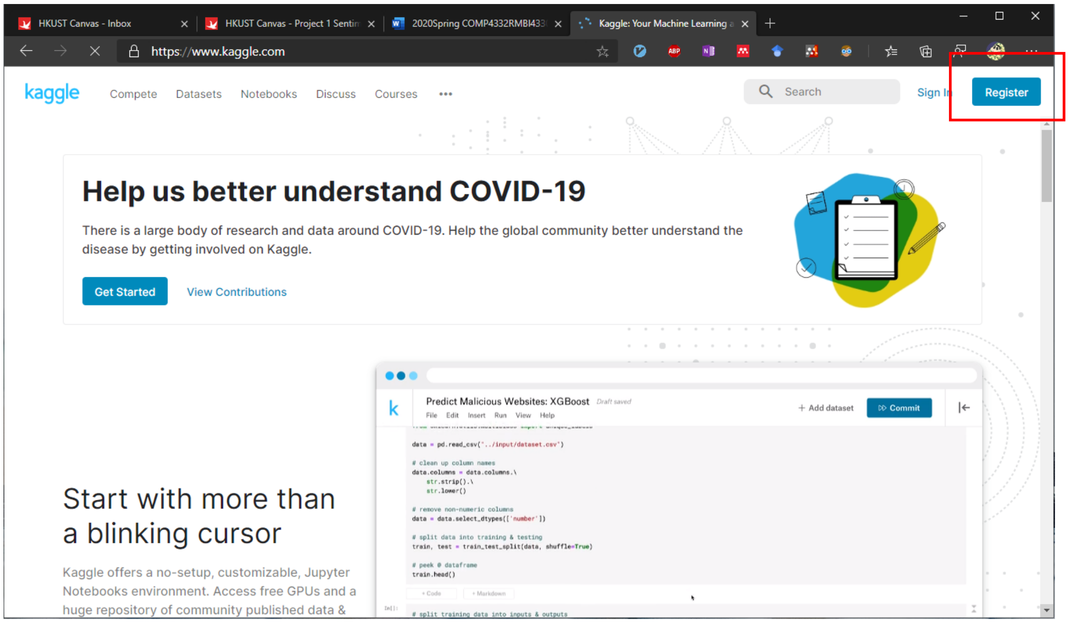
Kaggle Notebook

Enter the kaggle website and register your account.



Help us better understand COVID-19

There is a large body of research and data around COVID-19. Help the global community better understand the disease by getting involved on Kaggle.

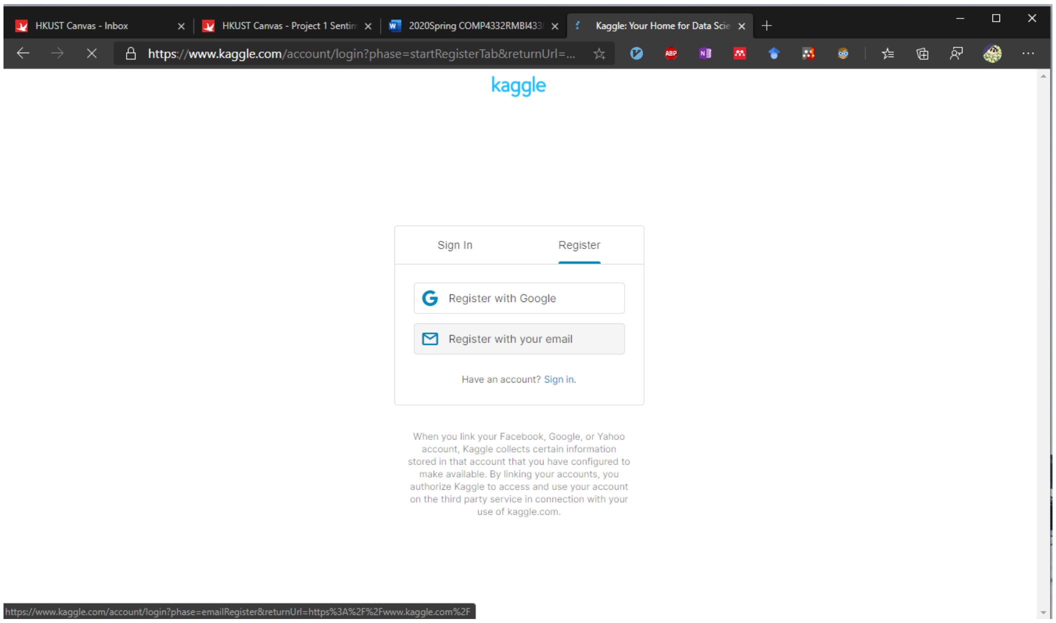
Predict Malicious Websites: XGBoost *Draft saved*

```
data = pd.read_csv('../input/dataset.csv')

# clean up column names
data.columns = data.columns.\
    str.strip().\
    str.lower()

# remove non-numeric columns
data = data.select_dtypes(['number'])

# split data into training & testing
train, test = train_test_split(data, shuffle=True)

# peek @ dataframe
train.head()
```

```
# split training data into inputs & outputs
```

Start with more than a blinking cursor

Kaggle offers a no-setup, customizable, Jupyter Notebooks environment. Access free GPUs and a huge repository of community published data &

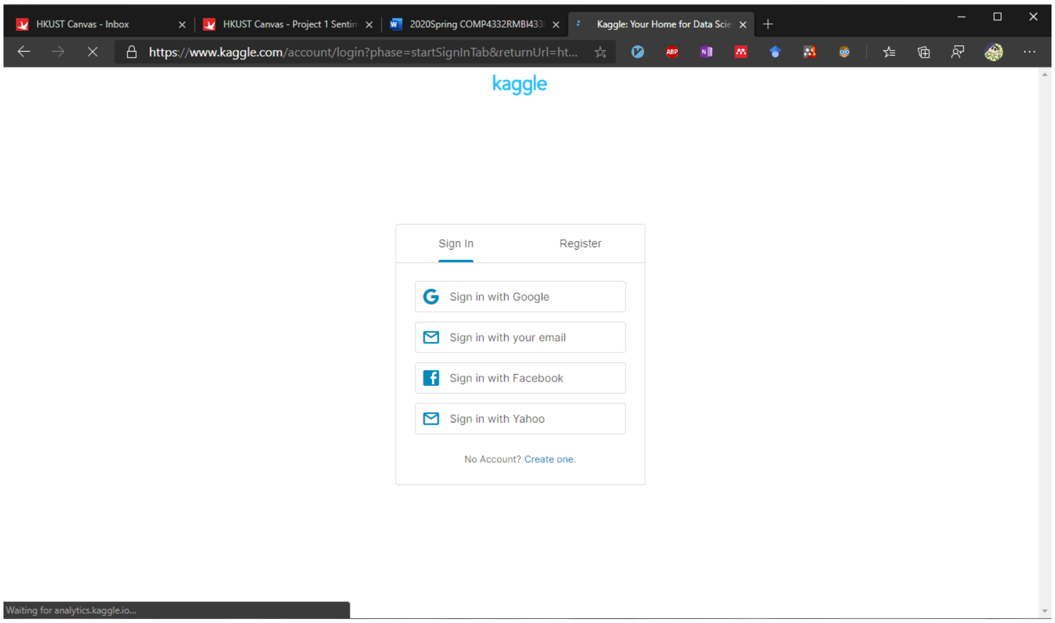

HKUST Canvas - Inbox
HKUST Canvas - Project 1 Sentir
2020Spring COMP4332RMBI433
Kaggle: Your Home for Data Scie
https://www.kaggle.com/account/login?phase=startRegisterTab&returnUrl=...
kaggle
Sign In
Register
Register with Google
Register with your email

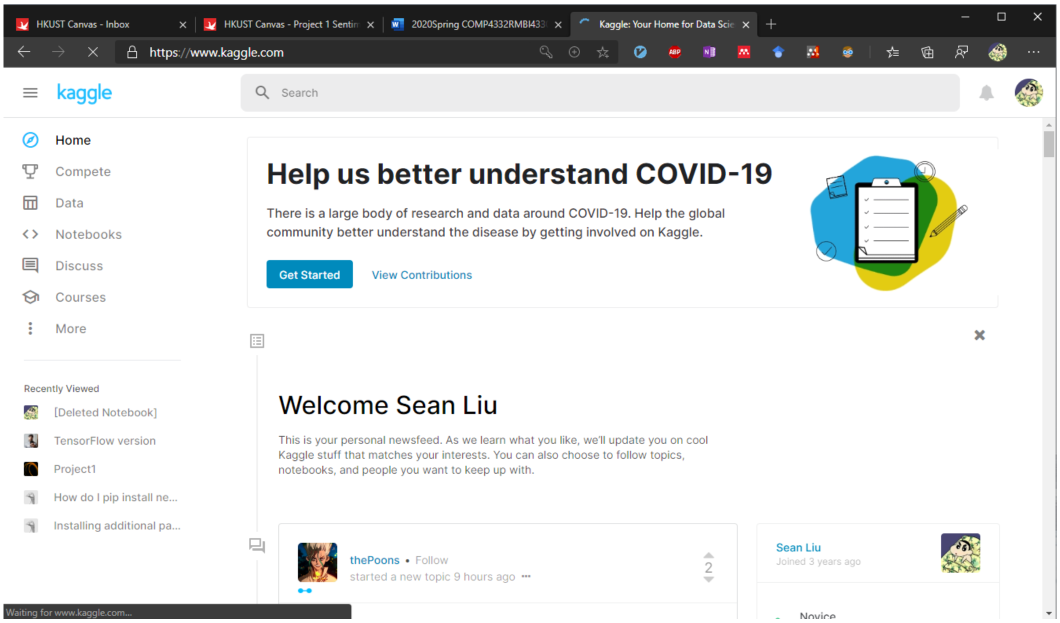
Have an account? Sign in.

When you link your Facebook, Google, or Yahoo account, Kaggle collects certain information stored in that account that you have configured to make available. By linking your accounts, you authorize Kaggle to access and use your account on the third party service in connection with your use of kaggle.com.

https://www.kaggle.com/account/login?phase=emailRegister&returnUrl=https%3A%2F%2Fwww.kaggle.com%2F

After the registration, you can sign in.


HKUST Canvas - Inbox
HKUST Canvas - Project 1 Sentim
2020Spring COMP4332RMBI433
Kaggle: Your Home for Data Scie
https://www.kaggle.com/account/login?phase=startSignInTab&returnUrl=ht...
kaggle
Sign In
Register
Sign in with Google
Sign in with your email
Sign in with Facebook
Sign in with Yahoo
No Account? Create one.
Waiting for analytics.kaggle.io...



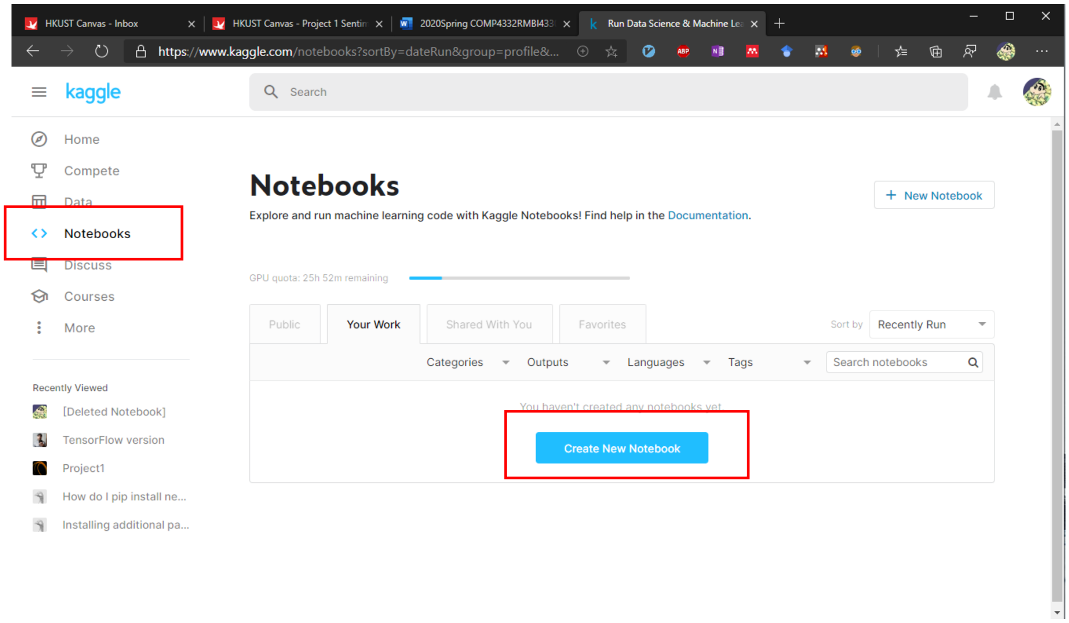
Help us better understand COVID-19

There is a large body of research and data around COVID-19. Help the global community better understand the disease by getting involved on Kaggle.

Welcome Sean Liu

This is your personal newsfeed. As we learn what you like, we'll update you on cool Kaggle stuff that matches your interests. You can also choose to follow topics, notebooks, and people you want to keep up with.

You can create notebooks. If you want to use a powerful GPU (P100 with 16 GB memory), you should choose GPU as the accelerator. You will get a free GPU quota of 30 hours per week.



HKUST Canvas - Inbox | HKUST Canvas - Project 1 Sentin | 2020Spring COMP4332RMBI433 | Run Data Science & Machine Le

https://www.kaggle.com/notebooks?sortBy=dateRun&group=profile&...

kaggle

Search

Home
Compete
Data
Notebooks
Discuss
Courses
More

Recently Viewed
[Deleted Notebook]
TensorFlow version
Project1
How do I pip install ne...
Installing additional pa...

Notebooks

Explore and run machine learning code with Kaggle Notebooks! Find help in the Documentation.

+ New Notebook

GPU quota: 25h 52m remaining

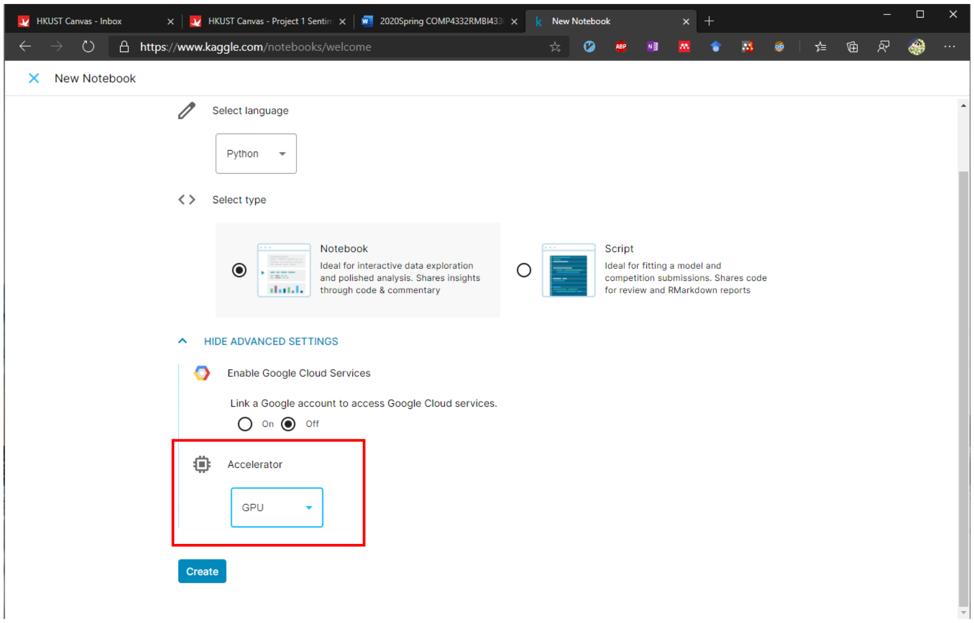
Public | Your Work | Shared With You | Favorites

Sort by Recently Run

Categories | Outputs | Languages | Tags | Search notebooks

You haven't created any notebooks yet

Create New Notebook



HKUST Canvas - Inbox | HKUST Canvas - Project 1 Sentin | 2020Spring COMP4332RMBI433 | New Notebook

https://www.kaggle.com/notebooks/welcome

New Notebook

Select language

Python

Select type

● Notebook
Ideal for interactive data exploration and polished analysis. Shares insights through code & commentary

○ Script
Ideal for fitting a model and competition submissions. Shares code for review and RMarkdown reports

HIDE ADVANCED SETTINGS

Enable Google Cloud Services

Link a Google account to access Google Cloud services.

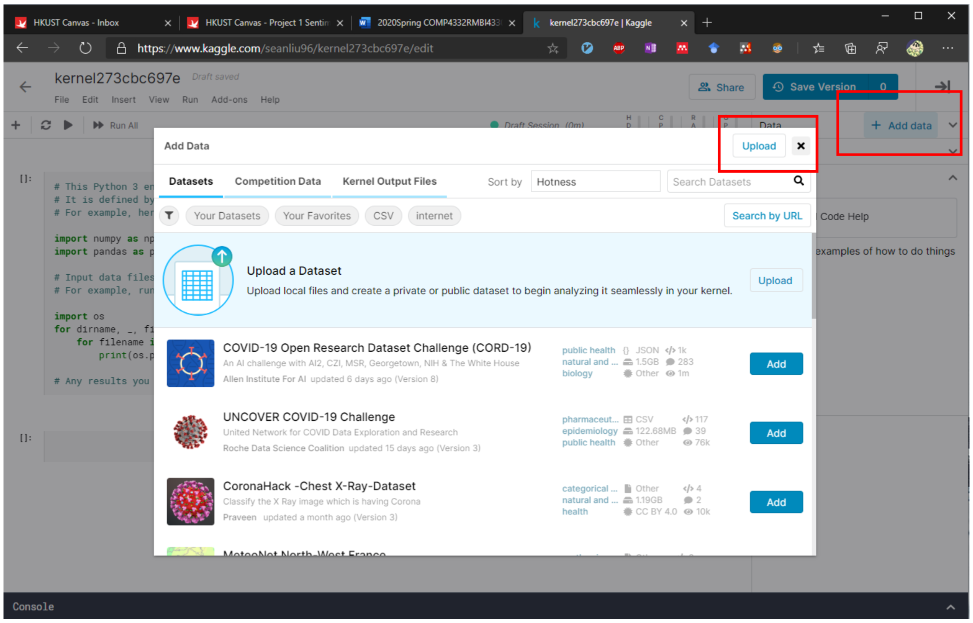
○ On ● Off

Accelerator

GPU

Create

After creating a notebook, you should upload your data.



Upload a Dataset

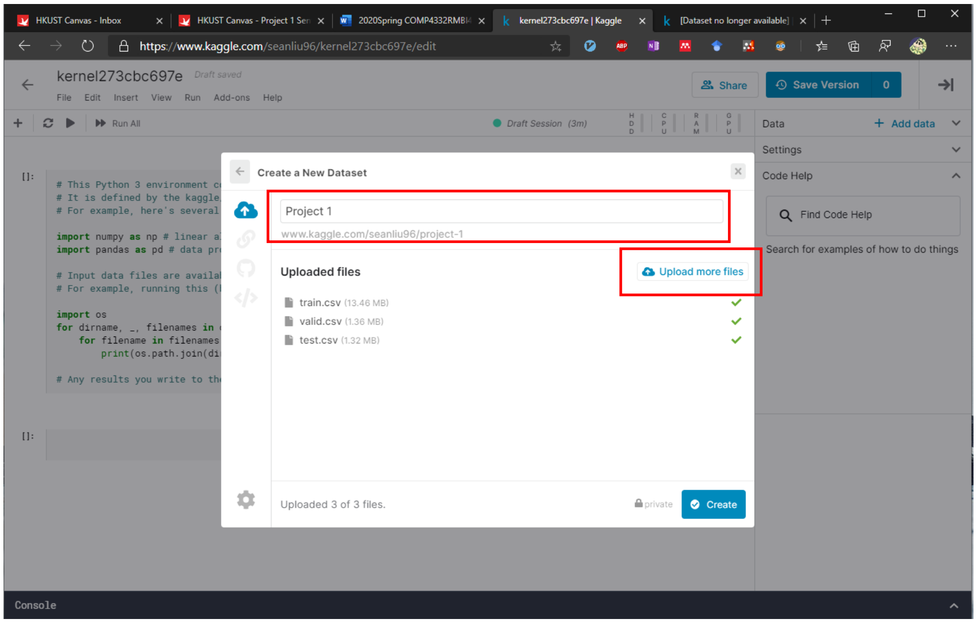
Upload local files and create a private or public dataset to begin analyzing it seamlessly in your kernel.

COVID-19 Open Research Dataset Challenge (CORD-19)
An AI challenge with AI2, CZI, MSR, Georgetown, NIH & The White House
Allen Institute For AI updated 6 days ago (Version 8)

UNCOVER COVID-19 Challenge
United Network for COVID Data Exploration and Research
Roche Data Science Coalition updated 15 days ago (Version 3)

CoronaHack -Chest X-Ray-Dataset
Classify the X Ray image which is having Corona
Praveen updated a month ago (Version 3)

You need to name your dataset.



Project 1

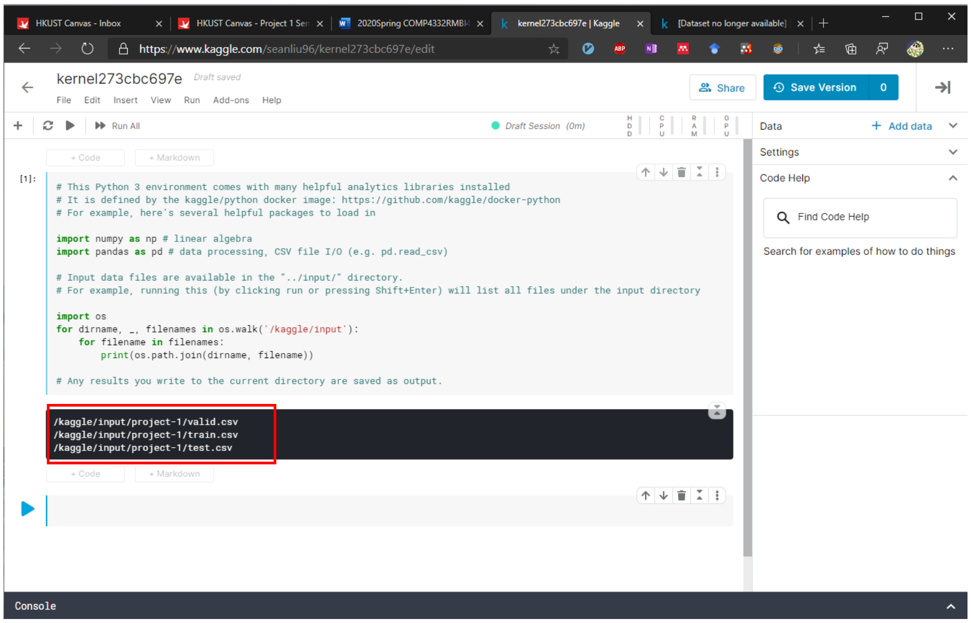
www.kaggle.com/seanliu96/project-1

Uploaded files

- train.csv (13.46 MB) ✓
- valid.csv (1.36 MB) ✓
- test.csv (1.32 MB) ✓

Uploaded 3 of 3 files.

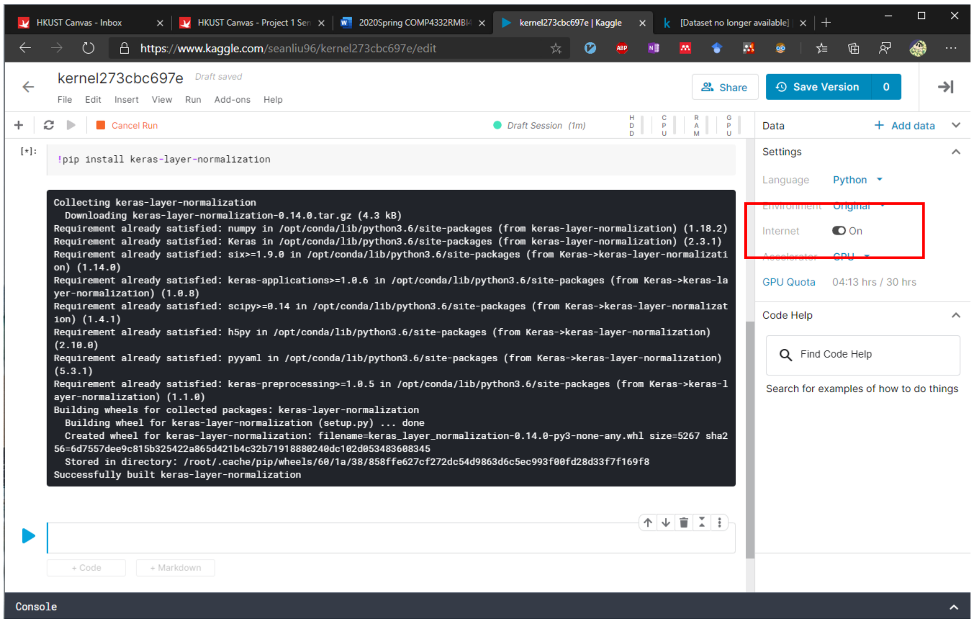
After that, you can run the first code to see the location of your data.



```
[1]:
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load in

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# Any results you write to the current directory are saved as output.
```

```
/kaggle/input/project-1/valid.csv
/kaggle/input/project-1/train.csv
/kaggle/input/project-1/test.csv
```

If you want to install the missing packages or download packages, you need to toggle on the internet.

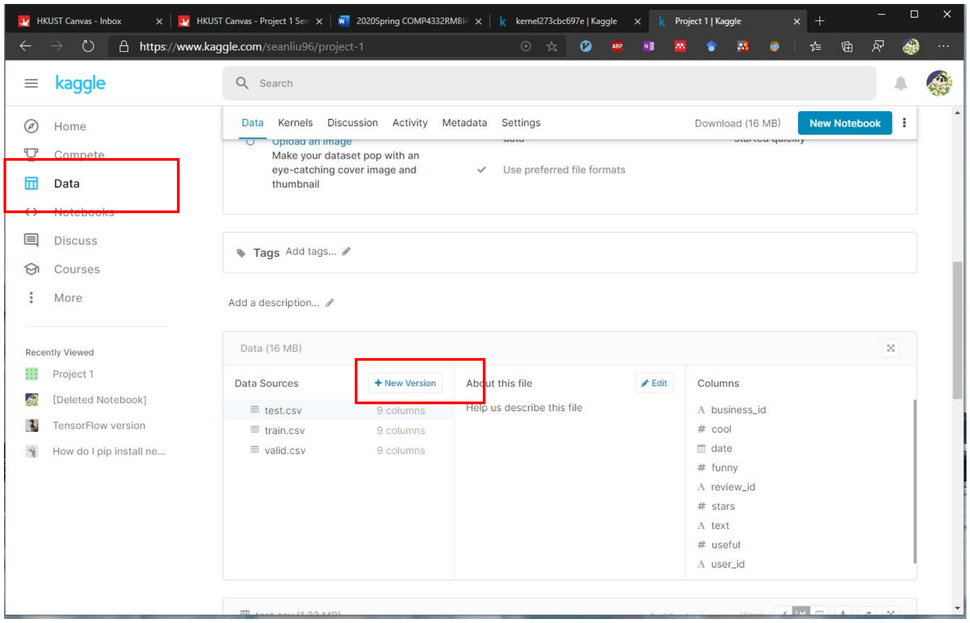


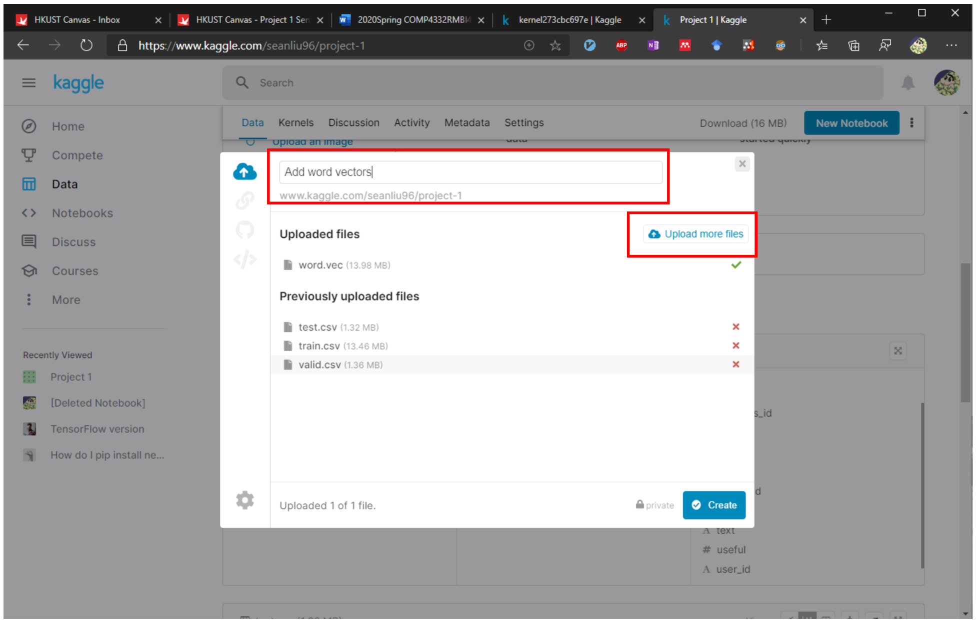
Language Python
Environment Original
Internet On
Accelerator GPU
GPU Quota 04:13 hrs / 30 hrs

```
[*]:
!pip install keras-layer-normalization
```

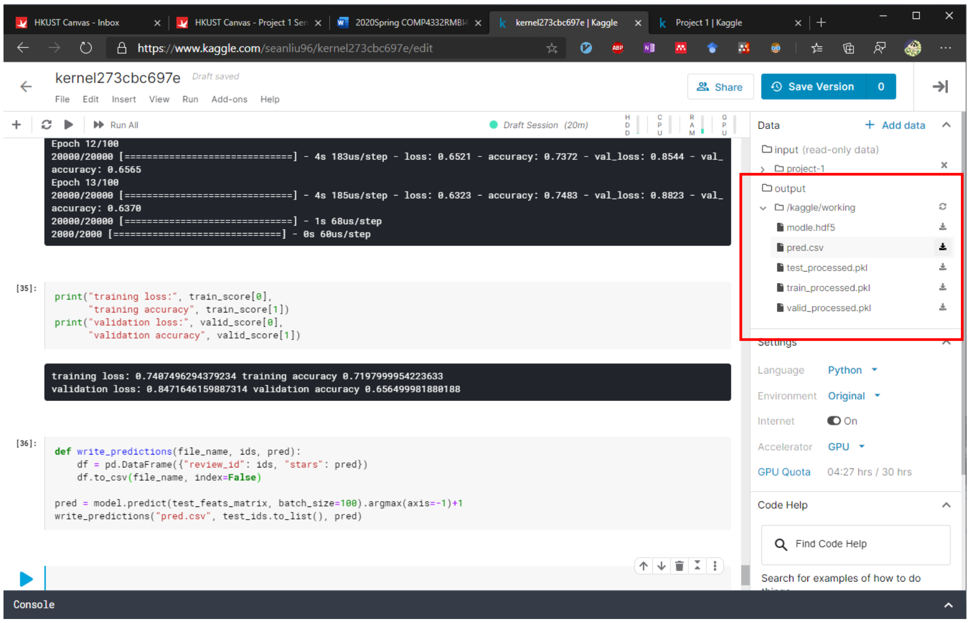
```
Collecting keras-layer-normalization
  Downloading keras-layer-normalization-0.14.0.tar.gz (4.3 kB)
Requirement already satisfied: numpy in /opt/conda/lib/python3.6/site-packages (from keras-layer-normalization) (1.18.2)
Requirement already satisfied: Keras in /opt/conda/lib/python3.6/site-packages (from keras-layer-normalization) (2.3.1)
Requirement already satisfied: six>=1.9.0 in /opt/conda/lib/python3.6/site-packages (from Keras->keras-layer-normalizati
on) (1.14.0)
Requirement already satisfied: keras-applications>=1.0.6 in /opt/conda/lib/python3.6/site-packages (from Keras->keras-la
yer-normalization) (1.0.8)
Requirement already satisfied: scipy>=0.14 in /opt/conda/lib/python3.6/site-packages (from Keras->keras-layer-normalizat
ion) (1.4.1)
Requirement already satisfied: h5py in /opt/conda/lib/python3.6/site-packages (from Keras->keras-layer-normalization)
(2.10.0)
Requirement already satisfied: pyyaml in /opt/conda/lib/python3.6/site-packages (from Keras->keras-layer-normalization)
(5.3.1)
Requirement already satisfied: keras-preprocessing>=1.0.5 in /opt/conda/lib/python3.6/site-packages (from Keras->keras-l
ayer-normalization) (1.1.0)
Building wheels for collected packages: keras-layer-normalization
  Building wheel for keras-layer-normalization (setup.py) ... done
  Created wheel for keras-layer-normalization: filename=keras_layer_normalization-0.14.0-py3-none-any.whl size=5267 sha2
56=6d7557dee9c815b325422a865d421b4c32b71918880240dc102d053483608345
  Stored in directory: /root/.cache/pip/wheels/60/1a/38/858ffe627cf272dc54d9863d6c5ec993f00fd28d33f7f169f8
Successfully built keras-layer-normalization
```

If you want to add more necessary data, you can go to the data part and select the corresponding dataset. Then click the new version button and upload data. Remember you need to add a description for this operation.





After saving your predictions into a file, you need to copy back to your own laptop.



```
Epoch 12/100
20000/20000 [==============================] - 4s 183us/step - loss: 0.6521 - accuracy: 0.7372 - val_loss: 0.8544 - val_
accuracy: 0.6565
Epoch 13/100
20000/20000 [==============================] - 4s 185us/step - loss: 0.6323 - accuracy: 0.7483 - val_loss: 0.8823 - val_
accuracy: 0.6370
20000/20000 [==============================] - 1s 68us/step
2000/2000 [==============================] - 0s 60us/step
```

[35]:

```
print("training loss:", train_score[0],
      "training accuracy", train_score[1])
print("validation loss:", valid_score[0],
      "validation accuracy", valid_score[1])
```

```
training loss: 0.7407496294379234 training accuracy 0.7197999954223633
validation loss: 0.8471646159887314 validation accuracy 0.656499981880188
```

[36]:

```
def write_predictions(file_name, ids, pred):
    df = pd.DataFrame({"review_id": ids, "stars": pred})
    df.to_csv(file_name, index=False)

pred = model.predict(test_feats_matrix, batch_size=100).argmax(axis=-1)+1
write_predictions("pred.csv", test_ids.to_list(), pred)
```

input (read-only data)
- project-1

output
- /kaggle/working
  - modle.hdf5
  - pred.csv
  - test_processed.pkl
  - train_processed.pkl
  - valid_processed.pkl

Settings

Language	Python
Environment	Original
Internet	On
Accelerator	GPU
GPU Quota	04:27 hrs / 30 hrs

Search for examples of how to do things