LINKSFOUNDATION.COM

FONDAZIONE links
PASSION FOR INNOVATION

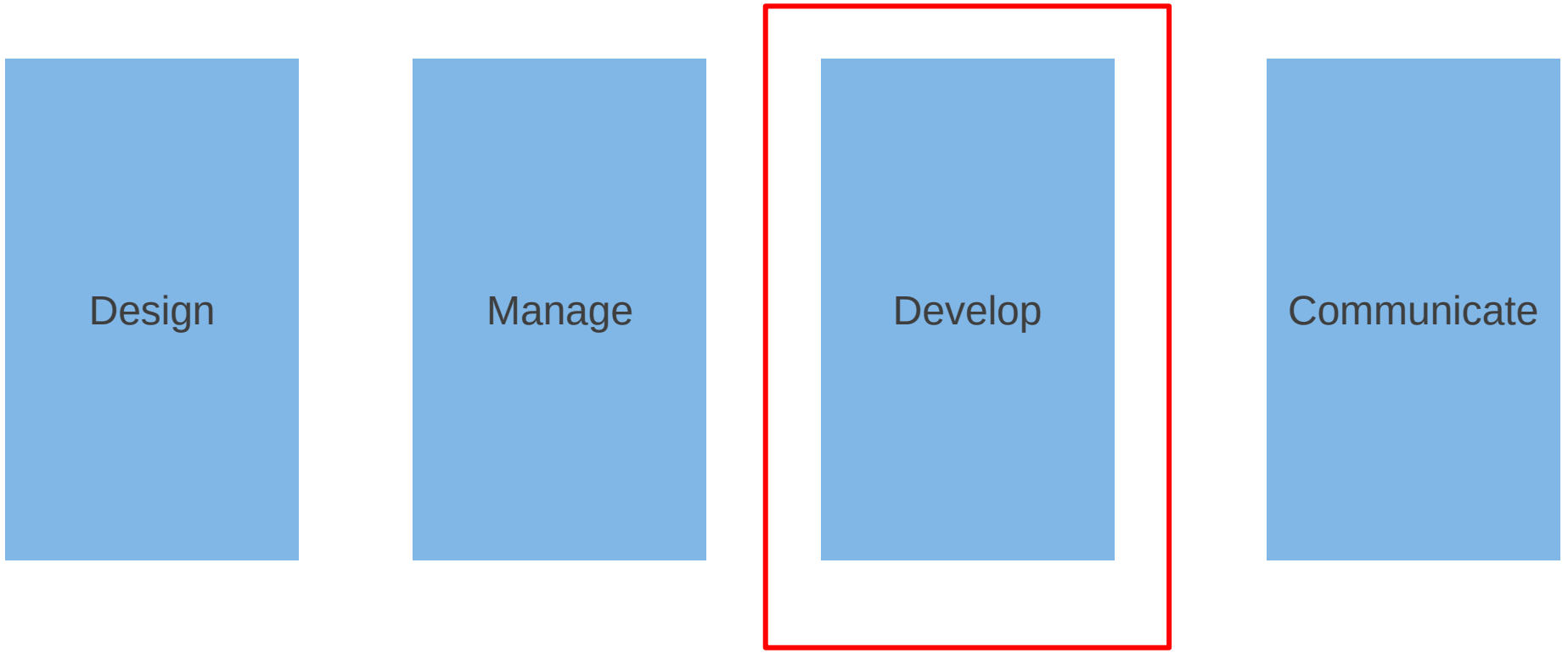
Applied Data Science Project

L10 – Colaboratory

Politecnico di Torino

ellis
European Laboratory for Learning and Intelligent Systems

Pillars



linksfoundation.com
COPYRIGHT ©2021 LINKS

Development

It is where the magic happens

An artificial intelligence system is created to generate the outputs that meet objectives and requirements with the involvement of a team tasked on activities with due dates

Objectives and requirements have been defined in the Design pillar

Activities and due dates in the Manage pillar

linksfoundation.com
COPYRIGHT ©2021 LINKS

Knowledge tools

Development environment

Foundation models

Version control

Internal communication (with team mates)

linksfoundation.com
COPYRIGHT ©2021 LINKS

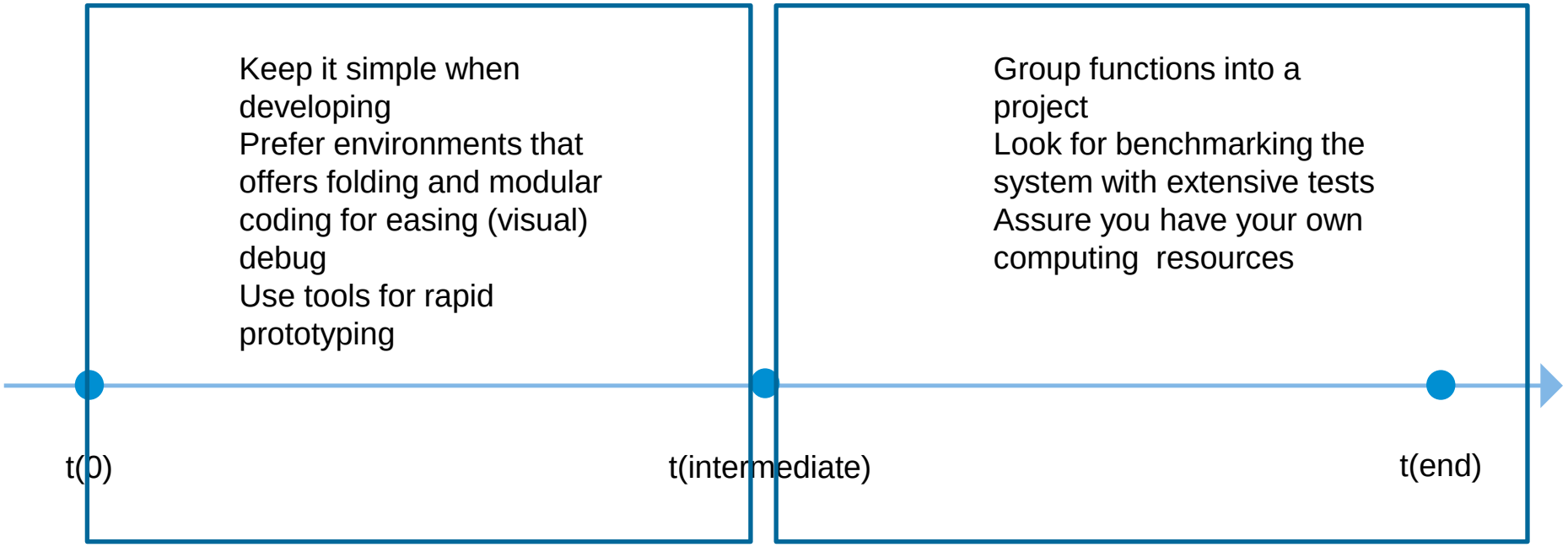
Development

- “Divide and rule”
- Tasks are mapped into modules
- One module has one lead developer and, eventually, contributors
- Choose the programming language according to:
  - ecosystem of software modules that you can utilize
  - easiness of model integration
  - familiarity. Do not be afraid to switch to another (similar programming language) since they share most of the features and development patterns

Python is considered the default language for developing machine intelligence nowadays

linksfoundation.com
COPYRIGHT ©2021 LINKS

Development timeline



Colab Notebook
Focus of the course

Hubs for coding (MLHub) or in IDE (VS Code)

linksfoundation.com
COPYRIGHT ©2021 LINKS

Colaboratory for rapid prototyping

https://colab.research.google.com

An application where to develop, share, and also test on dedicated hardware (GPU to speed up the computing)

linksfoundation.com
COPYRIGHT ©2021 LINKS

Overview

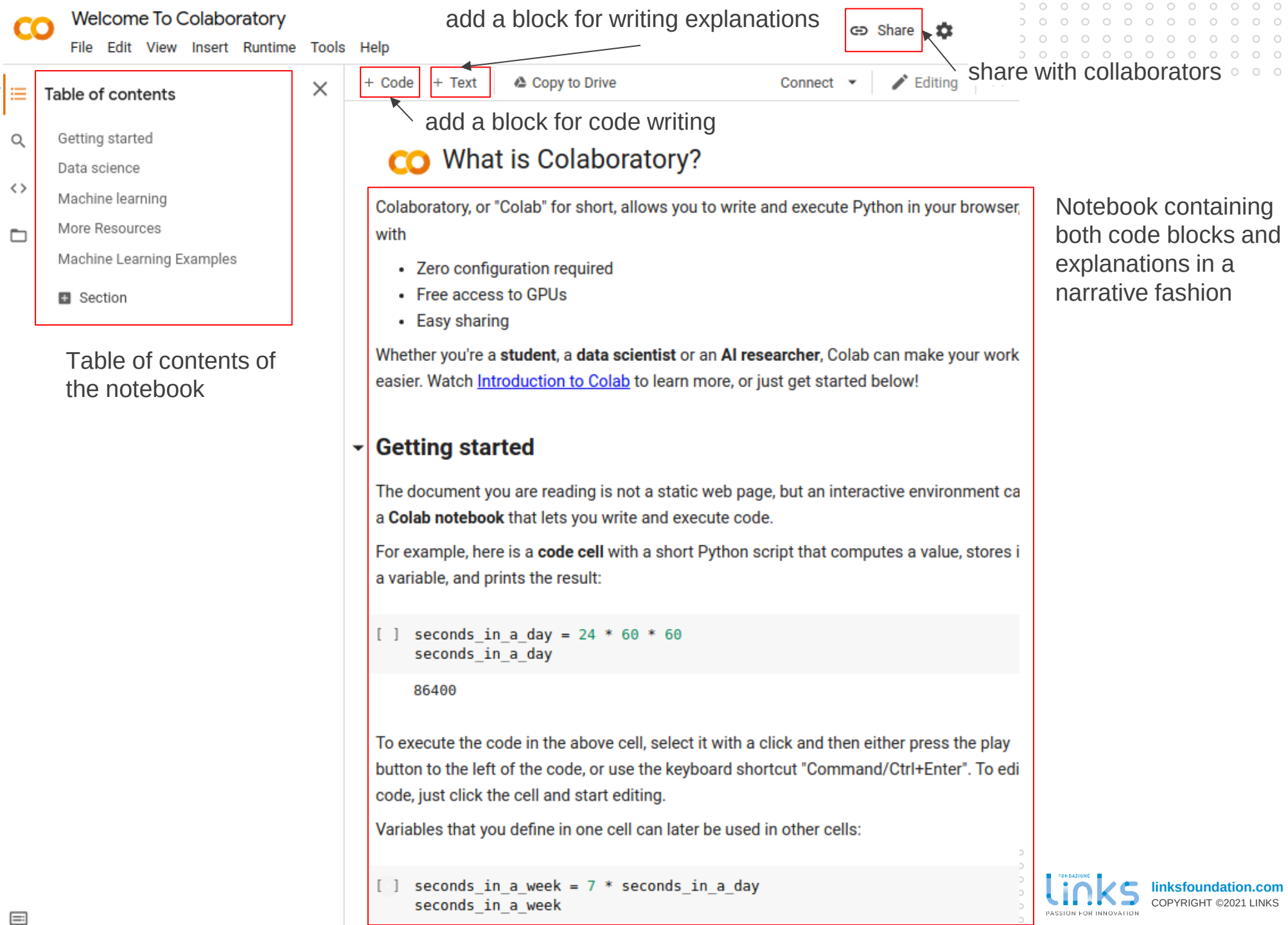
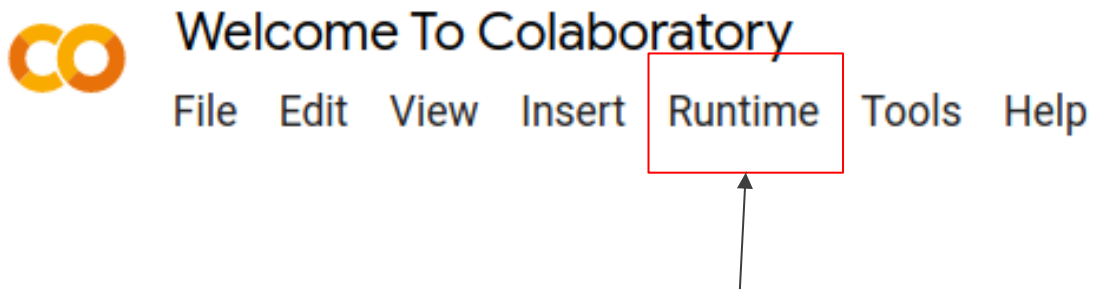


Table of contents of the notebook

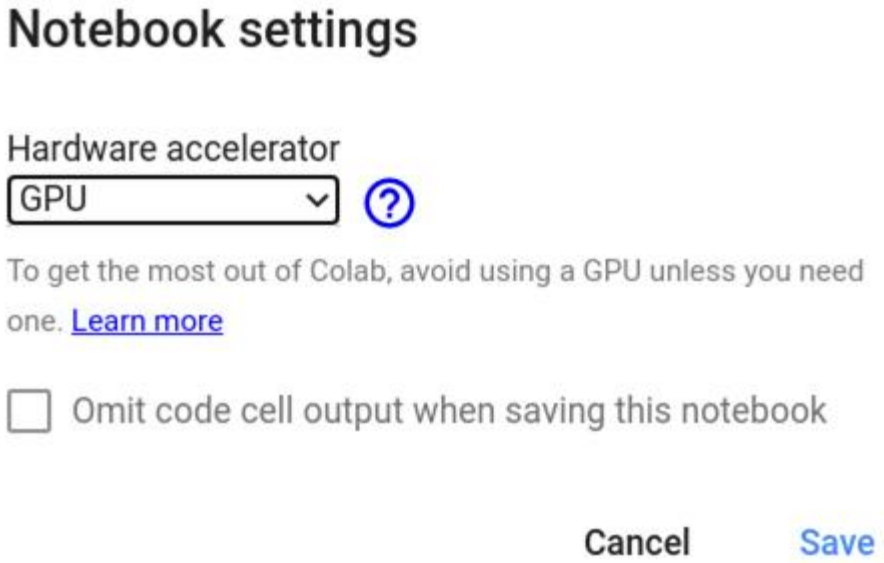
Notebook containing both code blocks and explanations in a narrative fashion

linksfoundation.com
COPYRIGHT ©2021 LINKS

Menu bar

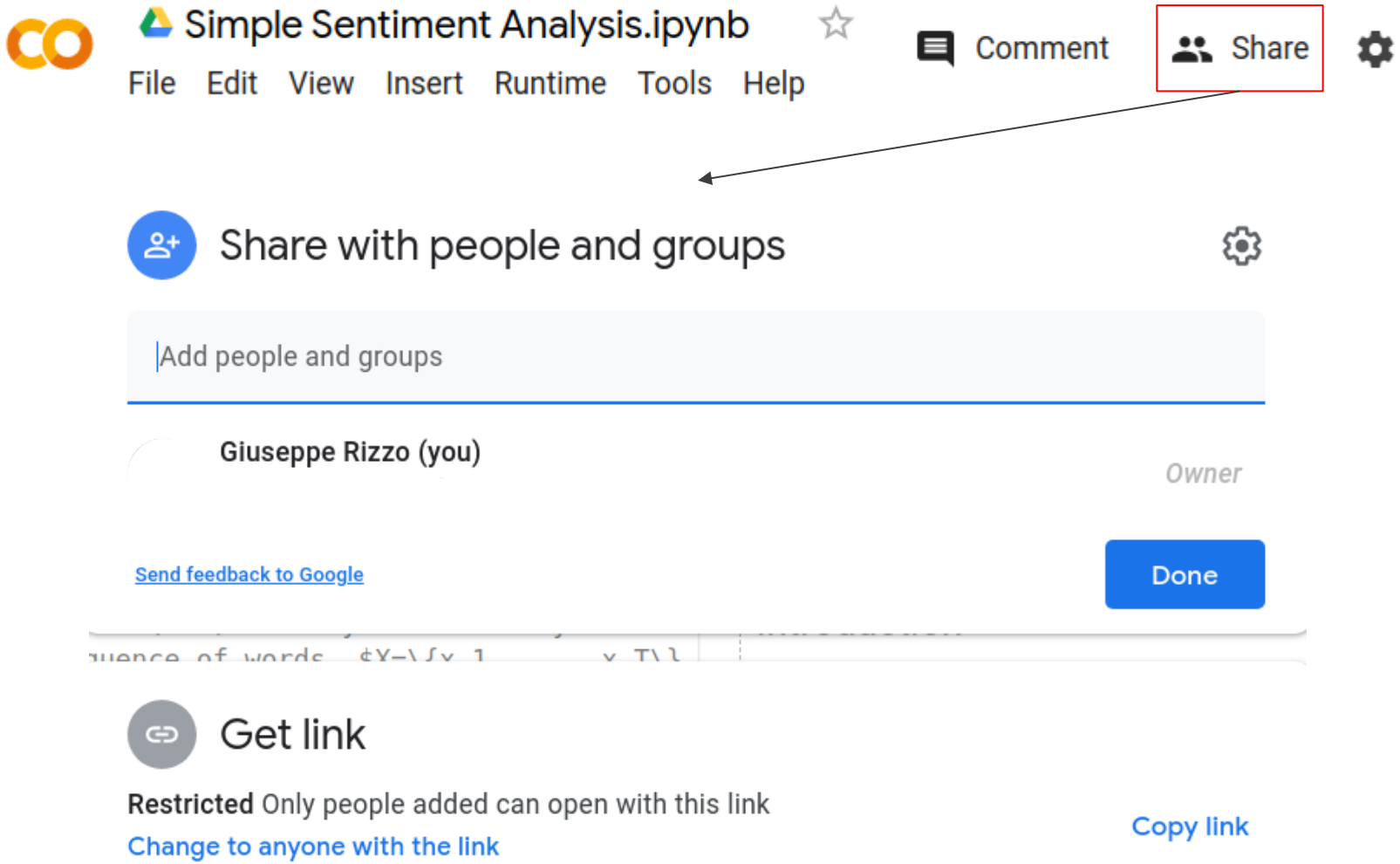


define the runtime configurations, among those the hardware acceleration if needed



linksfoundation.com
COPYRIGHT ©2021 LINKS

Share



linksfoundation.com
COPYRIGHT ©2021 LINKS

Connect your Google Drive

Enable authorization to import data from Google Drive

```
1  from google.colab import drive
2  drive.mount('/content/drive/')
```

List the files in your drive

```
1  !ls "/content/drive/My Drive/"
```


linksfoundation.com
COPYRIGHT ©2021 LINKS

Running with Google Colab

Upload YOUR_PYTHON_FILE.py to Google Drive & Run with Google Colab

```
1  !python3 "/content/drive/My Drive/Colab Notebooks/YOUR_PYTHON_FILE.py"
```

Run with Google Colab to Download YOUR_PYTHON_FILE.py from Google Drive

```
1  from google.colab import files

2  files.download('/content/drive/My Drive/Colab Notebooks/YOUR_PYTHON_FILE.py')
```


linksfoundation.com
COPYRIGHT ©2021 LINKS

Bash commands

Bash commands are executed with the environment “!”

Download an external file

```
1    !wget http://ai.stanford.edu/~amaas/data/sentiment/aclImdb_v1.tar.gz -P "/content/drive/My Drive/Colab Notebooks"
```

Clone a repository

```
1    !git clone https://github.com/pytorch/examples.git
```


linksfoundation.com
COPYRIGHT ©2021 LINKS

Colab == virtual environment

The environment can be customized with the addition of python packages

Install

```
1    !pip install -q datasets
```

Show a version

```
1    !pip show datasets
```


linksfoundation.com
COPYRIGHT ©2021 LINKS

Example

IMDb Sentiment Analysis.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Note: make sure that Runtime -> Change runtime type -> hardware accellerator -> GPU is selected

Install the **libraries** used by this Colab notebook in the virtual environment

```
[4]  !pip install -q datasets
     !pip install -q transformers
     !pip install -q simpletransformers
```

10s

Load IMDb review dataset

https://www.imdb.com/interfaces/

```
[5]  import pandas as pd
     from datasets import load_dataset

     dataset_train = load_dataset('imdb',split='train')
     dataset_train.rename_column('label', 'labels')
     train_df=pd.DataFrame(dataset_train)

     dataset_test = load_dataset('imdb',split='test')
     dataset_test.rename_column('label', 'labels')
     test_df=pd.DataFrame(dataset_test)
```

49s

```
Downloading builder script: 100% 4.31k/4.31k [00:00<00:00, 42.2kB/s]
Downloading metadata: 100% 2.17k/2.17k [00:00<00:00, 26.1kB/s]
Downloading and preparing dataset imdb/plain_text (download: 80.23 MiB, generated: 127.02 MiB, post-pr
Downloading data: 100% 84.1M/84.1M [00:11<00:00, 15.6MB/s]
Dataset imdb downloaded and prepared to /root/.cache/huggingface/datasets/imdb/plain_text/1.0.0/2fdd8b
WARNING:datasets.builder:Found cached dataset imdb (/root/.cache/huggingface/datasets/imdb/plain_text/
```


linksfoundation.com
COPYRIGHT ©2021 LINKS

FONDAZIONE links

PASSION FOR INNOVATION

Thank you for your attention.

Questions?

CONTACTS

Giuseppe Rizzo

Program Manager (LINKS Foundation) and
Adjunct Professor (Politecnico di Torino)

giuseppe.rizzo@polito.it

FONDAZIONE LINKS
Via Pier Carlo Boggio 61 | 10138 Torino
P. +39 011 22 76 150
LINKSFOUNDATION.COM