45 minutesFree

Getting Started with BQML

45 minutesFree
Rate Lab

GSP247

Google Cloud Self-Paced Labs

Overview

[BigQuery Machine Learning](#) (BQML, product in beta) enables users to create and execute machine learning models in BigQuery using SQL queries. The goal is to democratise machine learning by enabling SQL practitioners to build models using their existing tools and to increase development speed by eliminating the need for data movement.

There is a newly available [ecommerce dataset](#) that has millions of Google Analytics records for the [Google Merchandise Store](#) loaded into BigQuery. In this lab you will use this data to create a model that predicts whether a visitor will make a transaction.

What you'll learn

How to create, evaluate and use machine learning models in BigQuery

What you'll need

- A Browser, such as [Chrome](#) or [Firefox](#)

- Basic knowledge of SQL or BigQuery

Setup and requirements

Qwiklabs setup

Before you click the Start Lab button
Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need
To complete this lab, you need:

- Access to a standard internet browser (Chrome browser recommended).
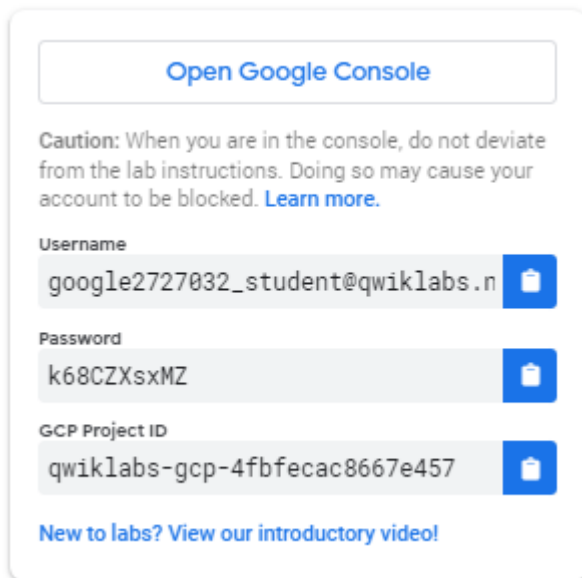- Time to complete the lab.
  Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

  Note: If you are using a Chrome OS device, open an Incognito window to run this lab.
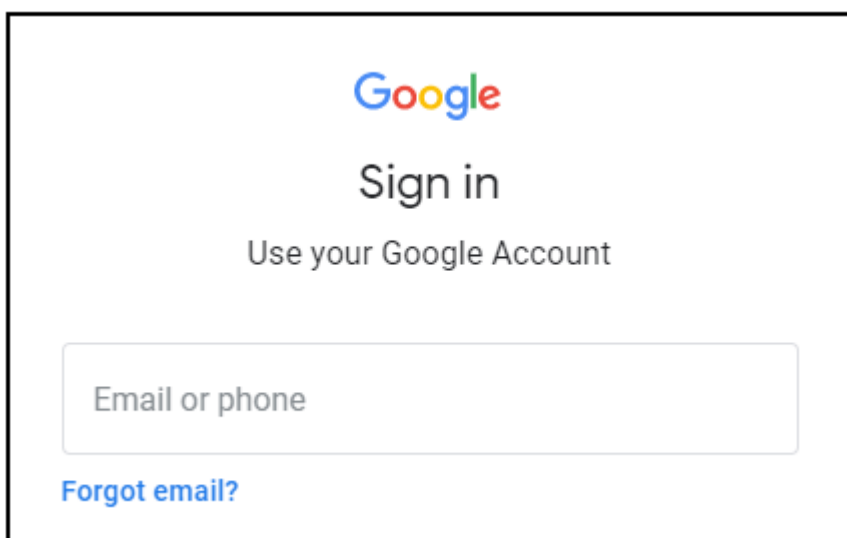
Cloud Console

How to start your lab and sign in to the Google Cloud Console

1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.
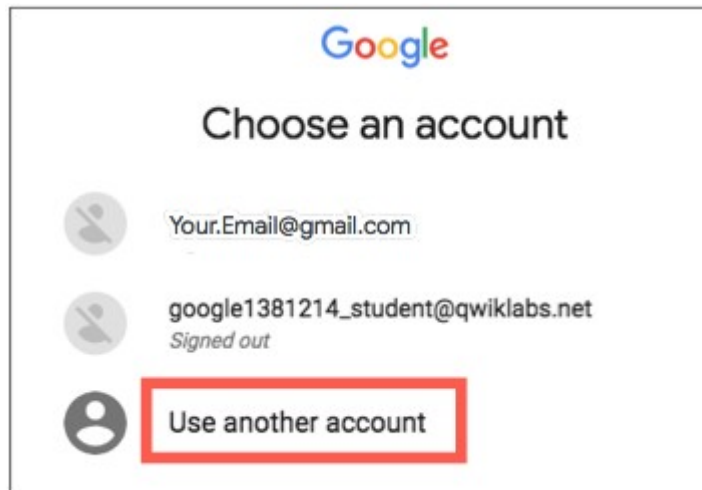

2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.

Tip: Open the tabs in separate windows, side-by-side.

Choose an account page, click **Use Another**

Account.

3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

   Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

4. Click through the subsequent pages:

   - Accept the terms and conditions.
   - Do not add recovery options or two-factor authentication (because this is a temporary account).
   - Do not sign up for free trials.

   After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-

left.
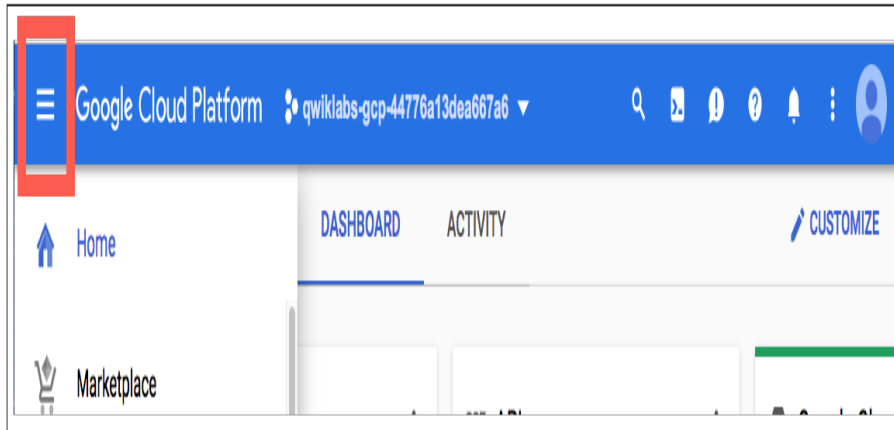
Open BigQuery Console

In the Google Cloud Console, select **Navigation menu** > **BigQuery**:
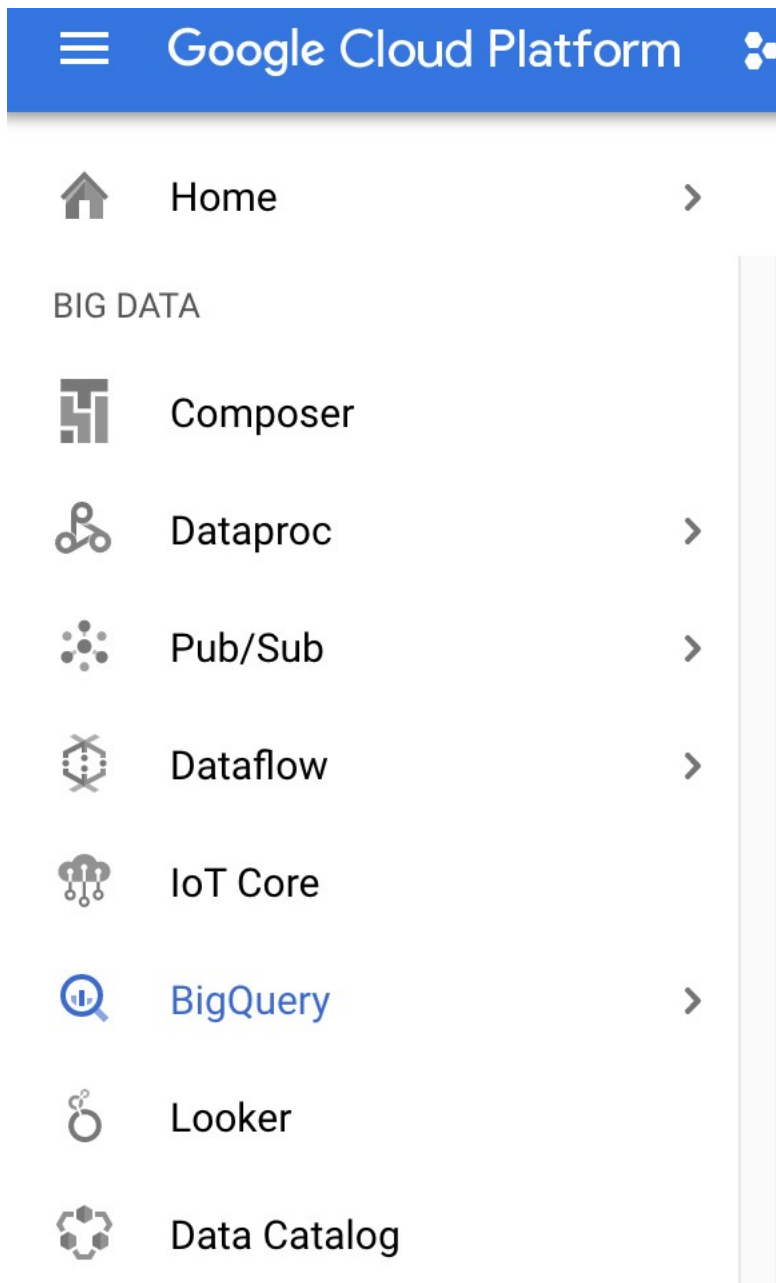
The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and the release notes.

Click **Done**.

The BigQuery console opens.

Explorer   + ADD DATA

🔍 EDITOR ▾ ✕   ⊞ COMPOSE NEW QUERY

▶ RUN   💾 SAVE ▾   🕐 SCHEDULE   ⚙ MORE ▾   ❶ Type a query to get started

🔍 Type to search   ❷

1

Viewing pinned projects.

▸ ⠿ qwiklabs-gcp-01-1e4f9c41ff3f

JOB HISTORY   QUERY HISTORY   SAVED QUERIES   ⌄

Job history   ↻ REFRESH

Personal history Project history

No queries yet

Compose a query to get started
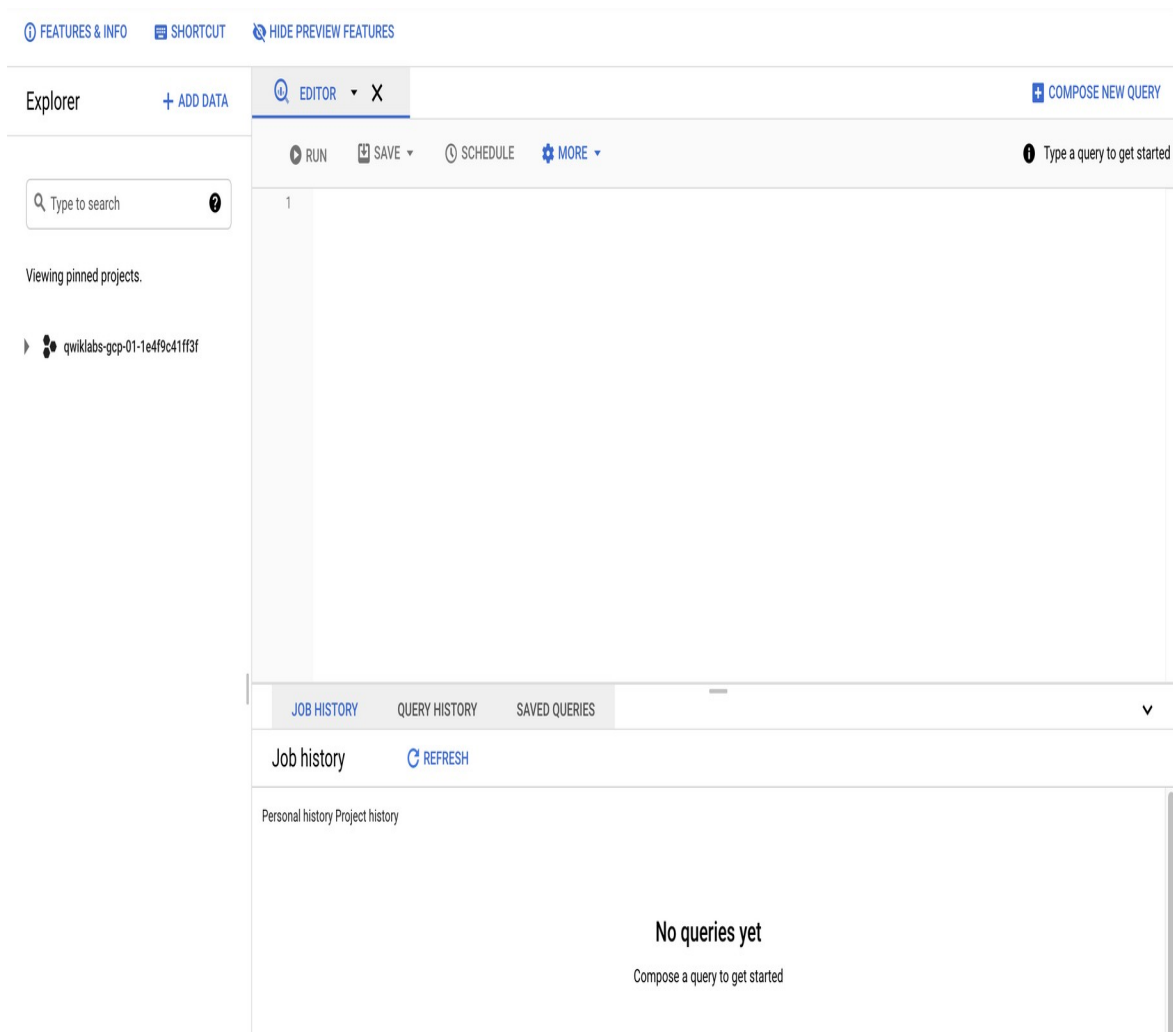
Create a dataset

To create a dataset, click on the **View actions** icon next to your project ID and select **Create dataset**
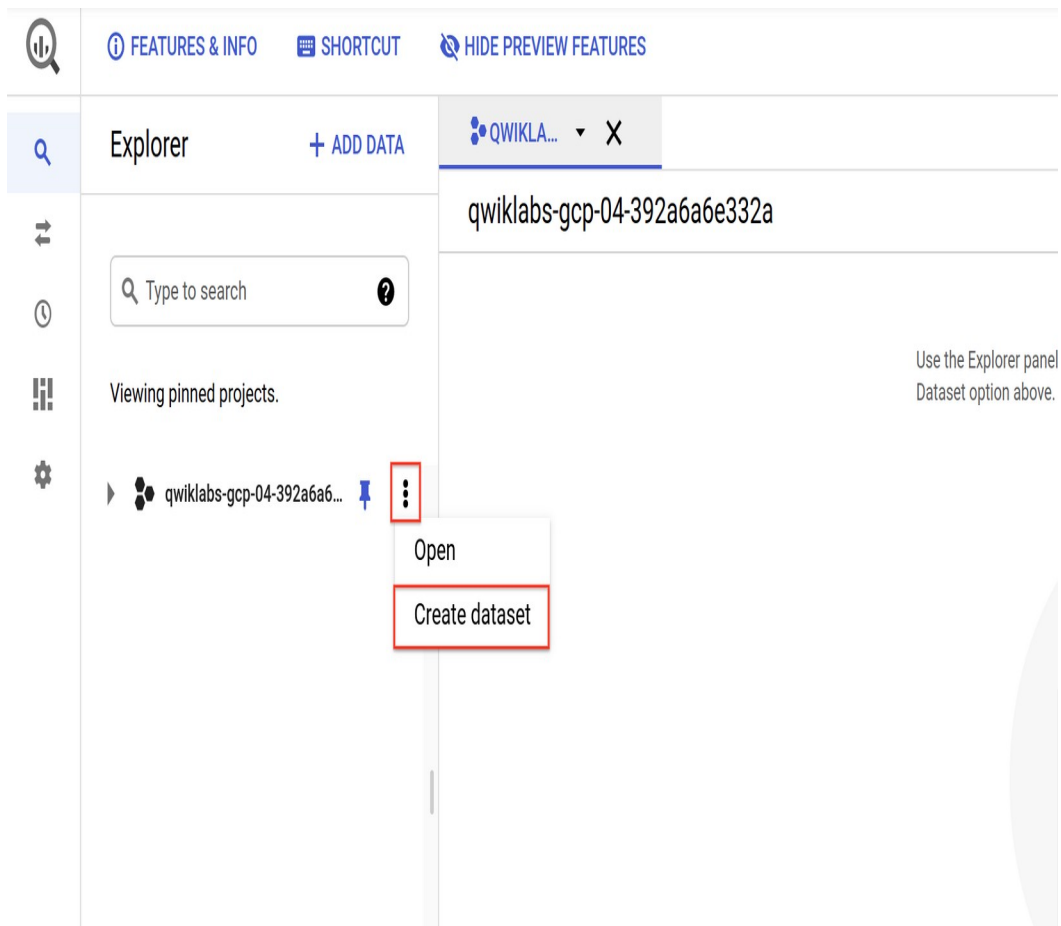
Next, name your Dataset ID **bqml_lab** and click **Create dataset**.

Create dataset

Dataset ID *

bqml_lab

Letters, numbers, and underscores allowed

Data location

Default ▼ ❓

Default table expiration

☐ Enable table expiration ❓

Default maximum table age   Days

Encryption

🔘 Google-managed encryption key
No configuration required

⚪ Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service

CREATE DATASET   CANCEL

Test Completed Task

Click **Check my progress** to verify your performed task. If you have completed the task successfully you will be granted with an assessment score.

Create a model

Now, move on to your task!

Go to BigQuery **EDITOR**, type or paste the following query to create a model that predicts whether a visitor will make a transaction:

```sql
#standardSQL
CREATE OR REPLACE MODEL `bqml_lab.sample_model`
OPTIONS(model_type='logistic_reg') AS
SELECT
  IF(totals.transactions IS NULL, 0, 1) AS label,
  IFNULL(device.operatingSystem, "") AS os,
  device.isMobile AS is_mobile,
  IFNULL(geoNetwork.country, "") AS country,
  IFNULL(totals.pageviews, 0) AS pageviews
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  _TABLE_SUFFIX BETWEEN '20160801' AND '20170631'
LIMIT 100000;content_copy
```
Click **RUN**.

Here the visitor's device's operating system is used, whether said device is a mobile device, the visitor's country and the number of page views as the criteria for whether a transaction has been made.

In this case, bqml_lab is the name of the dataset and sample_model is the name of the model. The model type specified is binary logistic regression. In this case, label is what you're trying to fit to.

Note: If you're only interested in 1 column, this is an alternative way to setting input_label_cols.
The training data is being limited to those collected from 1 August 2016 to 30 June 2017. This is done to save the last month of data for "prediction". It is further limited to 100,000 data points to save some time.

Running the CREATE MODEL command creates a Query Job that will run asynchronously so you can, for example, close or refresh the BigQuery UI window.

Test Completed Task

Click **Check my progress** to verify your performed task. If you have completed the task successfully you will be granted with an assessment score.

(Optional) Model information & training statistics

If interested, you can get information about the model by expanding bqml_lab dataset and then clicking the sample_model model in the UI. Under the **Details** tab you should find some basic model info and training options used to produce the model. Under **Training**, you should see a table either a table or graphs, depending on your View as settings:

sample_model

QUERY MODEL   DELETE MODEL   EXPORT MODEL

DETAILS   **TRAINING**   EVALUATION   SCHEMA

View as

◯ Graphs

◉ Table

Iteration	Training Data Loss	Evaluation Data Loss	Learn Rate	Duration (seconds)
10	0.0467	0.0342	25.6	4.63
9	0.0470	0.0343	12.8	4.70
8	0.0475	0.0350	25.6	5.31
7	0.0482	0.0354	25.6	5.03
6	0.0511	0.0393	12.8	5.05
5	0.0583	0.0471	6.4	6.10
4	0.0724	0.0624	3.2	6.96
3	0.1017	0.0934	1.6	5.93
2	0.1732	0.1673	0.8	6.01
1	0.3231	0.3197	0.4	6.33
0	0.5227	0.5214	0.2	5.04

Rows per page: 50 ▼   1 – 11 of 11   ‹   ›

sample_model

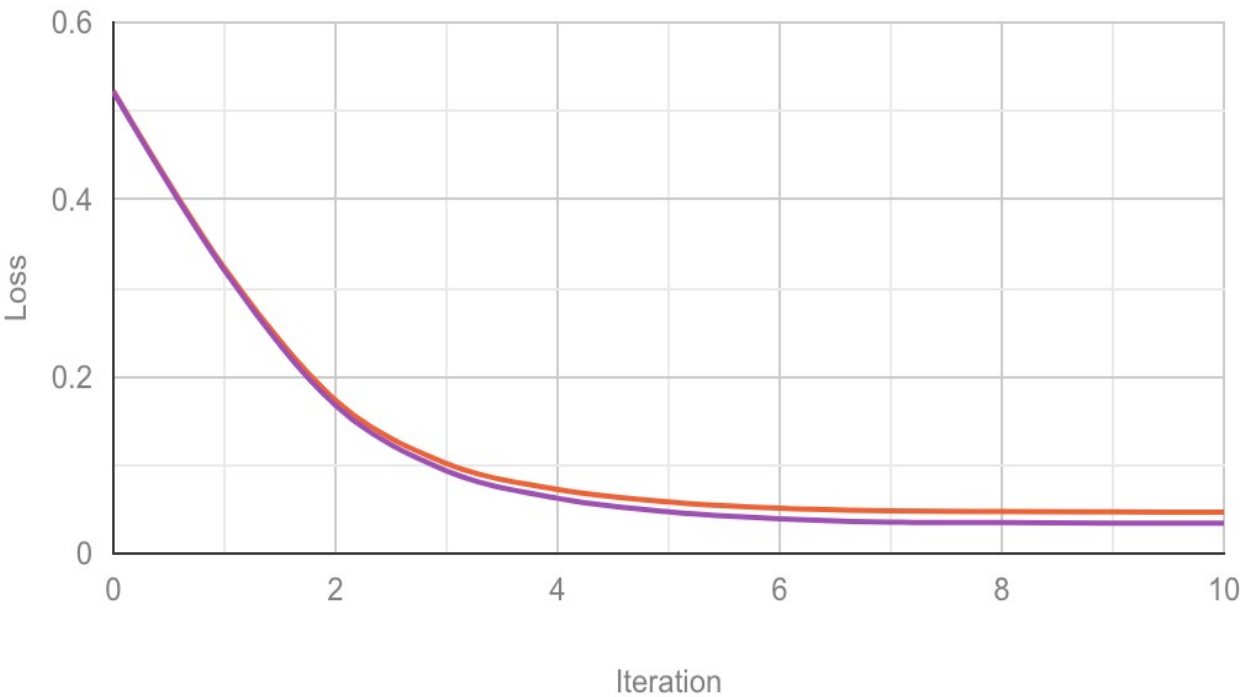**View as**

◉ Graphs

◯ Table

Loss

Evaluate the model

Replace the previous query with the following and then click **Run:**

```sql
#standardSQL
SELECT
  *
FROM
  ml.EVALUATE(MODEL `bqml_lab.sample_model`, (
SELECT
  IF(totals.transactions IS NULL, 0, 1) AS label,
  IFNULL(device.operatingSystem, "") AS os,
  device.isMobile AS is_mobile,
  IFNULL(geoNetwork.country, "") AS country,
  IFNULL(totals.pageviews, 0) AS pageviews
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  _TABLE_SUFFIX BETWEEN '20170701' AND '20170801'));content_copy
```

If used with a linear regression model, the above query returns the following

columns:

- mean_absolute_error, mean_squared_error, mean_squared_log_error,
- median_absolute_error, r2_score, explained_variance.
  If used with a logistic regression model, the above query returns the following

  columns:

- precision, recall
- accuracy, f1_score
- log_loss, roc_auc
  Please consult the machine learning glossary or run a Google search to

  understand how each of these metrics are calculated and what they mean.

You'll realize the SELECT and FROM portions of the query is identical to that used during training. The WHERE portion reflects the change in time frame and the FROM portion shows that you're calling ml.EVALUATE.

You should see a table similar to this:

Row	precision	recall	accuracy	f1_score	log_loss	roc_auc
1	0.47368421052631576	0.10893854748603352	0.9853834982788297	0.17713853141559424	0.04552280390355375	0.9773986013986014

Test Completed Task

Click **Check my progress** to verify your performed task. If you have completed the task successfully you will be granted with an assessment score.

Use the Model

Predict purchases per country

With this query you will try to predict the number of transactions made by visitors of each country, sort the results, and select the top 10 countries by purchases:

Replace the previous query with the following and then click **Run:**

```
#standardSQL
SELECT
  country,
  SUM(predicted_label) as total_predicted_purchases
FROM
```

```
    ml.PREDICT(MODEL `bqml_lab.sample_model`, (
SELECT
    IFNULL(device.operatingSystem, "") AS os,
    device.isMobile AS is_mobile,
    IFNULL(totals.pageviews, 0) AS pageviews,
    IFNULL(geoNetwork.country, "") AS country
FROM
    `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
    _TABLE_SUFFIX BETWEEN '20170701' AND '20170801'))
GROUP BY country
ORDER BY total_predicted_purchases DESC
LIMIT 10;content_copy
```

This query is very similar to the evaluation query demonstrated in the previous section. Instead of ml.EVALUATE, you're using ml.PREDICT and the BQML portion of the query is wrapped with standard SQL commands. For this lab you're interested in the country and the sum of purchases for each country, so that's why SELECT, GROUP BY and ORDER BY. LIMIT is used to ensure you only get the top 10 results.

You should see a table similar to this:

Row	country	total_predicted_purchases
1	United States	215
2	Canada	6
3	Taiwan	6
4	Singapore	2
5	Japan	2
6	India	2
7	Turkey	2
8	Indonesia	1
9	Venezuela	1
10	St. Lucia	1

Test Completed Task

Click **Check my progress** to verify your performed task. If you have completed the task successfully you will be granted with an assessment score.

Predict purchases per user

Here is another example. This time you will try to predict the number of transactions each visitor makes, sort the results, and select the top 10 visitors by transactions:

Replace the previous query with the following and then click **Run:**

```
#standardSQL
SELECT
  fullVisitorId,
  SUM(predicted_label) as total_predicted_purchases
FROM
  ml.PREDICT(MODEL `bqml_lab.sample_model`, (
SELECT
  IFNULL(device.operatingSystem, "") AS os,
  device.isMobile AS is_mobile,
  IFNULL(totals.pageviews, 0) AS pageviews,
  IFNULL(geoNetwork.country, "") AS country,
  fullVisitorId
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  _TABLE_SUFFIX BETWEEN '20170701' AND '20170801'))
GROUP BY fullVisitorId
ORDER BY total_predicted_purchases DESC
LIMIT 10;content_copy
```
You should see a table similar to this:

Row	fullVisitorId	total_predicted_purchases
1	9417857471295131045	4
2	112288330928895942	2
3	057693500927581077	2
4	2105122376016897629	2
5	7420300501523012460	2
6	0456807427403774085	2
7	2158257269735455737	2
8	489038402765684003	2
9	8388931032955052746	2
10	5073919761051630191	2

Test Completed Task

Click **Check my progress** to verify your performed task. If you have completed the task successfully you will be granted with an assessment score.

Test your Understanding

Below are multiple choice questions to reinforce your understanding of this lab's concepts. Answer them to the best of your abilities.

True

Command line tool
BigQuery REST API
Web UI

Congratulations!

This concludes the self-paced lab, Getting Started with BQML. You created a binary logistic regression model, evaluated the model, and used the model to make predictions.

Finish your Quest

This self-paced lab is part of the Qwiklabs Quest BigQuery for Machine Learning. A Quest is a series of related labs that form a learning path. Completing this Quest earns you the badge above, to recognize your achievement. You can make your badge (or badges) public and link to them in your online resume or social media account. Enroll in this Quest and get immediate completion credit if you've taken this lab. See other available Qwiklabs Quests.

Next steps / learn more

- For more information on BQML, see the [documentation](#).
- [Getting Started with BigQuery ML for Data Scientists](#)
- [Getting Started with BigQuery ML for Data Analysts](#)
- Already have a Google Analytics account and want to query your own datasets in BigQuery? Follow this [export guide](#).
- The complete BigQuery SQL reference guide is here as an additional resource: https://cloud.google.com/bigquery/docs/reference/standard-sql/query-syntax

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. [Our classes](#) include technical skills and best practices to help you get up to speed quickly and continue your learning journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. [Certifications](#) help you validate and prove your skill and expertise in Google Cloud technologies.