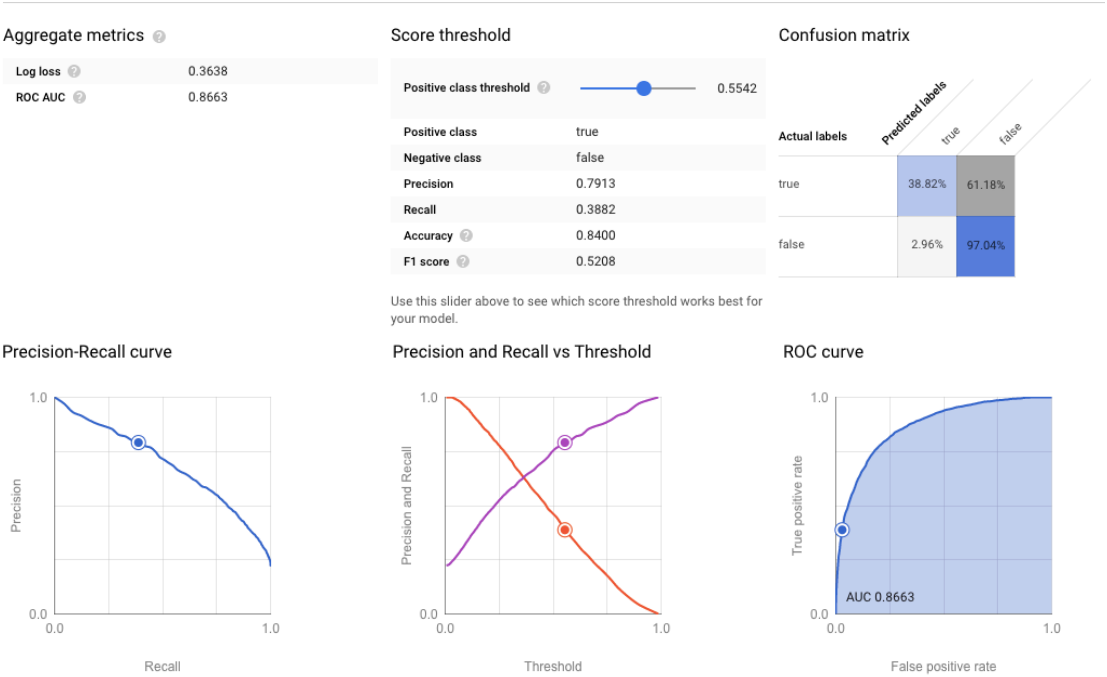
PREDICTING RAINFALL IN AUSTRALIA
MACHINE LEARNING USING BIGQUERY

- We used BigQuery to train a model, evaluate a model and make predictions.
- The data was loaded into BigQuery using command shell on Google Cloud Platform.
- The project id is *thematic-flash-266714*, dataset is *australia_weather* and the table is weather.
- The data was pre-processed, and EDA was performed using Python 3 in Jupyter Notebooks on AI Platform.
- After preprocessing the data, the data is loaded onto BigQuery and the name of the new table is *preprocessed_new*.
- We train the Logistic Regression model using data from 2011 to 2019 which constitutes about 78% of data. The model is named *log_reg_model*. The data forms the training set for the model.
- The model is then used to predict for data before 2011. This forms the testing set for the model.
- Then the model is evaluated and found to have an accuracy of 0.85.

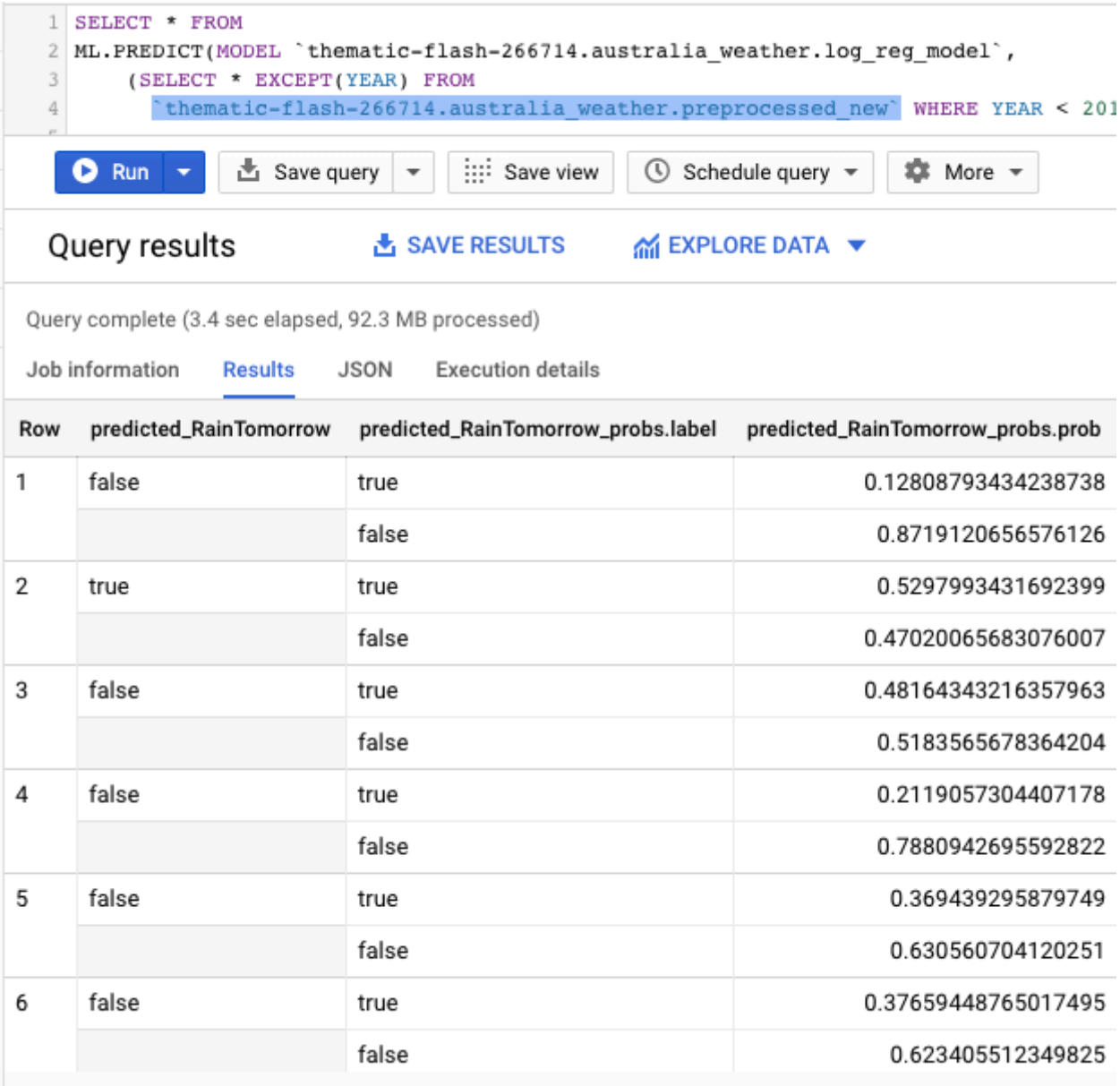
1. Query used to create the model is :

```
CREATE OR REPLACE MODEL `thematic-flash-266714.australia_weather.log_reg_model`
OPTIONS
( model_type="logistic_reg",
input_label_cols=["RainTomorrow"] ) AS SELECT * EXCEPT(YEAR) FROM
`thematic-flash-266714.australia_weather.preprocessed_new` WHERE YEAR >= 2011
```



2. Query used to predict on test data using the model is :

```
SELECT * FROM
ML.PREDICT(MODEL `thematic-flash-266714.australia_weather.log_reg_model`,
  (SELECT * EXCEPT(YEAR) FROM
    `thematic-flash-266714.australia_weather.preprocessed_new` WHERE YEAR < 2011 ))
```



```
1 SELECT * FROM
2 ML.PREDICT(MODEL `thematic-flash-266714.australia_weather.log_reg_model`,
3     (SELECT * EXCEPT(YEAR) FROM
4       `thematic-flash-266714.australia_weather.preprocessed_new` WHERE YEAR < 201
```

Run | Save query | Save view | Schedule query | More

Query results | SAVE RESULTS | EXPLORE DATA

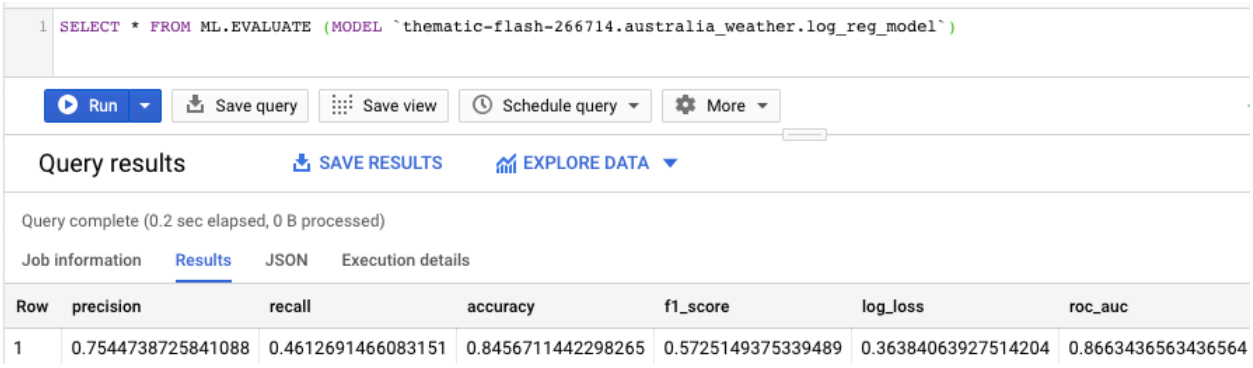
Query complete (3.4 sec elapsed, 92.3 MB processed)

Job information | Results | JSON | Execution details

Row	predicted_RainTomorrow	predicted_RainTomorrow_probs.label	predicted_RainTomorrow_probs.prob
1	false	true	0.12808793434238738
		false	0.8719120656576126
2	true	true	0.5297993431692399
		false	0.47020065683076007
3	false	true	0.48164343216357963
		false	0.5183565678364204
4	false	true	0.2119057304407178
		false	0.7880942695592822
5	false	true	0.369439295879749
		false	0.630560704120251
6	false	true	0.37659448765017495
		false	0.623405512349825

3. Query used to evaluate the logistic regression model. :

SELECT * FROM
ML.EVALUATE(MODEL `thematic-flash-266714.australia_weather.log_reg_model`,



Row	precision	recall	accuracy	f1_score	log_loss	roc_auc
1	0.7544738725841088	0.4612691466083151	0.8456711442298265	0.5725149375339489	0.36384063927514204	0.8663436563436564