

a) Custom training job and prediction using managed datasets

Reference: <https://codelabs.developers.google.com/codelabs/vertex-ai-custom-code-training#0>

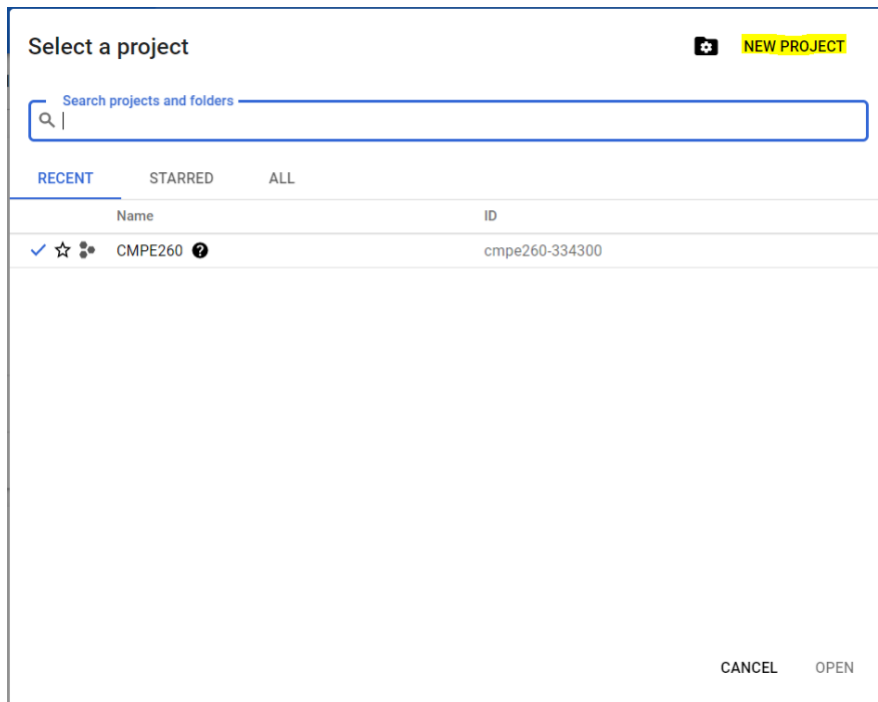
Objectives:

- Use Vertex AI to train and deploy a ML model
- Use Datasets for dataset creation and management, and custom model for training a Scikit Learn model.
- Deploy the trained model and get online predictions

Setup the environment

Create a project

- To create a project, check if our Role has the `resourcemanager.projects.create` permission
- Login to Cloud Console.
- On the **Select organization** drop-down list at the top of the page, select the organization in which you want to create a project. If you are a free trial user, skip this step, as this list does not appear.
- Click **Create Project**.



- In the **New Project** window that appears, enter a project name and select a billing account as applicable. A project name can contain only letters, numbers, single quotes, hyphens, spaces, or exclamation points, and must be between 4 and 30 characters.

Google Cloud Platform

Search products and resources

New Project

You have 22 projects remaining in your quota. Request an increase or delete projects. [Learn more](#)
[MANAGE QUOTAS](#)

Project name *
 My Project 8894

Project ID: acquired-subset-334316. It cannot be changed later. [EDIT](#)

Location *
 No organization [BROWSE](#)

Parent organization or folder

[CREATE](#) [CANCEL](#)

- Enter the parent organization or folder in the **Location** box. That resource will be the hierarchical parent of the new project.
- When you're finished entering new project details, click **Create**.

Load data in BigQuery

In order to train a Machine Learning model, you need access to data. BigQuery is a serverless, highly scalable, and cost-effective multi-cloud data warehouse and it is the perfect service for keeping your data.

Create dataset

1. Make sure that you select the right project from the top of console page
2. Navigate to Big Query
3. Select the project you want to create the Dataset in
4. Click Create Dataset

Upgrade your account to avoid a break in service (\$159.65 credit and 9 days left in your trial). [LEARN MORE](#) [UPGRADE](#)

Google Cloud Platform CMPE260

Navigation menu: RESOURCES & INFO, SHORTCUT, DISABLE EDITOR TABS

Explorer + ADD DATA

cmpe260-334300

COMPOSE NEW QUERY

CREATE DATASET PIN PROJECT

Create dataset

Dataset ID
 Letters, numbers, and underscores allowed

Data location (Optional)
 Default

Default table expiration
☒ Never
☐ Number of days after table creation:

Encryption
 Data is encrypted automatically. Select an encryption key management solution.
☒ Google-managed key
 No configuration required
☐ Customer-managed key
 Manage via Google Cloud Key Management Service

[Create dataset](#) [Cancel](#)

Resources in this project

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Upgrade your account to avoid a break in service (\$159.65 credit and 8 days left in your trial).

Google Cloud Platform CMPE260 Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

Type to search

Viewing pinned projects.

cmpe260-334300

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Create dataset

Project ID
cmpe260-334300 [CHANGE](#)

Dataset ID *
titanic
Letters, numbers, and underscores allowed

Data location
us-east1 (South Carolina)

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](#)

5. Select the 'titanic' dataset created above
6. Click 'Create Table'

From the Sidebar select the following:

7. Create table from: Upload
8. Select file: Use the downloaded titanic dataset
9. File Format: CSV
10. Table Name: survivors
11. Auto-detect: Select auto-detect checkbox - Schema and input parameters
12. Click 'Create Table' button.

Upgrade your account to avoid a break in service (\$159.65 credit and 8 days left in your trial).

Google Cloud Platform CMPE260

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

Type to search

Viewing pinned projects.

cmpe260-334300

titanic

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Create table

Source

Create table from
Upload

Select file *
titanic.csv [X](#) [BROWSE](#) ?

File format
CSV

Destination

Project *
cmpe260-334300 [BROWSE](#)

Dataset *
titanic

Table *
survivors
Unicode letters, marks, numbers, connectors, dashes or spaces allowed.

Table type
Native table

Schema

☒ Auto detect

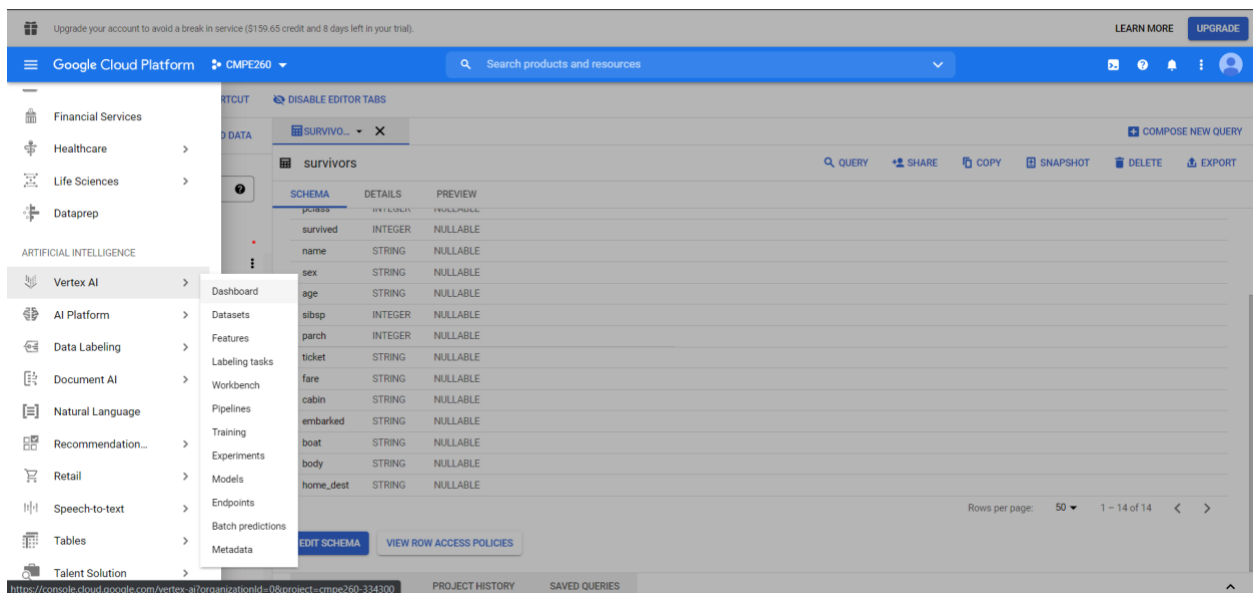
[CREATE TABLE](#) [CANCEL](#)

Create a dataset

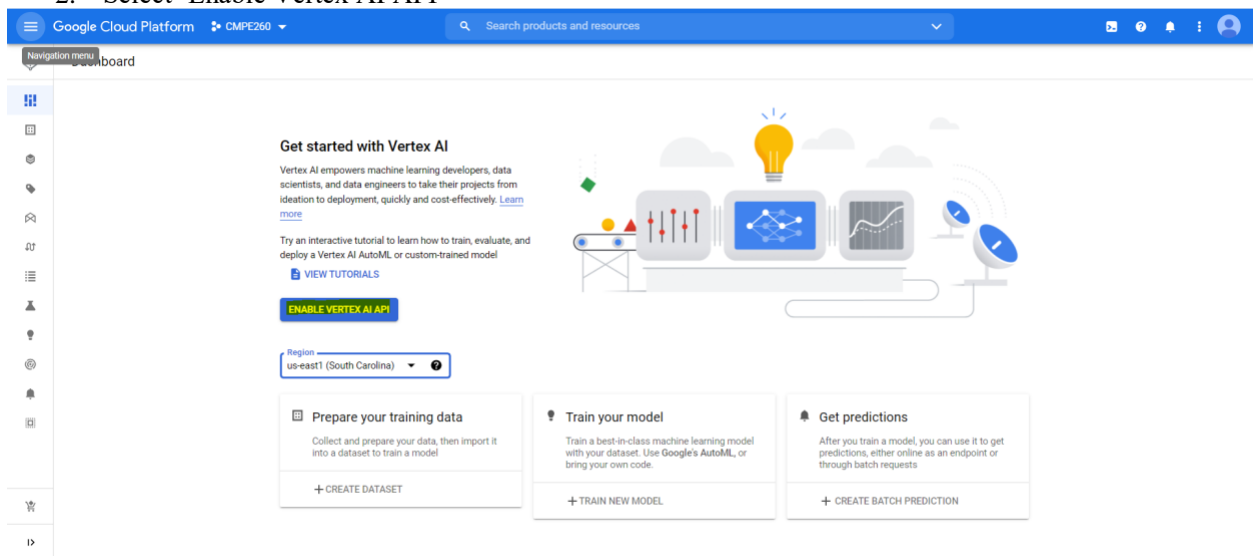
Datasets in Vertex AI allow you to create datasets for your Machine Learning workloads. You can create datasets for structured data (CSV files or BigQuery tables) or unstructured data such as Images and Text.

Create ML dataset

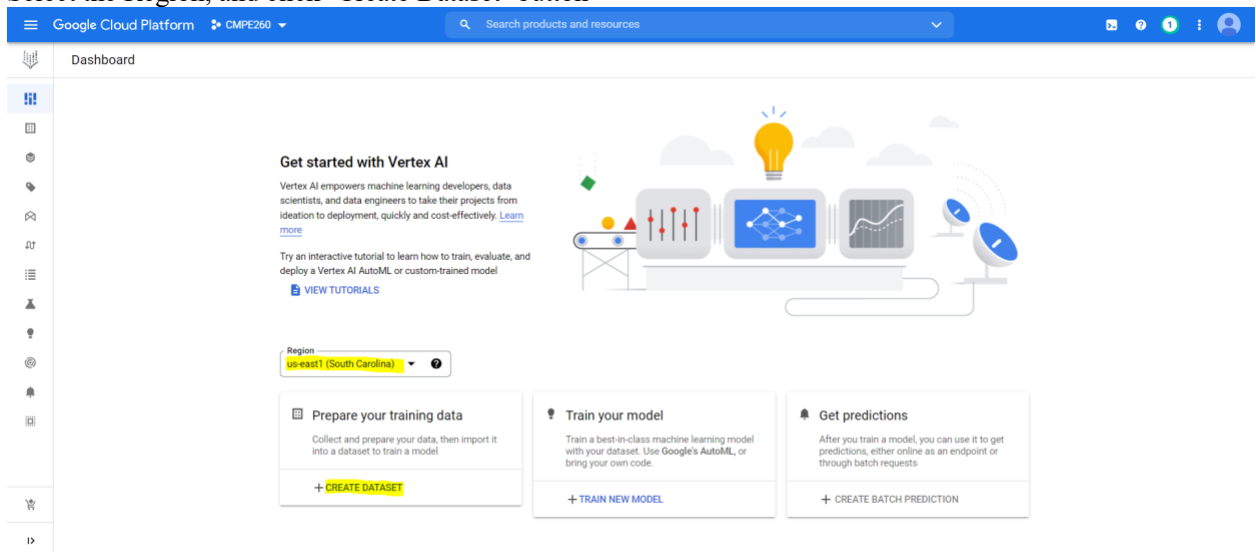
1. Find Vertex AI on the GCP side menu, under Artificial Intelligence



2. Select 'Enable Vertex AI API'

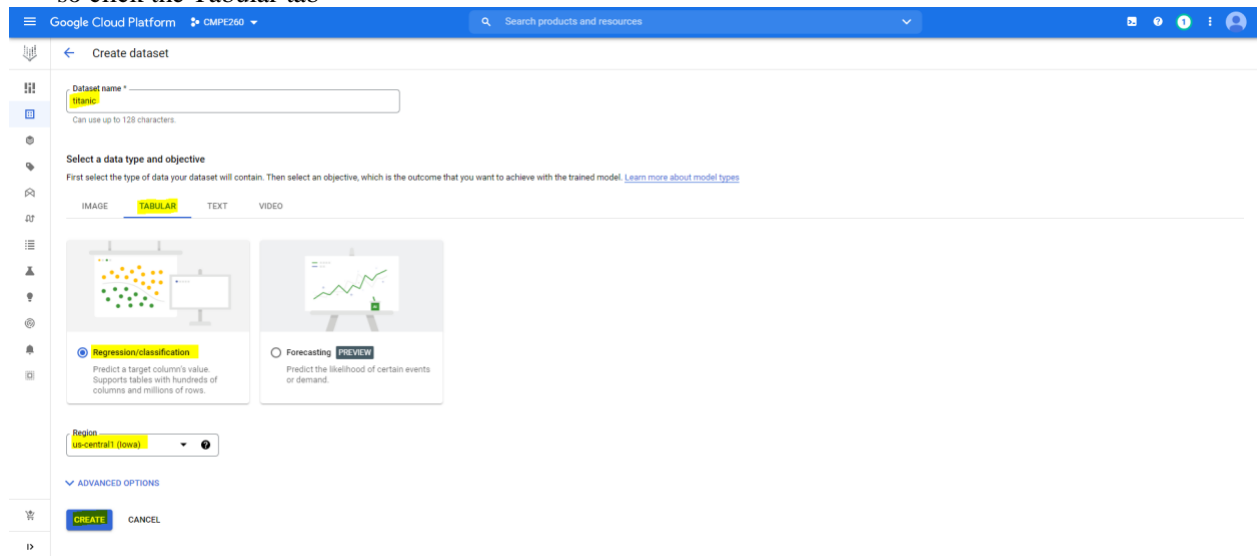


3. Select the Region, and click 'Create Dataset' button



4. Enter dataset name as 'titanic'

5. We can create datasets for images, text or videos as well as tabular data. The 'titanic' dataset is tabular so click the Tabular tab

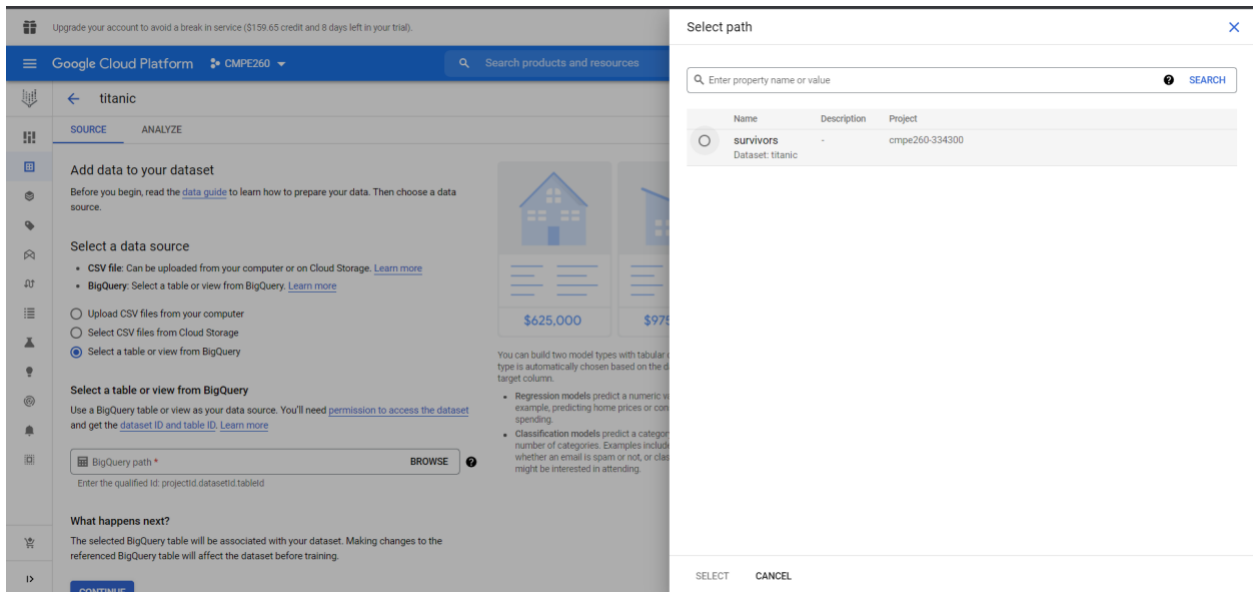


6. Select 'Tabular' and 'Regression/Classification'

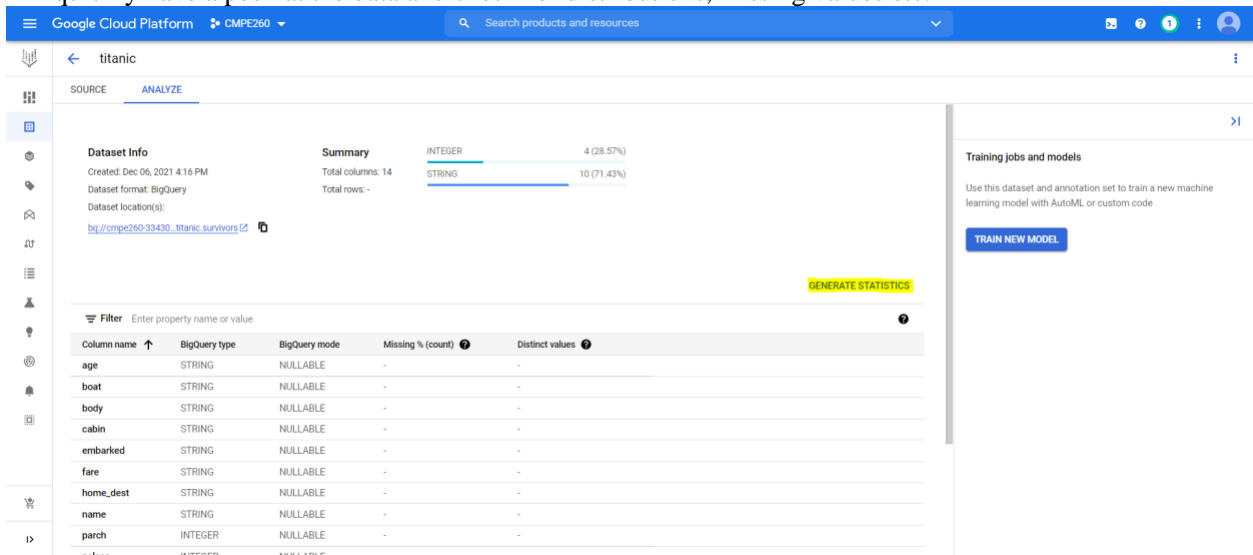
7. Click the 'Create' button

Select datasource

1. As we had already loaded the titanic dataset in BigQuery, we can connect our ML dataset to our BigQuery table.
2. Select 'Select a table or view from BigQuery'
3. Select the BigQuery Path
4. Select the 'survivors' table.



- Click 'Continue'
- In the 'Analyze' tab we can generate statistics regarding your data. This gives you the ability to quickly have a peek at the data and check for distributions, missing values etc.



- Click on 'Generate Statistics'

Google Cloud Platform CMPE260

Search products and resources

titanic

SOURCE ANALYZE

General statistics generated by Dec 06, 2021 4:35 PM [GENERATE STATISTICS](#)

Filter Enter property name or value

Column name	BigQuery type	BigQuery mode	Missing % (count)	Distinct values
age	STRING	NULLABLE	-	99
boat	STRING	NULLABLE	-	28
body	STRING	NULLABLE	-	122
cabin	STRING	NULLABLE	-	187
embarked	STRING	NULLABLE	-	4
fare	STRING	NULLABLE	-	282
home_dest	STRING	NULLABLE	-	370
name	STRING	NULLABLE	-	1307
parch	INTEGER	NULLABLE	-	8
pclass	INTEGER	NULLABLE	-	3
sex	STRING	NULLABLE	-	2
sibsp	INTEGER	NULLABLE	-	7
survived	INTEGER	NULLABLE	-	2
ticket	STRING	NULLABLE	-	929

Rows per page: 50 1 - 14 of 14

Training jobs and models

Use this dataset and annotation set to train a new machine learning model with AutoML or custom code

[TRAIN NEW MODEL](#)

8. Statistics are generated.

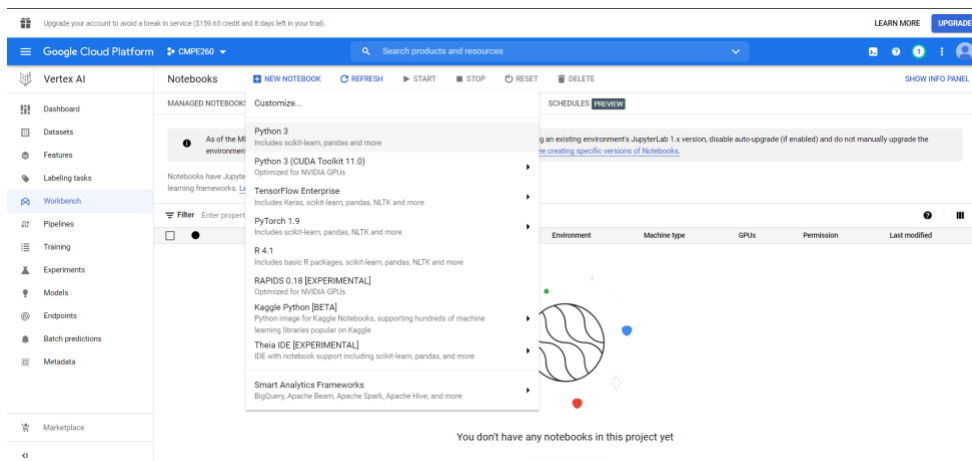
Custom training package using Notebooks

It is a good practice to package and parameterise your code so that it becomes a portable asset.

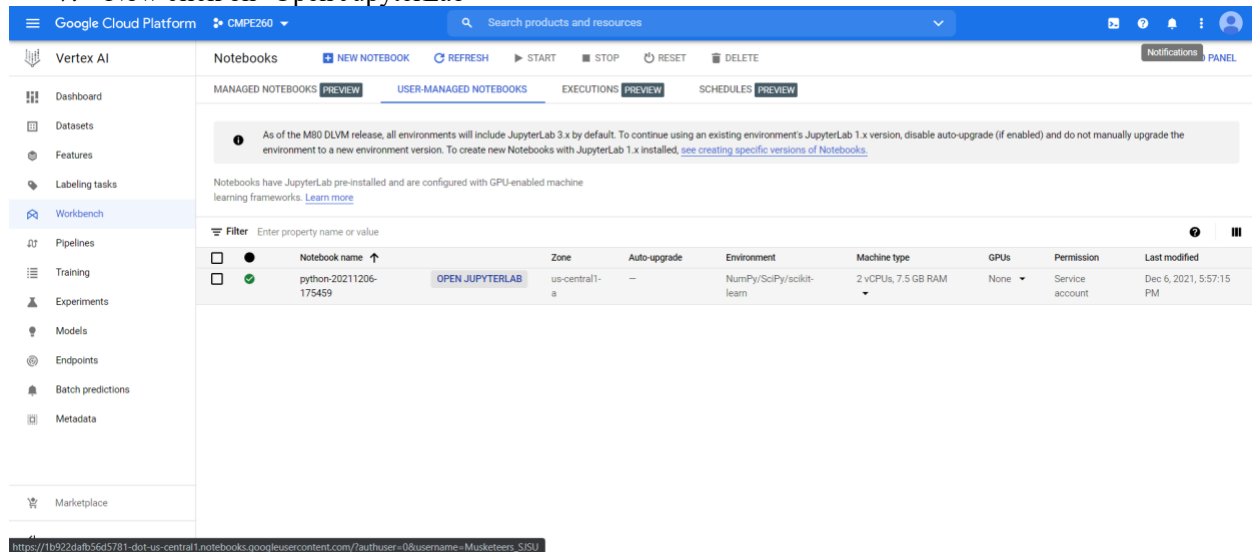
1. Create a training package with custom code using Notebooks

Create your notebook instance

1. From the Vertex AI navigate to notebooks and start an instance with **Python 3**, which includes scikit-learn
2. Select 'Vertex AI', Workbench
3. Enable Notebooks
4. On the top, select 'New Notebook'



5. Select 'Python3', with scikit learn module.
6. The Notebook instance is created.
7. Now click on 'Open JupyterLab'



Google Cloud Platform CMPE260

Search products and resources

Vertex AI

Notebooks

NEW NOTEBOOK REFRESH START STOP RESET DELETE

MANAGED NOTEBOOKS PREVIEW USER-MANAGED NOTEBOOKS EXECUTIONS PREVIEW SCHEDULES PREVIEW

As of the M80 DLVM release, all environments will include JupyterLab 3.x by default. To continue using an existing environment's JupyterLab 1.x version, disable auto-upgrade (if enabled) and do not manually upgrade the environment to a new environment version. To create new Notebooks with JupyterLab 1.x installed, see creating specific versions of Notebooks.

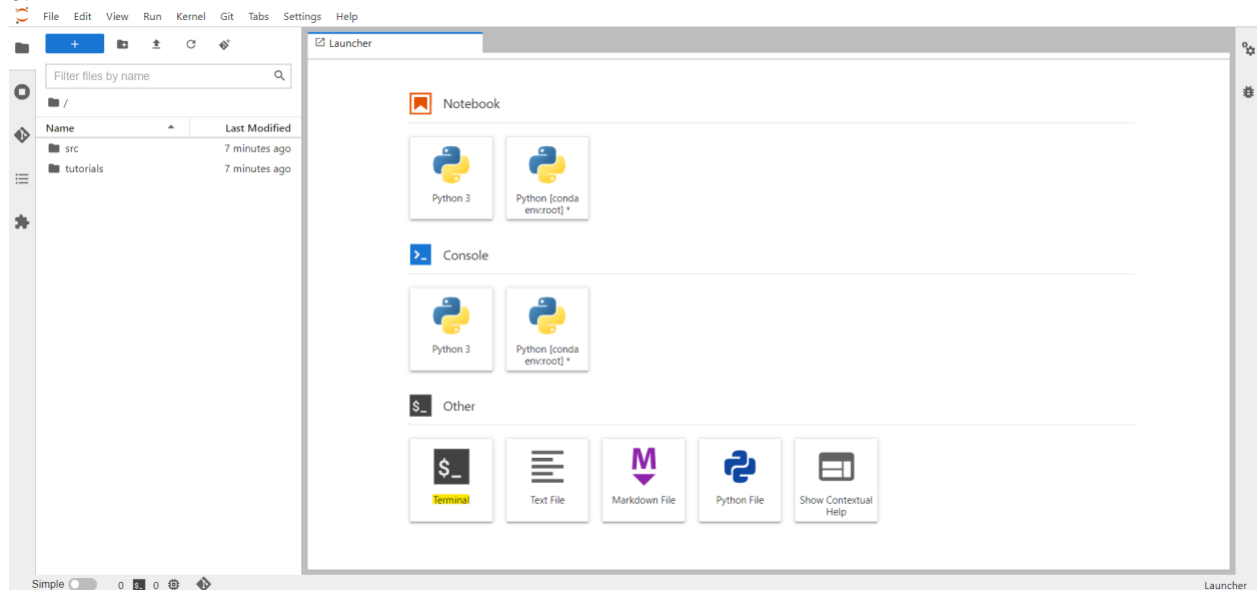
Notebooks have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

Filter Enter property name or value

Notebook name	Zone	Auto-upgrade	Environment	Machine type	GPUs	Permission	Last modified
python-20211206-175459	us-central1-a	—	NumPy/SciPy/scikit-learn	2 vCPUs, 7.5 GB RAM	None	Service account	Dec 6, 2021, 5:57:15 PM

https://1b9224dfb56d5781-dot-us-central1.notebooks.googleusercontent.com/?authuser=0&username=Musketiers_5JSU

8. Select 'Terminal' in this window.



File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name	Last Modified
src	7 minutes ago
tutorials	7 minutes ago

Launcher

Notebook

Python 3 Python [conda envroot] *

Console

Python 3 Python [conda envroot] *

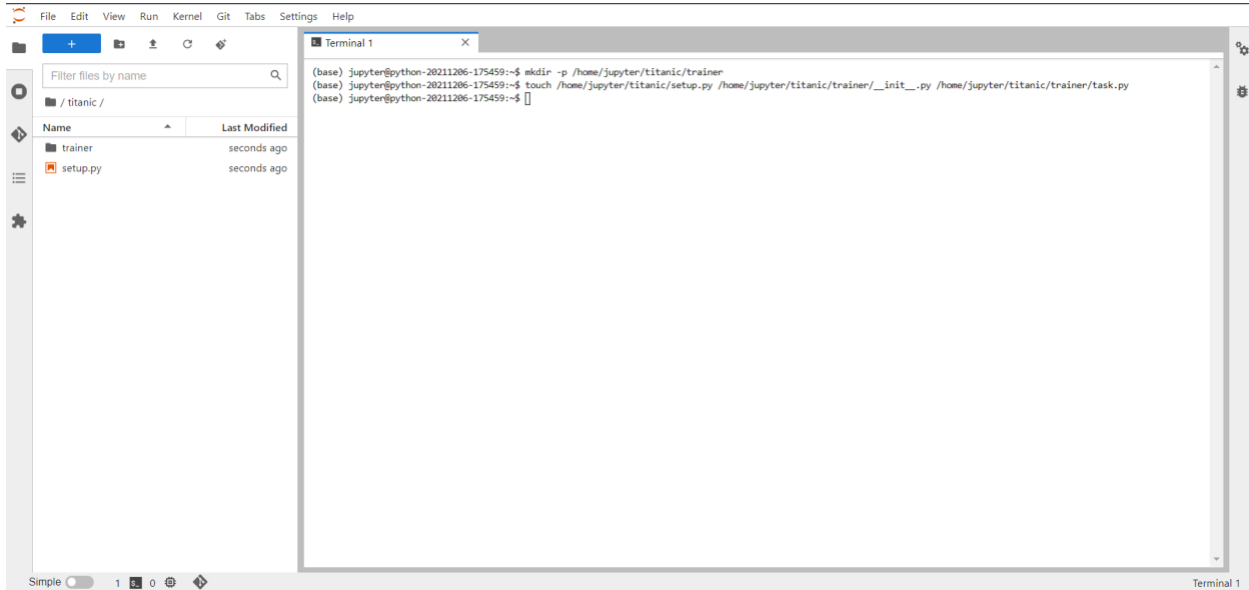
Other

Terminal Text File Markdown File Python File Show Contextual Help

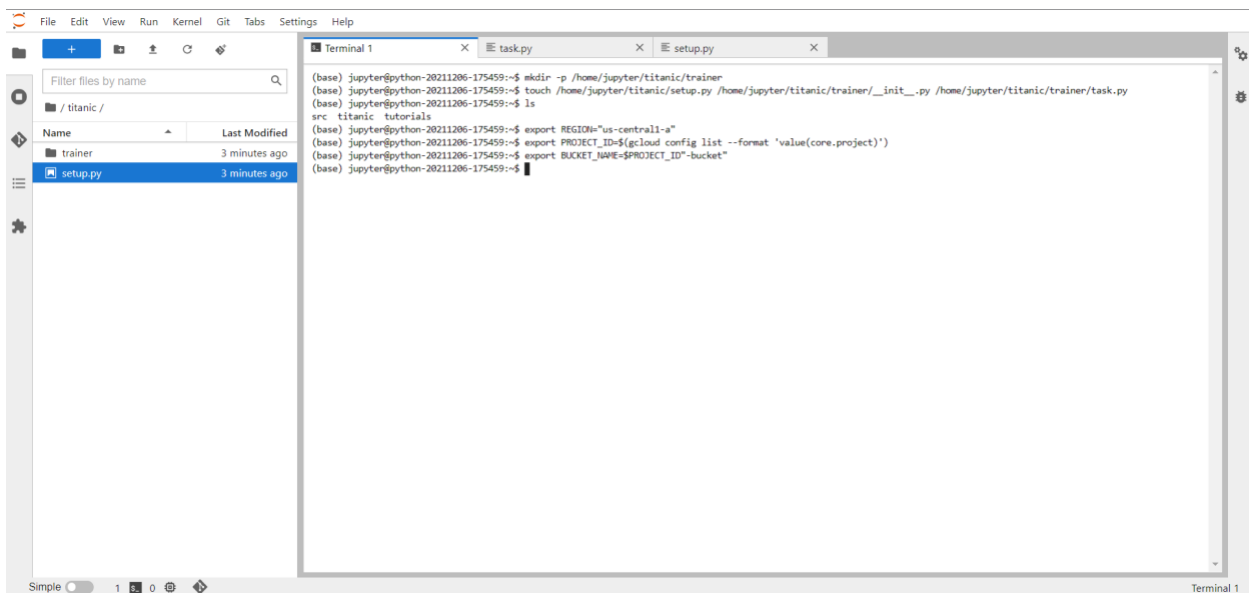
Simple Launcher

9. Run the following command to create the directory structure for our project.

```
mkdir -p /home/jupyter/titanic/trainer
touch /home/jupyter/titanic/setup.py /home/jupyter/titanic/trainer/__init__.py
/home/jupyter/titanic/trainer/task.py
```

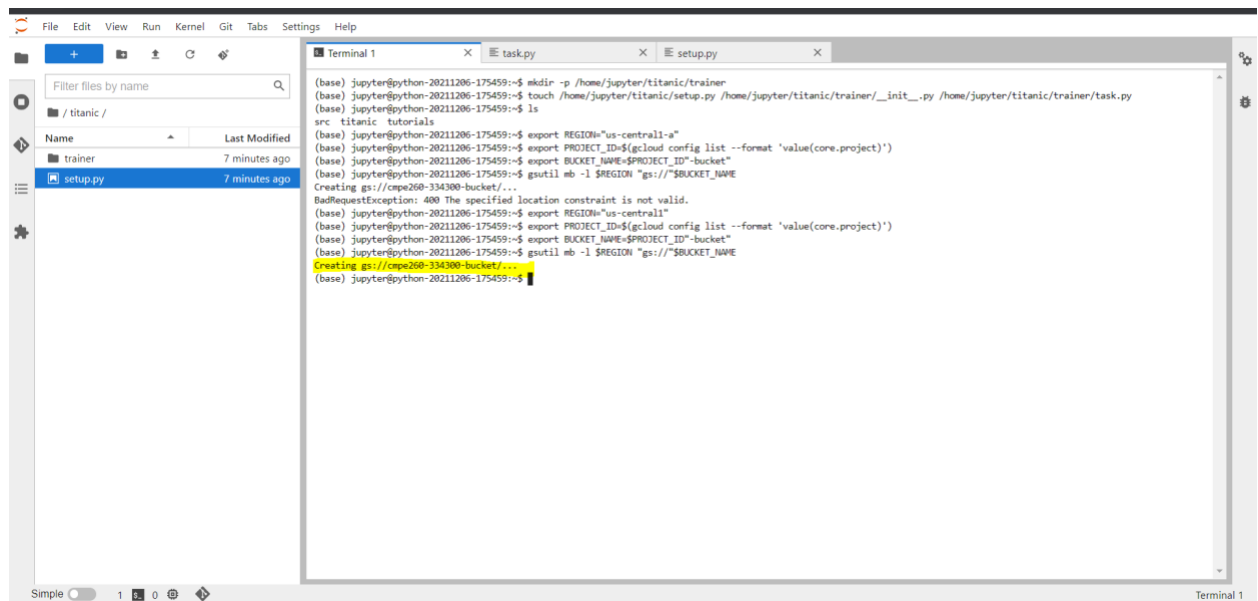


10. Create the code for 'task.py' and 'setup.py', to train and run our model
11. Setup the environment variables, by running the below commands:
export REGION="us-central1"
export PROJECT_ID=\$(gcloud config list --format 'value(core.project)')
export BUCKET_NAME=\$PROJECT_ID"-bucket"



12. Create a bucket where you want to export your trained model

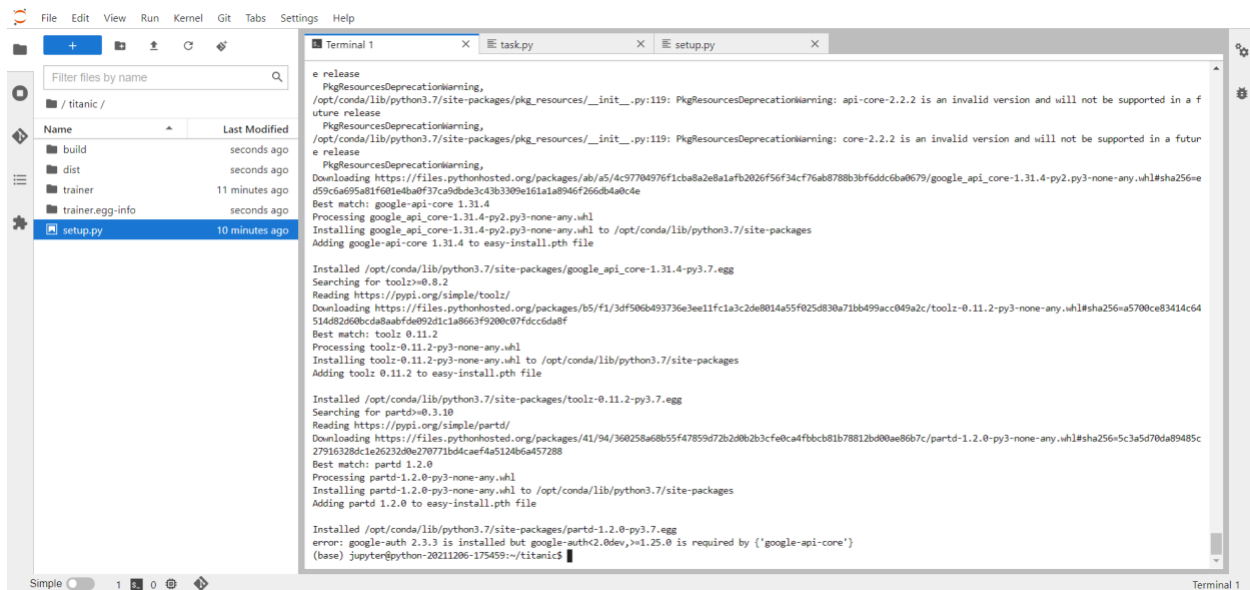
Run the command: `gsutil mb -l $REGION "gs://"$BUCKET_NAME`



```
(base) jupyter@python-20211206-175459:~$ mkdir -p /home/jupyter/titanic/trainer
(base) jupyter@python-20211206-175459:~$ touch /home/jupyter/titanic/trainer/__init__.py /home/jupyter/titanic/trainer/task.py
(base) jupyter@python-20211206-175459:~$ ls
src  titanic  tutorials
(base) jupyter@python-20211206-175459:~$ export REGION="us-central1-a"
(base) jupyter@python-20211206-175459:~$ export PROJECT_ID=$(gcloud config list --format 'value(core.project)')
(base) jupyter@python-20211206-175459:~$ export BUCKET_NAME=$PROJECT_ID-"bucket"
(base) jupyter@python-20211206-175459:~$ gsutil mb -l $REGION "gs://"$BUCKET_NAME
Creating gs://comp260-334300-bucket/...
BadRequestException: 400 The specified location constraint is not valid.
(base) jupyter@python-20211206-175459:~$ export REGION="us-central1"
(base) jupyter@python-20211206-175459:~$ export PROJECT_ID=$(gcloud config list --format 'value(core.project)')
(base) jupyter@python-20211206-175459:~$ export BUCKET_NAME=$PROJECT_ID-"bucket"
(base) jupyter@python-20211206-175459:~$ gsutil mb -l $REGION "gs://"$BUCKET_NAME
Creating gs://comp260-334300-bucket/...
(base) jupyter@python-20211206-175459:~$
```

13. Install the required libraries

```
cd /home/jupyter/titanic
pip install setuptools
python setup.py install
```



```
e release
PkgResourcesDeprecationWarning:
/opt/conda/lib/python3.7/site-packages/pkg_resources/_init_.py:119: PkgResourcesDeprecationWarning: api-core-2.2.2 is an invalid version and will not be supported in a future release
/opt/conda/lib/python3.7/site-packages/pkg_resources/_init_.py:119: PkgResourcesDeprecationWarning: core-2.2.2 is an invalid version and will not be supported in a future release
e release
PkgResourcesDeprecationWarning:
/opt/conda/lib/python3.7/site-packages/pkg_resources/_init_.py:119: PkgResourcesDeprecationWarning: core-2.2.2 is an invalid version and will not be supported in a future release
Downloading https://files.pythonhosted.org/packages/ab/a5/4c9704976f1c8a2e8a1af2026f56f34cf76ab8788b3bf6ddc8ba0679/google_api_core-1.31.4-py2.py3-none-any.whl#sha256=ed59c6a695a81f601e4ba9f37ca9dbd3c43b3309e161a1a8946f266db4a0cde
Best match: google-api-core 1.31.4
Processing google_api_core-1.31.4-py2.py3-none-any.whl
Installing google_api_core-1.31.4-py2.py3-none-any.whl to /opt/conda/lib/python3.7/site-packages
Adding google-api-core 1.31.4 to easy-install.pth file

Installed /opt/conda/lib/python3.7/site-packages/google_api_core-1.31.4-py3.7.egg
Searching for toolz>=0.8.2
Reading https://pypi.org/simple/toolz/
Downloading https://files.pythonhosted.org/packages/b5/f1/3df506b493736e3ee11fca3c2de8014a55f025d830a71bb499acc049a2c/toolz-0.11.2-py3-none-any.whl#sha256=a5700ce83414c64514d82d696cda8a8fde92d1ca8663f9208c07f6cc6da8f
Best match: toolz 0.11.2
Processing toolz-0.11.2-py3-none-any.whl
Installing toolz-0.11.2-py3-none-any.whl to /opt/conda/lib/python3.7/site-packages
Adding toolz 0.11.2 to easy-install.pth file

Installed /opt/conda/lib/python3.7/site-packages/toolz-0.11.2-py3.7.egg
Searching for partd>=0.3.10
Reading https://pypi.org/simple/partd/
Downloading https://files.pythonhosted.org/packages/41/94/360258a68855f47859d72b2d8b2b3cfe0ca4fbcb81b78812b00ae86b7c/partd-1.2.0-py3-none-any.whl#sha256=5c3a5d78da99485c2791632841c62328b270771bdcae4a51246a457288
Best match: partd 1.2.0
Processing partd-1.2.0-py3-none-any.whl
Installing partd-1.2.0-py3-none-any.whl to /opt/conda/lib/python3.7/site-packages
Adding partd 1.2.0 to easy-install.pth file

Installed /opt/conda/lib/python3.7/site-packages/partd-1.2.0-py3.7.egg
error: google-auth 2.3.3 is installed but google-auth2.0dev,>=1.25.0 is required by ('google-api-core')
(base) jupyter@python-20211206-175459:~/titanic$
```

14. Run the training code to verify that it executes without issues

```
python -m trainer.task -v \
--model_param_kernel=linear \
--model_dir="gs:/// $BUCKET_NAME /titanic/trial" \
--data_format=bigquery \
--training_data_uri="bq:/// $PROJECT_ID .titanic.survivors" \
--test_data_uri="bq:/// $PROJECT_ID .titanic.survivors" \
--validation_data_uri="bq:/// $PROJECT_ID .titanic.survivors"
```

```
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
titanic /
Name Last Modified
build 2 minutes ago
dist 2 minutes ago
trainer seconds ago
trainer.egg-info 2 minutes ago
setup.py 12 minutes ago

Terminal 1 task.py setup.py
Searching for partd=0.3.10
Reading https://pypi.org/simple/partd/
Downloading https://files.pythonhosted.org/packages/41/94/360258a6855f47859d72b2d0b2b3cfe0ca4fbbcb81b78812b00ae86b7c/partd-1.2.0-py3-none-any.whl#sha256=5c3a5d76da89485c27916328dc1e262320be270771bd4cae4a5124b6a457288
Best match: partd 1.2.0
Processing partd-1.2.0-py3-none-any.whl
Installing partd-1.2.0-py3-none-any.whl to /opt/conda/lib/python3.7/site-packages
Adding partd 1.2.0 to easy-install.pth file

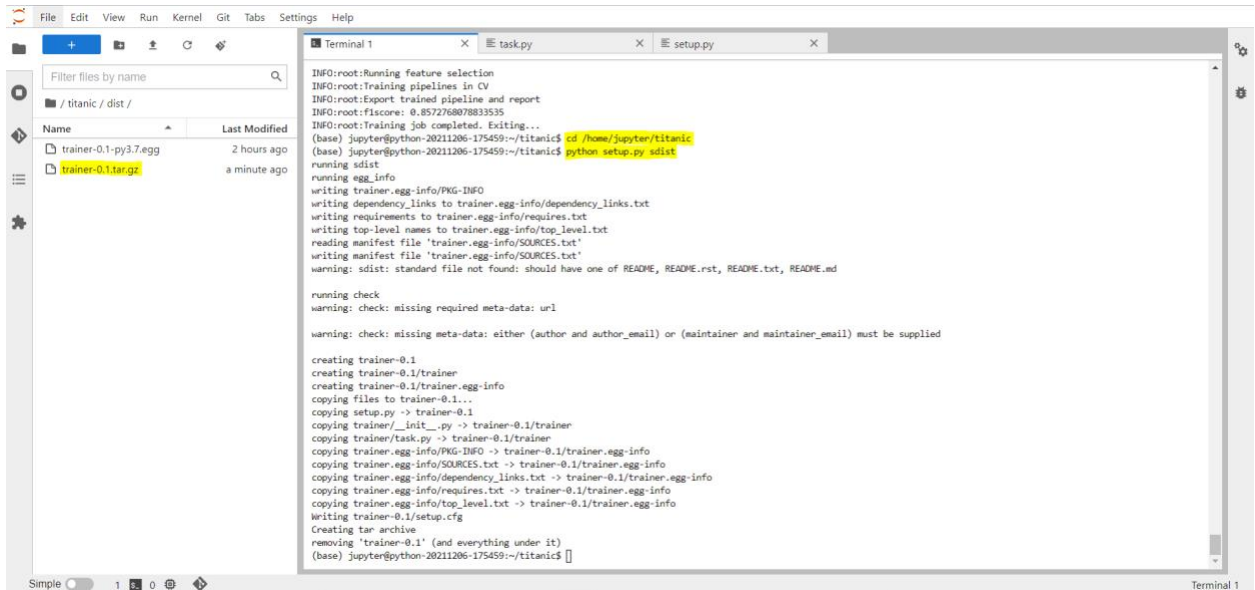
Installed /opt/conda/lib/python3.7/site-packages/partd-1.2.0-py3.7.egg
error: google-auth 2.3.3 is installed but google-auth2.0dev>v1.25.0 is required by ['google-api-core']
(base) jupyter@python-20211206-175459:~/titanic$ python -m trainer.task -v \
> --model_param_kernel=linear \
> --model_dir="gs:/// $BUCKET_NAME /titanic/trial" \
> --data_format=bigquery \
> --training_data_uri="bq:/// $PROJECT_ID .titanic.survivors" \
> --test_data_uri="bq:/// $PROJECT_ID .titanic.survivors" \
> --validation_data_uri="bq:/// $PROJECT_ID .titanic.survivors" \
INFO:root:Model artifacts will be exported here: gs://cmpe260-334300-bucket/titanic/trial
INFO:root:Data format: bigquery
INFO:root:Training data uri: bq://cmpe260-334300.titanic.survivors
INFO:root:Validation data uri: bq://cmpe260-334300.titanic.survivors
INFO:root:Test data uri: bq://cmpe260-334300.titanic.survivors
INFO:root>Loading bigquery data
INFO:root>Loading bigquery data
INFO:root:reading bq data: bq://cmpe260-334300.titanic.survivors
INFO:root:reading bq data: bq://cmpe260-334300.titanic.survivors
INFO:root:reading bq data: bq://cmpe260-334300.titanic.survivors
INFO:root:Defining model parameters
/home/jupyter/titanic/trainer/task.py:104: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
df = df.fillna(df.mean())
INFO:root:Running feature selection
INFO:root:Training pipelines in CV
INFO:root:Export trained pipeline and report
INFO:root:f1score: 0.8572768078813539
INFO:root:Training job completed. Exiting...
(base) jupyter@python-20211206-175459:~/titanic$
```

15. The two lines indicate the f1 score which is around 0.85 and the last line indicating that the training job completed successfully

16. Run the below commands to create the distributable file.

```
cd /home/jupyter/titanic
python setup.py sdist
```

17. Check for the tar.gz file created.



```
INFO:root:Running feature selection
INFO:root:Training pipelines in CV
INFO:root:Export trained pipeline and report
INFO:root:score: 0.857276807883535
INFO:root:Training job completed. Exiting...
(base) jupyter@python-20211206-175459:~/titanic$ cd /home/jupyter/titanic
(base) jupyter@python-20211206-175459:~/titanic$ python setup.py sdist
running sdist
running egg_info
writing trainer.egg-info/PKG-INFO
writing dependency_links to trainer.egg-info/dependency_links.txt
writing requirements to trainer.egg-info/requirements.txt
writing top-level names to trainer.egg-info/top_level.txt
reading manifest file 'trainer.egg-info/SOURCES.txt'
writing manifest file 'trainer.egg-info/SOURCES.txt'
warning: sdist: standard file not found: should have one of README, README.rst, README.txt, README.md

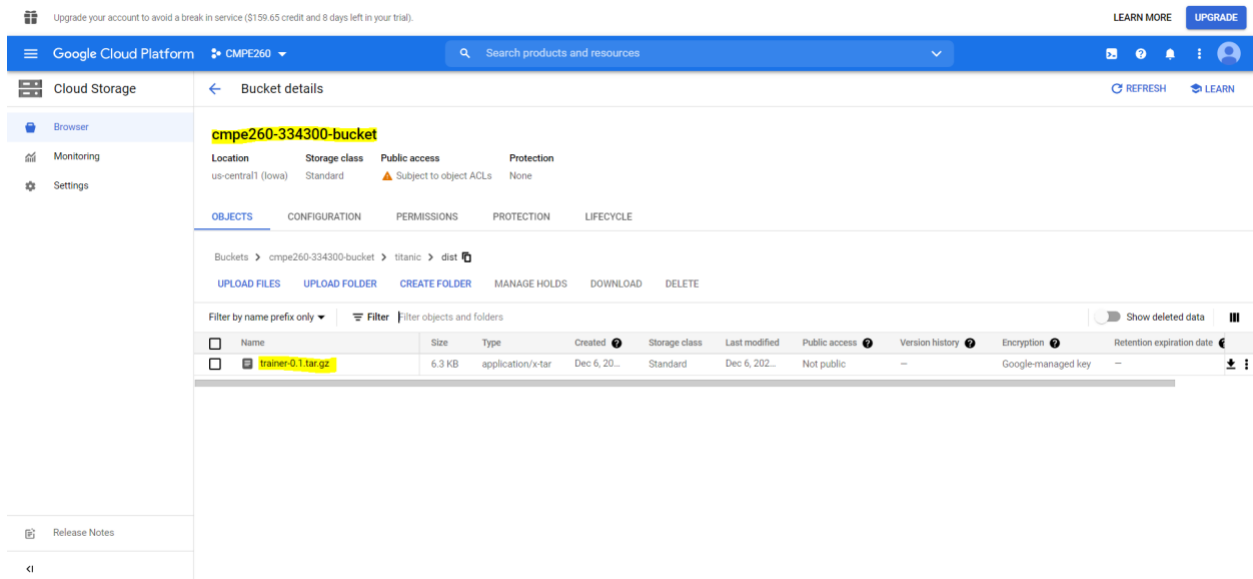
running check
warning: check: missing required meta-data: url

warning: check: missing meta-data: either (author and author_email) or (maintainer and maintainer_email) must be supplied

creating trainer-0.1
creating trainer-0.1/trainer
creating trainer-0.1/trainer.egg-info
copying files to trainer-0.1...
copying setup.py -> trainer-0.1
copying trainer/__init__.py -> trainer-0.1/trainer
copying trainer/task.py -> trainer-0.1/trainer
copying trainer.egg-info/PKG-INFO -> trainer-0.1/trainer.egg-info
copying trainer.egg-info/SOURCES.txt -> trainer-0.1/trainer.egg-info
copying trainer.egg-info/dependency_links.txt -> trainer-0.1/trainer.egg-info
copying trainer.egg-info/requirements.txt -> trainer-0.1/trainer.egg-info
copying trainer.egg-info/top_level.txt -> trainer-0.1/trainer.egg-info
writing trainer-0.1/setup.cfg
Creating tar archive
removing 'trainer-0.1' (and everything under it)
(base) jupyter@python-20211206-175459:~/titanic$
```

18. Copy the tar.gz file to GCS so that the training service can use it to train a new model

```
gsutil cp dist/trainer-0.1.tar.gz "gs://"$BUCKET_NAME"/titanic/dist/trainer-0.1.tar.gz"
```

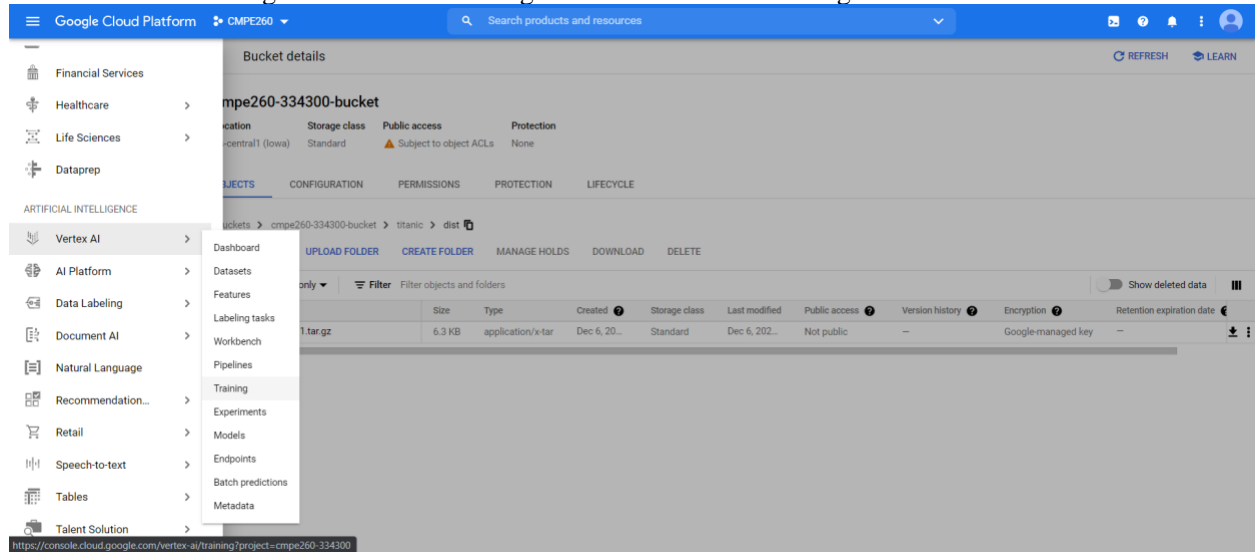


19. Check if the dist file is copied into the bucket.

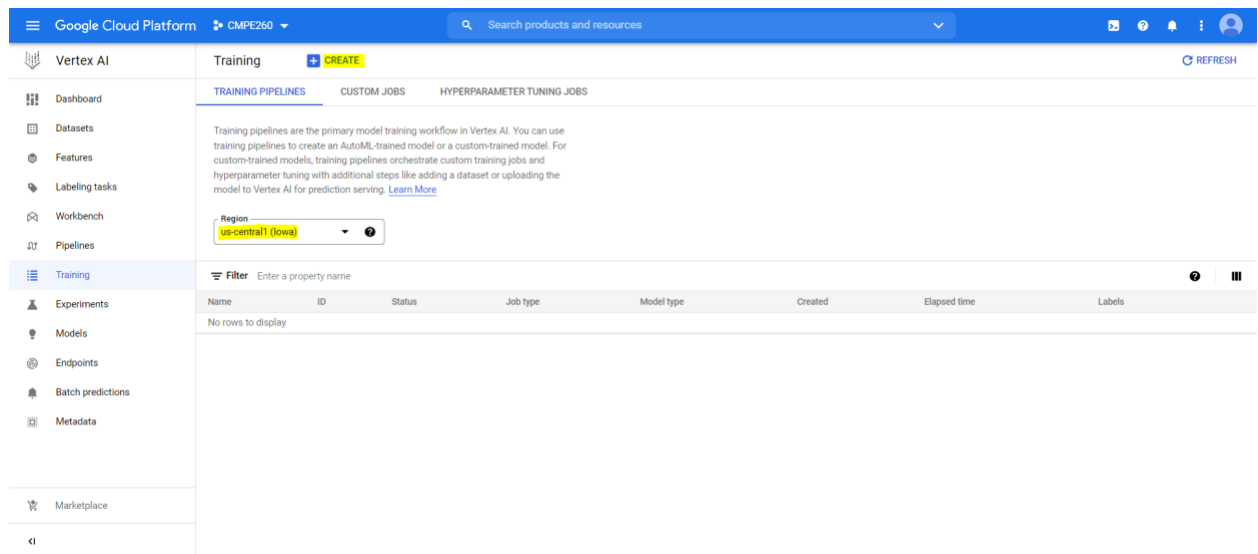
Model Training

In this, we will train a model on Vertex AI

1. From the Google Cloud console navigate to Vertex AI -> Training



2. Select the Region and select 'Create'



Step 1: Training method

Select the DataSet as 'titanic', Objective as 'Classification'

The screenshot shows the 'Train new model' interface in the Google Cloud Platform. The left sidebar lists various Vertex AI services, with 'Training' selected. The main panel is titled 'Train new model' and has a progress bar with six steps: 1. Training method (selected), 2. Model details, 3. Training container, 4. Hyperparameters (optional), 5. Compute and pricing, and 6. Prediction container (optional). Below the progress bar are 'START TRAINING' and 'CANCEL' buttons. The right panel shows the 'Dataset' dropdown set to 'titanic' and the 'Objective' dropdown set to 'Classification'. Below these, there is a note about pricing and two options: 'AutoML' (with a description and a 'Learn more' link) and 'Custom training (advanced)' (with a description and a 'Learn more' link). A 'CONTINUE' button is at the bottom.

Step 2: Model details

Enter a Model name, and Data split to 'Random Assignment'

The screenshot shows the 'Train new model' interface in the Google Cloud Platform, specifically the 'Model details' step. The left sidebar is the same as in the previous screenshot. The main panel has the progress bar with 'Model details' selected. Below the progress bar are 'START TRAINING' and 'CANCEL' buttons. The right panel shows the 'Model name' dropdown set to 'titanic_202112744834'. Below this is the 'Data split' section, which has three options: 'Random assignment' (selected, with a description and a 'Learn more' link), 'Manual' (with a description and a 'Learn more' link), and 'Chronological assignment' (with a description and a 'Learn more' link). Below the options is a donut chart showing the data split: Training 80%, Validation 10%, and Testing 10%. Below the chart is a horizontal bar chart showing the timeline from 'Start time' to 'End time'. Below the bar chart is the 'Encryption' section, which has a checkbox for 'Use a customer-managed encryption key (CMK)' and is currently unchecked. Below this is the 'Service account' section, which has a text input field and a 'BROWSE' button.

Step 3: Training container

Define your training environment

1. **Pre-built container:** Google cloud offers a set of prebuilt containers that make it easy to train your models. Those containers support frameworks such as Scikit-Learn, Tensorflow and XGBoost. Your model is based on scikit-learn and prebuilt container already exists.
2. **Model framework:** **Scikit-learn**. This is the library you used for model training.
3. **Model framework version:** Your code is compatible with **0.23**.
4. **Package location:** You can browse to the location of your training package. This is the bucket location where the training-0.1.tar.gz got uploaded.
5. **Python Module:** The python module you created in Notebooks. It will correspond to the folder that has your training code/module and the name of the entry file. This should be trainer.task
6. **BigQuery project for exporting data:**

Step 4: Hyperparameter tuning

Upgrade your account

Google Cloud

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbenches

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

<

Train new model

✓ Training method

✓ Model details

✓ Training container

4 Hyperparameters (optional)

5 Compute and pricing

6 Prediction container (optional)

START TRAINING

CANCEL

Hyperparameter tuning optimizes your model through multiple trials in one training job, but will increase the cost of this job. **After training finishes, the best-performing model will be saved to your Model List.** [Learn more](#)

☐ Enable hyperparameter tuning

CONTINUE

Step 5: Compute and pricing

Upgrade your account

Google Cloud

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbenches

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

<

Train new model

✓ Training method

✓ Model details

✓ Training container

✓ Hyperparameters (optional)

5 Compute and pricing

6 Prediction container (optional)

START TRAINING

CANCEL

Model training pricing is based on the length of time spent training, machine types, and any accelerators used. [Learn more](#)

Region

us-central1 (Iowa)

Compute settings

Select the type of virtual machine to use for your worker pool. You can add up to 4 worker pools. To learn about compute costs and how to map your ML framework's roles to specific worker pools, consult the [documentation](#)

Worker pool 0

Machine type *

n1-standard-4, 4 vCPUs, 15 GiB memory

Worker count

1

Disk type

SSD

Disk size (GB)

100

ADD MORE WORKER POOLS (OPTIONAL)

CONTINUE

Step 6: Prediction container

Upgrade your account to avoid a break in service

Google Cloud Platform

Vertex AI

Train new model

- Training method
- Model details
- Training container
- Hyperparameters (optional)
- Compute and pricing
- Prediction container (optional)**

START TRAINING CANCEL

You can associate your custom-trained model with a container in order to serve prediction requests using Vertex AI. [Learn more about getting predictions.](#)

☐ No prediction container
You can always import your model artifact later to serve prediction requests

☒ **Pre-built container**
View the list of [supported runtimes](#) including TensorFlow, scikit-learn and PyTorch versions

☐ Custom container
Build a custom Docker container. Must be stored in [Container Registry](#) or [Artifact Registry](#)

Pre-built container settings

Vertex AI provides Docker container images for serving predictions. To use a pre-built container, your trained model code must be in Python 3.7. [Learn more about pre-built containers](#)

In order to run in a pre-built container, your code needs to be in Python 3.7

Model framework * **scikit-learn**

Model framework version * **0.23**

Model directory * **gs://cmpe260-234200-bucket/titanic/assets/** BROWSE
Cloud Storage location containing the model artifact and any supporting files

Predict schemata

Optional. [Learn more about the predict schemata](#)

gs://instances BROWSE
Cloud Storage location to a YAML file that defines the format of a single instance used in prediction and evaluation requests

Click on 'Start Training'.

Google Cloud Platform CMPE260

Search products and resources

Vertex AI

Training CREATE

TRAINING PIPELINES CUSTOM JOBS HYPERPARAMETER TUNING JOBS

Training pipelines are the primary model training workflow in Vertex AI. You can use training pipelines to create an AutoML-trained model or a custom-trained model. For custom-trained models, training pipelines orchestrate custom training jobs and hyperparameter tuning with additional steps like adding a dataset or uploading the model to Vertex AI for prediction serving. [Learn More](#)

Region: us-central1 (Iowa)

Filter: Enter a property name

Name	ID	Status	Job type	Model type	Created	Elapsed time	Labels
titanic_202112744834	1504731140634705920	Finished	Training pipeline	Custom	Dec 6, 2021, 9:17:57 PM	12 min 39 sec	—

Check for the training to be 'Finished'

Model Evaluation

1. After the training job completion artifacts will be exported under gs://YOUR-BUCKET-NAME/training/assets
2. Check for the report.txt, under this folder.

Bucket details

cmpe260-334300-bucket

Location: us-central1 (Iowa) | Storage class: Standard | Public access: Subject to object ACLs | Protection: None

OBJECTS | CONFIGURATION | PERMISSIONS | PROTECTION | LIFECYCLE

Buckets > cmpe260-334300-bucket > training > assets > model

UPLOAD FILES | UPLOAD FOLDER | CREATE FOLDER | MANAGE HOLDS | DOWNLOAD | DELETE

Filter by name prefix only | Filter | Filter objects and folders | Show deleted data

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption	Retention expiration date	Holds
model.pkl	88 KB	text/plain	Dec 6, 2021, 9:30:35 ...	Standard	Dec 6, 2021, 9:30:35 ...	Not public	—	Google-managed key	—	None
report.txt	977 B	text/plain	Dec 6, 2021, 9:30:35 ...	Standard	Dec 6, 2021, 9:30:35 ...	Not public	—	Google-managed key	—	None

Model Deployment

1. Last step is model deployment
2. Click on the trained model and **DEPLOY TO ENDPOINT**

Vertex AI | Training | CREATE | REFRESH

TRAINING PIPELINES | CUSTOM JOBS | HYPERPARAMETER TUNING JOBS

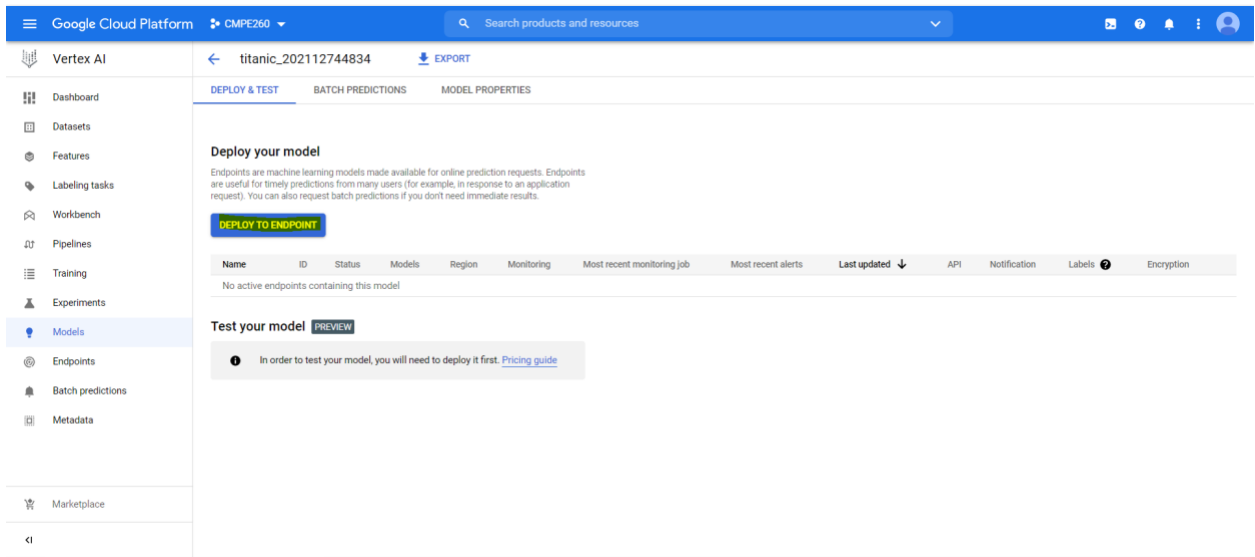
Training pipelines are the primary model training workflow in Vertex AI. You can use training pipelines to create an AutoML-trained model or a custom-trained model. For custom-trained models, training pipelines orchestrate custom training jobs and hyperparameter tuning with additional steps like adding a dataset or uploading the model to Vertex AI for prediction serving. [Learn More](#)

Region: us-central1 (Iowa)

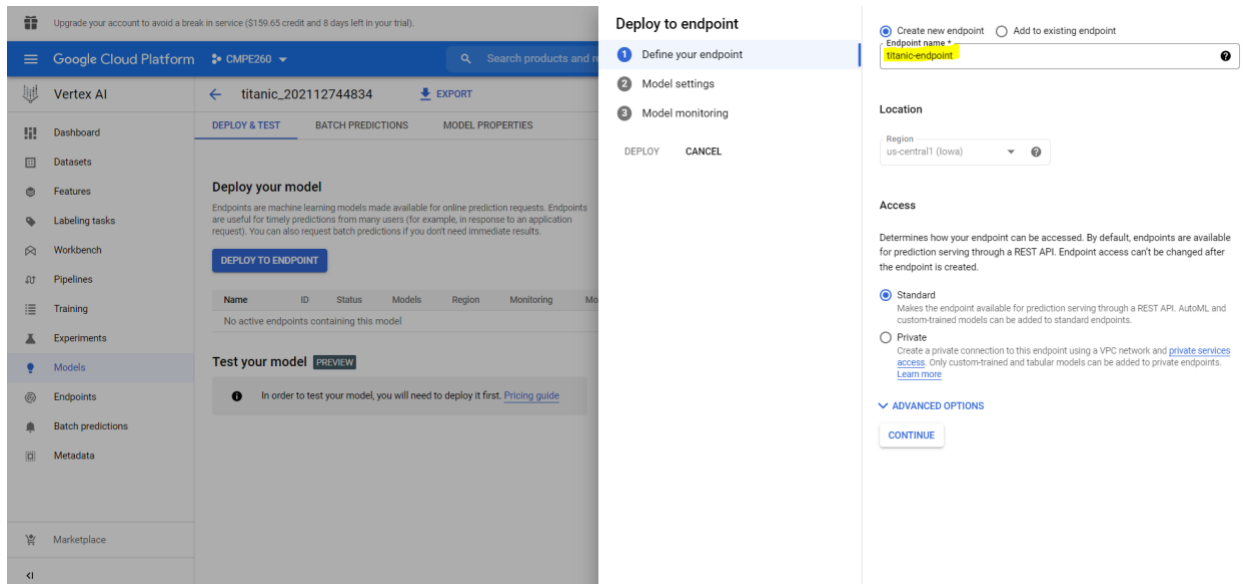
Filter | Enter a property name

Name	ID	Status	Job type	Model type	Created	Elapsed time	Labels
tfmc_202112744826	1504731140634705920	Finished	Training pipeline	Custom	Dec 6, 2021, 9:17:57 PM	12 min 39 sec	—

<https://console.cloud.google.com/vertex-ai/locations/us-central1/models/4465534534816800880/training/trainingpipelines/1504731140634705920/view?cmpe260-334300>



- **Endpoint name:** titanic-endpoint Endpoint URL where the model is served.
- **Traffic split:** Defines the percentage of traffic that you want to direct to this model. An endpoint can have multiple models and you can decide how to split the traffic among them. In this case you are deploying a single model so the traffic has to be 100 percent.
- **Minimum number of compute nodes:** The minimum number of nodes required to serve model predictions. Start with 1. Additionally the prediction service will autoscale in case there is traffic
- **Maximum number of compute nodes:** In case of autoscaling, this variable defines the upper limit of nodes. It helps protecting against unwanted costs that autoscaling might result in. Set this variable to 2
- **Machine type:** Google cloud offers a set of machine types you can deploy your model to. Each machine has its own memory and vcpu specs. Your model is simple so serving on an n1-standard-4 instance will do the job



Upgrade your account to avoid a break in service (\$159.65 credit and 8 days left in your trial).

Google Cloud Platform CMPE260

Search products and resources

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

titanic_202112744834

EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

Deploy your model

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.

DEPLOY TO ENDPOINT

No active endpoints containing this model

Test your model PREVIEW

In order to test your model, you will need to deploy it first. [Pricing guide](#)

Deploy to endpoint

Define your endpoint

Model settings

Model monitoring

DEPLOY CANCEL

titanic_202112744834

Traffic split 100%

Compute resources

Choose how compute resources will serve prediction traffic to your model

Autoscaling: If you set a minimum and maximum, compute nodes will scale to meet traffic demand within those boundaries

No scaling: If you only set a minimum, then that number of compute nodes will always run regardless of traffic demand (the maximum will be set to minimum)

Once scaling settings are set, they can't be changed unless you redeploy the model. [Pricing guide](#)

Minimum number of compute nodes 1

Default is 1. If set to 1 or more, then compute resources will continuously run even without traffic demand. This can increase cost but avoid dropped requests due to node initialization.

Maximum number of compute nodes (optional) 2

Enter a number equal to or greater than the minimum nodes. Can reduce costs but may cause reliability issues for high traffic.

ADVANCED SCALING OPTIONS

Machine type mt-standard-4, 4 vCPUs, 15 GiB memory

Service account

A service account determines what Google Cloud resources your service code can access.

3. Click CONTINUE and DEPLOY

Google Cloud Platform CMPE260

Search products and resources

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

titanic_202112744834

EXPORT

DEPLOY & TEST BATCH PREDICTIONS MODEL PROPERTIES

Deploy your model

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.

DEPLOY TO ENDPOINT

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Endpoint key
titanic-endpoint	432424729064767488	Deploying model	0	us-central1	—	—	—	Dec 6, 2021, 9:37:57 PM	Sample request			Google Cloud

Test your model PREVIEW

Your model must be successfully deployed to an endpoint before you can test it.

Your JSON request must contain an instances field and an optional parameters field if you're using a custom container. No other fields can be present in the JSON request. [Learn how to format your JSON request.](#)

JSON request

Response

Google Cloud Platform

CMPE260

Search products and resources

Vertex AI

titanic_202112744834

EXPORT

DEPLOY & TEST

BATCH PREDICTIONS

MODEL PROPERTIES

Deploy your model

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.

DEPLOY TO ENDPOINT

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated ↓	API	Notification	Labels ⓘ	Encryption
titanic-endpoint	432424729064767488	Active	1	us-central1	Disabled	—	—	Dec 6, 2021, 9:49:36 PM	Sample request			Google-managed key

Test your model PREVIEW

Your JSON request must contain an `instances` field and an optional `parameters` field if you're using a custom container. No other fields can be present in the JSON request. [Learn how to format your JSON request.](#)

JSON request

Response

```

{
  "instances": [
    {
      "sample_key": "sample_value"
    }
  ]
}

```

Model Prediction

Under **Models** test the model prediction endpoint.

Try for different payloads:

Request:

```

{
  "instances": [
    ["male", 29.8811345124283, 26.0, 1, "S", "New York, NY", 0, 0],
    ["female", 48.0, 39.6, 1, "C", "London / Paris", 0, 1]]
}

```

Vertex AI

titanic_202112744834

EXPORT

DEPLOY & TEST

BATCH PREDICTIONS

MODEL PROPERTIES

Test your model PREVIEW

Your JSON request must contain an `instances` field and an optional `parameters` field if you're using a custom container. No other fields can be present in the JSON request. [Learn how to format your JSON request.](#)

JSON request

Response

```

{
  "instances": [
    ["male", 29.8811345124283, 26.0, 1, "S", "New York, NY", 0, 0],
    ["female", 48.0, 39.6, 1, "C", "London / Paris", 0, 1]]
}

```

PREDICT

```

{
  "predictions": [
    0,
    1
  ],
  "deployedModelId": "5093685527765319680",
  "model": "projects/196373151126/locations/us-central1/models/4465534534816",
  "modelDisplayName": "titanic_202112744834"
}

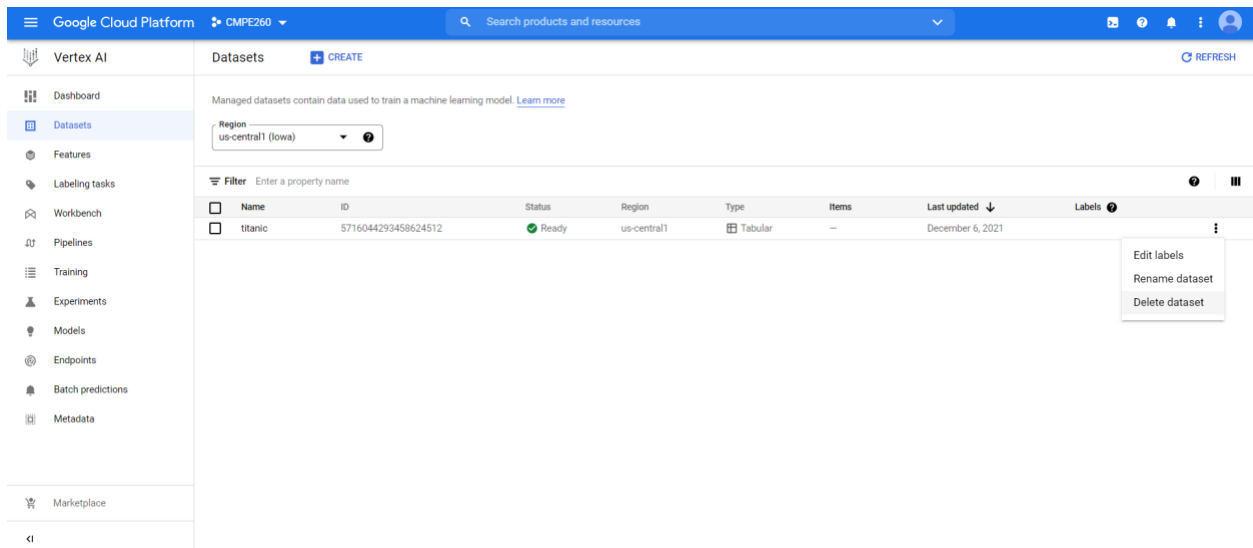
```

Response:

```
{
  "predictions": [
    0,
    1
  ],
  "deployedModelId": "5093685527765319680",
  "model": "projects/196373151126/locations/us-central1/models/4465534534816890880",
  "modelDisplayName": "titanic_202112744834"
}
```

Cleaning up

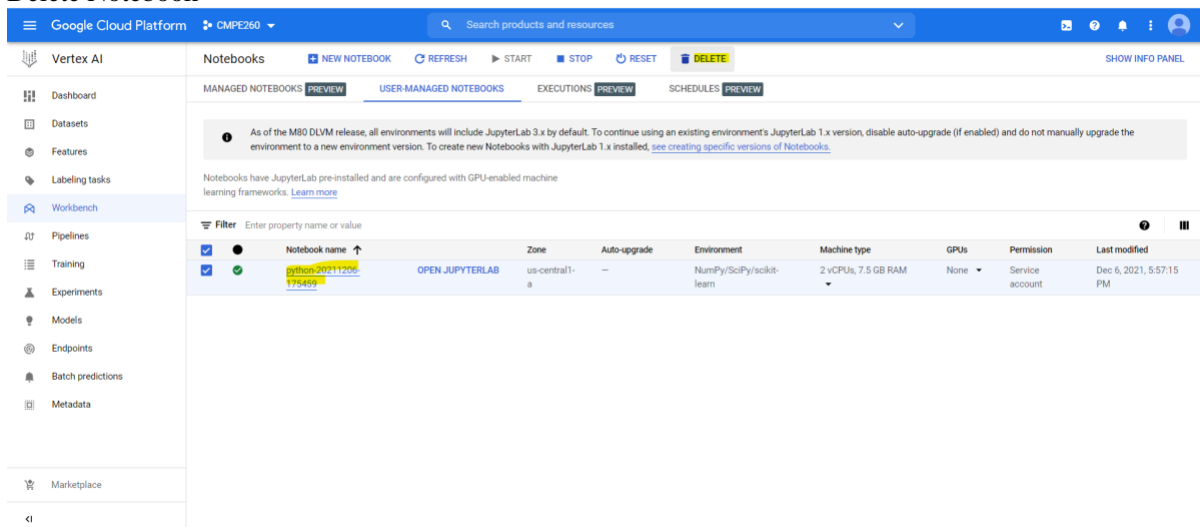
Delete ML Dataset



The screenshot shows the Google Cloud Platform Vertex AI Datasets page. The left sidebar contains navigation links for Vertex AI, Dashboard, Datasets, Features, Labeling tasks, Workbench, Pipelines, Training, Experiments, Models, Endpoints, Batch predictions, and Metadata. The main content area displays a list of datasets. A filter is applied for the region 'us-central1 (Iowa)'. The dataset list has columns for Name, ID, Status, Region, Type, Items, Last updated, and Labels. One dataset named 'titanic' is listed with ID 5716044293458624512, status 'Ready', and last updated on December 6, 2021. A context menu is open for the 'titanic' dataset, showing options: Edit labels, Rename dataset, and Delete dataset.

Name	ID	Status	Region	Type	Items	Last updated	Labels
titanic	5716044293458624512	Ready	us-central1	Tabular	—	December 6, 2021	

Delete Notebook



The screenshot shows the Google Cloud Platform Vertex AI Notebooks page. The left sidebar contains navigation links for Vertex AI, Dashboard, Datasets, Features, Labeling tasks, Workbench, Pipelines, Training, Experiments, Models, Endpoints, Batch predictions, and Metadata. The main content area displays a list of notebooks. A filter is applied for the region 'us-central1 (Iowa)'. The notebook list has columns for Notebook name, Zone, Auto-upgrade, Environment, Machine type, GPUs, Permission, and Last modified. One notebook named 'python-20211206-194549' is listed with ID 5716044293458624512, status 'Ready', and last updated on December 6, 2021, 5:57:15 PM.

Notebook name	Zone	Auto-upgrade	Environment	Machine type	GPUs	Permission	Last modified
python-20211206-194549	us-central1-a	—	NumPy/SciPy/scikit-learn	2 vCPUs, 7.5 GB RAM	None	Service account	Dec 6, 2021, 5:57:15 PM

Delete Endpoint

Google Cloud Platform CMPE260

Search products and resources

Deploy & Test BATCH PREDICTIONS MODEL PROPERTIES

Deploy your model

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.

[DEPLOY TO ENDPOINT](#)

Name	ID	Status	Models	Region	Monitoring	Most recent monitoring job	Most recent alerts	Last updated	API	Notification	Labels	Encryption
titanic_202112744834	432424729064757488	Active	1	us-central1	Disabled	—	—	Dec 6, 2021, 9:49:36 PM	Sample request			Google-managed key

Test your model [REVIEW](#)

Your JSON request must contain an `instances` field and an optional `parameters` field if you're using a custom container. No other fields can be present in the JSON request. [Learn how to format your JSON request.](#)

JSON request

```
{
  "instances": [
    {
      "sample_key": "sample_value"
    }
  ]
}
```

Response

- View logs
- Edit labels
- Un-deploy model
- Rename endpoint
- Remove endpoint

Delete Model

Google Cloud Platform CMPE260

Search products and resources

Models [CREATE](#) [IMPORT](#) [REFRESH](#)

Models are built from your datasets or unmanaged data sources. There are many different types of machine learning models available on Vertex AI, depending on your use case and level of experience with machine learning. [Learn more](#)

Region: us-central1 (Iowa)

Filter: Enter a property name

Name	ID	Status	Data	Endpoints	Region	Type	Created	Notifications	Labels
titanic_202112744834	4465534534816890880	Ready	—	0	us-central1	Custom trained Custom training	Dec 6, 2021, 9:17:57 PM		

- Add to endpoint
- Edit labels
- Rename model
- Delete model

Delete Objects, GCP Bucket

Delete BigQuery dataset