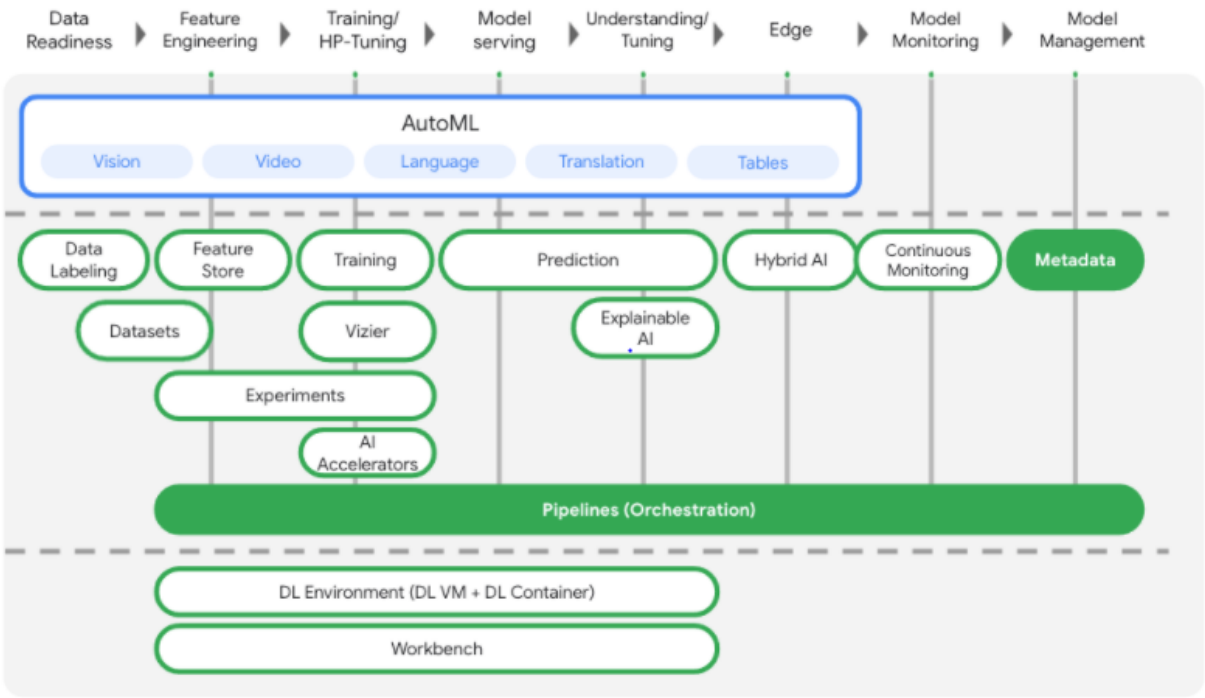
Assignment 7: optional catchup assignment 2 - VERTEX AI - for midterm and quiz - this will catch up midterm.

1) Using Vertex ML Metadata with Pipelines
   Reference: https://codelabs.developers.google.com/vertex-mlmd-pipelines#0

Objectives:

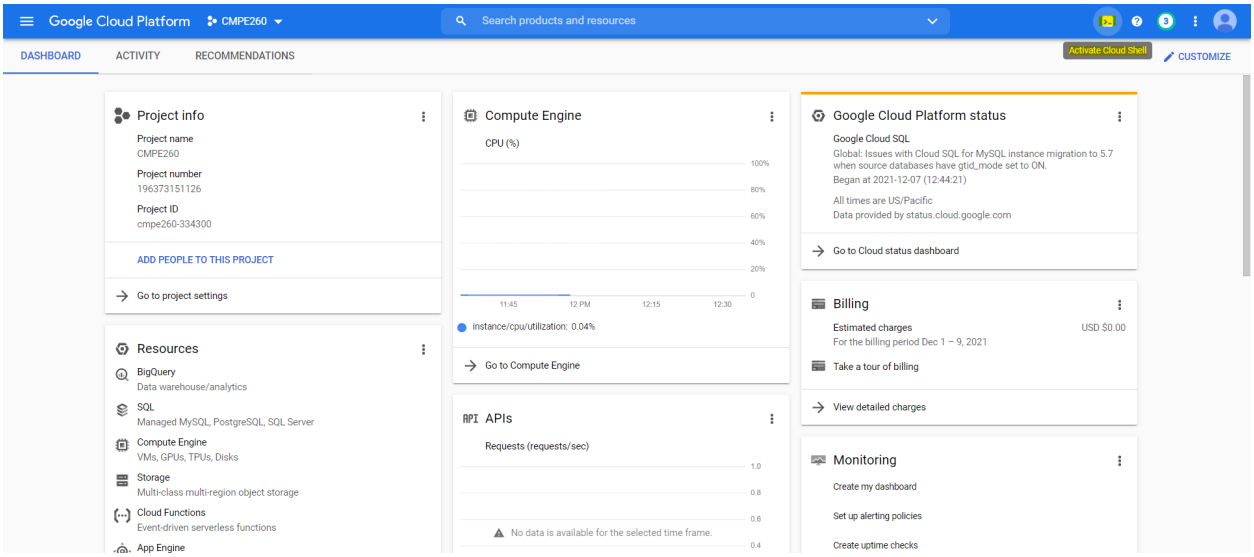
- Use the Kubeflow Pipelines SDK to build an ML pipeline that creates a dataset in Vertex AI, and trains and deploys a custom Scikit-learn model on that dataset
- Write custom pipeline components that generate artifacts and metadata
- Compare Vertex Pipelines runs, both in the Cloud console and programmatically
- Trace the lineage for pipeline-generated artifacts
- Query your pipeline run metadata

Vertex AI also includes a variety of MLOps products, including Vertex Pipelines, ML Metadata, Model Monitoring, Feature Store, and more. This lab focuses on **Vertex Pipelines** and **Vertex ML Metadata**.

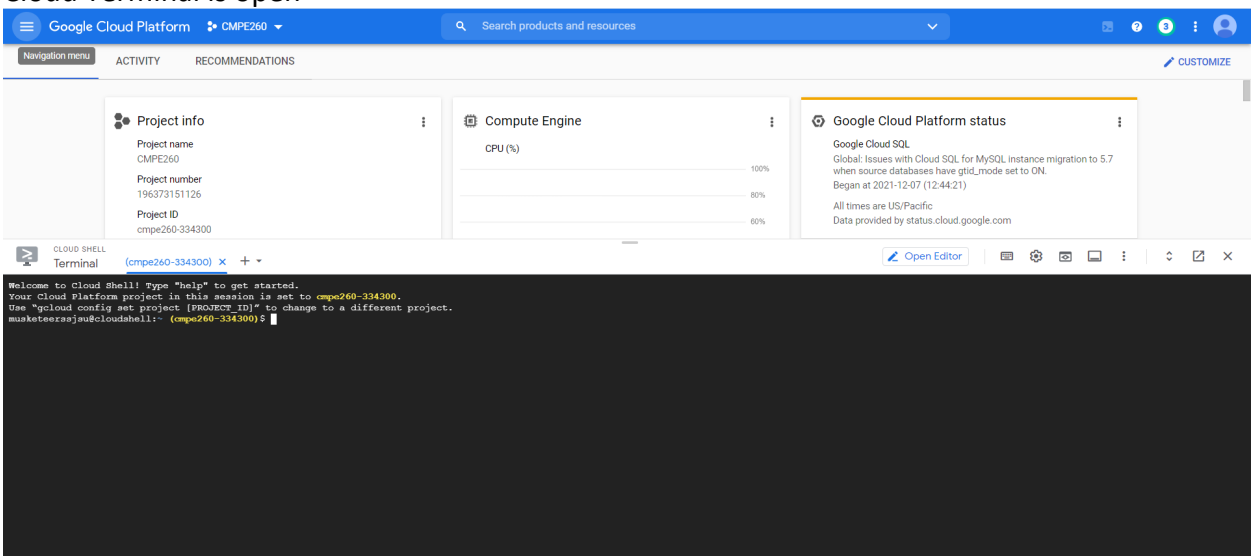


Cloud environment setup

Start Cloud Shell

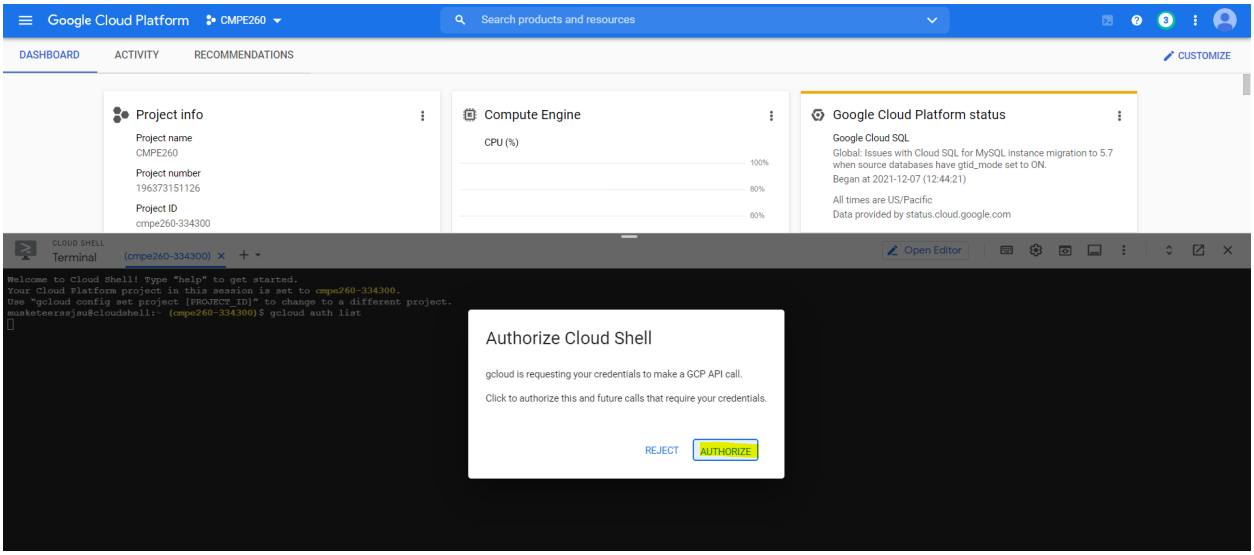


Cloud Terminal is open

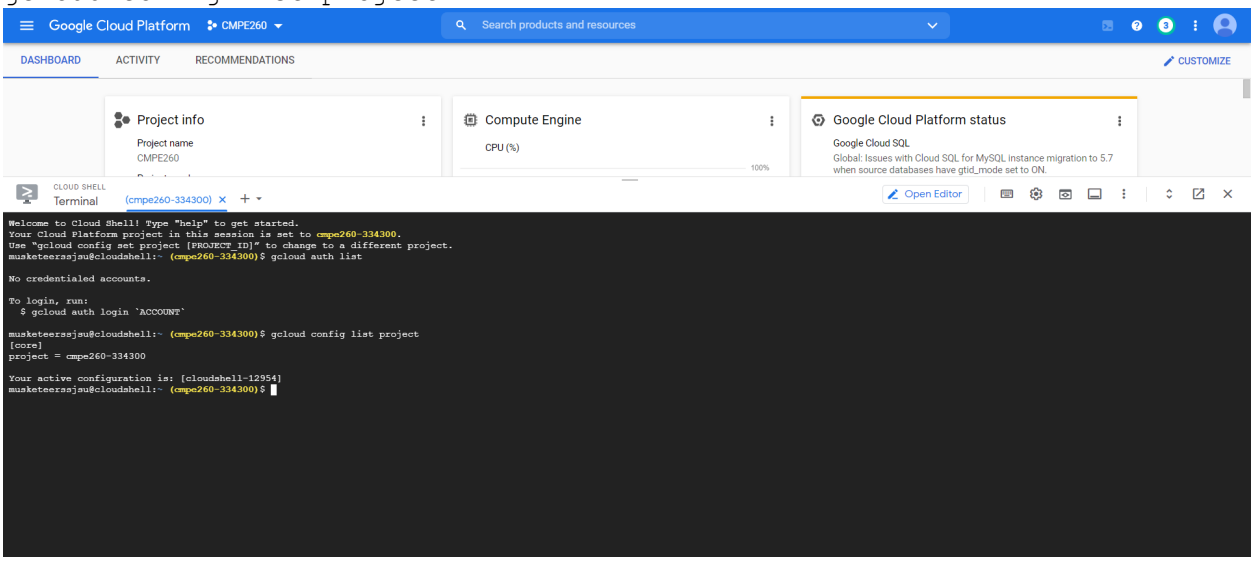


Run the following commands:

```
gcloud auth list
```



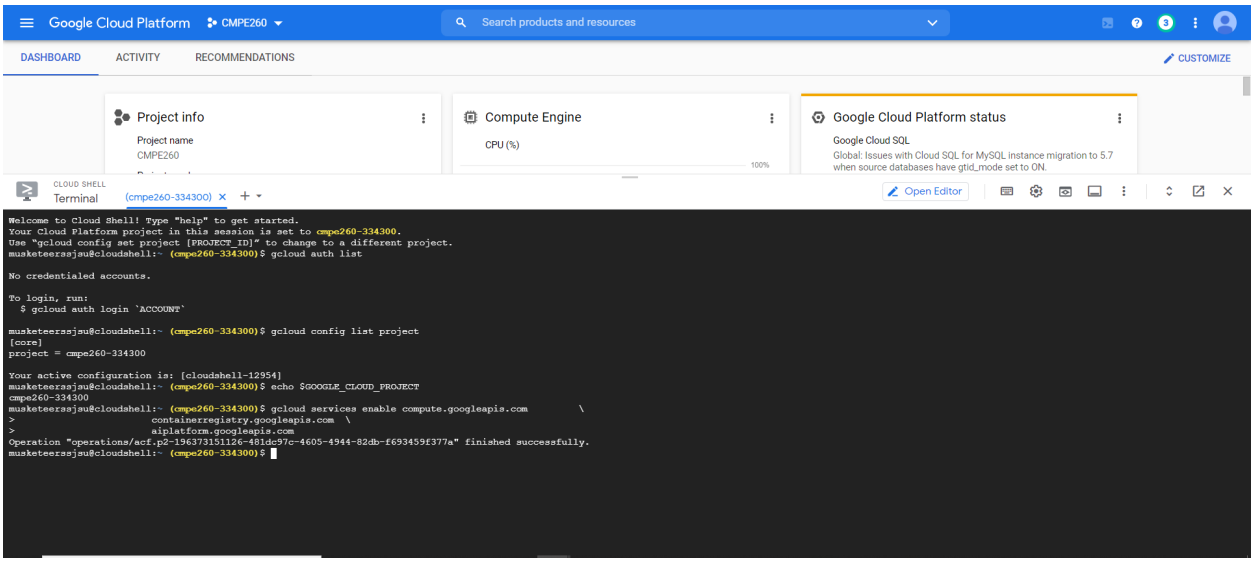

```
gcloud config list project
```



Enable APIs

```
echo $GOOGLE_CLOUD_PROJECT
```

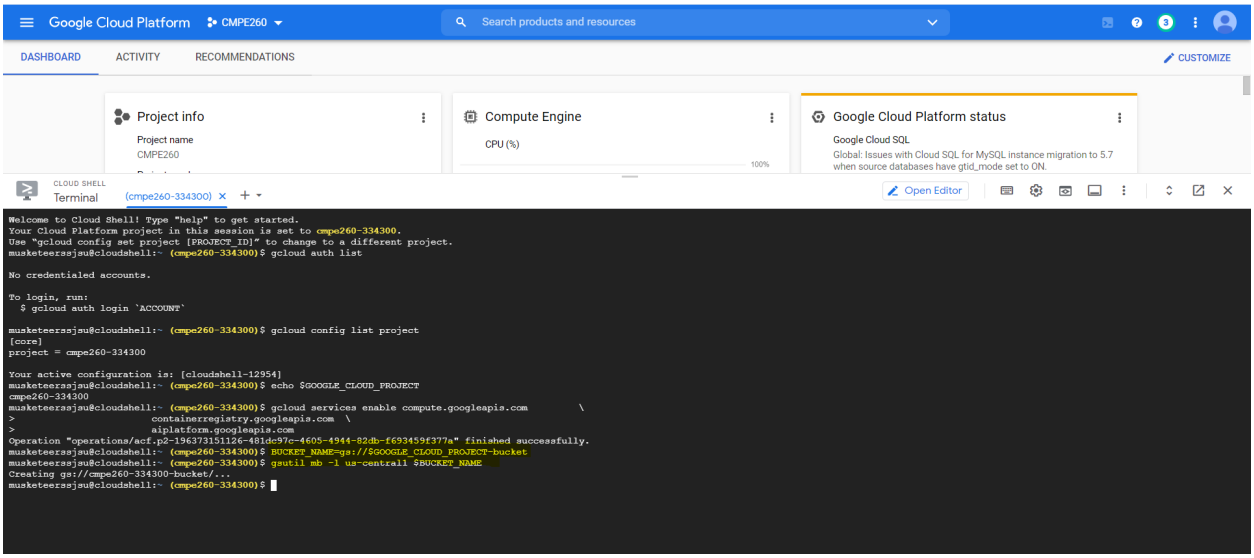
```
gcloud services enable compute.googleapis.com          \
                       containerregistry.googleapis.com  \
                       aiplatform.googleapis.com
```



Create a Cloud Storage Bucket

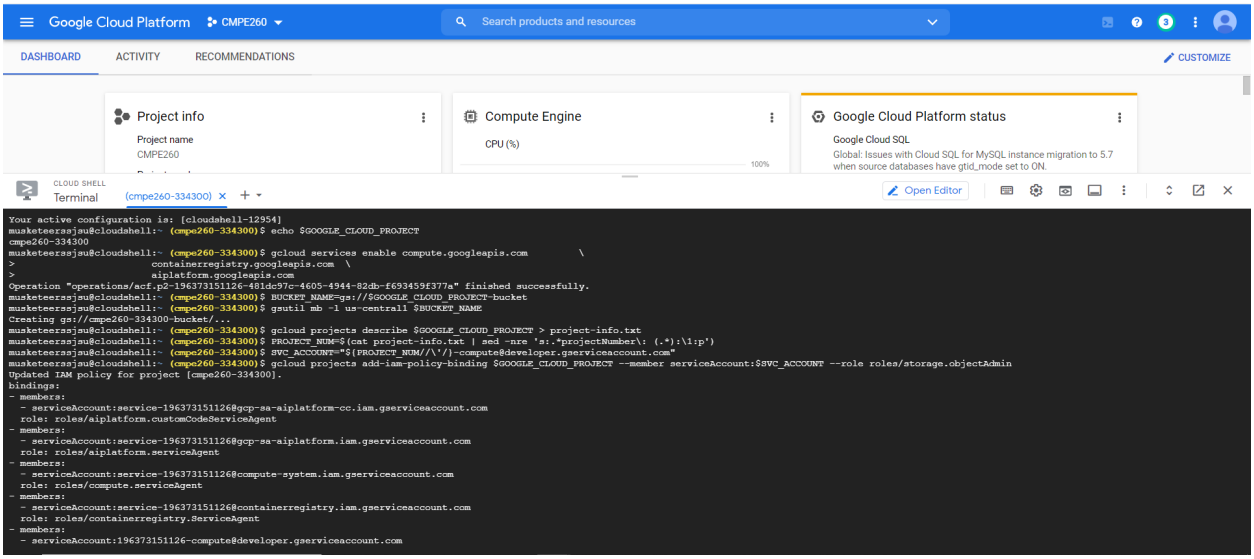
Run the following commands:

```
BUCKET_NAME=gs://$GOOGLE_CLOUD_PROJECT-bucket
gsutil mb -l us-central1 $BUCKET_NAME
```



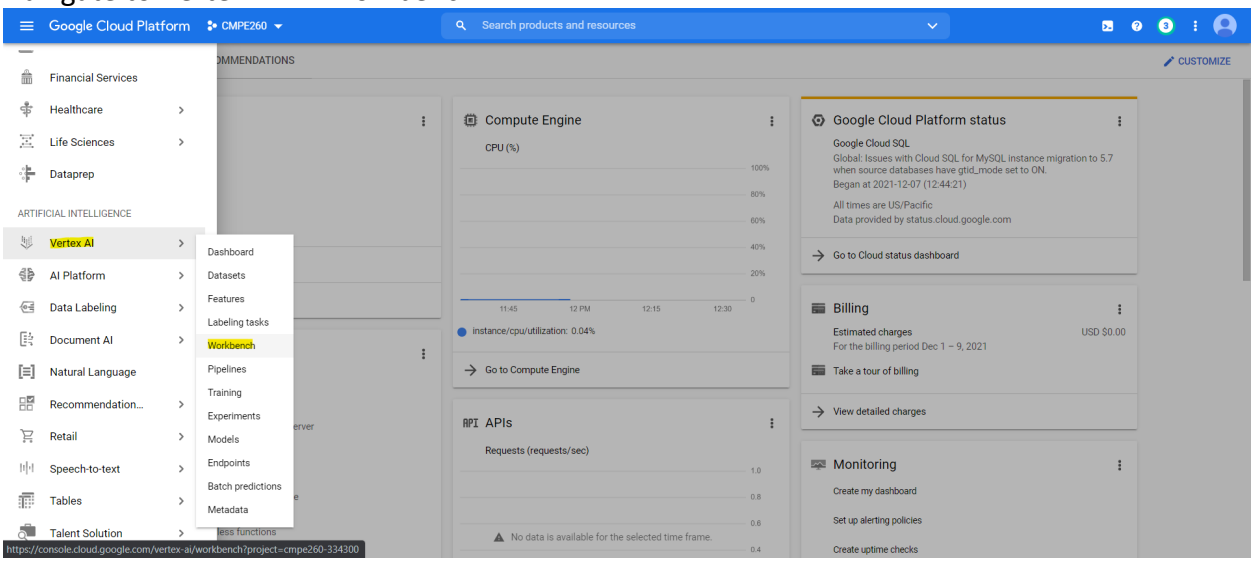
Run the following command to add the permission for compute service account access to this Bucket

```
gcloud projects describe $GOOGLE_CLOUD_PROJECT > project-info.txt
PROJECT_NUM=$(cat project-info.txt | sed -nre 's:.*projectNumber\: (.*):\1:p')
SVC_ACCOUNT="${PROJECT_NUM//\'/}-compute@developer.gserviceaccount.com"
gcloud projects add-iam-policy-binding $GOOGLE_CLOUD_PROJECT --member
serviceAccount:$SVC_ACCOUNT --role roles/storage.objectAdmin
```

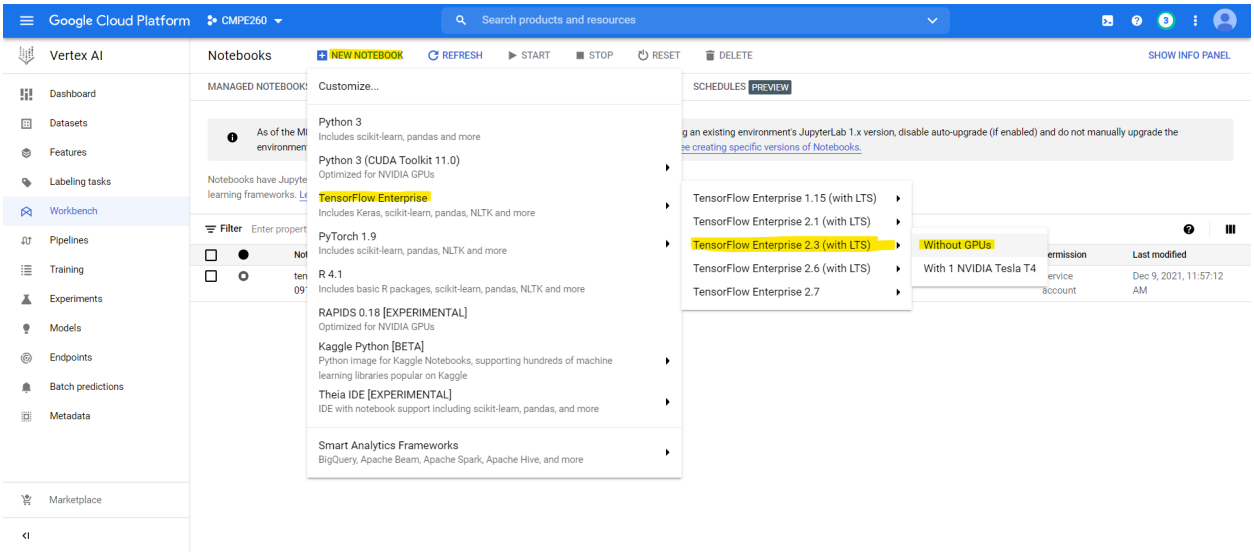


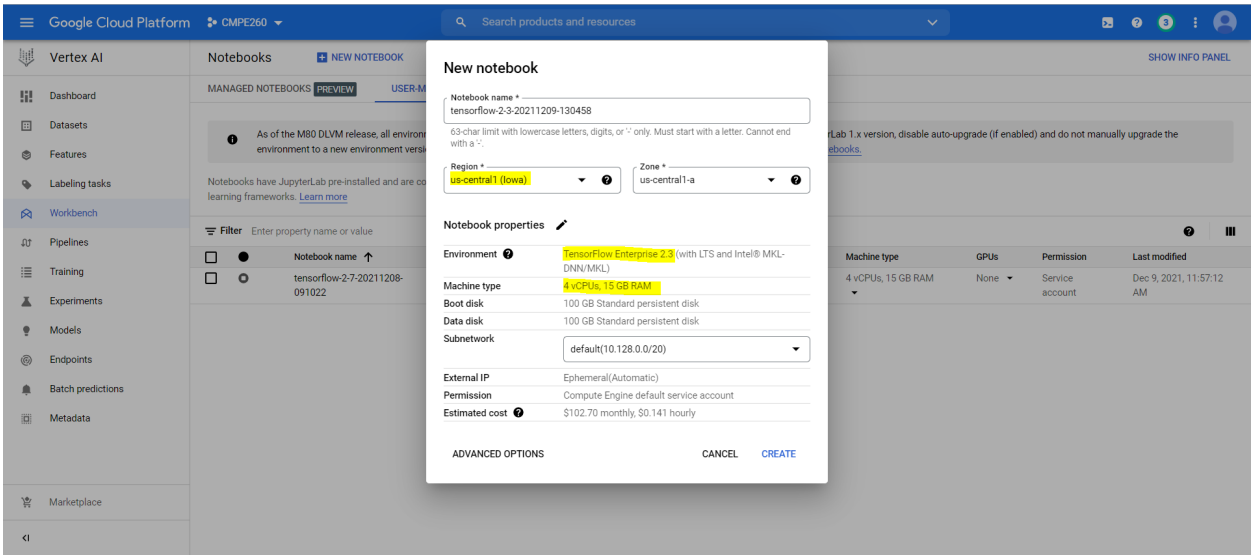
Create a Vertex AI Workbench instance

Navigate to Vertex AI → Workbench



Select **TensorFlow Enterprise 2.3 (with LTS)** instance type **without GPUs**:



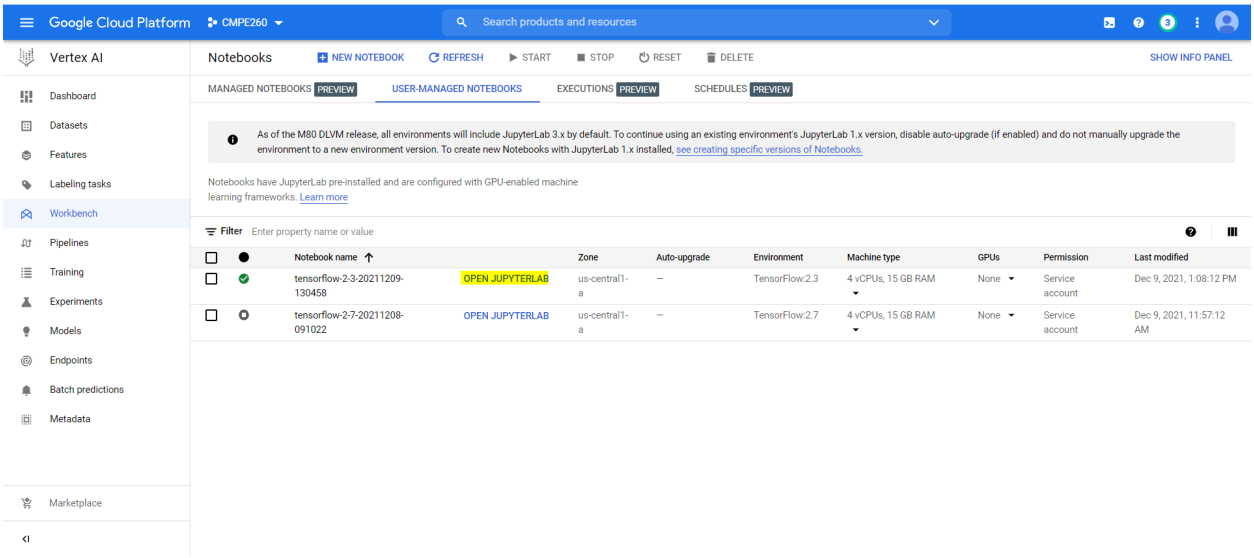


Vertex Pipelines setup

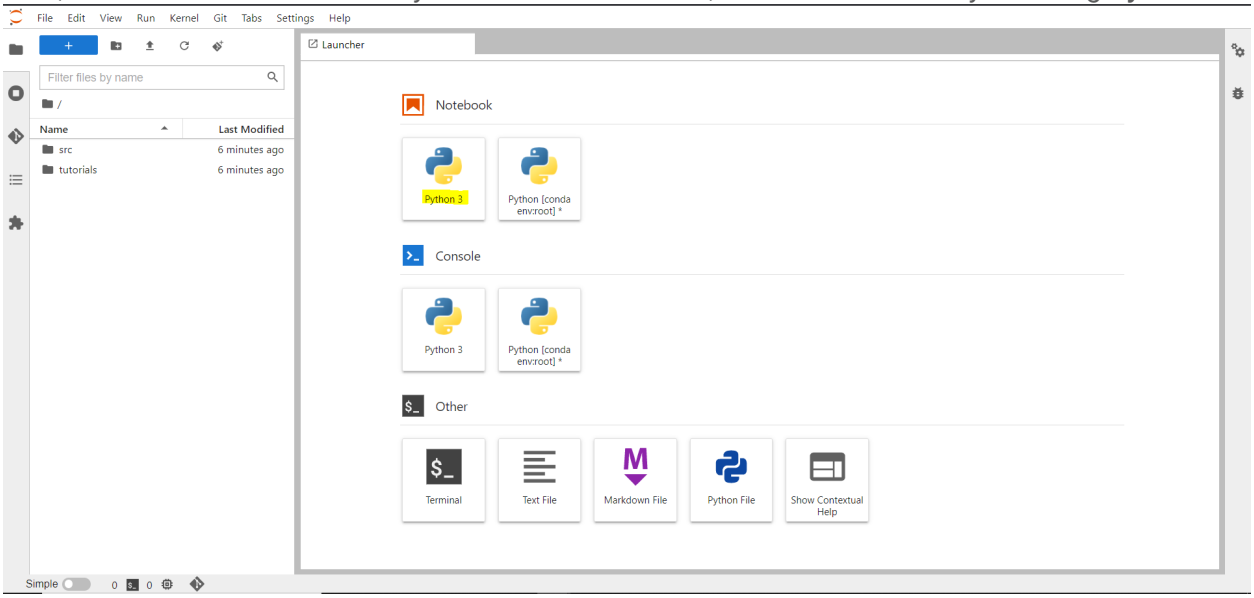
There are a few additional libraries we'll need to install in order to use Vertex Pipelines:

- **Kubeflow Pipelines**: This is the SDK we'll be using to build our pipeline. Vertex Pipelines supports running pipelines built with both Kubeflow Pipelines or TFX.
- **Vertex AI SDK**: This SDK optimizes the experience for calling the Vertex AI API. We'll use it to run our pipeline on Vertex AI.

Create Python notebook and install libraries

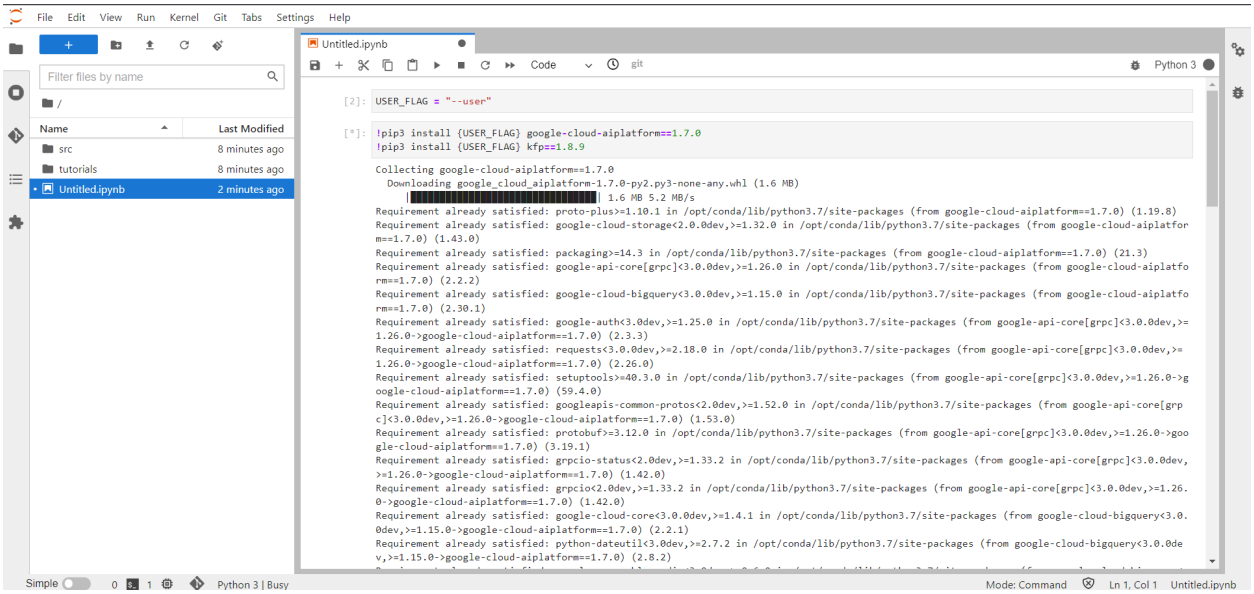


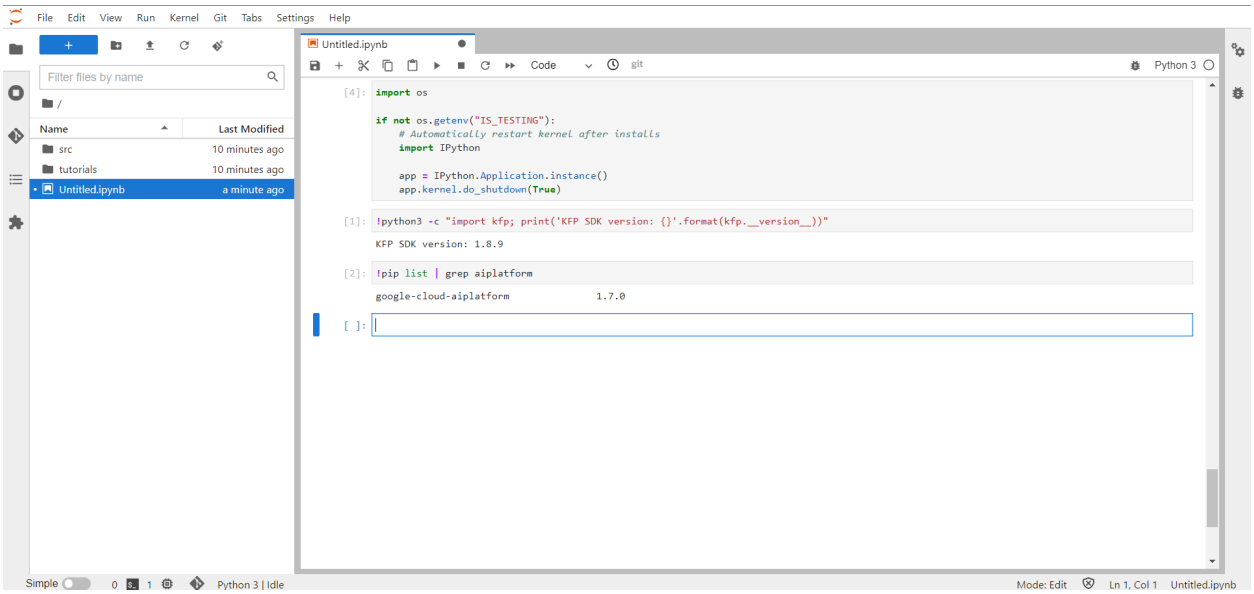
First, from the Launcher menu in your Notebook instance, create a notebook by selecting **Python 3**:



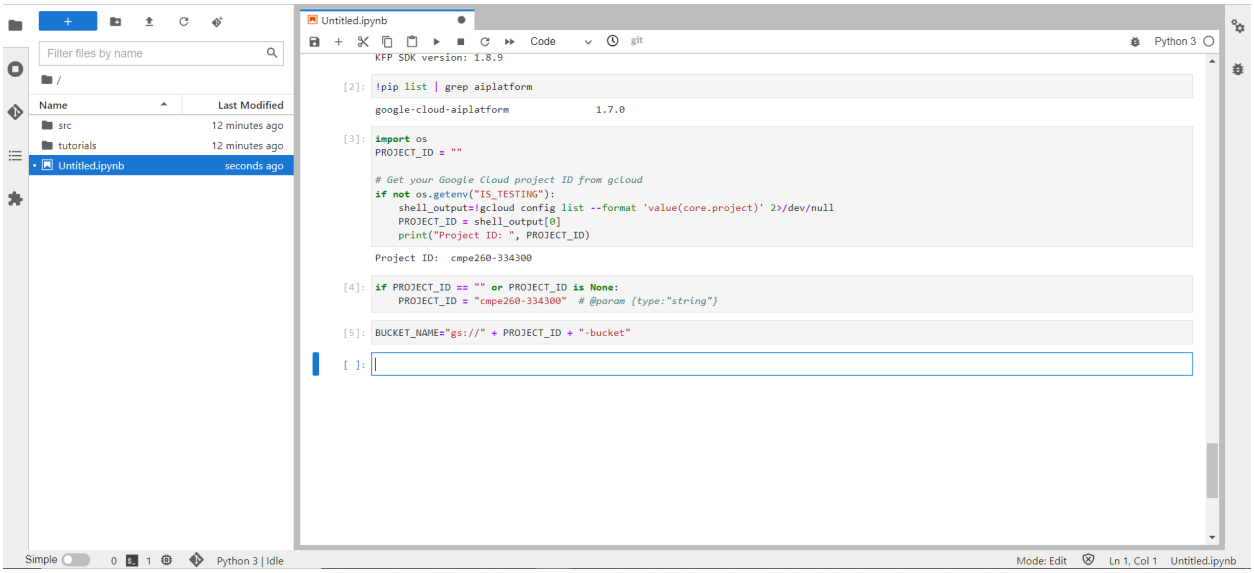
To install both services we'll be using in this lab, first set the user flag in a notebook cell:

```
USER_FLAG = "--user"
!pip3 install {USER_FLAG} google-cloud-aiplatform==1.7.0
!pip3 install {USER_FLAG} kfp==1.8.9
```

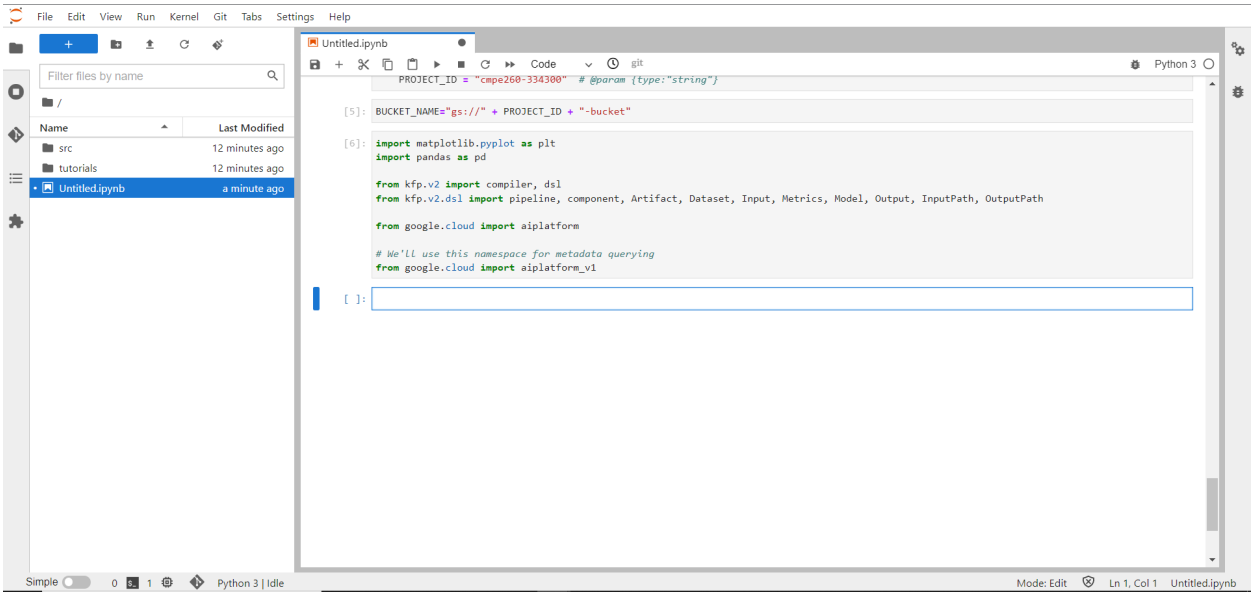




Set your project ID and bucket

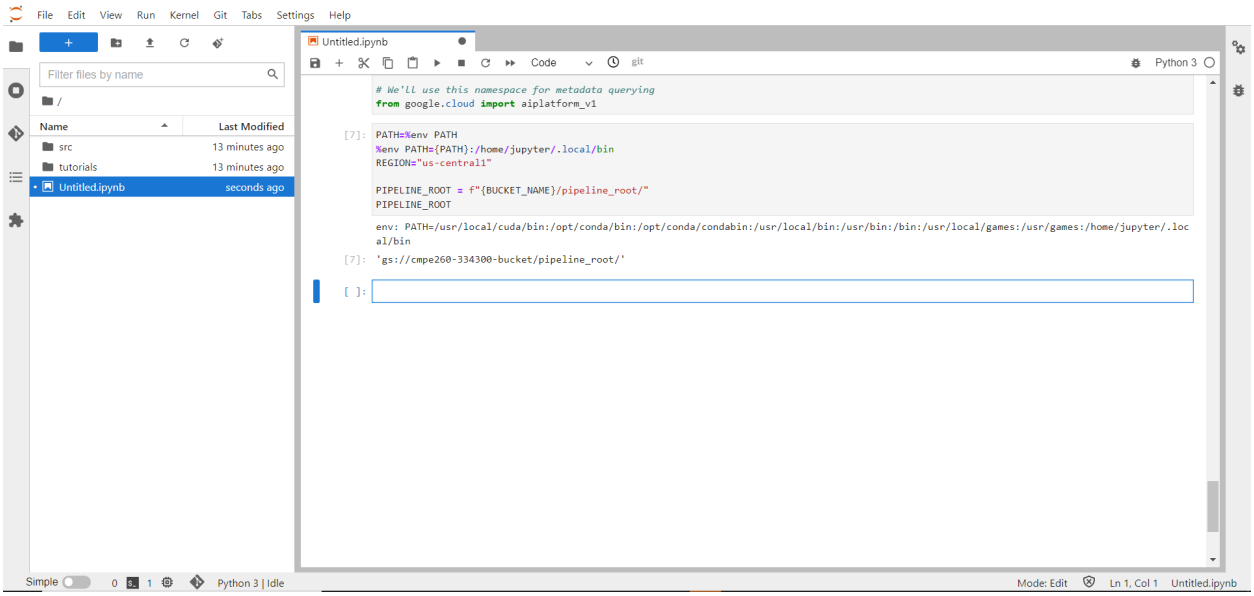


Import libraries



Define constants

`PIPELINE_ROOT` is the Cloud Storage path where the artifacts created by our pipeline will be written.



Creating a 3-step pipeline with custom components

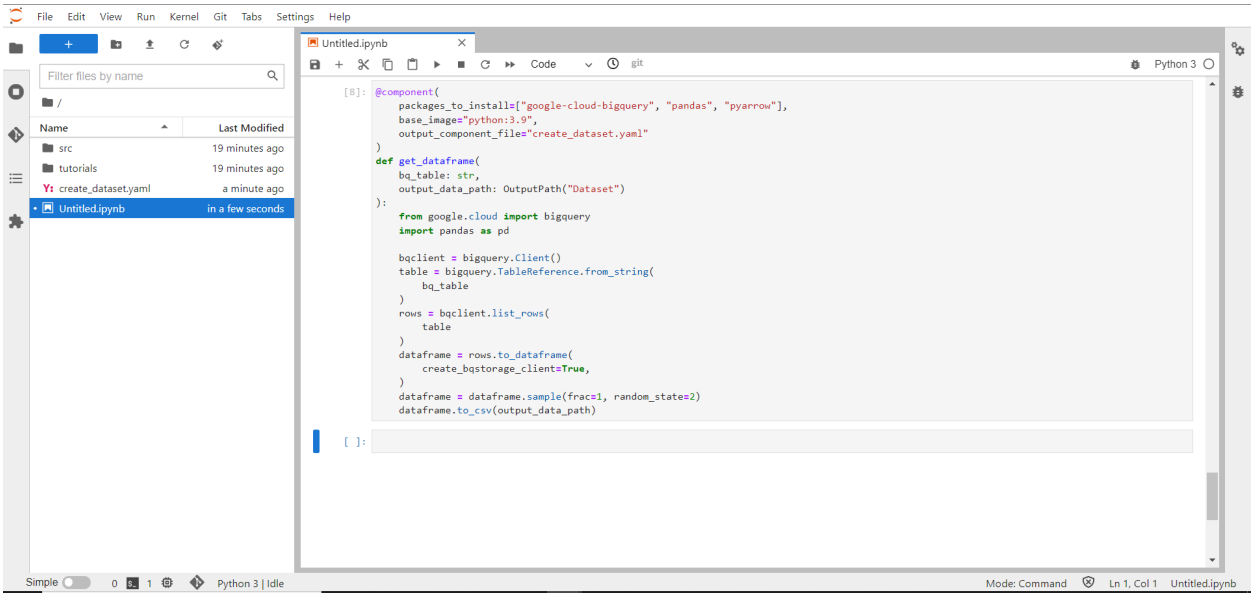
The focus of this lab is on understanding *metadata* from pipeline runs. In order to do that, we'll need a pipeline to run on Vertex Pipelines, which is where we'll start. Here we'll define a 3-step pipeline with the following custom components:

- `get_dataframe`: Retrieve data from a BigQuery table and convert it into a Pandas DataFrame
- `train_sklearn_model`: Use the Pandas DataFrame to train and export a Scikit Learn model, along with some metrics
- `deploy_model`: Deploy the exported Scikit Learn model to an endpoint in Vertex AI

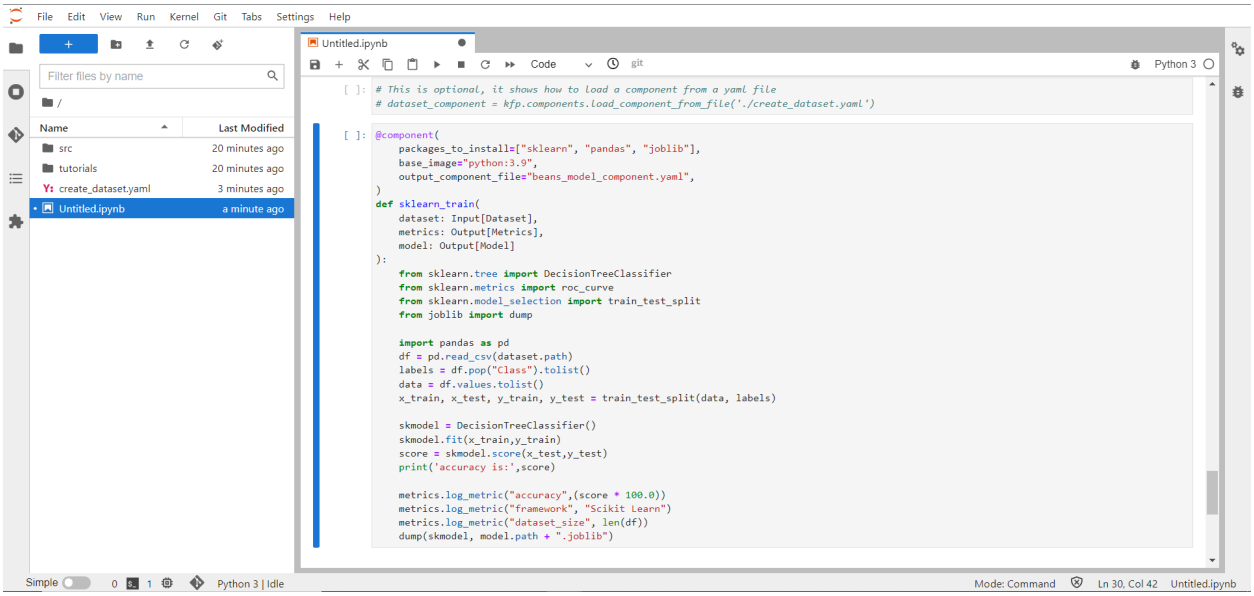
Create Python function based components

Using the KFP SDK, we can create components based on Python functions. We'll use that for the 3 components in this pipeline.

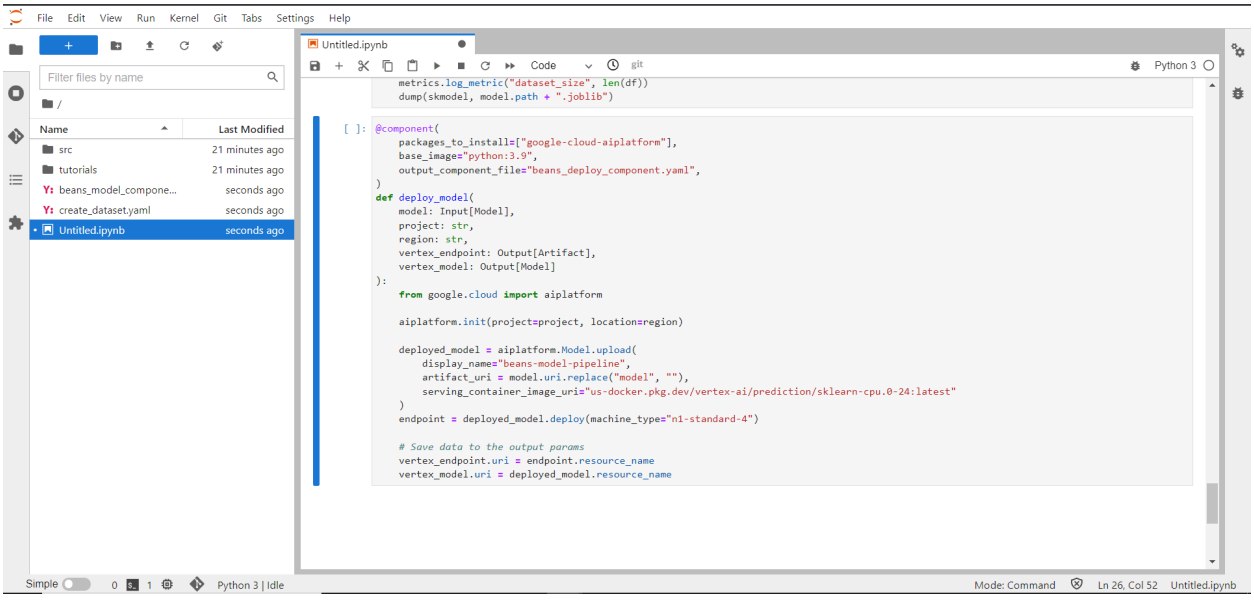
`get_dataframe` component:



This component exports the resulting Scikit model, along with a `Metrics` artifact that includes our model's accuracy, framework, and size of the dataset used to train it

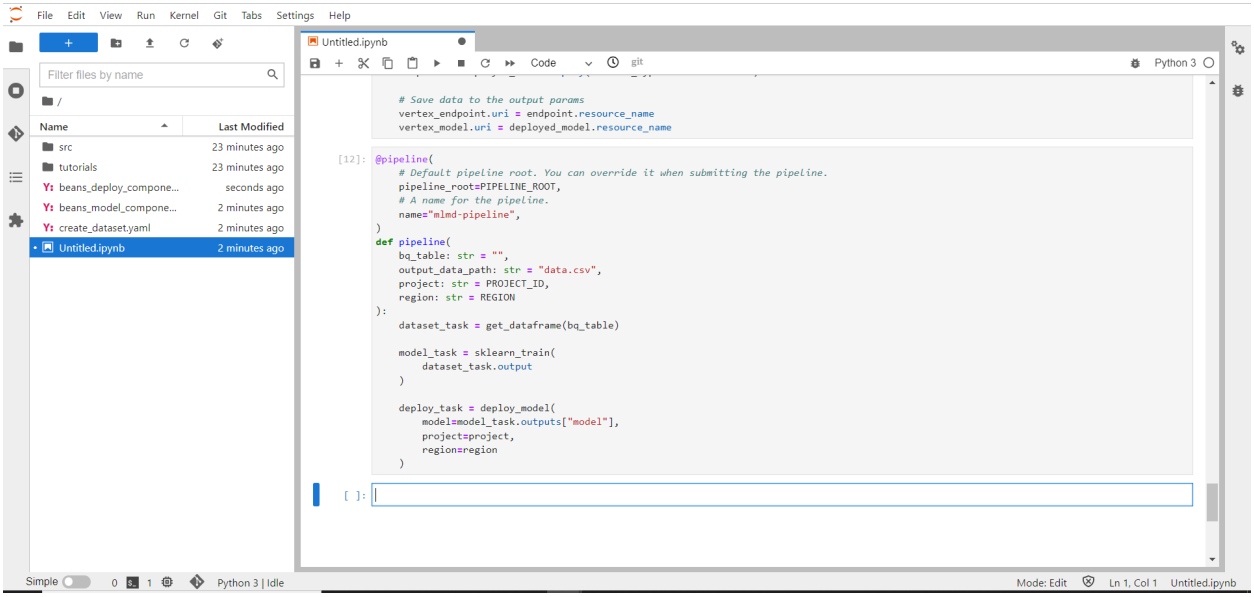


Define a component to upload and deploy the model to Vertex AI

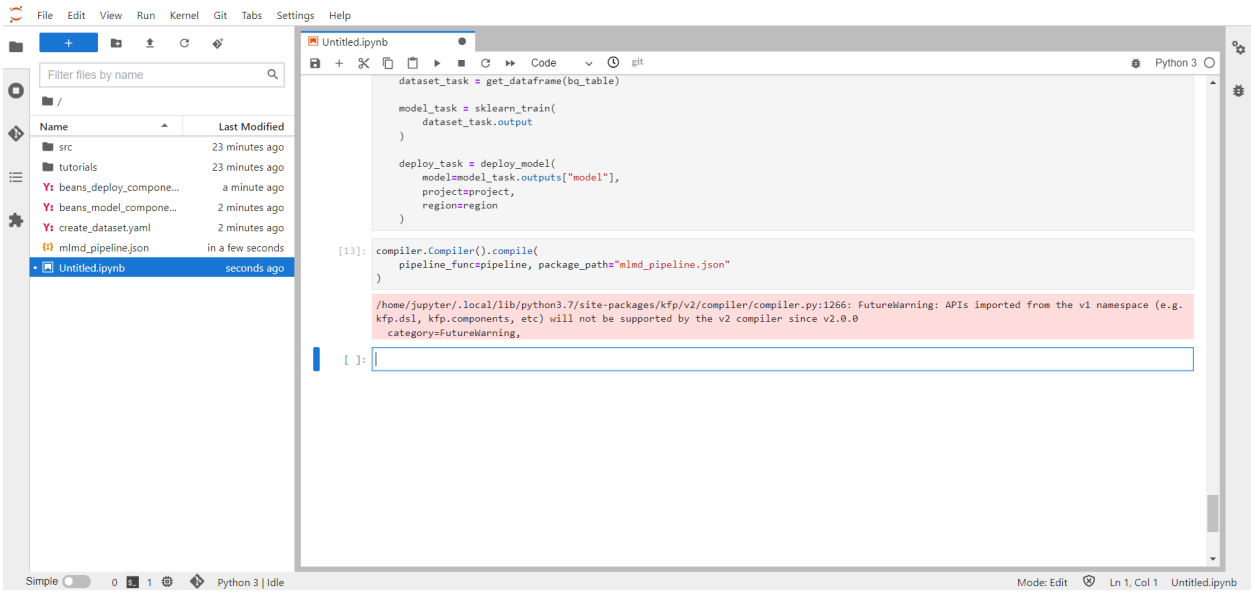


Define and compile the pipeline

Now that we've defined our three components, next we'll create our *pipeline definition*.

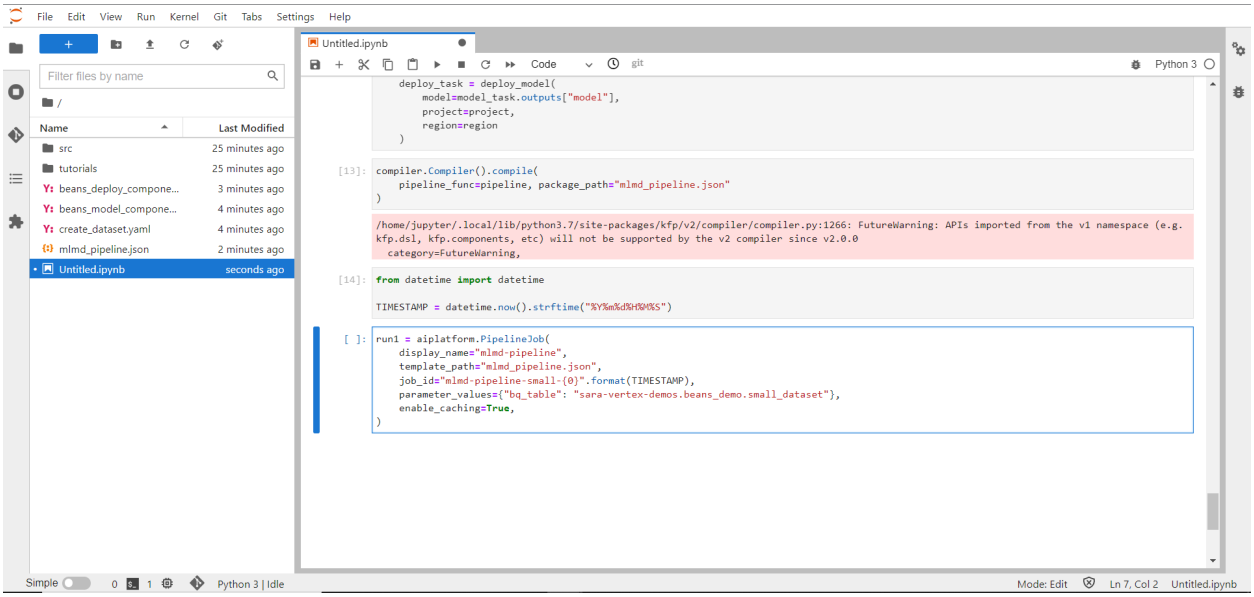


The following will generate a JSON file that you'll use to run the pipeline:

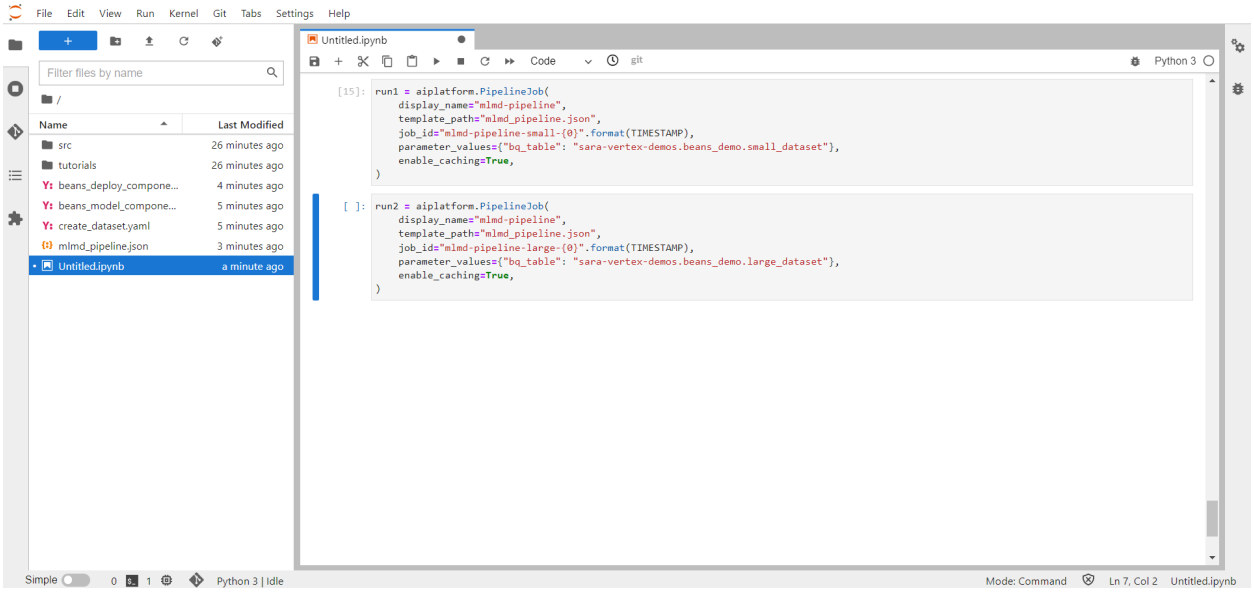


Start two pipeline runs

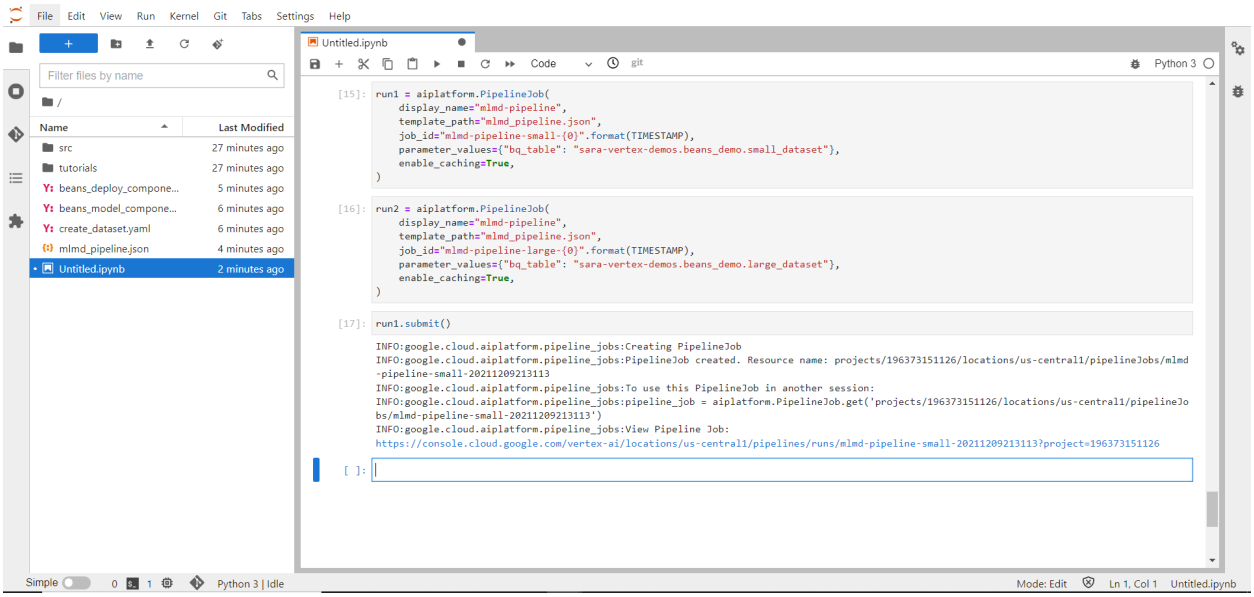
Remember that our pipeline takes one parameter when we run it: the `bq_table` we want to use for training data. This pipeline run will use a smaller version of the beans dataset:



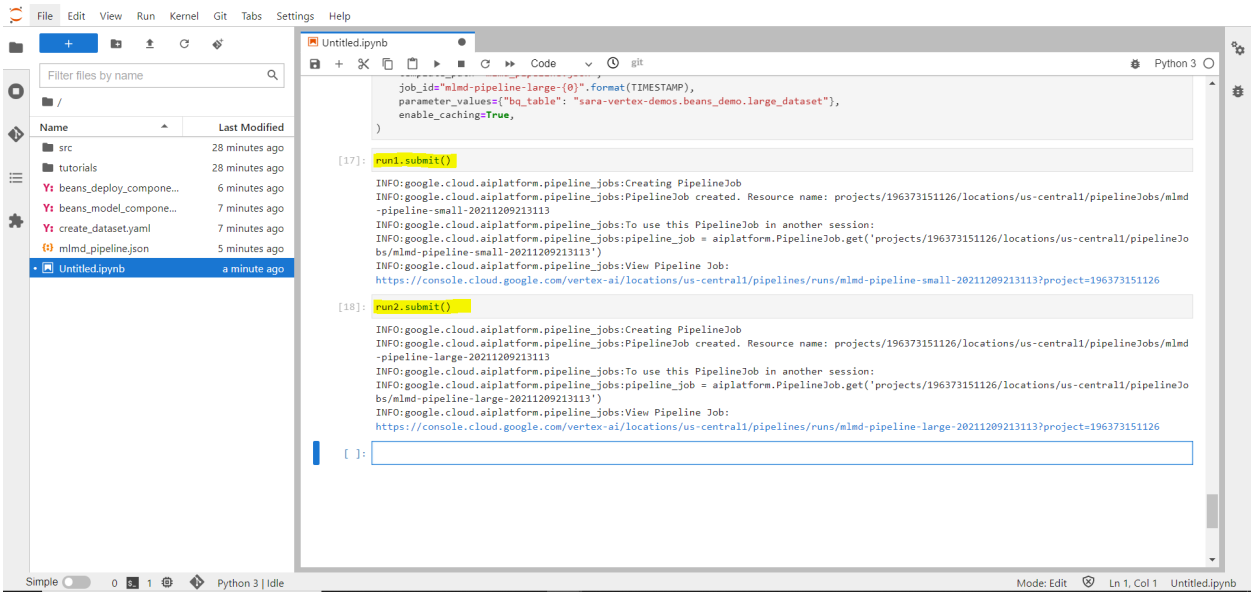
Next, create another pipeline run using a larger version of the same dataset.



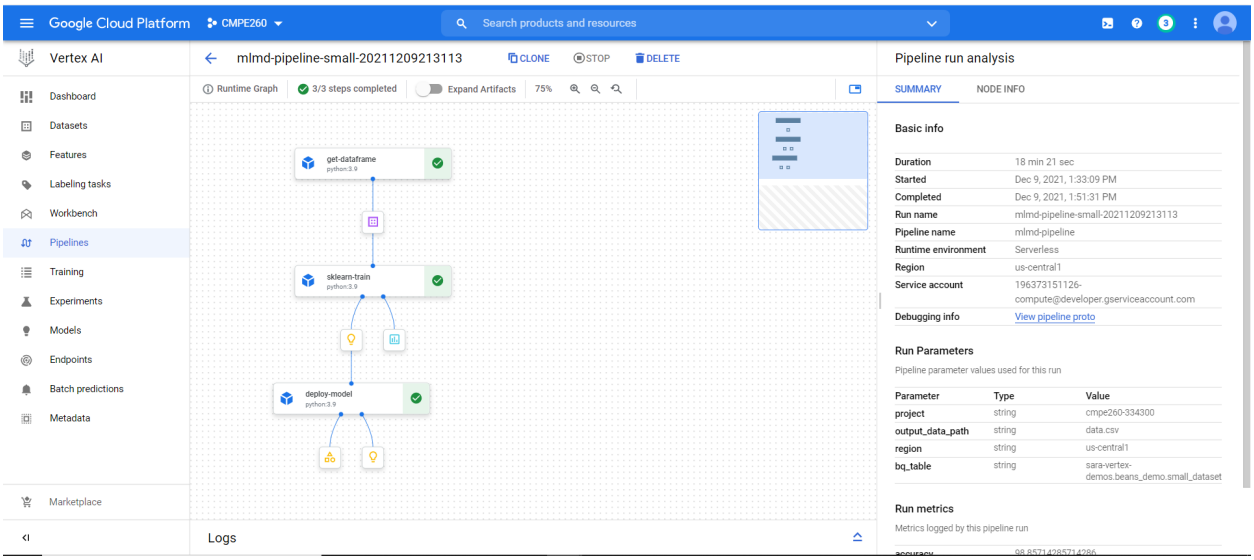
Kick off pipeline executions for both runs

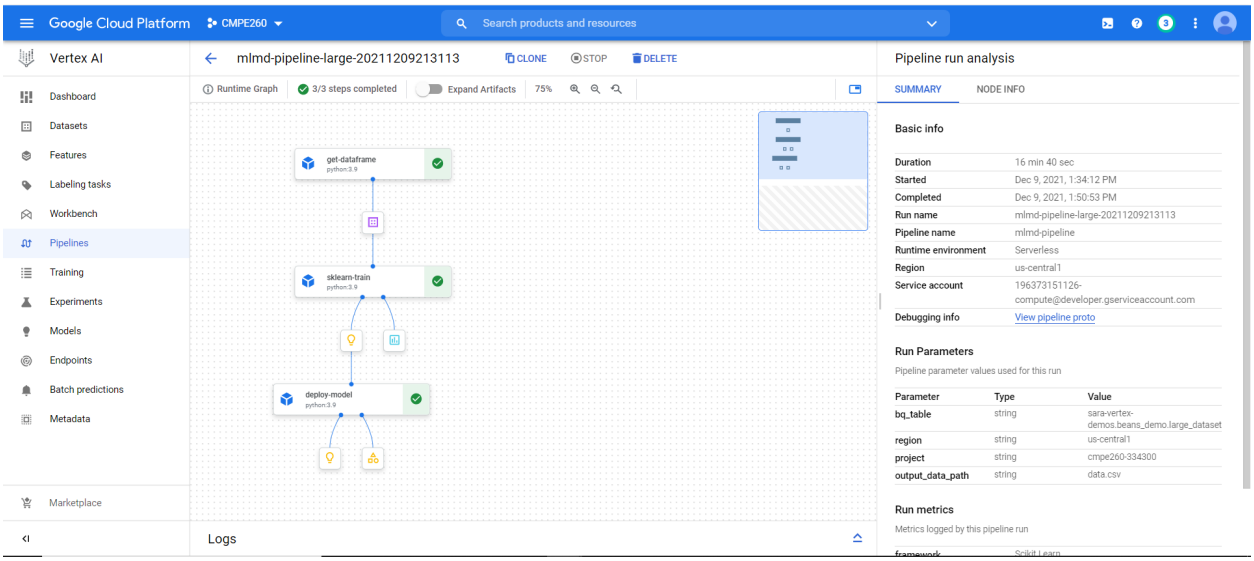


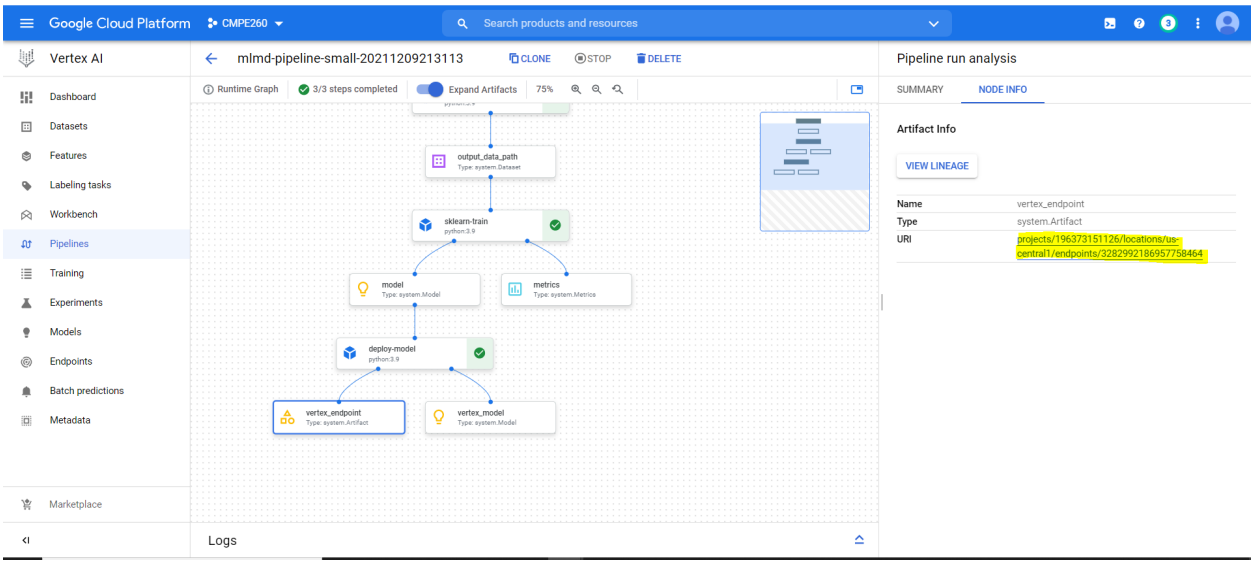
kick off the second run:

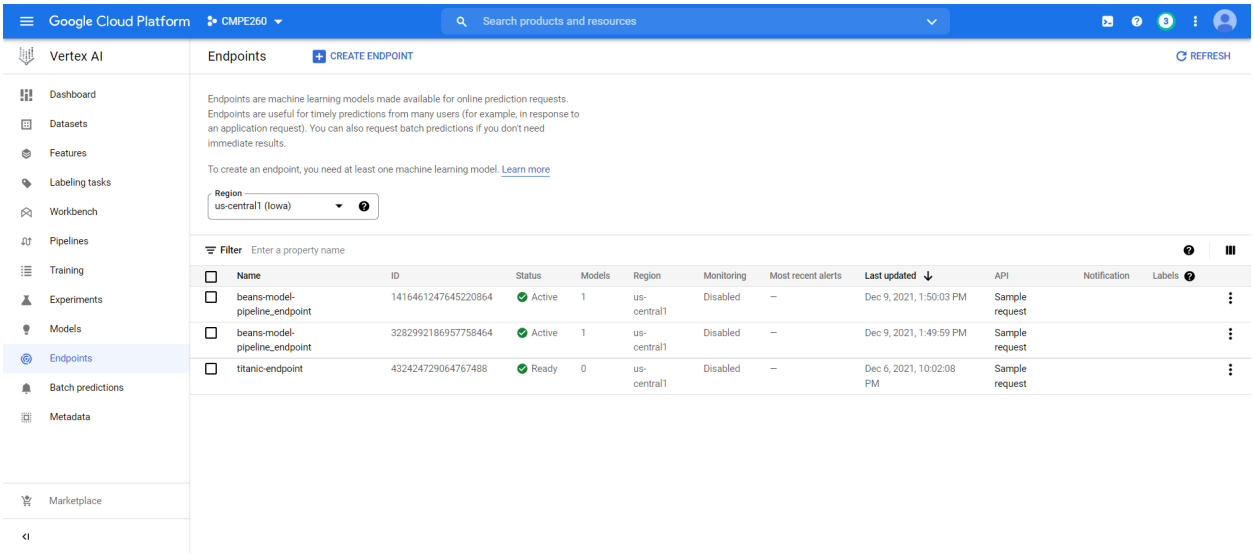


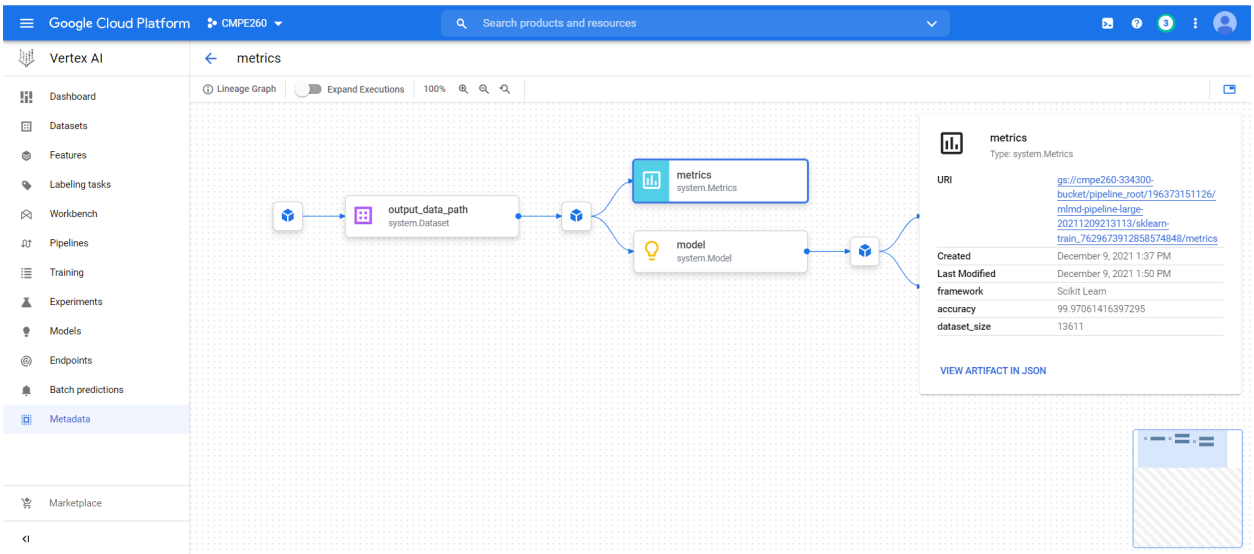
Results:


Google Cloud Platform
CMPE260
Search products and resources
Vertex AI
mlmd-pipeline-small-20211209213113
CLONE
STOP
DELETE
Runtime Graph
3/3 steps completed
Expand Artifacts
75%
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
get-dataframe
python:3.9
sklearn-train
python:3.9
deploy-model
python:3.9
Logs
Pipeline run analysis
SUMMARY
NODE INFO
Basic info
Duration 18 min 21 sec
Started Dec 9, 2021, 1:33:09 PM
Completed Dec 9, 2021, 1:51:31 PM
Run name mlmd-pipeline-small-20211209213113
Pipeline name mlmd-pipeline
Runtime environment Serverless
Region us-central1
Service account 196373151126-compute@developer.gserviceaccount.com
Debugging info View pipeline proto
Run Parameters
Pipeline parameter values used for this run
Parameter Type Value
project string cmpe260-334300
output_data_path string data.csv
region string us-central1
bq_table string sara-vertex-demos.beans_demo.small_dataset
Run metrics
Metrics logged by this pipeline run

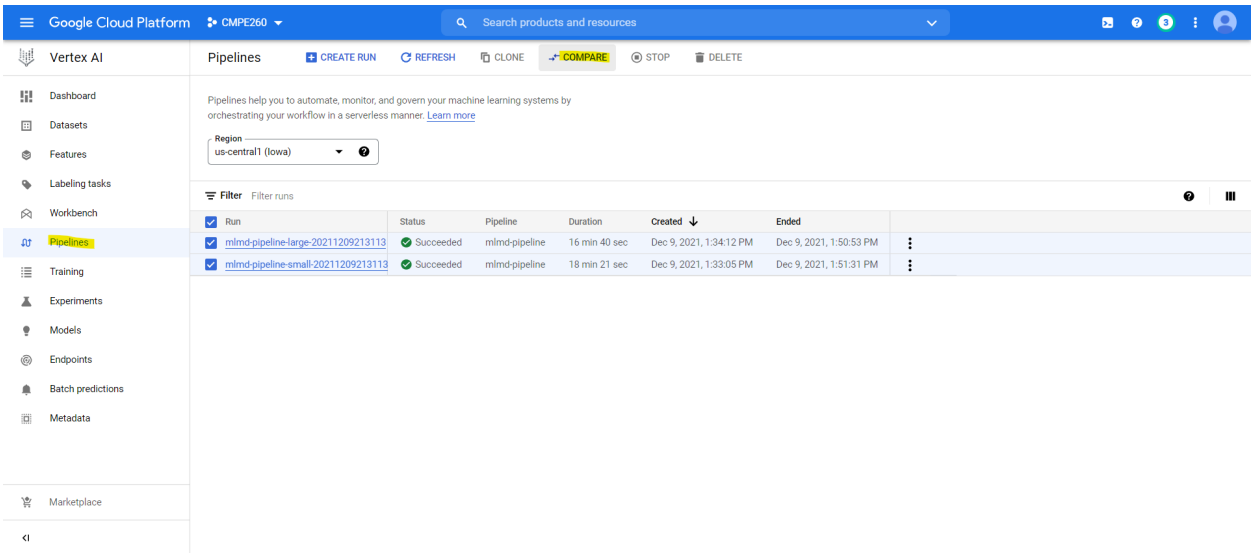

Google Cloud Platform
CMPE260
Search products and resources
Vertex AI
mlmd-pipeline-large-20211209213113
CLONE
STOP
DELETE
Runtime Graph
3/3 steps completed
Expand Artifacts
75%
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
get-dataframe
python:3.9
sklearn-train
python:3.9
deploy-model
python:3.9
Logs
Pipeline run analysis
SUMMARY
NODE INFO
Basic info
Duration 16 min 40 sec
Started Dec 9, 2021, 1:34:12 PM
Completed Dec 9, 2021, 1:50:53 PM
Run name mlmd-pipeline-large-20211209213113
Pipeline name mlmd-pipeline
Runtime environment Serverless
Region us-central1
Service account 196373151126-compute@developer.gserviceaccount.com
Debugging info View pipeline proto
Run Parameters
Pipeline parameter values used for this run
Parameter Type Value
bq_table string sara-vertex-demos.beans_demo.large_dataset
region string us-central1
project string cmpe260-334300
output_data_path string data.csv
Run metrics
Metrics logged by this pipeline run

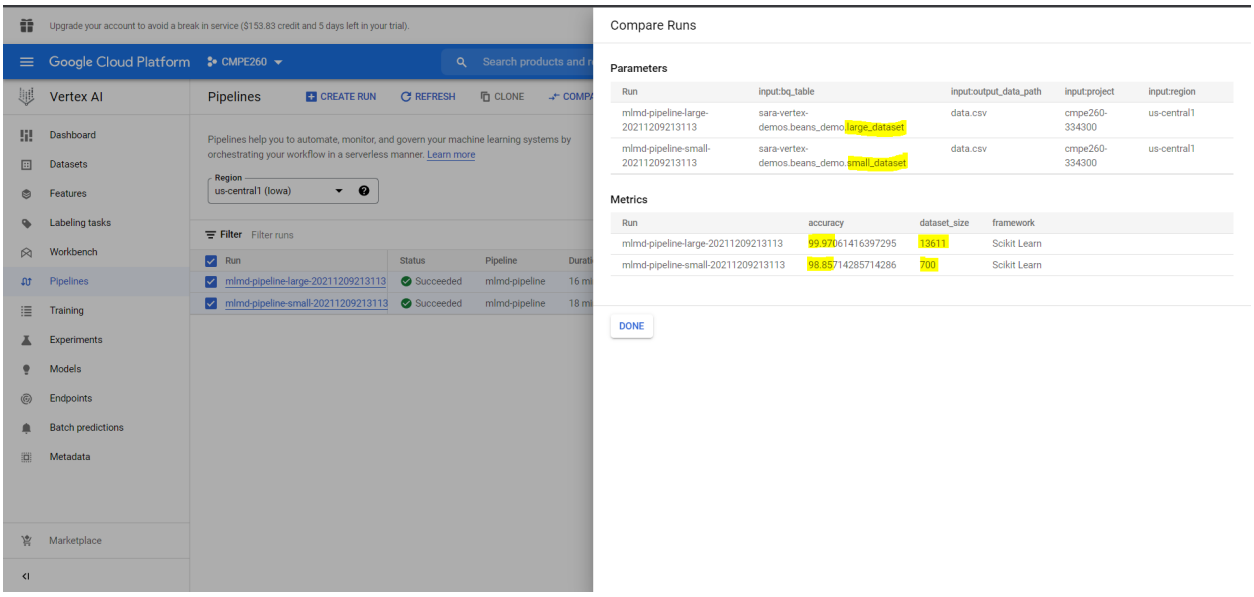

Google Cloud Platform
CMPE260
Search products and resources
Vertex AI
mlmd-pipeline-small-20211209213113
CLONE
STOP
DELETE
Runtime Graph
3/3 steps completed
Expand Artifacts
75%
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
output_data_path
Type: system.Dataset
sklearn-train
python:3.9
model
Type: system.Model
metrics
Type: system.Metrics
deploy-model
python:3.9
vertex_endpoint
Type: system.Artifact
vertex_model
Type: system.Model
Logs
Pipeline run analysis
SUMMARY
NODE INFO
Artifact Info
VIEW LINEAGE
Name
vertex_endpoint
Type
system.Artifact
URI
projects/196373151126/locations/us-central1/endpoints/3282992186957758464


Google Cloud Platform
CMPE260
Search products and resources
Vertex AI
Endpoints
CREATE ENDPOINT
REFRESH
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.
To create an endpoint, you need at least one machine learning model. Learn more
Region
us-central1 (Iowa)
Filter
Enter a property name
Name
ID
Status
Models
Region
Monitoring
Most recent alerts
Last updated
API
Notification
Labels
beans-model-pipeline_endpoint
1416461247645220864
Active
1
us-central1
Disabled
–
Dec 9, 2021, 1:50:03 PM
Sample request
beans-model-pipeline_endpoint
3282992186957758464
Active
1
us-central1
Disabled
–
Dec 9, 2021, 1:49:59 PM
Sample request
titanic-endpoint
432424729064767488
Ready
0
us-central1
Disabled
–
Dec 6, 2021, 10:02:08 PM
Sample request

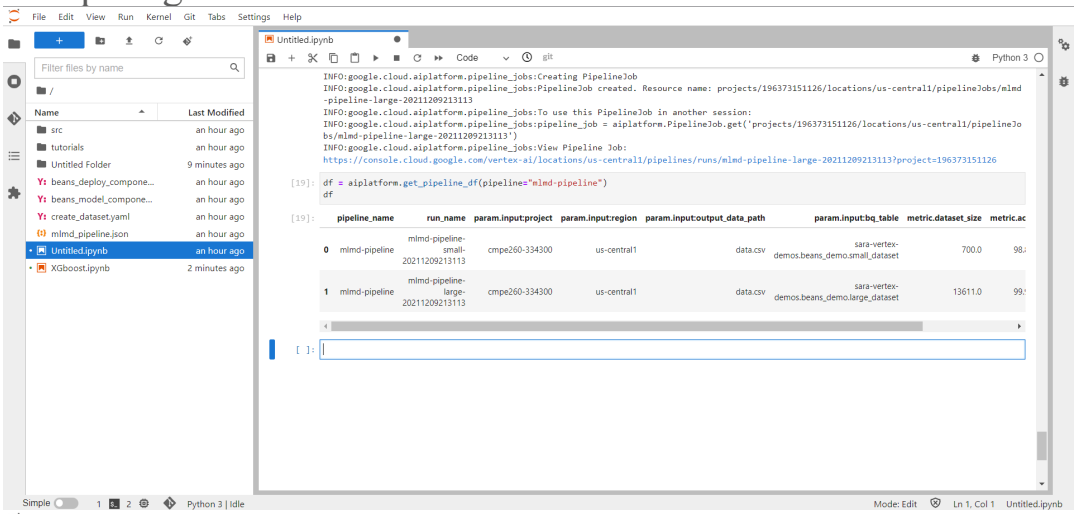


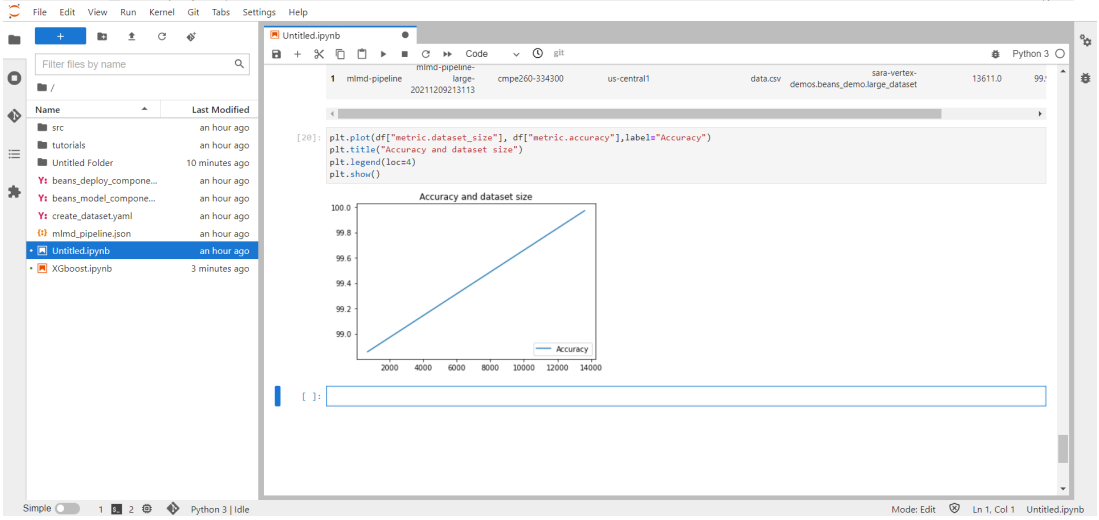
Comparing pipeline runs



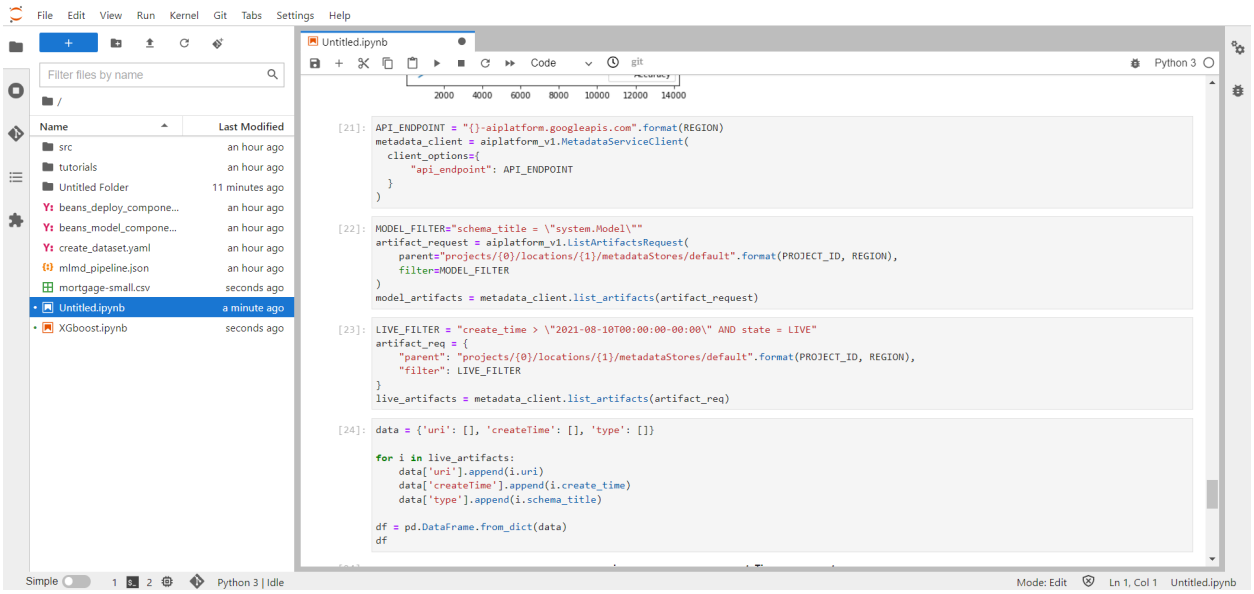


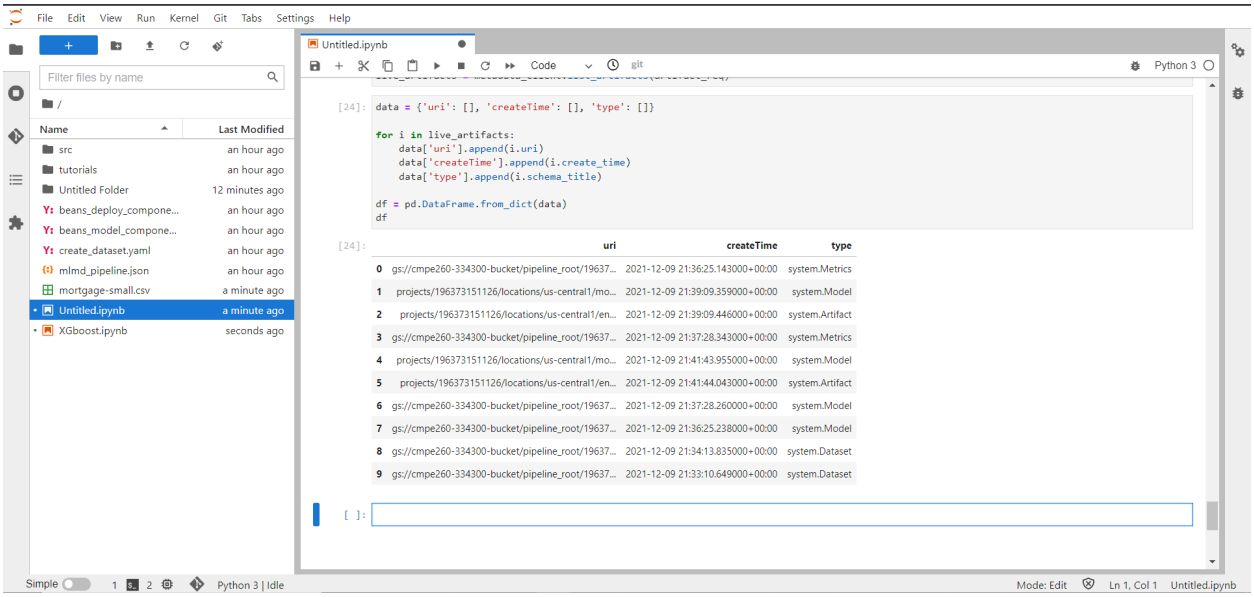
Comparing runs with the Vertex AI SDK





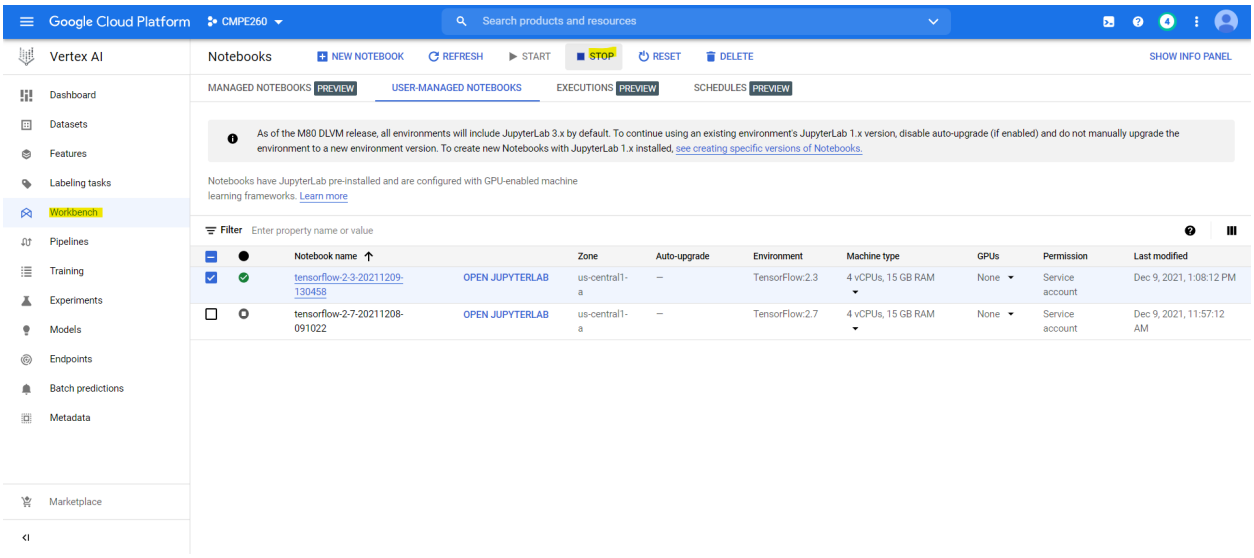
Querying pipeline metrics



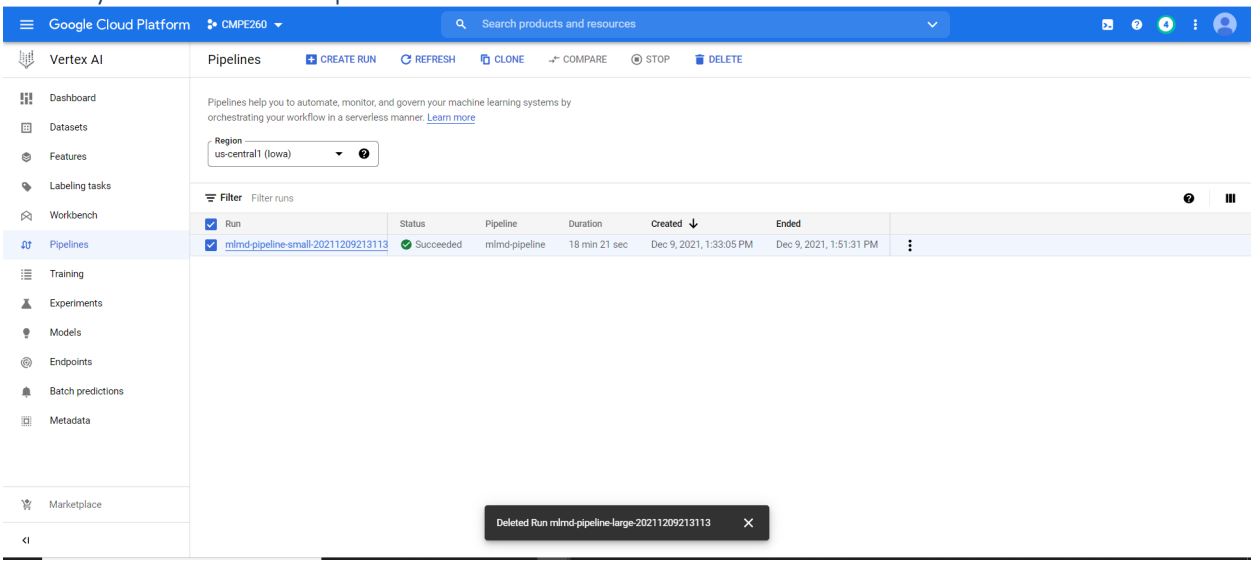


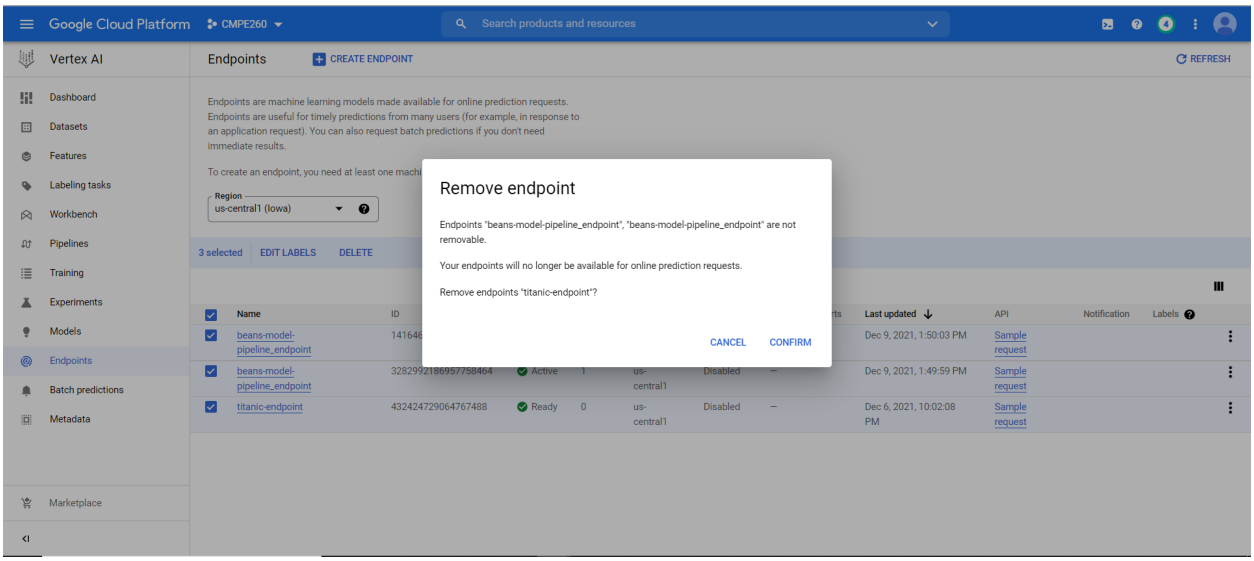
Cleanup

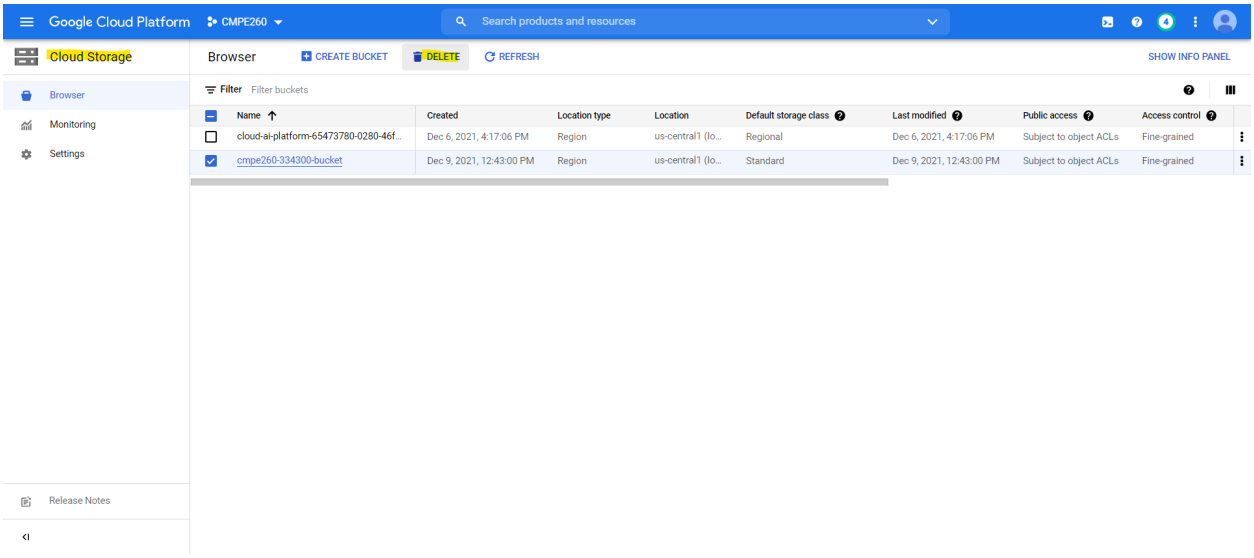
Stop or delete your Notebooks instance



Delete your Vertex AI endpoints




Google Cloud Platform
CMPE260
Search products and resources
Vertex AI
Endpoints
CREATE ENDPOINT
REFRESH
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Endpoints are machine learning models made available for online prediction requests. Endpoints are useful for timely predictions from many users (for example, in response to an application request). You can also request batch predictions if you don't need immediate results.
To create an endpoint, you need at least one machi
Region
us-central1 (Iowa)
3 selected
EDIT LABELS
DELETE
Remove endpoint
Endpoints "beans-model-pipeline_endpoint", "beans-model-pipeline_endpoint" are not removable.
Your endpoints will no longer be available for online prediction requests.
Remove endpoints "titanic-endpoint"?
CANCEL
CONFIRM
Name
ID
Last updated
API
Notification
Labels
beans-model-pipeline_endpoint
Dec 9, 2021, 1:50:03 PM
Sample request
beans-model-pipeline_endpoint
Active
us-central1
Disabled
Dec 9, 2021, 1:49:59 PM
Sample request
titanic-endpoint
432424729064767488
Ready
0
us-central1
Disabled
Dec 6, 2021, 10:02:08 PM
Sample request


Google Cloud Platform
CMPE260
Search products and resources
Cloud Storage
Browser
CREATE BUCKET
DELETE
REFRESH
SHOW INFO PANEL
Browser
Monitoring
Settings
Release Notes
Filter
Filter buckets
Name
Created
Location type
Location
Default storage class
Last modified
Public access
Access control
cloud-ai-platform-65473780-0280-46f...
Dec 6, 2021, 4:17:06 PM
Region
us-central1 (lo...
Regional
Dec 6, 2021, 4:17:06 PM
Subject to object ACLs
Fine-grained
cmpe260-334300-bucket
Dec 9, 2021, 12:43:00 PM
Region
us-central1 (lo...
Standard
Dec 9, 2021, 12:43:00 PM
Subject to object ACLs
Fine-grained