with different models.
* Then predicting the trained models and checking the accuracy of model and check the accuracy difference.

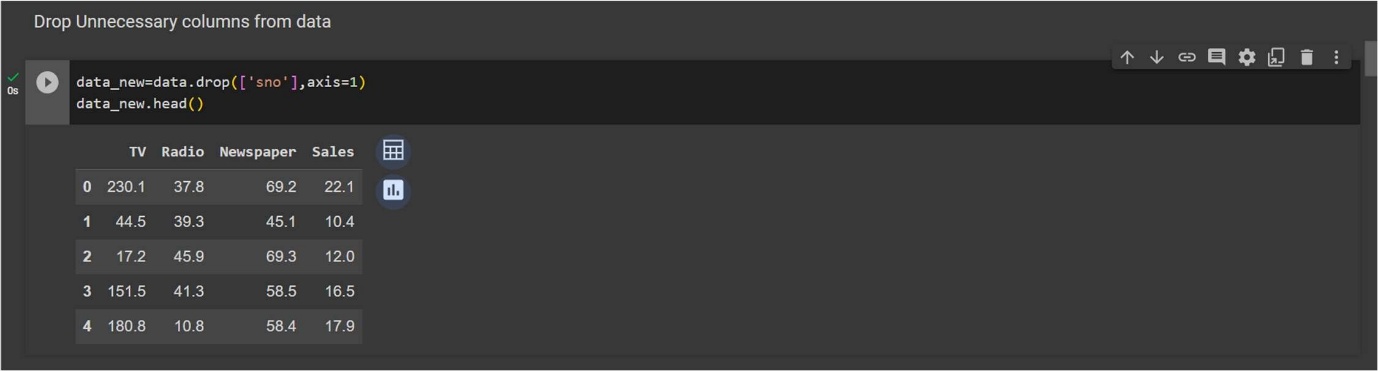
* Step-1 – Loading Necessary Libraries used in machine learning.



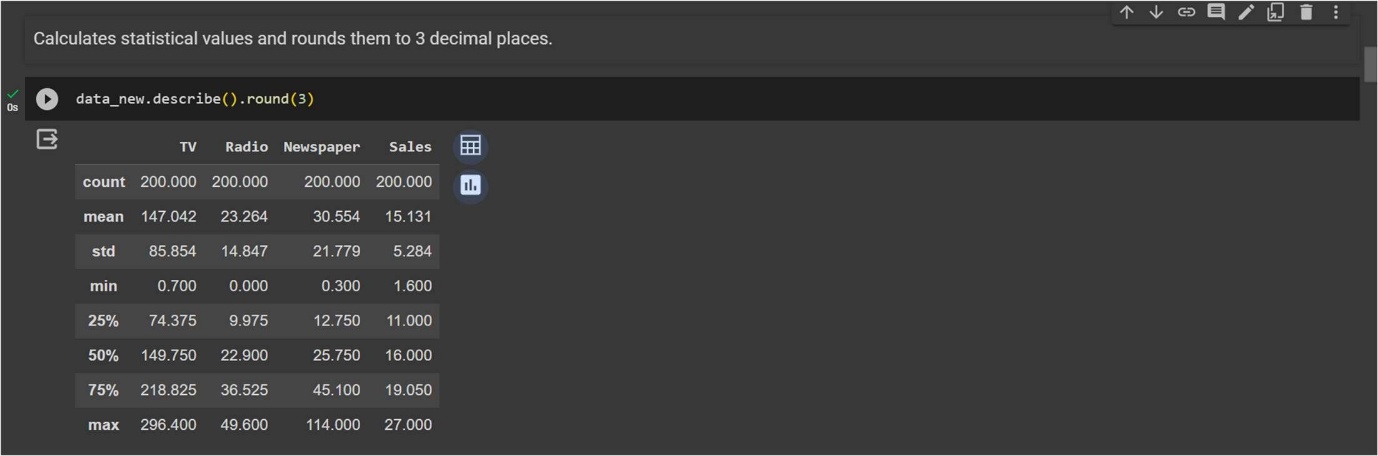
* Step-2 – Loading the dataset as csv file and showing first ten rows.



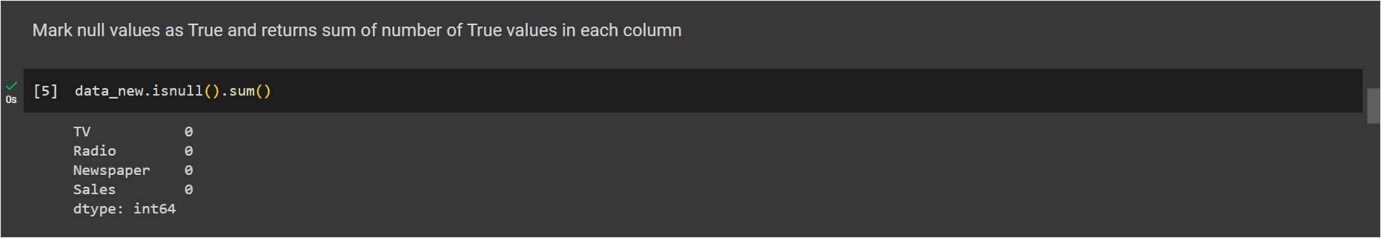
* Step-3 – Drop the unnecessary columns from dataset.



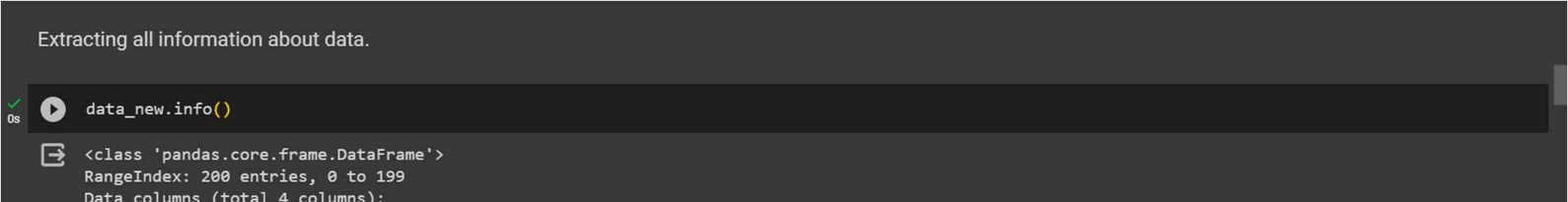
* Step-4 – Calculate statistical values and round them up to 3 decimal places.



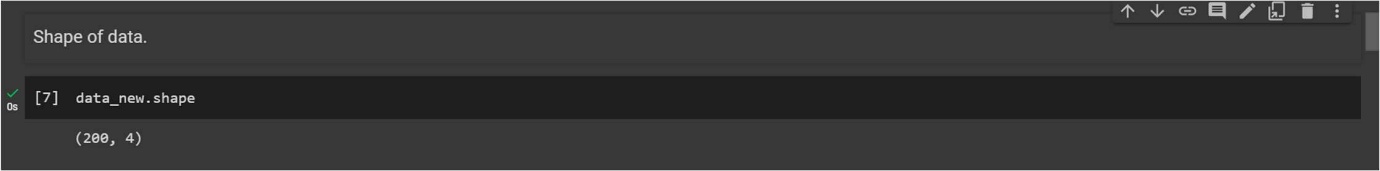
* Step-5 – Checking for null values and return their sum of numbers of true values in each column.



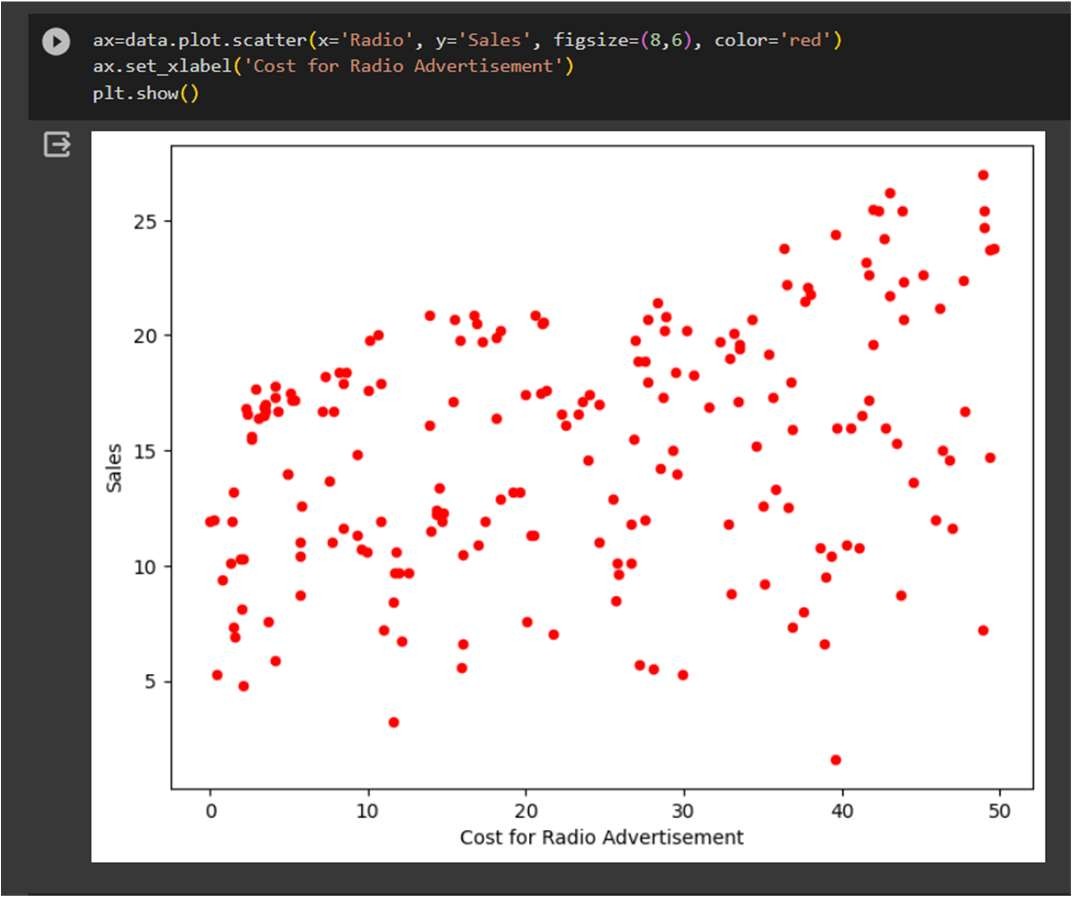
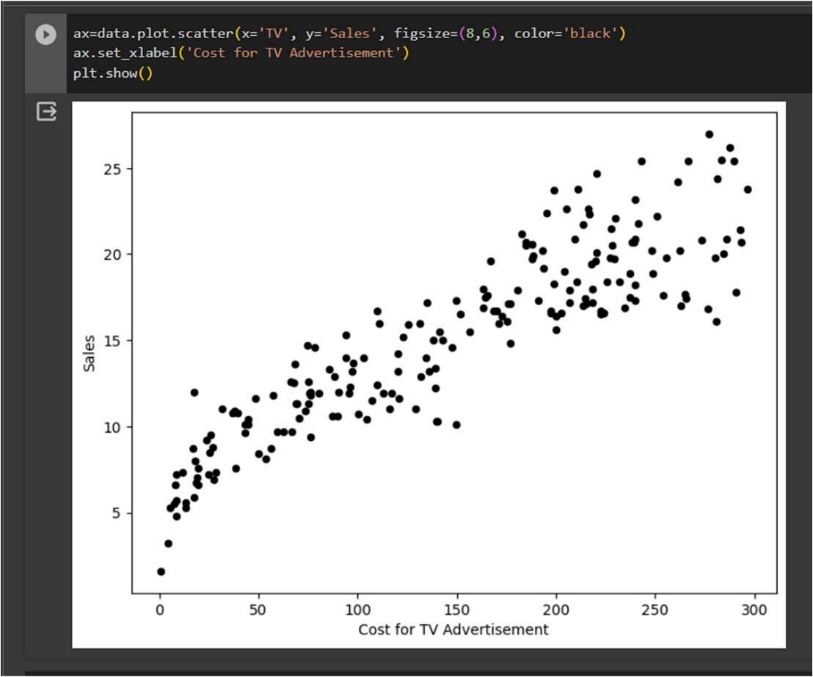
* Step-6 – Extracting all information about data.

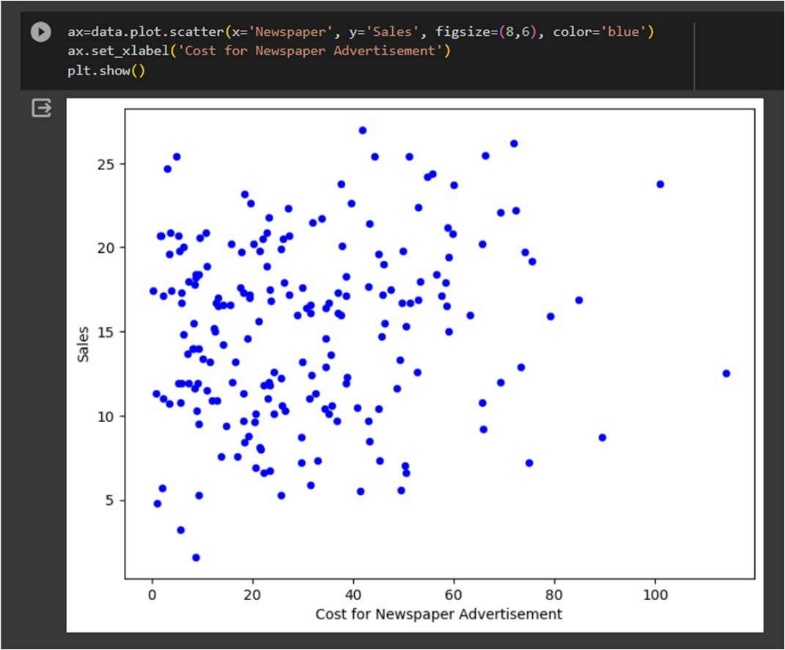


* Step-7 – Checking shape of data.

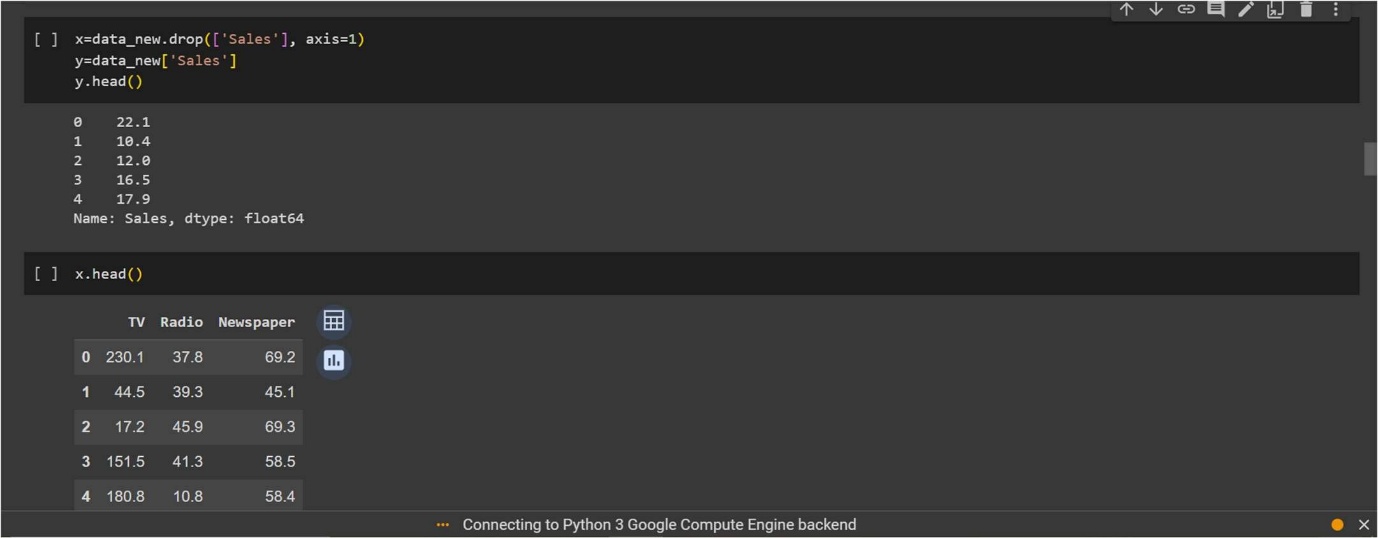


* Step-8 – Visualization of Sales by different source of Advertisement cost using Python data visualization.





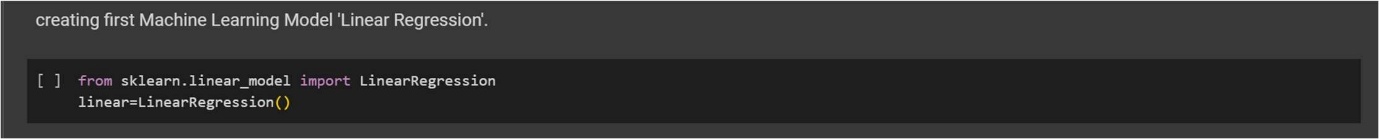
* Step-9 – Splitting the cleaned data into dependent and independent variables.



* Step-10 – Splitting the data into train and test sets with train\_test\_split using sklearn library.



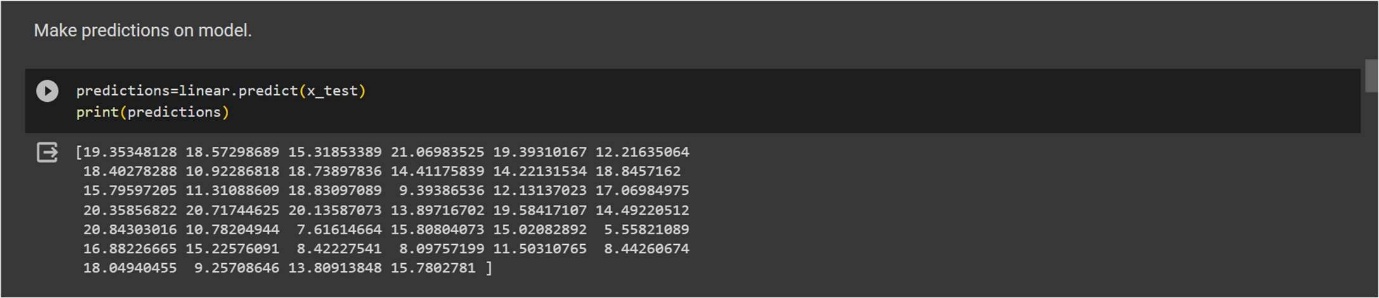
* Step-11 – Import first regression model ‘Linear Regression’.



* Step-12 – Train the model using .fit() function.



* Step-13 – Make predictions on model.



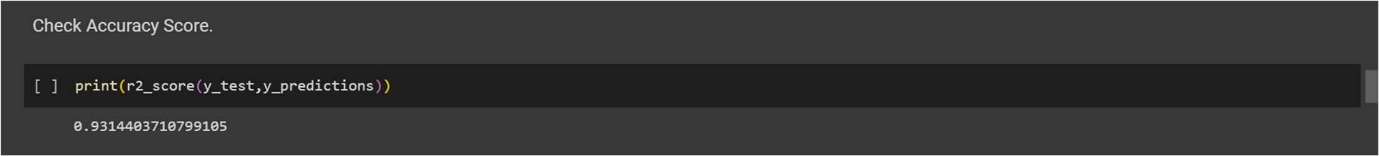
* Step-14 – Check the accuracy of model by r2 score.



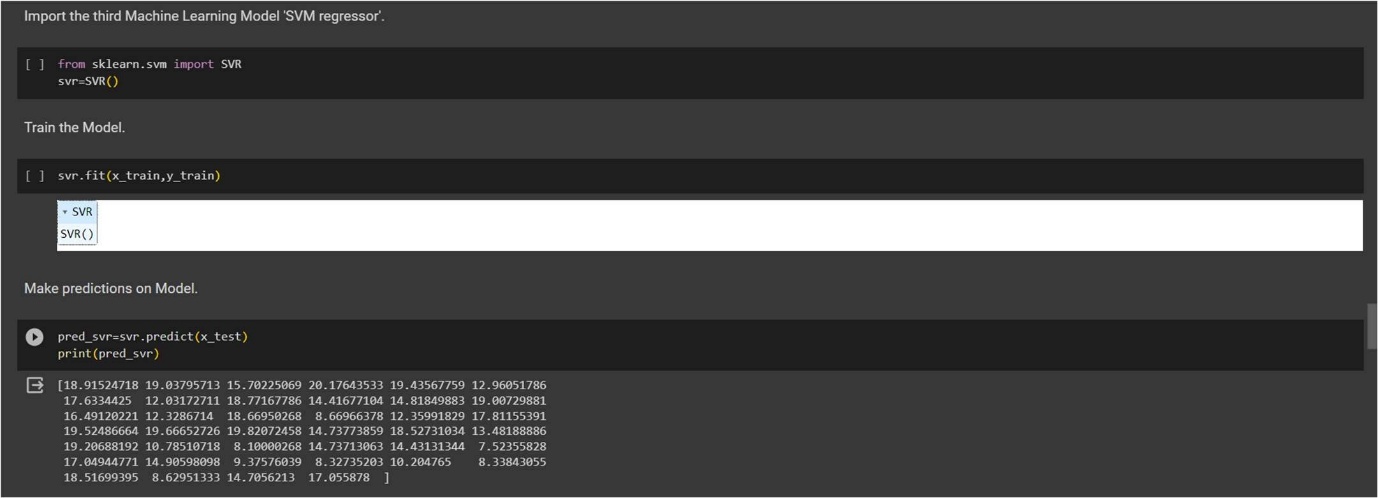
* Step-15 – Import the Second Machine Learning Model Polynomial Regression and train model and then make prediction.



* Step-16 – Check the accuracy of model by r2 score.



* Step-17 – Import the Third Machine Learning Model Support Vector Regressor and train model and then make prediction.



* Step-18 – Check the accuracy score.



Conclusion – In this project, we have demonstrated in detail how to apply linear regression, polynomial regression and support vector regressor model for predicting sales from data of spend cost for advertising. By carefully selecting the right variables, preparing and cleaning the data, and selecting an appropriate regression models, businesses can accurately predict sales from advertising cost ads.

project analysis resulted in a good R-squared value of 0.87100, 0.93144 and 0.89639 respectively, which indicates that the linear regression model has a decent fit for the data and gave 87%, 93% and 89% accuracy respectively. This level of accuracy can provide businesses with valuable insights into the effectiveness of their advertising campaigns and enable them to make informed decisions about how to allocate their resources.

**Credit Card Fraud Detection using Machine Learning Model.**

Problem Statement - For many banks, retaining high profitable customers is the number one business goal. Banking fraud, however, poses a significant threat to this goal for different banks. In terms of substantial financial losses, trust and credibility, this is a concerning issue to both banks and customers alike.

In the banking industry, credit card fraud detection using machine learning is not only a trend but a necessity for them to put proactive monitoring and fraud prevention mechanisms in place. Machine learning is helping these institutions to reduce timeconsuming manual reviews, costly chargebacks and fees as well as denials of legitimate transactions.

The problem statement chosen for this project is to predict fraudulent credit card transactions with the help of machine learning models.

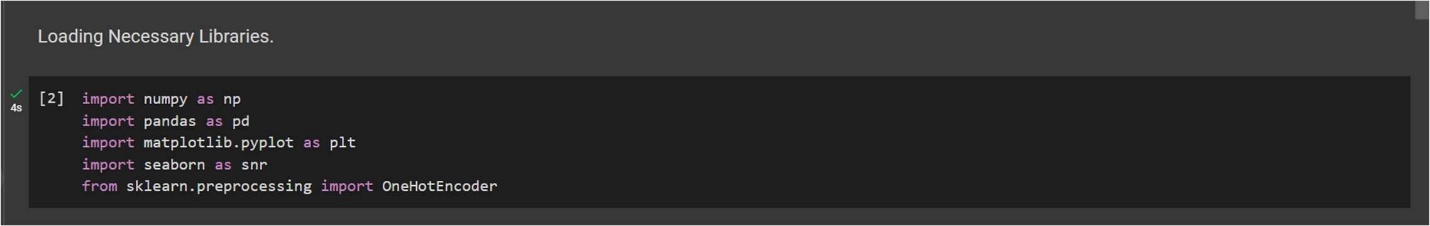
In this project, we will analyze customer-level data that has been collected and analyzed during a research collaboration of Worldline and the Machine Learning Group.

Steps to be taken in the Project is sub-divided into the following sections. These are:

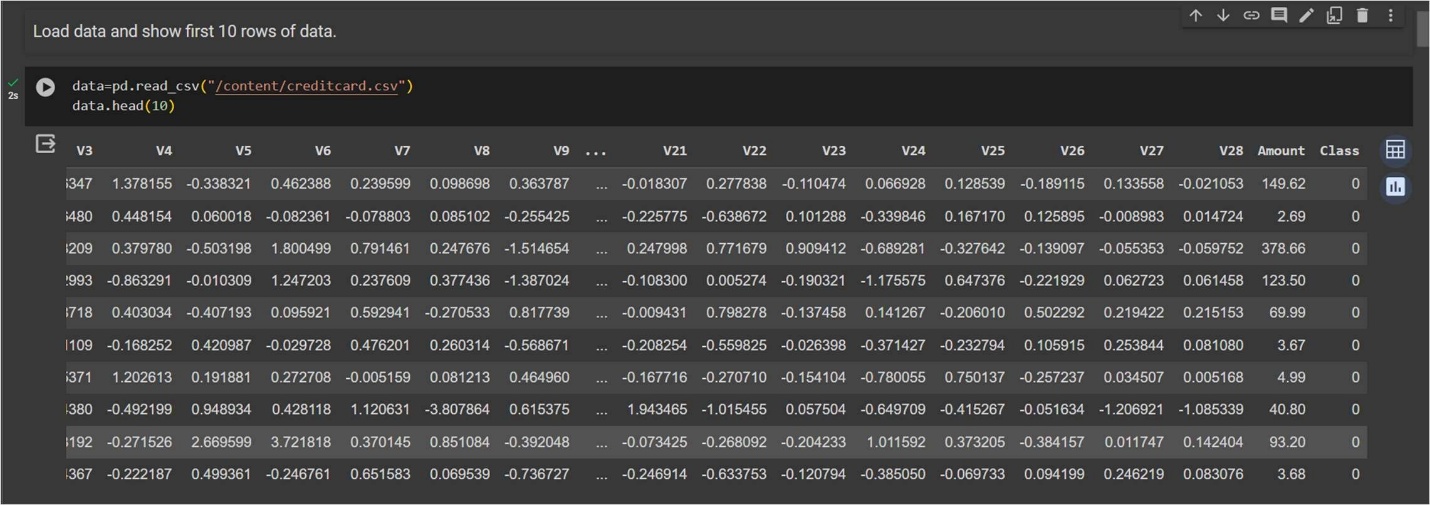


* Load the necessary libraries such as Numpy , Pandas , sklearn.model etc.
* Loading the dataset as csv file and showing first ten rows.
* Drop the unnecessary columns from the data.
* Calculate statistical values and round them up to 3 decimal places.
* Checking for null values and return their sum of numbers of true values in each column.
* Handle the null by mean of all values fill into them.
* Extracting all information about data.
* Checking shape of data.
* Visualization on different species of credit cards transaction information using python data visualization.
* Data preprocessing or (Data cleaning) performed by the one hot encoding in this process we change categorical data into numerical data and the technique is called feature Engineering.
* Splitting the cleaned data into dependent and independent variables.
* Splitting the data into train and test sets with train\_test\_split using sklearn library.
* Import different kind of Classification Models and Train that model with the help of .fit().
* Predicting the trained models and then checking their accuracy score and confusion metrics of the model using confusion metrics & accuracy score.
* Then recall the train\_test\_split and split the data into training and testing set with different models.
* Then predicting the trained models and checking the accuracy of model and check the accuracy difference.
* And finally predict whether the classification (or detection) of Credit cards is generated or not.

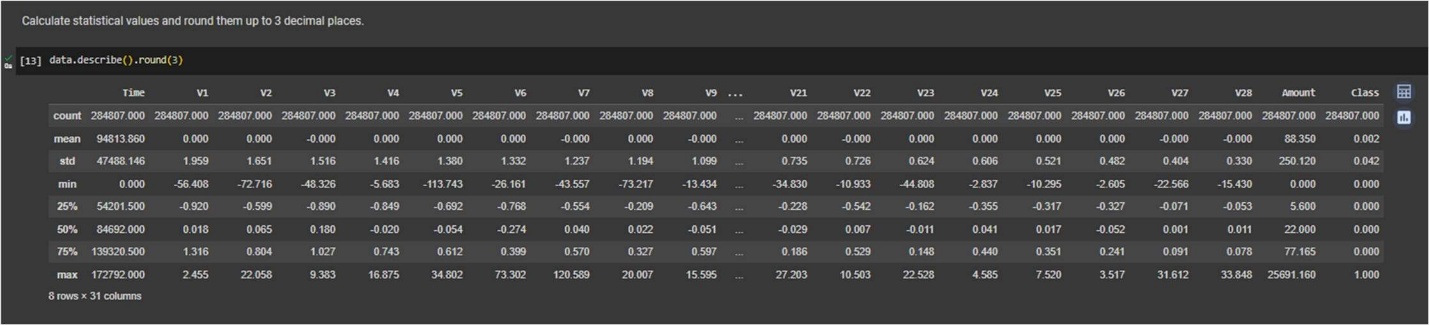
* Step-1 – Import necessary libraries.



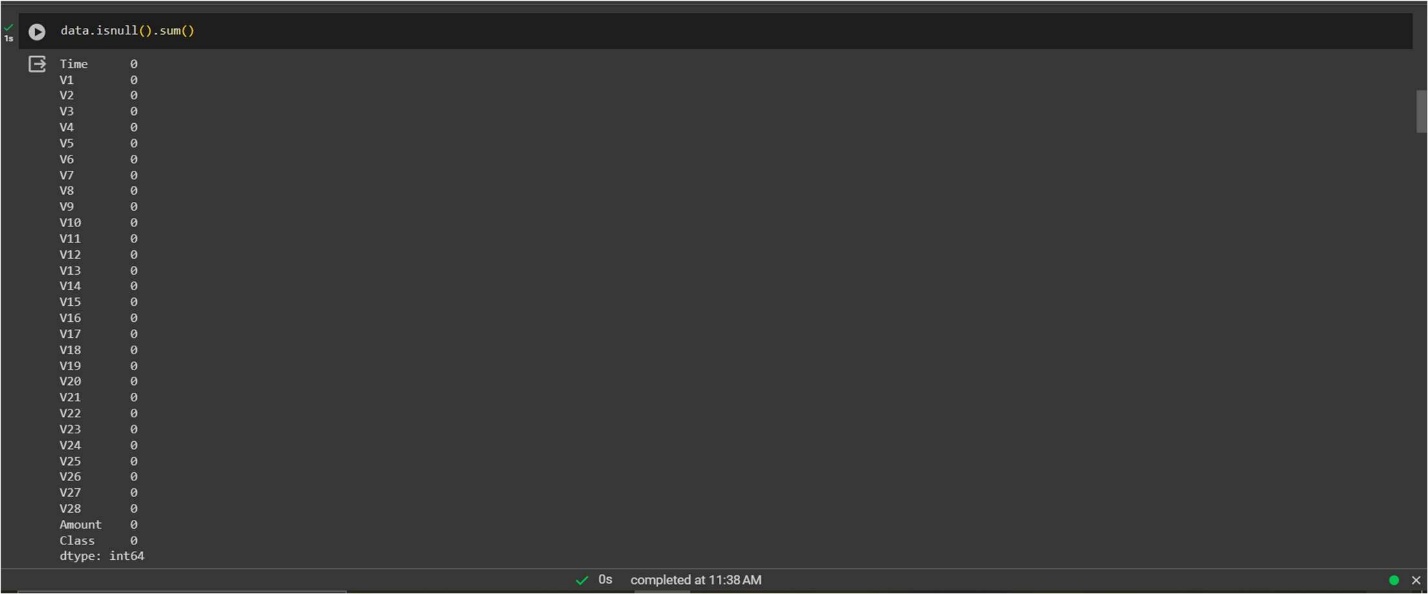
* Step-2 – Loading the dataset as csv file and showing first ten rows.



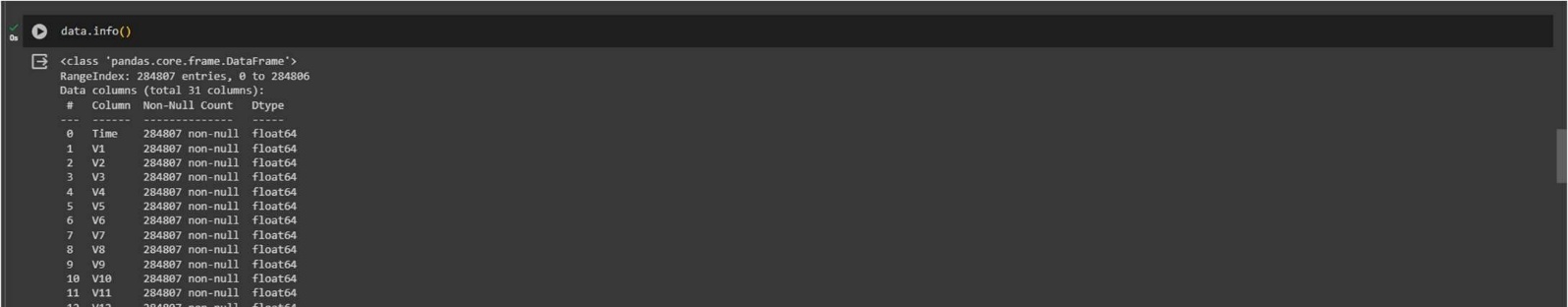
* Step-3 – Calculate statistical values and round them up to 3 decimal places.



* Step-4 – Checking for null values and return their sum of numbers of true values in each column.



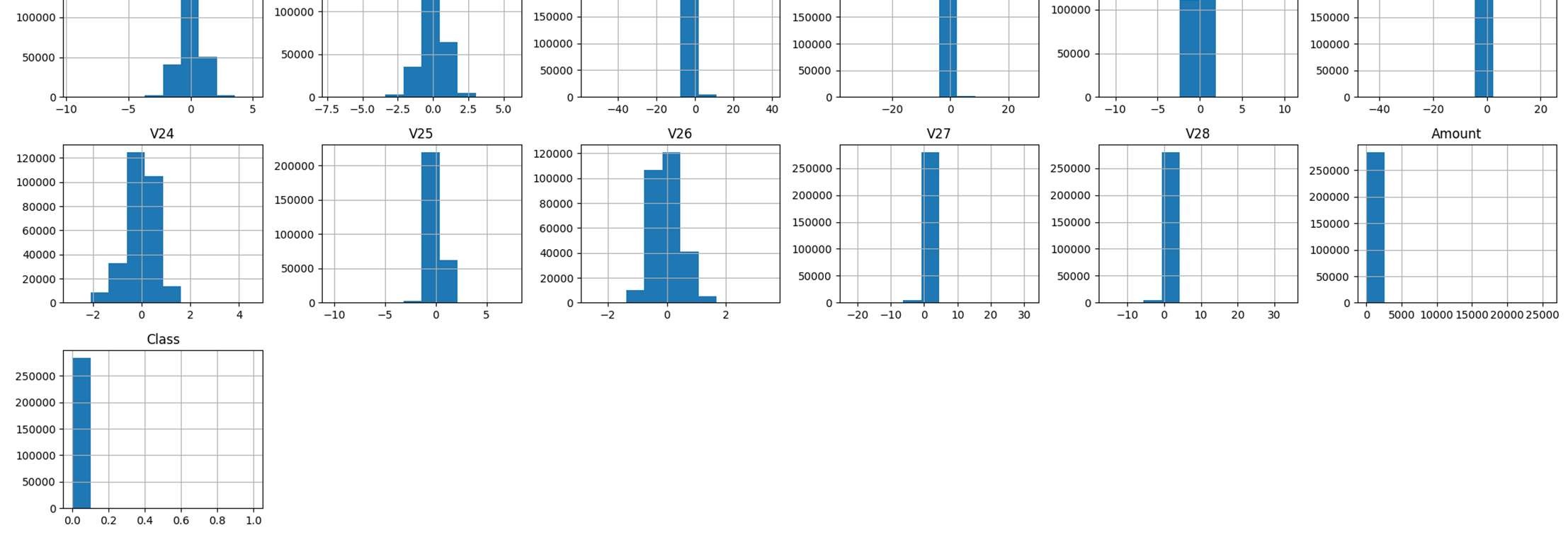
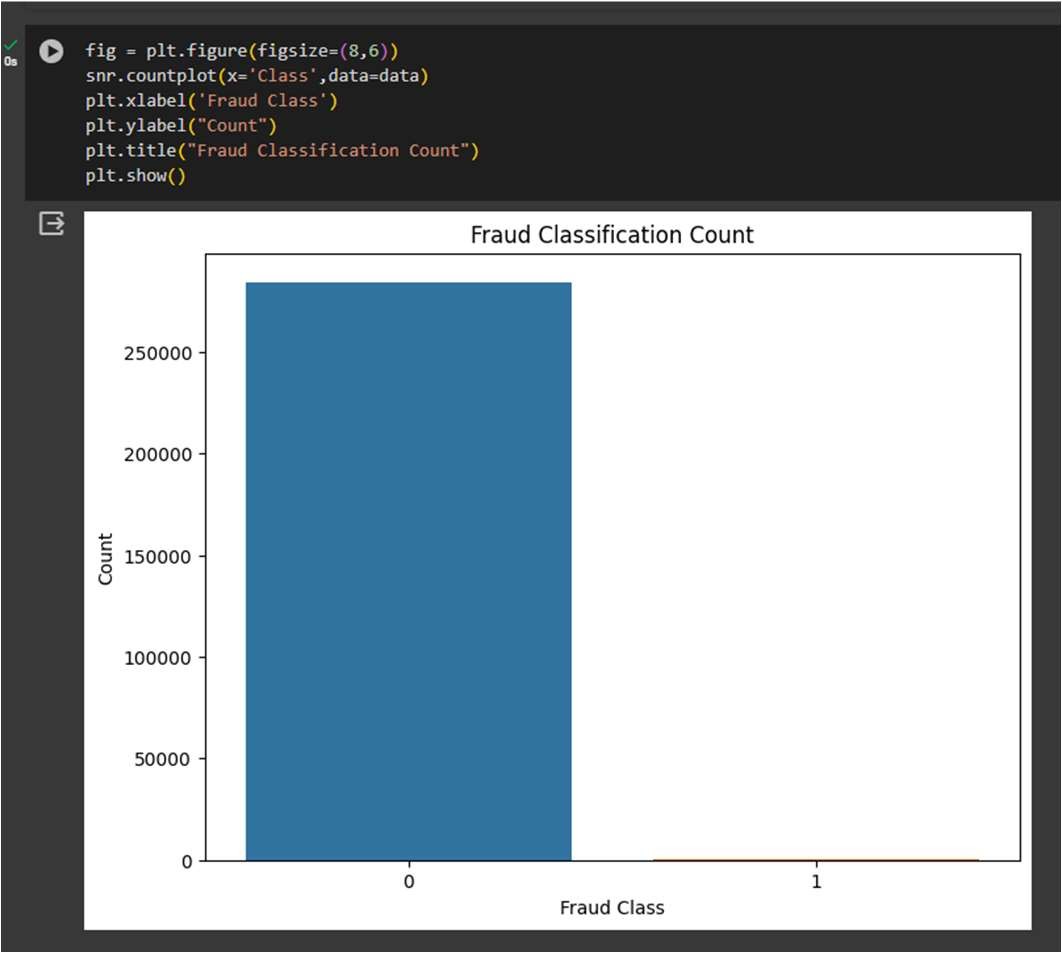
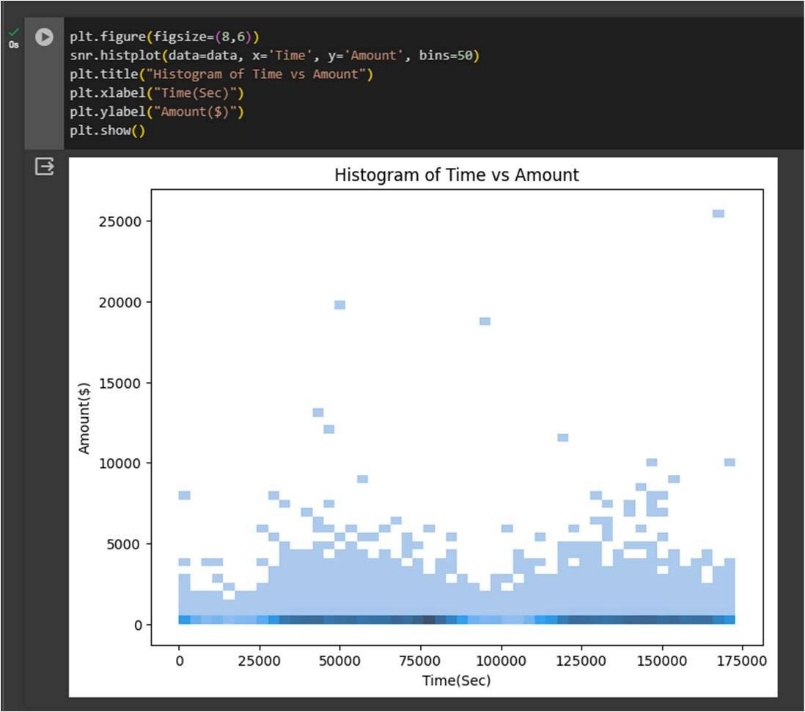
Step-5 – Extracting all information about data.



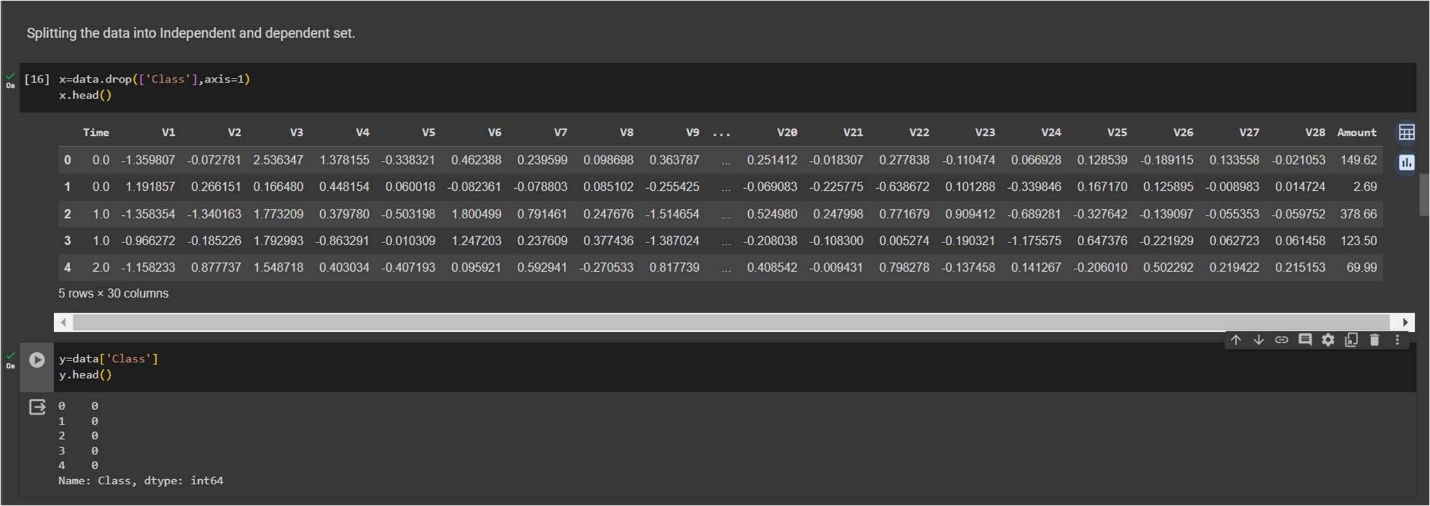
* Step-6 – Checking shape of data.



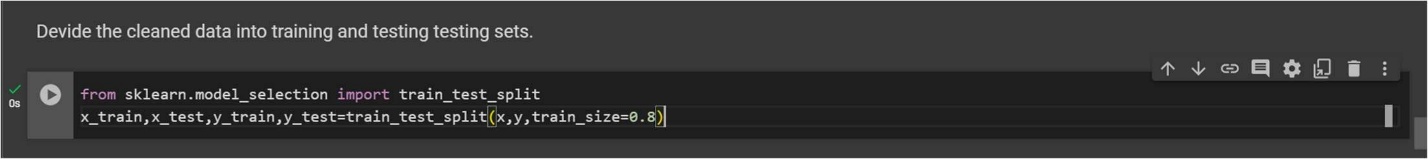
* Step-7 – Visualization on different species of credit cards transaction information using python data visualization.



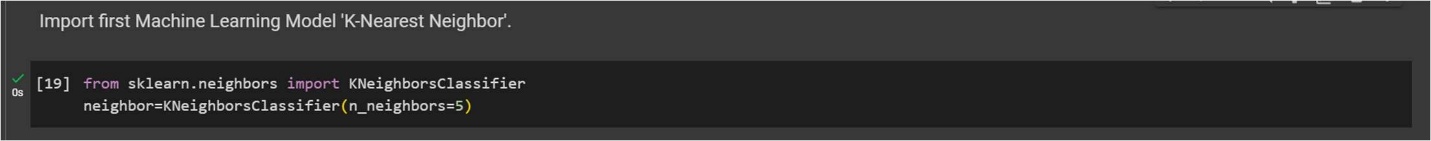
Step-8 – Splitting the data into dependent and independent variables.



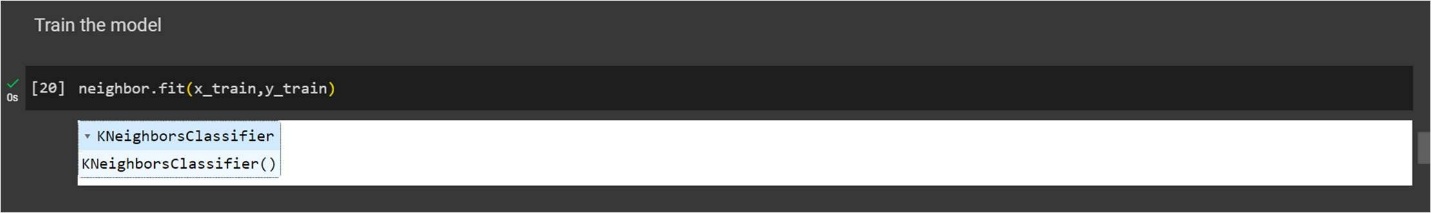
* Step-9 – Splitting the data into training and testing sets.



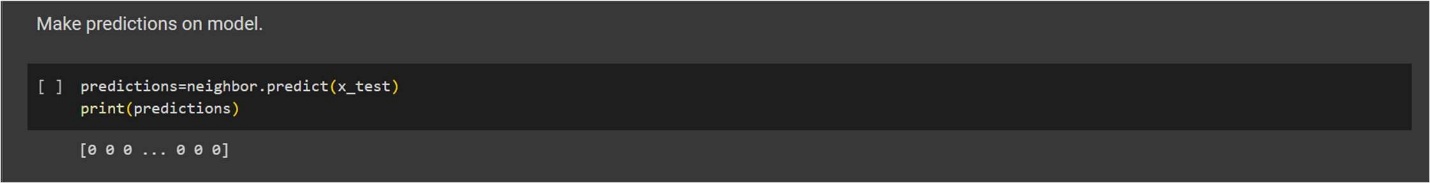
* Step-10 – Import first machine learning model K-Nearest neighbor taking n\_neighbor=5.



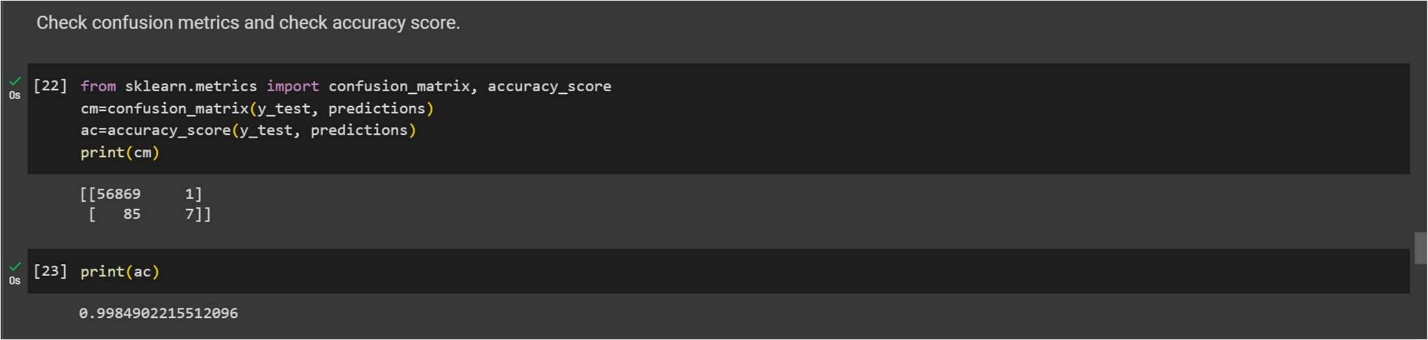
* Step-11 – Train the model using .fit() function.



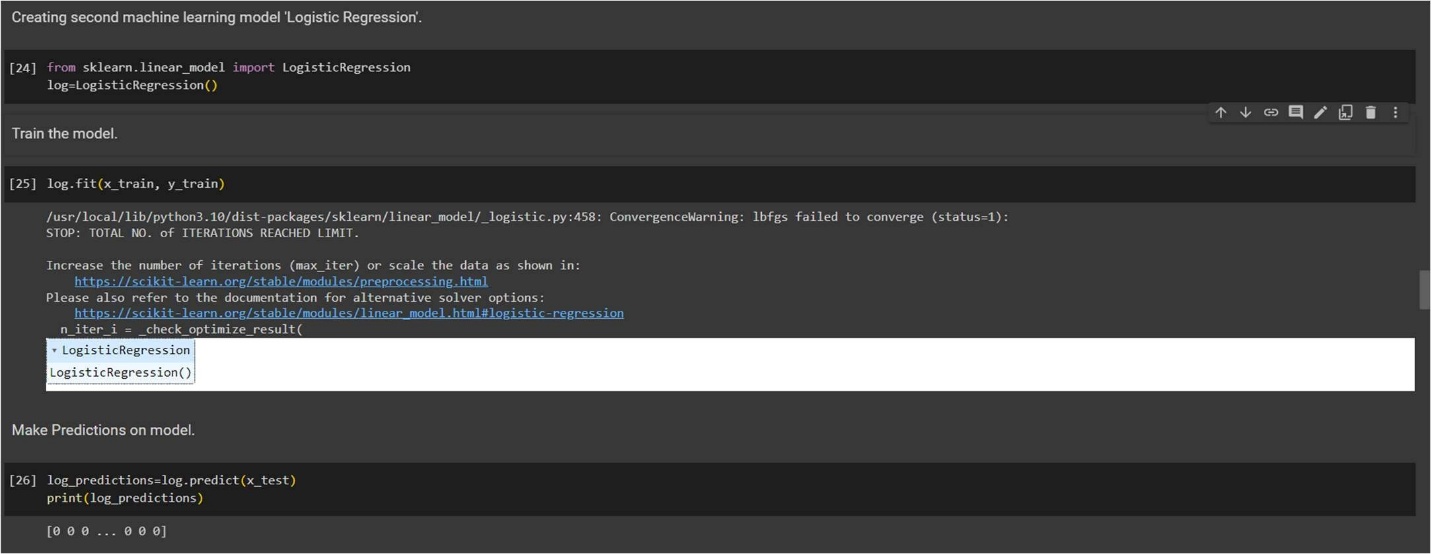
* Step-12 – Making predictions on model.



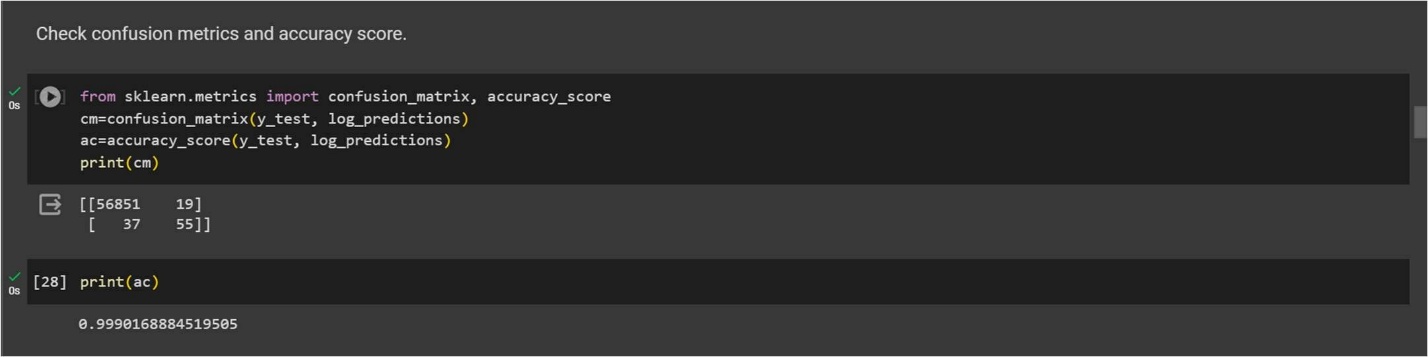
* Step-13 – Checking confusion metrics and accuracy score of model.



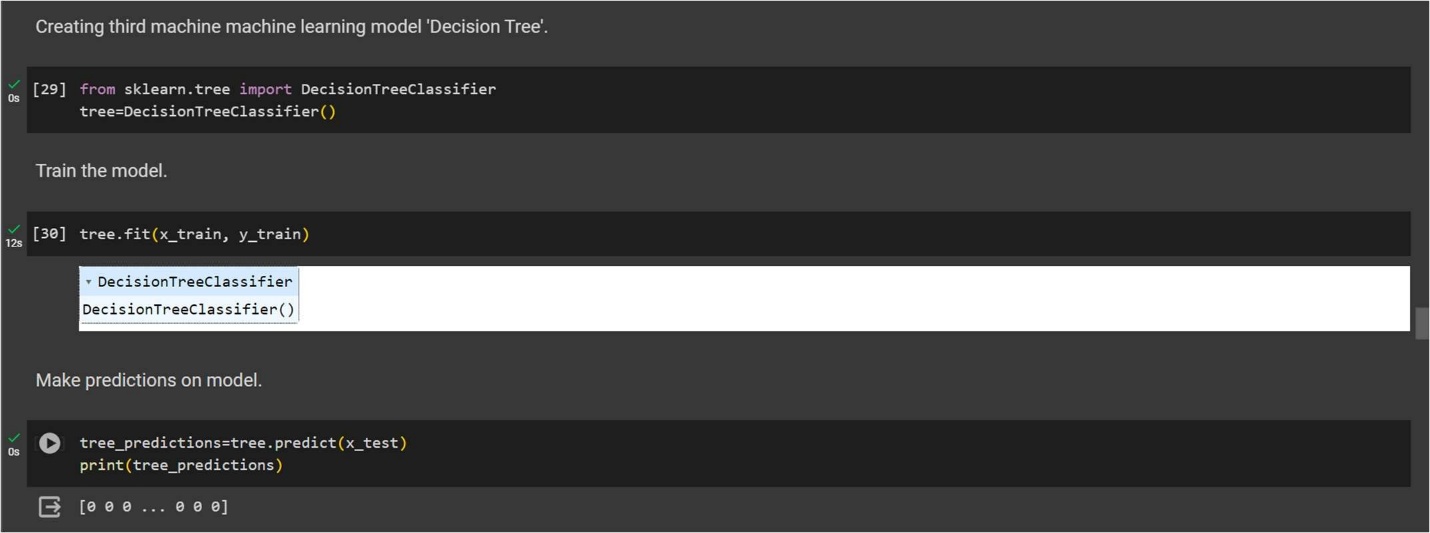
* Step-14 – Import the Second Machine Learning Model ‘Logistic Regression’ and train model and then make prediction.



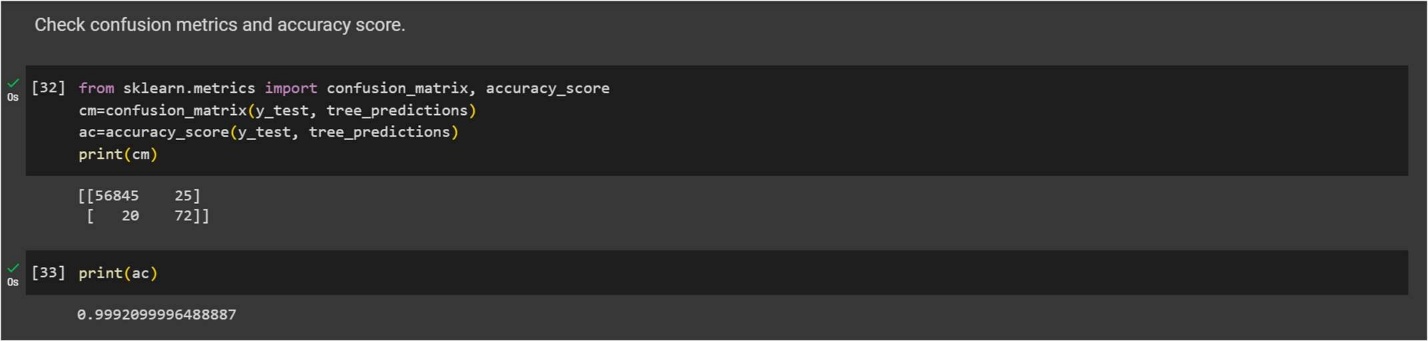
Step-15 – Print a confusion metrics and check accuracy score for Logistic Regression Model.



* Step-16 – Import the Third Machine Learning Model Decision Tree and train model and then make prediction.



* Step-17 – Print a confusion metrics and check accuracy score for Support Vector Machine Model.



Conclusion – Here, we have to focus on a high recall in order to detect actual fraudulent transactions in order to save the banks from high-value fraudulent transactions,

After performing several models, we have seen that on performing Machine Learning models I got 99% of accuracy of model by using K-Nearest Neighbor algorithm of Machine Learning, I got 99% of accuracy of model by using Logistic Regression algorithm of Machine Learning and I got 99% of accuracy of model by using Decision Tree algorithm of machine Learning. Overall I got 99% of accuracy for the complete Machine Learning Model of Credit Card Fraud Detection.