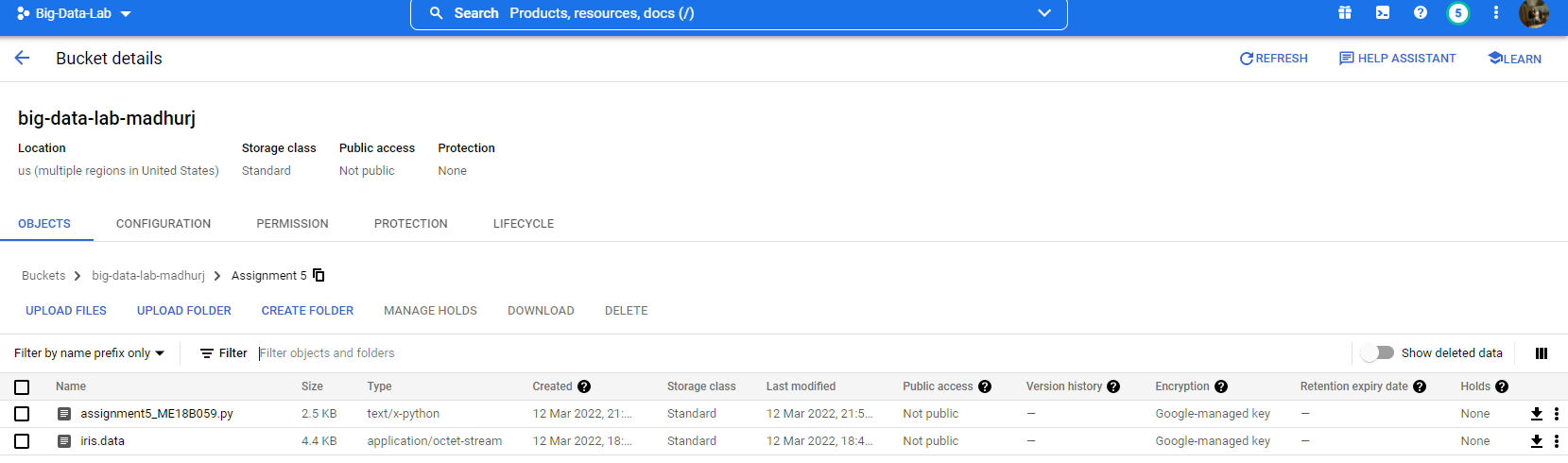
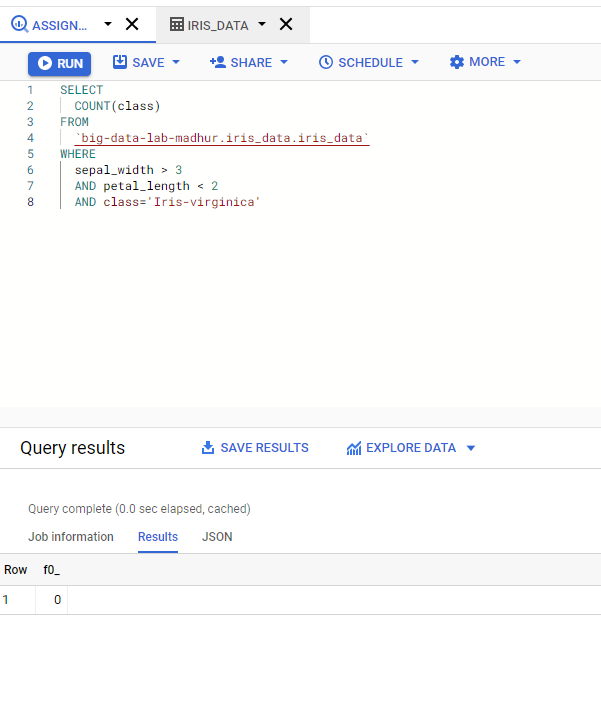
1. PFA the screenshot to location of the data file and the code for task 3 in the bucket.



2. PFA the screenshot to the query for task 2 along with the answer as 0. Hence there do not exist any flowers of the given configuration - Iris Virginica flowers which have sepal width greater than 3 cm and petal length smaller than 2 cm.



3. The code for the Task 3 is given in the zip named as assignment5\_ME18B059.py file. We try to use different classification models along with different pre-processing techniques on the given IRIS dataset and try to predict the class of the iris plant.

1. We start with importing the required libraries and setting up the Spark Context for connection with spark and also setting up the Spark Session using sc for setting up the connection with sql engine.
2. Then we read the table using spark sql from our saved tables as an rdd.
3. Then we transform the data in the class column to encoded format and also use Vector.dense to transform the other columns to a features column. Finally, we create a Spark DataFrame from the rdd after applying the transform.
4. We use randomSplit method to create a 4:1 train test split and use MultiClassClassificationEvaluator for calculating accuracy.
5. Following that we try training different models with no pre-processing on the raw data and also using Standard Scaler for transformation.
6. The different models that were trained are:
   1. Logistic Regression
   2. Decision Tree Classifier
   3. Random Forest Classifier
   4. Naïve Bayes Classifier
7. The training and test accuracies for the combinations have been provided.



