

How to Download Kaggle Datasets on Ubuntu

In Kaggle, find the dataset you want to download, and check the name of the dataset and the user that uploaded the dataset. You can find this in the URL of the dataset https://www.kaggle.com/<USER_NAME>/<DATASET_NAME>.

For example, if your dataset is located in <https://www.kaggle.com/Cornell-University/arxiv>,

- its **DATASET_NAME** is **arxiv**, and
- its **USER_NAME** is **Cornell-University**.

You should also have Kaggle account.

Step 1 - Download Kaggle API

Kaggle has a [command-line API](#) that can be installed using **pip**.

```
python3 -m pip install --user kaggle
```

```
1
```

```
or
```

```
pip install --user kaggle
```

pip will install Kaggle API and any required dependencies to your machine.

Step 2 - Setup API Credentials

Navigate to the Accounts page of Kaggle at https://www.kaggle.com/<USER_NAME>/account. Go to the “API” section and select the “Create New API Token”. This will trigger the download of **kaggle.json**, a file that contains your API credentials. The JSON has a single line of below format:

```
1{"username":<USER_NAME>,"key":<API_KEY>}
```

Make a directory **.kaggle** at root **~**, and place **kaggle.json** in that directory. (**~/.kaggle/kaggle.json**)

```
1mkdir ~/.kaggle
```

```
2mv kaggle.json ~/.kaggle
```

You can verify that the JSON was saved correctly by printing it using the `cat` command:

```
1 cat ~/.kaggle/kaggle.json
```

For safety, edit the file permission to ensure that other users cannot read this file. You can use the `chmod` command to change the permission:

```
1 chmod 600 ~/.kaggle/kaggle.json
```

Step 3 - Download Dataset

Now, you can download the dataset using Kaggle's `kaggle datasets download` API. Navigate to the directory that you want to download the dataset to. Then, Check the `USER_NAME` and the `DATASET_NAME` that you noted in [Prerequisites](#) section of this tutorial, and paste it in the below template:

```
1 $ kaggle datasets download <USER_NAME>/<DATASET_NAME>
```

For example, if your dataset came from <https://www.kaggle.com/Cornell-University/arxiv>, you should execute the following line:

```
1$ kaggle datasets download Cornell-University/arxiv
```

Kaggle API will display a progress bar and start downloading the dataset. Depending on the dataset size and your internet connection, you will have to wait a few seconds to a few hours to download the dataset.

```
1 Downloading arxiv.zip to ~
1 100%|████████████████████████████████████████████████████████████████████████████████| 877M/877M [00:28<00:00, 32.4MB/s]
2
```

Conclusion

Now that you have the dataset downloaded, you have many options to explore the data. Try using Jupyter Notebook with Pandas for exploratory data analysis (EDA).

Note:-

To download any python package in IT Server:

```
python3 -m pip install --user sklearn
```