## Pre-requisites:

1. Firstly, we create an Amazon S3 bucket.

Graphical user interface, text, application, Word, email

Description automatically generated

1. Upload the CSV training and test data files to the bucket.

Graphical user interface, text

Description automatically generated

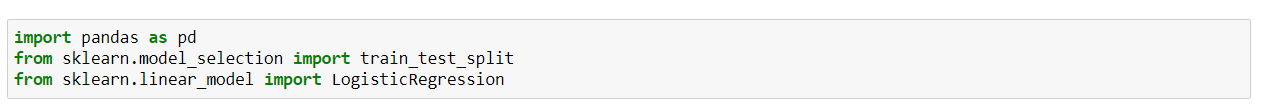
Note: The objects permission must be public.

Graphical user interface, application

Description automatically generated

## Logistic Regression with AWS \_Task 1

1. Firstly, we import all the libraries needed for regression.



1. Now, we load the data from the files that we have uploaded to the AWS S3 bucket into a panda data frame. Here the object URL of the file is added. The size of the data frame is printed.

Table

Description automatically generated

1. Next, we clean the data. If there are any unnecessary columns or rows, we eliminate them.

Text

Description automatically generated

1. Next, we create training and testing data by using the train\_test\_split() function. We are running logistic regression and fitting the training data into the logistic regression model. We predict the values for the test data. Here, we are using Albury weather reports to predict if it will rain or not tomorrow.A picture containing graphical user interface

   Description automatically generated
2. Finally, we evaluate the model and print the accuracy of the model.

Rectangle

Description automatically generated with medium confidence

1. Now make predict the outputs for the inputs that are loaded by reading the test file from the amazon S3 bucket. We print the predicted values. Here, we predict if it will rain tomorrow or not based on the inputs given. Zero represents ‘no’ and one represents ‘yes’.

Graphical user interface, text

Description automatically generated

## Linear Regression with AWS \_ Task 4

1. Firstly, we import all the libraries needed for regression.



1. Now, we load the data from the files that we have uploaded to the AWS S3 bucket into a panda data frame. Here the object URL of the file is added. The size of the data frame is printed.

Text

Description automatically generated

1. Next, we clean the data. If there are any unnecessary columns or rows, we eliminate them.

Rectangle

Description automatically generated with low confidence

1. Next, we create training and testing data by using the train\_test\_split() function. We are running linear regression and fitting the training data into the linear regression model. We predict the values for the test data. Here, we are using real estate data to predict trends in the real estate data based on the input variables like transaction data, house age, distance to MRT, number of convenience stores, latitude, and longitude.

Graphical user interface, text, application

Description automatically generated

1. Finally, we evaluate the model and print the accuracy of the model. Here, we use r2\_score to calculate accuracy.

Application

Description automatically generated with medium confidence

1. Now make predict the outputs for the inputs that are loaded by reading the test file from the amazon S3 bucket. We print the predicted values. Here we predict the house price per unit based on the inputs.

Graphical user interface, application

Description automatically generated