

# 2020/21 CSC 5741

## Jupyter Notebook—Python for Machine Learning

Lighton Phiri  
<lighton.phiri@unza.zm>

March 29 2021

### Contents

### Introduction

During these “hands-on” activities, we will explore and experiment the following: 1. Jupyter Notebooks—Quick walkthrough of Jupyter Notebooks 2. Google Colab—Introduction to the interface and useful tips 3. Python 3—Crash course introduction to Python 3 4. Core Python Modules—Quick walkthrough of some core Python modules that will be used in the course.

In all instances, you are encouraged to make reference to online documentation for the various tools. Additionally, you can exploit tools like [Zeal Offline Documentation Browser](#) to download and search through offline documentation. You are also encouraged to look up and explore other libraries, especially as you work towards the Mini Projects.

### General Notebook Configuration

```
[1]: # Aesthetics for pandas cell output
import pandas as pd

pd.set_option('display.latex.repr', True)
pd.set_option('display.latex.longtable', True)
pd.set_option('max_colwidth', 30)

# Show all output
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

### Jupyter Notebooks

- Install Jupyter Notebooks on your local machine
- Using Shell Commands
- Using Cell Magic Commands

### Installing Jupyter Notebook

- Go to <https://jupyter.org>

- Follow instructions specific to the platform you are using

## Using Shell Commands

- You want to ensure that the cell is set to be a code cell
- Prefix all shell commands with an !
  - You can then run all normal shell commands as you normally would using your preferred shell, e.g. bash
  - You can pipe/chain commands to get a desired outcome

```
[2]: # List files in current directory
!ls -ltr
```

```
total 44908
-rw-rw-r-- 1 lightonphiri lightonphiri    6640 Mar 30 10:49 db-
unza20-csc5741-ict1110_2018_19-ca_scores.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    4177 Mar 30 10:49 db-
unza20-csc5741-2018_19-student_demographics.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    6640 Mar 30 10:49 db-
unza19-ict1110_ca_scores_2018_19.csv
-rw-rw-r-- 1 lightonphiri lightonphiri  5458321 Mar 30 10:49 db-unza20-dspace_unza_zm-
json_records.json
-rw-rw-r-- 1 lightonphiri lightonphiri  8459526 Mar 30 21:56
dspace_unza_zm-20190417-2.csv
-rw-r--r-- 1 lightonphiri lightonphiri 11292672 Mar 30 22:07 dspace.sqlite
-rw-rw-r-- 1 lightonphiri lightonphiri   914796 Mar 31 08:23 zwdPageFBcrawl.pagePosts.csv
-rw-rw-r-- 1 lightonphiri lightonphiri   137156 Apr 11 05:44 _code-
unza21-csc5741-notebooks-python_for_data_mining.pdf
-rw-rw-r-- 1 lightonphiri lightonphiri    6714 Apr 11 05:50 pdftk-code-
unza21-csc5741-notebooks-python_for_data_mining.txt
-rw-rw-r-- 1 lightonphiri lightonphiri   165584 Apr 11 05:50 code-
unza21-csc5741-notebooks-python_for_data_mining.pdf
-rw-rw-r-- 1 lightonphiri lightonphiri    51407 Apr 12 04:47 img-
unza21-csc5741-postgraduate_grading.png
-rw-rw-r-- 1 lightonphiri lightonphiri    1041 Apr 12 05:55 unzapostgrad.py
-rw-rw-r-- 1 lightonphiri lightonphiri     72 Apr 12 18:21 Untitled.ipynb
-rw-rw-r-- 1 lightonphiri lightonphiri 12980043 Apr 19 19:36 db-unza19-dspace_unza_zm.csv
-rw-rw-r-- 1 lightonphiri lightonphiri   15361 May 24 15:06 db-
unza21-csc5741-ict1110_2018_19-preliminary_survey.csv
-rw-rw-r-- 1 lightonphiri lightonphiri  4751334 May 24 15:52 db-
unza21-csc5741-dspace_unza_zm.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    26252 May 30 15:50 db-
unza21-csc5741-ict1110_2019_20-preliminary_survey.csv
-rw-rw-r-- 1 lightonphiri lightonphiri   13296 May 30 15:59 db-
unza21-csc5741-ict1110_student_demographics.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    2147 May 30 16:06 db-
unza21-csc5741-ict1110_assessment_scores-final_examination.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    2196 May 30 16:07 db-
unza21-csc5741-ict1110_assessment_scores-test1.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    2124 May 30 16:07 db-
unza21-csc5741-ict1110_assessment_scores-test2.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    2074 May 30 16:07 db-
```

```

unza21-csc5741-ict1110_assessment_scores-test3.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      2088 May 30 16:07 db-
unza21-csc5741-ict1110_assessment_scores-test4.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1986 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz09.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1990 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz08.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1920 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz07.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      2067 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz06.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      2104 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz05.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1966 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz04.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      2009 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz03.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      2028 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz02.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1949 May 30 16:08 db-
unza21-csc5741-ict1110_assessment_scores-quiz01.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1748 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz19.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1646 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz18.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1676 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz17.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1729 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz16.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1987 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz15.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1983 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz14.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1969 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz13.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1981 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz12.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1955 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz11.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1905 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz10.csv
-rw-rw-r-- 1 lightonphiri lightonphiri      1862 May 31 15:06 db-
unza21-csc5741-ict1110_assessment_scores-quiz20.csv
-rw-rw-r-- 1 lightonphiri lightonphiri     40957 May 31 19:14 db-
unza21-csc5741-ict1110_preliminary_survey.csv
-rw-rw-r-- 1 lightonphiri lightonphiri     77238 Jun  1 08:38 code-
unza21-csc5741-notebook-data_preprocessing.ipynb
-rw-rw-r-- 1 lightonphiri lightonphiri    1016652 Jun  1 09:20 code-
unza21-csc5741-notebook-exploratory_data_analysis.ipynb
-rw-rw-r-- 1 lightonphiri lightonphiri      485 Jun  5 12:30 var_ict1110_grades.csv
-rw-rw-r-- 1 lightonphiri lightonphiri    401611 Jun  5 12:34 code-
unza21-csc5741-notebooks-python_for_data_mining.ipynb

```

```
[3]: # View first few lines of file contents using cat and head commands
```

```
!cat -n db-unza20-csc5741-ict1110_2018_19-ca_scores.csv | head -n 5
```

```
1 StudentID|StudentName|Quiz1|Quiz2|Quiz3|ClassTest1|Quiz4|Quiz5|Quiz6|Quiz7|ClassT
est2|Quiz8|Quiz9|Quiz10|Quiz11|ClassTest3|Quiz12|Quiz13|Quiz14|Quiz15|ClassTest4|MakeupTe
st|Quiz16|Quiz17|Quiz18|Quiz19|Quiz20
2 575b9408b6daa2ddcefbcf6d81c9b4c9|Algeria|6|1|3|20|3.5|5.5|10|2|24|3|8.5|6|4|22|10
|0.5|5|7|31.5|0|7|9|9|9|10
3 232bf11cb81bcdb269f76a08fde8b947|Angola|6|5|3|22.5|5|6|9|6|17|7|8|6|5|17|10|1|7|9
|20.5|0|2.5|9|9|9|10
4 9e7002d53d4db7bfad4f5cf419b0c126|Benin|3|4|2.5|20.5|4.5|5|9|1|12.5|5|5.5|0|3|22|1
0|0.5|5|9|22.5|34|1|8|10|8|9
5 6cd50fb3091b0a9d3c1ac2cf52441390|Botswana|10|3|2|16|5|4|8|1|21|5|9|0|0|14|10|0|3|
0|25|26|1|0|0|10|9
```

```
[4]: # Count the number of records in the file
```

```
# Remember that the first line is the header
```

```
!wc -l db-unza20-csc5741-ict1110_2018_19-ca_scores.csv
```

```
62 db-unza20-csc5741-ict1110_2018_19-ca_scores.csv
```

## Built-in Cell Magic Commands

```
[5]: # Use the lsmagic magic to determine the available magics
```

```
[6]: %lsmagic
```

```
[6]: Available line magics:
```

```
%alias %alias_magic %autoawait %autocall %automagic %autosave %bookmark %cat %cd
%clear %colors %conda %config %connect_info %cp %debug %dhist %dirs
%doctest_mode %ed %edit %env %gui %hist %history %killbgscripts %ldir %less
↳%lf
%lk %ll %load %load_ext %loadpy %logoff %logon %logstart %logstate %logstop
↳%ls
%lsmagic %lx %macro %magic %man %matplotlib %mkdir %more %mv %notebook %page
%pastebin %pdb %pdef %pdoc %pfile %pinfo %pinfo2 %pip %popd %pprint %precision
%prun %psearch %psource %pushd %pwd %pycat %pylab %qtconsole %quickref %recall
%rehashx %reload_ext %rep %rerun %reset %reset_selective %rm %rmdir %run %save
%sc %set_env %store %sx %system %tb %time %timeit %unalias %unload_ext %who
%who_ls %whos %xdel %xmode
```

```
Available cell magics:
```

```
%%! %%HTML %%SVG %%bash %%capture %%debug %%file %%html %%javascript %%js
%%latex %%markdown %%perl %%prun %%pypy %%python %%python2 %%python3 %%ruby
%%script %%sh %%svg %%sx %%system %%time %%timeit %%writefile
```

```
Automagic is ON, % prefix IS NOT needed for line magics.
```

```
[7]: # Use the time magic to determine execution time
```

```
[8]: %%time
print (5+5)
```

```
10
CPU times: user 198 µs, sys: 72 µs, total: 270 µs
Wall time: 176 µs
```

```
[9]: # Use the who magic to list all variables defined in the Notebook
```

```
[10]: %who
```

```
InteractiveShell      pd
```

```
[11]: # Use the HTML magic to render the PDF document
```

```
[12]: %%HTML
<embed src="https://www.jctr.org.zm/uploads/1/1/8/1/118170975/lusaka.pdf" width="950"
↳height="400">
```

```
<IPython.core.display.HTML object>
```

```
[13]: # Use the javascript magic to execute JavaScript code
```

```
[14]: %%javascript
function js_fxn_add (var_one, var_two) {
    return var_one + var_two;
}

console.log(js_fxn_add(5, 9))
```

```
<IPython.core.display.Javascript object>
```

## Google Colab

### Google Colab Interface

- Go to <https://colab.research.google.com>
- Explore the Google Colab UI

### Working with Datasets

#### Using Datasets on Local Drive

```
[15]: #from google.colab import files
#file.upload()
```

## Using Datasets Hosted on Google Drive

```
[16]: #from google.colab import drive
      #drive.mount('/content/drive')

      # Alternatively, you can mount the drive manually
```

Ensure that all datasets to be incorporated into pipeline are uploaded onto Google Drive

1. Mount Drive
2. Prefix all paths with “/content/drive”
  - Alternatively, right-click desired file and copy the path
3. Run shell commands and Python scripts as you would with Jupyter Notebook Classic

## Python Crash Introduction

- Installation and Setup
- Basics
- Useful Data Structures
- Flow Control
- Functions
- Modules
- Libraries

### Installation and Setup

#### Basic

```
[17]: import keyword
```

```
[18]: # type(keyword.kwlist)
      # age = 19
      # type (age)

      var_age = 19 # This variable is an interger
```

```
[19]: print?
```

```
[20]: def fxn_csc5741 (var_input1):
      """
      This is an example of a docstrig
      -----
      param: var_input1 --- Input value
      """
      return var_input1, type(var_input1)
```

```
[21]: help(fxn_csc5741)
```

Help on function fxn\_csc5741 in module \_\_main\_\_:

```
fxn_csc5741(var_input1)
  This is an example of a docstrig
```