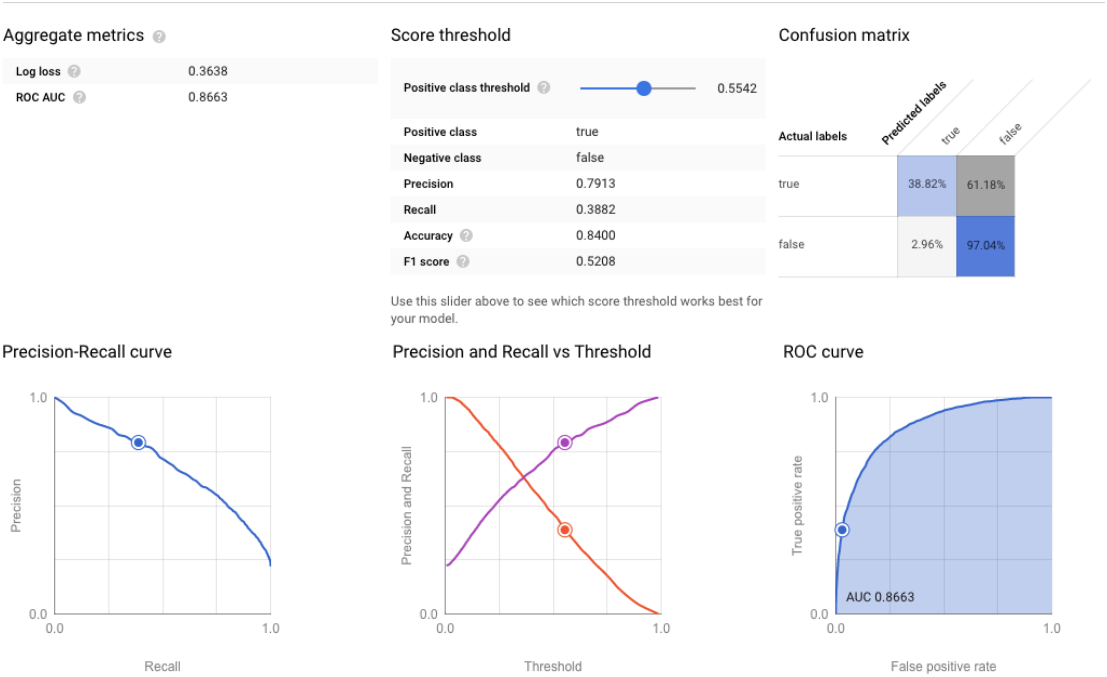
PREDICTING RAINFALL IN AUSTRALIA
MACHINE LEARNING USING BIGQUERY

- 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.

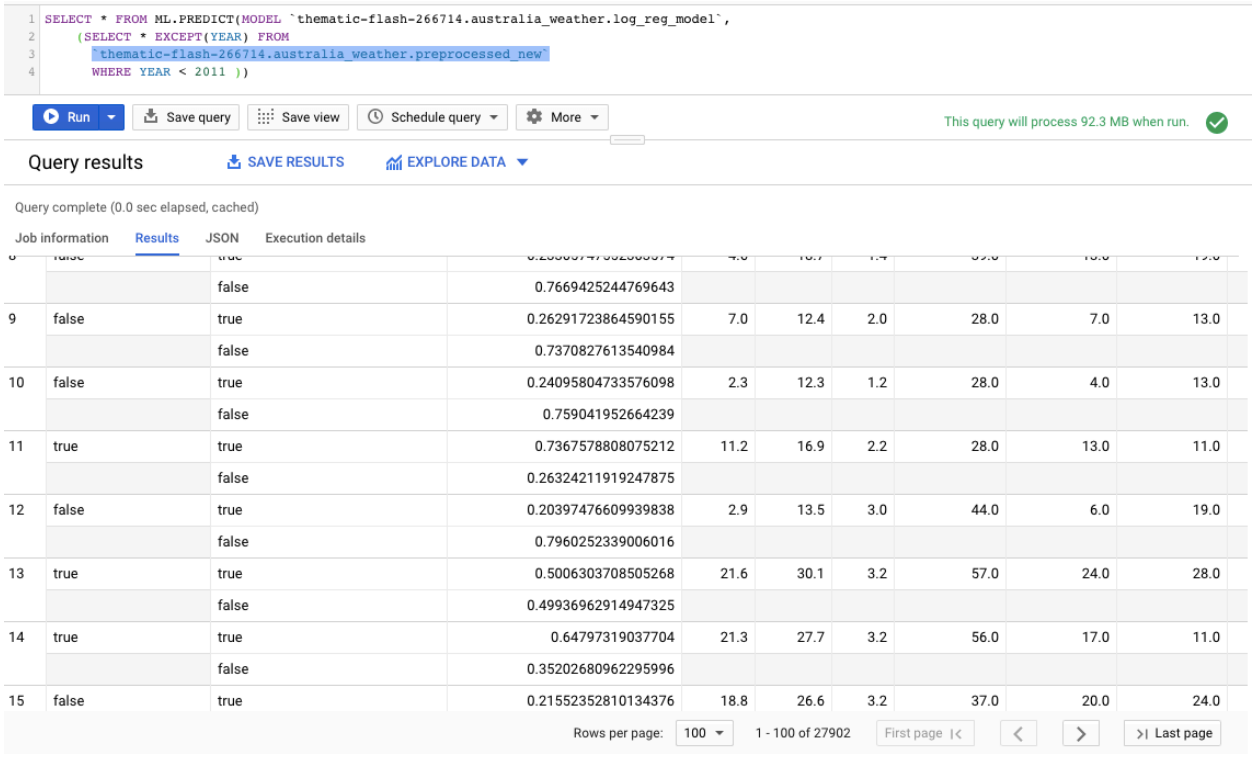
1. Query used to create the model is :

```
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 ))
```



```
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 ))
```

Run | Save query | Save view | Schedule query | More — This query will process 92.3 MB when run.

Query results — SAVE RESULTS — EXPLORE DATA

Query complete (0.0 sec elapsed, cached)

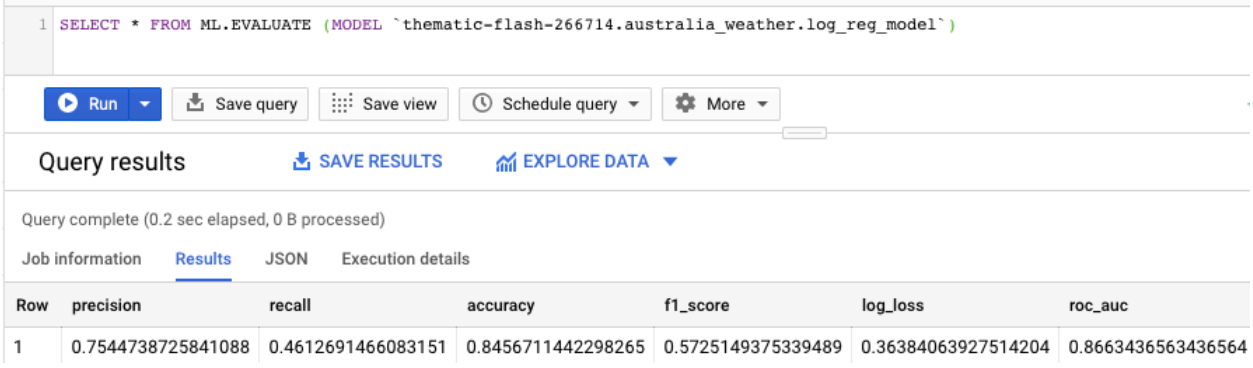
Job information | Results | JSON | Execution details

		false	0.7669425244769643						
9	false	true	0.26291723864590155	7.0	12.4	2.0	28.0	7.0	13.0
		false	0.7370827613540984						
10	false	true	0.24095804733576098	2.3	12.3	1.2	28.0	4.0	13.0
		false	0.759041952664239						
11	true	true	0.7367578808075212	11.2	16.9	2.2	28.0	13.0	11.0
		false	0.26324211919247875						
12	false	true	0.20397476609939838	2.9	13.5	3.0	44.0	6.0	19.0
		false	0.7960252339006016						
13	true	true	0.5006303708505268	21.6	30.1	3.2	57.0	24.0	28.0
		false	0.49936962914947325						
14	true	true	0.64797319037704	21.3	27.7	3.2	56.0	17.0	11.0
		false	0.35202680962295996						
15	false	true	0.21552352810134376	18.8	26.6	3.2	37.0	20.0	24.0

Rows per page: 100 — 1 - 100 of 27902 — First page | Last page

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

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



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

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

Query results — SAVE RESULTS — EXPLORE DATA

Query complete (0.2 sec elapsed, 0 B processed)

Job information | Results | JSON | Execution details

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