Dataset.info()  
It gives the raw information about the dataset

Dataset.describe()  
to get the statistical information about the dataset like the quartiles of the columns

pd.get\_dummies() #one-hot encoding  
one-hot encoding is a popular technique used in ml and data processing to represent categorical variables or features as binary vectors.

Drop\_first = True  
this is used to prevent multicollinearity

Multicollinearity = if 2 or more independent variables have a exact linear relationship between them then it is perfect multicollinearity

This helps in reducing columns  
in ours, if Gender\_male = 0 then it is understood that it is a female. Another column need not be present fir this information.

If this is not done then it leads to uncertainty and undefined coefficient estimation and loss of model interpretability. It becomes difficult to distinguish the individual effects of each variable on the dependent variable.   
Because of this our prediction fails.

dataset2.corrwith(dataset['Exited']).plot.bar(figsize=(16,9), title='Correlated with Exited column', rot=45, grid = True)

correlating dataset2 with dataset having the value of only the exited column   
the plotting the bar with the figure size, title and rotation as 45 and grids.  
Rotation = 45 is tilting the naming by 450

random\_state is used to maintain the consistency of the samples.

To scale the dataset we use standardscaler

Standardsacler is a commonly used technique in ml for standardizing or scaling numerical features before fitting a model. It transforms the data by subtracting the mean and dividing by the standard deviation resulting in a distribution with mean of 0 and standard deviation of 1

Logistic regression is used for predicting categorical dependent variable using a given set of independent variables

Accuracy score finds the accuracy of the preduicted data and the test data

F1 score is also used to test the accuracy   
if value of f1 score is 1 then it is best and if it is 0the it is the worst

In precision score also the best value is 1 and the worst value is 0 and so is with the recall value

The confusion matrix gives us the true positive, true negative, false positive and false negative values.

Random forest is a classifier that contains a no of decision trees on various subsets of the given dataset and it takes the average to improve the predictive accuracy of that dataset

So in a nutshell random forest classifier averages the decisions from different decision trees and it predicts the final output

Array([0]) = staying with the bank