## **Overview**

In this lab, you go from exploring a taxicab dataset to training and deploying a high-accuracy distributed model with AI-Platform.

## **Objectives**

In these labs, you will perform the following tasks:

Lab 1: Explore dataset, create ML datasets, create benchmark

* Sample the dataset and create training, validation, and testing datasets for local development of TensorFlow models
* Create a benchmark to evaluate the performance of ML

Lab 2: Getting Started with TensorFlow

* Explore the TensorFlow Python API
* Building a graph
* Running a graph
* Feeding values into a graph
* Find area of a triangle using TensorFlow

Lab 3: Machine Learning using tf.estimator

* Read from Pandas Dataframe into tf.constant
* Create feature columns for estimator
* Linear Regression with tf.Estimator framework
* Deep Neural Network regression
* Benchmark dataset

Lab 4: Refactoring to add batching and feature-creation

* Refactor the input
* Refactor the way the features are created
* Create and train the model
* Evaluate model

Lab 5: Distributed training and monitoring

* Create features out of input data
* Train and evaluate
* Monitor with Tensorboard

Lab 6: Scaling up ML using Cloud ML Engine

* Package up the code
* Find absolute paths to data
* Run the Python module from the command line
* Run locally using gcloud
* Submit training job using gcloud
* Deploy model
* Prediction
* Train on a larger dataset
* 1-million row dataset

Lab 7: Feature Engineering

* Working with feature columns
* Adding feature crosses in TensorFlow
* Reading data from BigQuery
* Creating datasets using Dataflow
* Using a wide-and-deep model

Lab 8: Demonstration of Hyper-Parameter Tuning and Training

## **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, img/time.png 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.

1. When ready, click img/start_lab.png.
2. Note your lab credentials. You will use them to sign in to Cloud Platform Console. 
3. Click **Open Google Console**.
4. 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**.

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

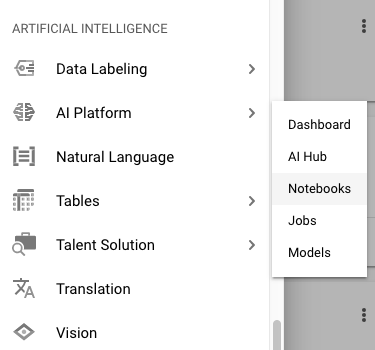
Do not click **End Lab** unless you are finished with the lab or want to restart it. This clears your work and removes the project.

## **Launch AI Platform Notebooks**

To launch AI Platform Notebooks:

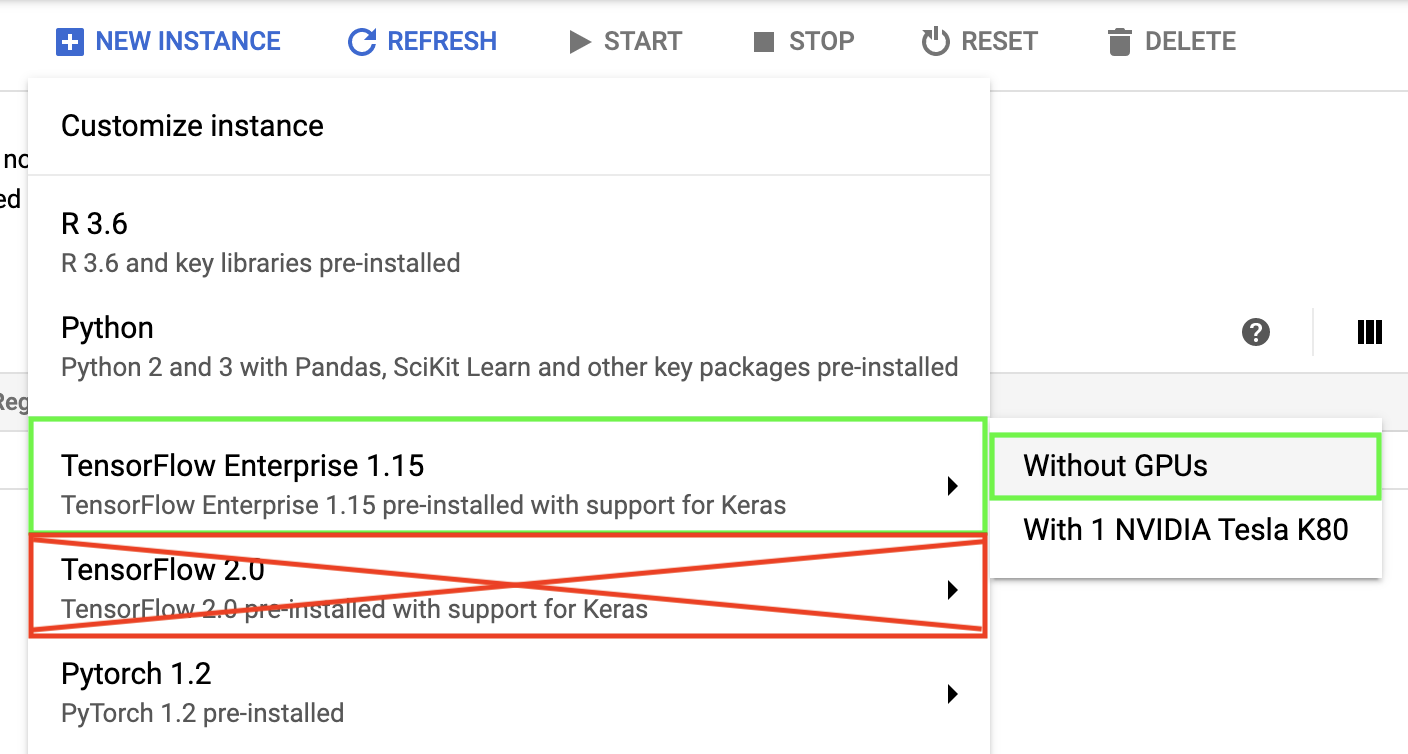
**Step 1**

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



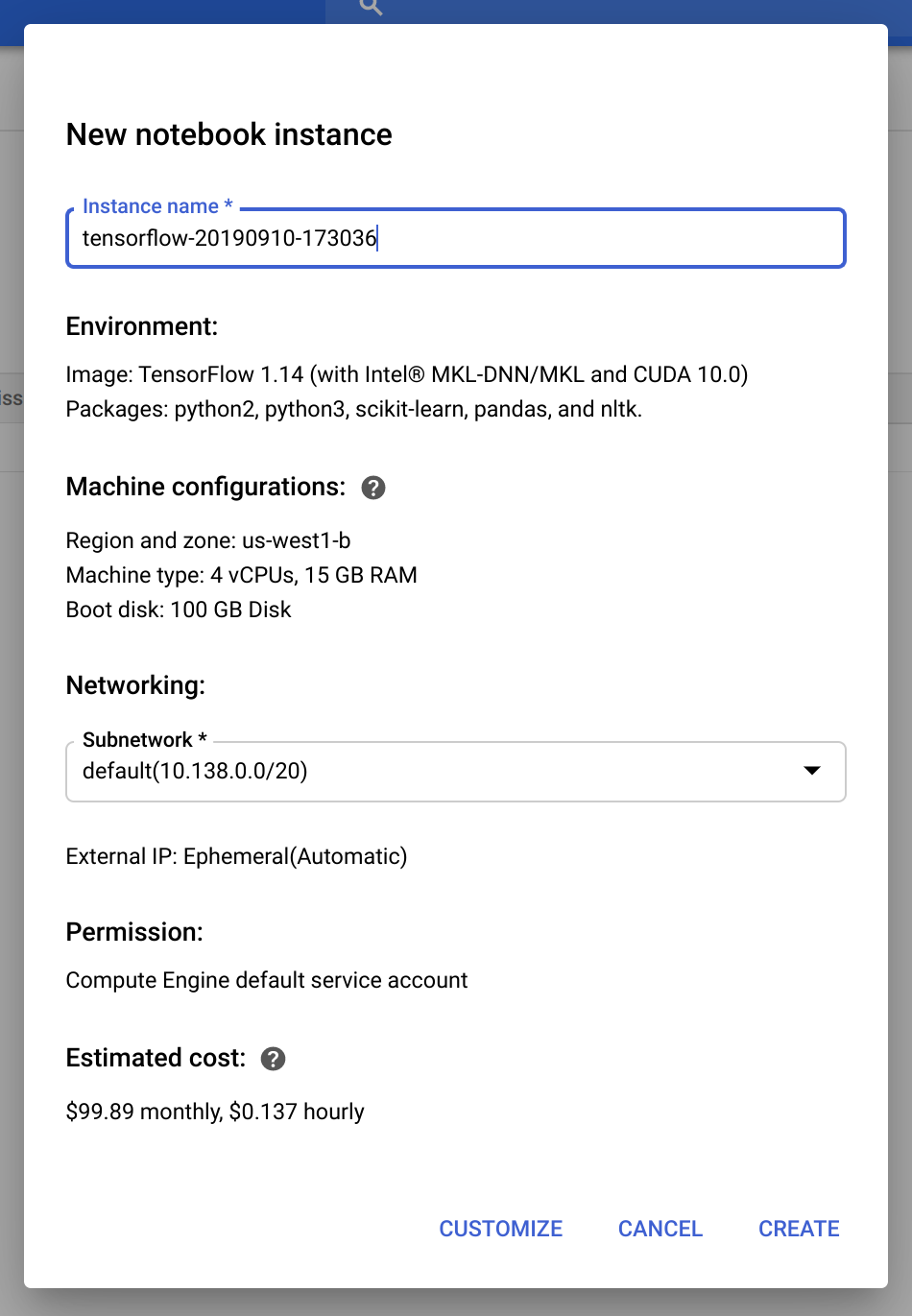
**Step 2**

On the Notebook instances page, click NEW INSTANCE. Select a 1.XX version of TensorFlow (not a 2.0) *without GPUs*. In the following example, you would select **Tensorflow Enterprise 1.15** > **Without GPUs**:



Tensorflow 1.XX versions change semi-frequently, so the version you pick may be different.

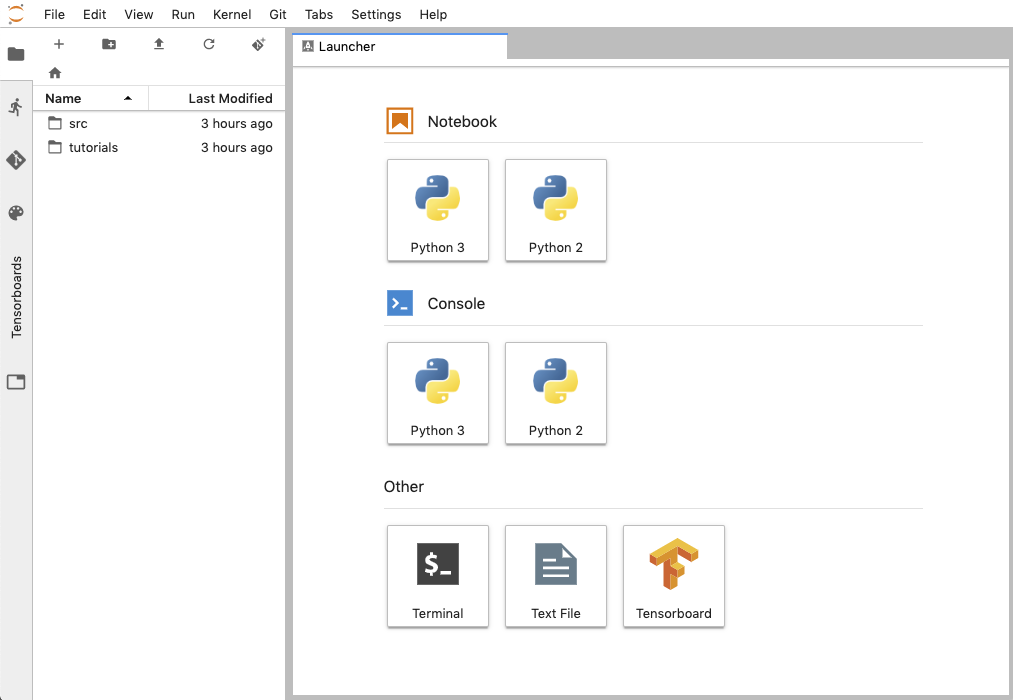
In the pop-up, confirm the name of the deep learning VM and click **Create**.



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

**Step 3**

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

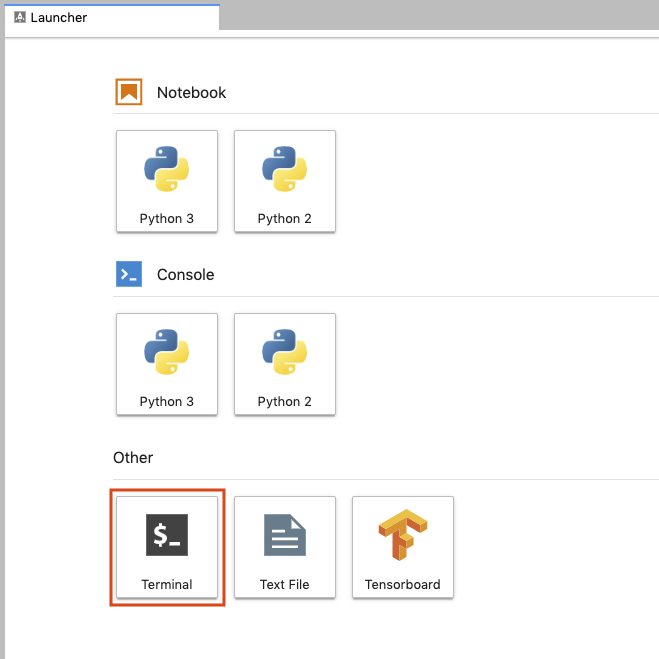


## **Clone course repo within your AI Platform Notebooks instance**

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

**Step 1**

In JupyterLab, click the Terminal icon to open a new terminal.



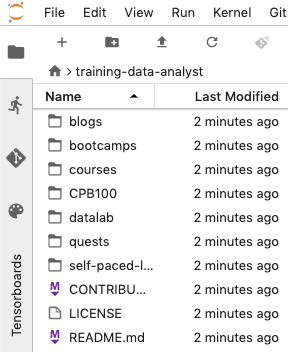
**Step 2**

At the command-line prompt, type in the following command and press Enter.

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

**Step 3**

Confirm that you have cloned the repository by double clicking on the training-data-analyst directory and ensuring that you can see its contents. The files for all the Jupyter notebook-based labs throughout this course are available in this directory.



## **Enable APIs**

In the Console, on the **Navigation menu** (7a91d354499ac9f1.png), select **APIs and Services** > **Library**.

For each of the products listed below, search the library and click Enable if they are not already enabled.

* AI Platform Training & Prediction API
* BigQuery API

## **Lab 1. Explore dataset, create ML datasets, create benchmark**

In this lab, you explore a dataset using BigQuery and Jupyterlab.

* Sample the dataset and create training, validation, and testing datasets for local development of TensorFlow models
* Create a benchmark to evaluate the performance of ML

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/datasets/** and open **create\_datasets.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).

Now read the narrative and execute each cell in turn.

## **Lab 2: Getting Started with TensorFlow**

In this lab, you will learn how the TensorFlow Python API works.

* Explore the TensorFlow Python API
* Building a graph
* Running a graph
* Feeding values into a graph
* Find area of a triangle using TensorFlow

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/tensorflow** and open **a\_tfstart.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).

Now read the narrative and execute each cell in turn.

## **Lab 3: Machine Learning using tf.estimator**

In this lab, you will implement a machine learning model using tf.estimator.

* Read from Pandas Dataframe into tf.constant
* Create feature columns for estimator
* Linear Regression with tf.Estimator framework
* Deep Neural Network regression
* Benchmark dataset

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/tensorflow** and open **b\_estimator.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).

Now read the narrative and execute each cell in turn.

## **Lab 4: Refactoring to add batching and feature-creation**

In this lab, you will:

* Refactor the input
* Refactor the way the features are created
* Create and train the model
* Evaluate model

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/tensorflow** and open **c\_batched.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).

Now read the narrative and execute each cell in turn.

## **Lab 5: Distributed training and monitoring**

In this lab, you will learn how to:

* Create features out of input data
* Train and evaluate
* Monitor with Tensorboard

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/tensorflow** and open **d\_traineval.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).

Now read the narrative and execute each cell in turn.

## **Lab 6: Scaling up ML using Cloud ML Engine**

In this lab, you will:

* Package up the code
* Find absolute paths to data
* Run the Python module from the command line
* Run locally using gcloud
* Submit training job using gcloud
* Deploy model
* Prediction
* Train on a larger dataset
* 1-million row dataset

## **Task 1: Verify that you have a Cloud Storage bucket**

The next lab notebook requires you to provide a **Project ID**, **Bucket Name**, and **Bucket Region**.

You can recall the Project ID in Cloud Shell by entering the following:

echo $DEVSHELL\_PROJECT\_ID

You should have a bucket from the previous lab. If you don't you can follow these instructions to create a bucket.

The bucket must be regional. A Multi-Regional bucket will not work with this lab. If your existing buckets are Multi-Regional, you will need to create a new one for this lab.

**Step 1**

In the Console, on the **Navigation menu** (7a91d354499ac9f1.png), click **Home**.

Select and copy the **Project ID**. For simplicity, you will use the Qwiklabs Project ID, which is already globally unique, as the bucket name. If that is already taken, create a globally unique bucket name.

**Step 2**

In the Console, on the **Navigation menu** (7a91d354499ac9f1.png), click **Storage** > **Browser**.

**Step 3**

Click **Create Bucket**.

**Step 4**

Specify the following, and leave the remaining settings as their defaults:

|  |  |
| --- | --- |
| **Property** | **Value (type value or select option as specified)** |
| **Name** | <your unique bucket name (Project ID)> |
| **Default storage class** | [x] Regional |
| **Location** | <Your location> **Valid Regions for this lab are** asia-east1, europe-west1, us-central1, us-east1, us-west1 |

**Step 5**

Click **Create**. Record the name of your bucket and the location. You will need it in subsequent tasks.

## **Task 2: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/cloudmle** and open **cloudmle.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).

Now read the narrative and execute each cell in turn.

## **Lab 7: Feature Engineering**

In this lab, you will improve the ML model using feature engineering. In the process, you will learn how to:

* Working with feature columns
* Adding feature crosses in TensorFlow
* Reading data from BigQuery
* Creating datasets using Dataflow
* Using a wide-and-deep model

## **Task 1: Continue the lab in the notebook**

**Step 1**

In the notebook interface, navigate to **training-data-analyst/courses/machine\_learning/feateng** and open **feateng.ipynb**.

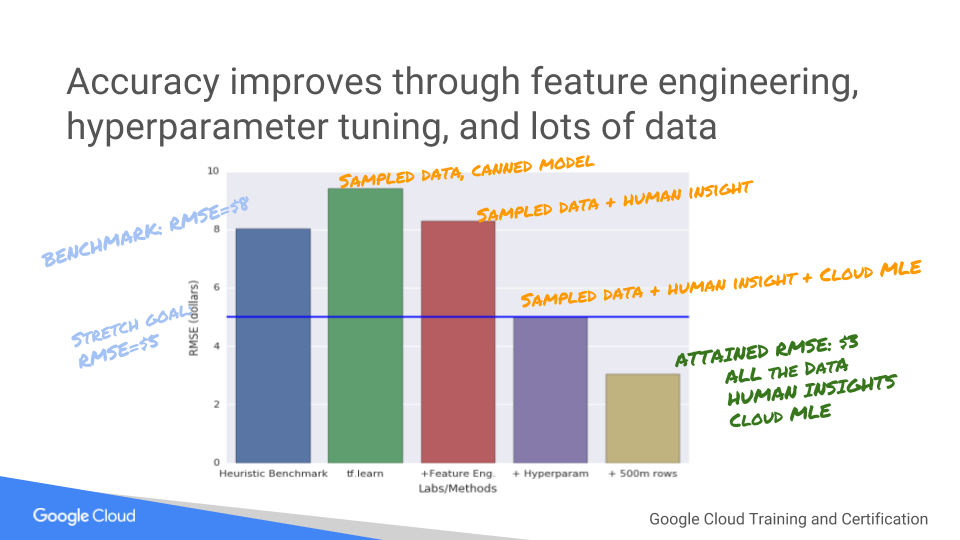
**Step 2**

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

Now read the narrative and execute each cell in turn.

## **Lab 8: Demonstration of Hyper-Parameter Tuning and Training**

Your instructor will demonstrate notebooks that contain hyper-parameter tuning and training on 500 million rows of data. The changes to the model are minor -- essentially just command-line parameters, but the impact on model accuracy is huge.



## **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.

###### Manual Last Updated: November 22, 2019

###### Lab Last Tested: November 22, 2019

Copyright 2019 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.