[ML on GCP C6] Explore dataset

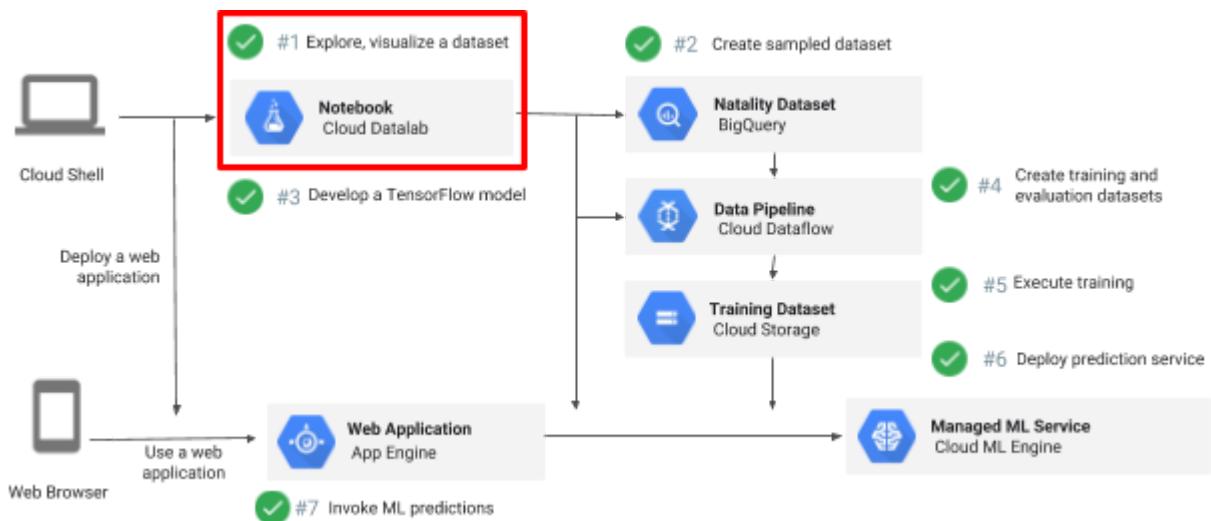2 hours   Free   Rate Lab

Overview

Duration is 1 min

This lab is part of a lab series where you train, evaluate, and deploy a machine learning model to predict a baby's weight.

In this lab #1, you explore and visualize a BigQuery dataset.

What you learn

In this lab, you will learn to:

- Explore the natality dataset in BigQuery using Datalab

- Use Python package to execute the query and convert the result into a Pandas dataframe for visualizations

Setup

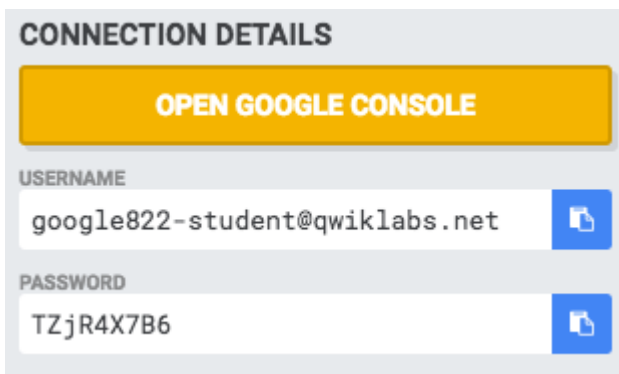For each lab, you get a new GCP project and set of resources for a fixed time at no cost.

1. Make sure you signed into Qwiklabs using an **incognito window**.

2. Note the lab's access time (for example, 02:00:00 and make sure you can finish in that time block.

   There is no pause feature. You can restart if needed, but you have to start at the beginning.

3. When ready, click START LAB .

4. Note your lab credentials. You will use them to sign in to Cloud Platform Console.

CONNECTION DETAILS

OPEN GOOGLE CONSOLE

USERNAME
google822-student@qwiklabs.net

PASSWORD
TZjR4X7B6

5. Click **Open Google Console**.

6. Click **Use another account** and copy/paste credentials for **this**lab into the prompts.

If you use other credentials, you'll get errors or **incur charges**.

7. Accept the terms and skip the recovery resource page.

Do not click **End** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

Create Storage Bucket

Duration is 2 min

Create a bucket using the GCP console:

Step 1

In your GCP Console, click on the **Navigation menu** (≡), and

select **Storage**.

Step 2

Click on **Create bucket**.

Step 3

Choose a Regional bucket and set a unique name (use your project ID because it is unique). Then, click **Create**.

Launch Cloud Datalab

To launch Cloud Datalab:

Step 1

Open Cloud Shell. The Cloud Shell icon is at the top right of the Google Cloud Platform [web console](#).

Step 2

In Cloud Shell, type:

```
gcloud compute zones list
```

> **Note**: Please pick a zone in a geographically close region from the following: **us-east1, us-central1, asia-east1, europe-west1**. These are the regions that currently support Cloud ML Engine jobs. Please verify [here](#)since this list may have changed after this lab was last updated. For example, if you are in the US, you may choose **us-east1-c** as your zone.

Step 3

In Cloud Shell, type:

```
datalab create mydatalabvm --zone <ZONE>
```

> Replace with a zone name you picked from the previous step.

> **Note**: follow the prompts during this process.

Datalab will take about 5 minutes to start.

Step 4

Look back at Cloud Shell and follow any prompts. If asked for an ssh passphrase, hit return (for no passphrase).
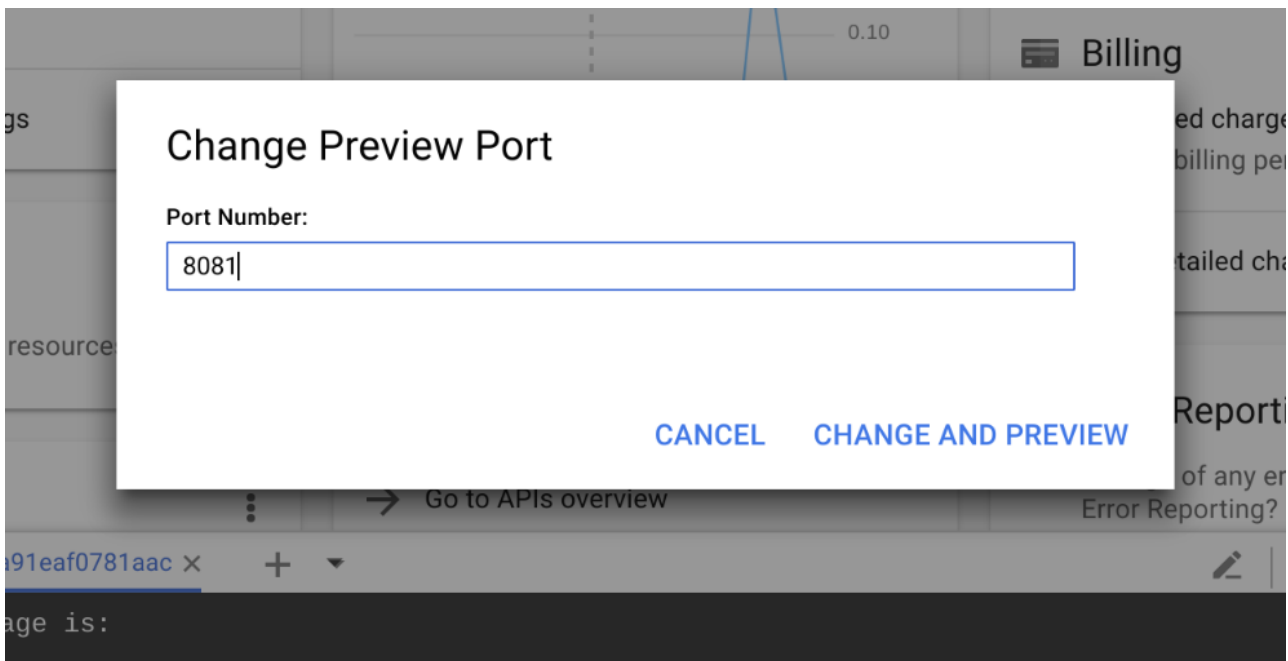
Step 5

If necessary, wait for Datalab to finishing launching. Datalab is ready when you see a message prompting you to do a **Web Preview**.

Step 6

Click on **Web Preview** icon on the top-right corner of the Cloud Shell ribbon.
Click **Change Port** and enter the port **8081** and click **Change and Preview**.
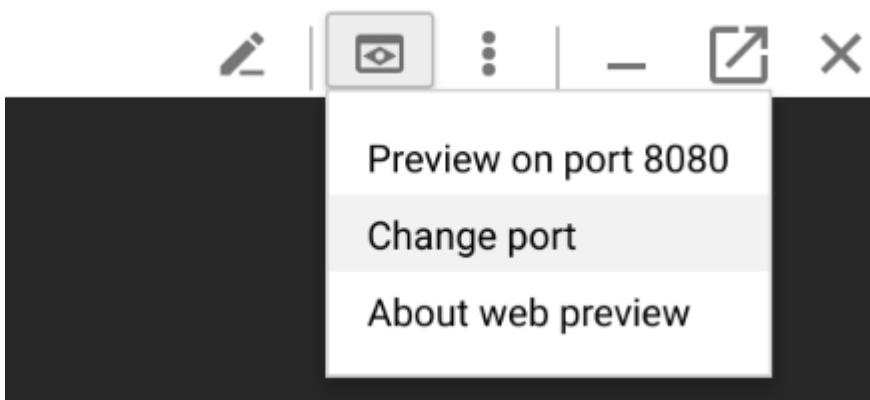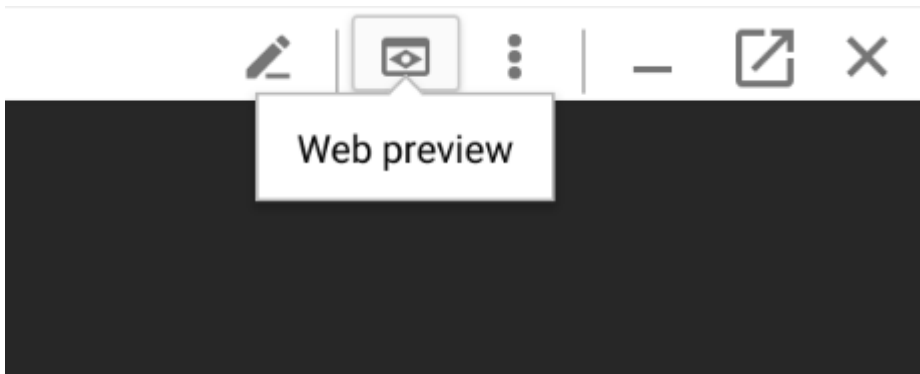
Note: If the cloud shell used for running the datalab command is closed or interrupted, the connection to your Cloud Datalab VM will terminate. If that happens, you may be able to reconnect using the command **datalab connect mydatalabvm** in your new Cloud Shell.

Clone course repo within your Datalab instance

To clone the course repo in your datalab instance:

Step 1

In Cloud Datalab home page (browser), navigate into **notebooks** and add a new notebook using the icon on the top left.
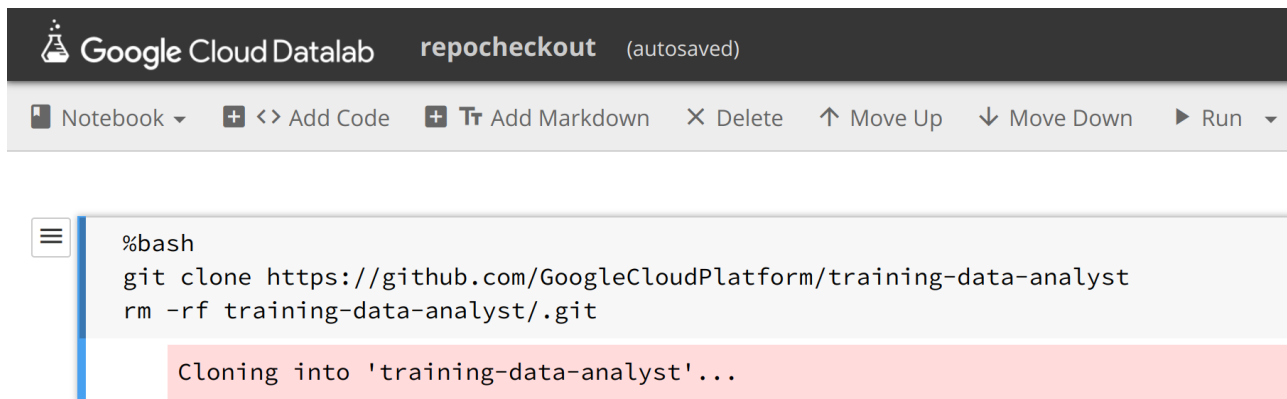
Step 2

Rename this notebook as **repocheckout**.

Step 3

In the new notebook, enter the following commands in the cell, and click on **Run** (on the top navigation bar) to run the commands:

```
%bash
git clone
https://github.com/GoogleC
data-analyst
rm -rf training-data-
analyst/.git
```


Step 4

Confirm that you have cloned the repo by going back to Datalab browser, and ensure you see the **training-data-analyst** directory. All the files for all labs throughout this course are available in this directory.
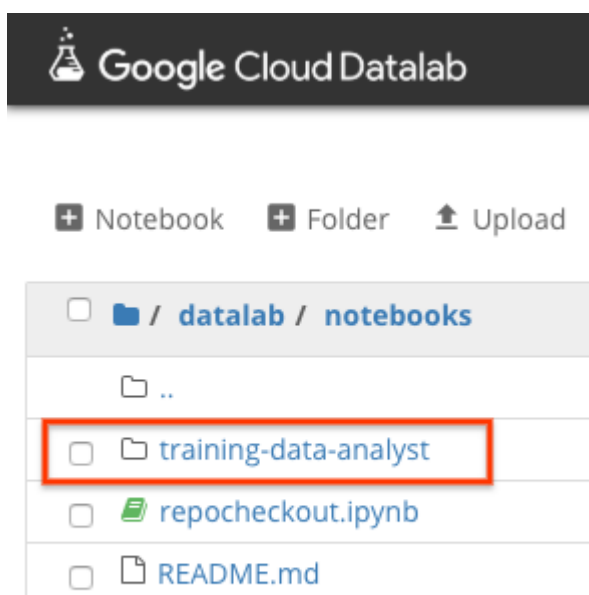
Explore dataset

Duration is 15 min

Explore a BigQuery dataset to find features to use in an ML model.

Step 1

In Cloud Datalab, click on the **Home** icon, and then navigate to **datalab > notebooks > training-data-analyst > courses > machine_learning > deepdive > 06_structured > labs** and open **1_explore.ipynb**.

> **Note:** If the cloud shell used for running the datalab command is closed or interrupted, the connection to your Cloud Datalab VM will terminate. If that happens, you may be able to reconnect using the command '**datalab connect mydatalabvm**' in your new Cloud Shell. Once connected, try the above step again.

Step 2

In Datalab, click on **Clear | Clear all Cells**. Now read the narrative and execute each cell in turn:

- If you notice sections marked "Lab Task", you will need to create a new code cell and write/complete code to achieve the task.
- Some lab tasks include starter code. In such cells, look for lines marked #TODO.
- Hints may also be provided for the tasks to guide you along. Highlight the text to read the hints (they are in white text).
- If you need more help, you may take a look at the complete solution by navigating to : **datalab > notebooks > training-data-analyst > courses > machine_learning > deepdive > 06_structured**and open **1_explore.ipynb**.

> **Note:** When doing copy/paste of python code, please be careful about indentation.

End your lab

When you have completed your lab, click **End Lab**. Qwiklabs removes the resources you've used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

- 1 star = Very dissatisfied
- 2 stars = Dissatisfied
- 3 stars = Neutral
- 4 stars = Satisfied
- 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.

Last Tested Date: 12-10-2018

Last Updated Date: 12-18-2018