End Lab 01:33:24

Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. Learn more.

Open Google Console

Username
student-04-685e271e8b8!

Password
dIcoobIq3G34

GCP Project ID
qwiklabs-gcp-00-841304(

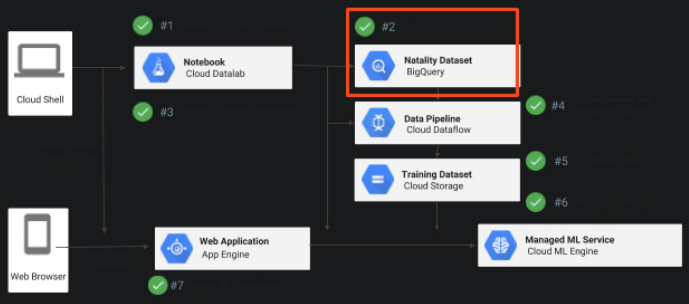
Create a sample dataset

2 hours Free ★★★★

Overview

In this lab we'll read data from BigQuery into our notebook to preprocess the data within a Pandas dataframe for a small, repeatable sample.

In this lab, you sample the full BigQuery dataset to create a smaller dataset for model development and local training. In the real world, it's much better to start out simple and develop your TensorFlow code locally on a small subset of data, then scale it out to the cloud. Developing on a smaller subset of data speeds up model development and makes debugging easier.



What you learn

In this lab, you will learn how to:

- Setup up the environment
- Sample the natality dataset to create train, eval, test sets
- Preprocess the data in Pandas dataframe

Setup

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

1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, 02:00:00), and make sure you can finish within that time. 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 (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.

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 receive errors or **incur charges**.

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

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

Create Storage Bucket

1. In the Google Cloud Console, on the **Navigation menu** ( ), click **Cloud Storage**.

2. Click **Create bucket**.

3. Type a unique name, such as your project ID.

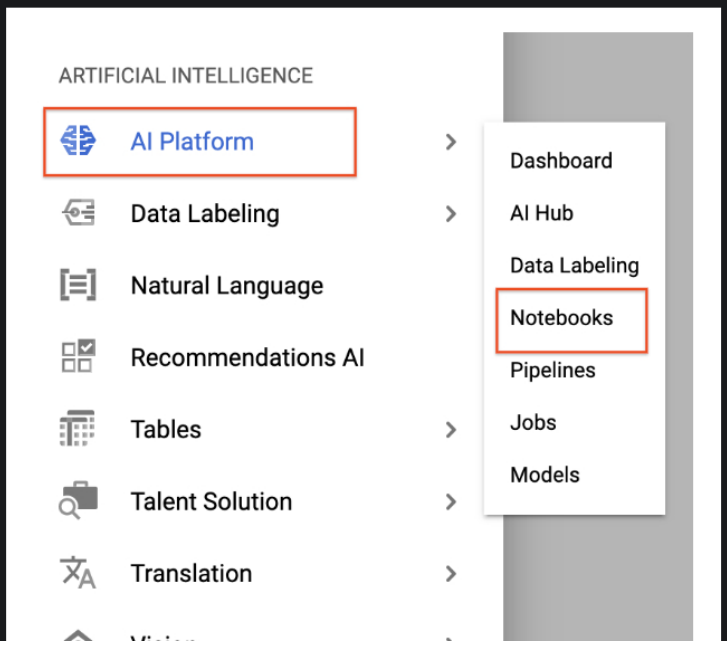
4. Click **Create**.

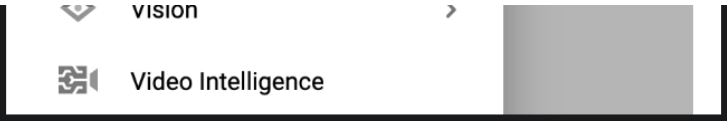
Launch AI Platform Notebooks

To launch AI Platform Notebooks:

Step 1

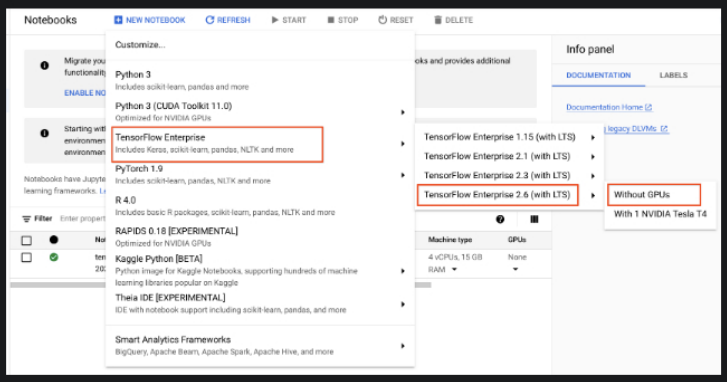
Click on the **Navigation Menu**. Navigate to **AI Platform**, then to **Notebooks**.



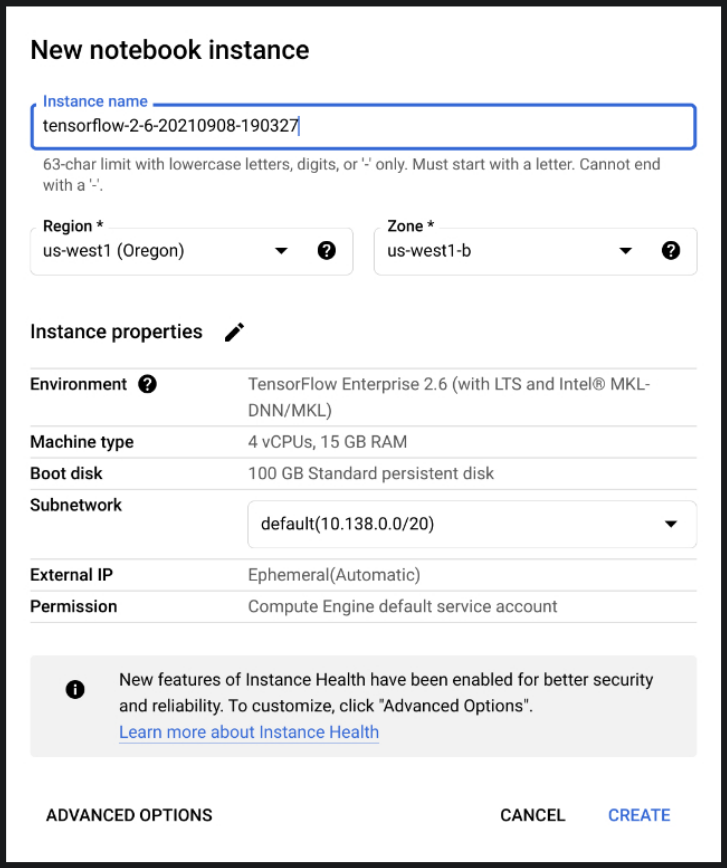


Step 2

On the Notebook instances page, click NEW NOTEBOOK. Select **TensorFlow Enterprise** and choose the latest version of **TensorFlow Enterprise 2.6 (with LTS)** > **Without GPUs**.



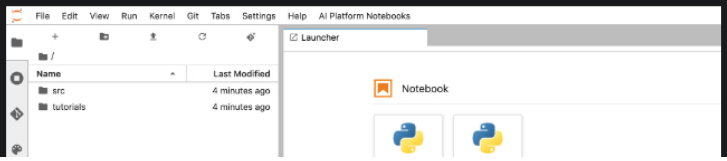
In the pop-up, confirm the name of the deep learning VM, move to the bottom of the window and click **Create**.

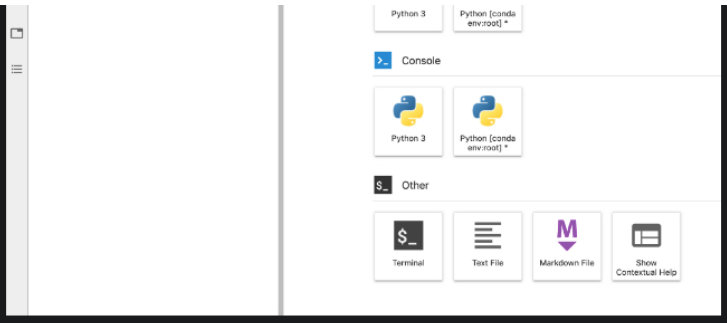


The new VM will take 2-3 minutes to start.

Step 3

Click **Open JupyterLab**. A JupyterLab window will open in a new tab.





Clone course repo within your AI Platform Notebooks instance

To clone the training-data-analyst notebook in your JupyterLab instance:

1. In JupyterLab, to open a new terminal, click the **Terminal** icon.
2. At the command-line prompt, run the following command:

```
git clone https://github.com/GoogleCloudPlatform/training-data-analyst
```

3. To confirm that you have cloned the repository, double-click on the training-data-analyst directory and ensure that you can see its contents. The files for all the Jupyter notebook-based labs throughout this course are available in this directory.

Create sampled dataset

Step 1

In the notebook interface, navigate to **training-data-analyst > courses > machine_learning > deepdive2 > end_to_end_ml > labs** and open **sample_babyweight.ipynb**.

Step 2

In the notebook interface, click on **Edit > Clear All Outputs** (click on Edit, then in the drop-down menu, select Clear All Outputs).

Carefully read through the notebook instructions and fill in lines marked with #TODO where you need to complete the code as needed

Tip: To run the current cell you can click the cell and hit **shift+enter**. Other cell commands are found in the notebook UI under **Run**.

- 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 **training-data-analyst > courses > machine_learning > deepdive2 > end_to_end_ml > solutions** and open **sample_babyweight.ipynb**.

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.