

Introduction to Python

This notebook is primarily focused on introducing the specifics of using Python in an interactive environment such as Datalab. It is not intended to provide a complete tutorial to Python as a language. If you're completely new to Python, no problem! Python is quite straightforward, and there are lots of resources. The interactive step-by-step material at [Codecademy](https://www.codecademy.com/tracks/python) (<https://www.codecademy.com/tracks/python>) might be of interest.

To get started, below is a code cell that contains a Python statement. You can run it by pressing Shift+Enter or clicking the the Run toolbar button with the cell selected.

```
print("Hello World")
```

Hello World

You can edit the cell above and re-execute it to iterate over it. You can also add additional code cells to enter new blocks of code.

```
import sys

number = 10

def square(n):
    return n * n
```

The cell above created a variable named `number` and a function named `square`, and placed them into the *global namespace*. It also imported the `sys` module into the same namespace. This global namespace is shared across all the cells in the notebook.

As a result, the following cell should be able to access (as well as modify) them.

```
print('The number is currently %d' % number)
number = 11
sys.stderr.write('And now it is %d' % number)

square(number)
```

The number is currently 10

And now it is 11

121

By now, you've probably noticed a few interesting things about code cells:

- Upon execution, their results are shown inline in the notebook, after the code that produced the results. These results are included into the saved notebook. Results include outputs of print statements (text that might have been written out to stdout as well as stderr) and the final result of the cell.
- Some code cells do not have any visible output.
- Code cells have a distinguishing border on the left. This border is a washed out gray color when the notebook is first loaded, indicating that a cell has not been run yet; the border changes to a filled blue border after the cell runs.

Getting Help

Python APIs are usually accompanied by documentation. You can use `?` to invoke help on a class or a method. For example, execute the cells below:

```
str?
```

```
g = globals()  
g.get?
```

When run, these cells produce docstring content that is displayed in the help pane within the sidebar.

The code cells also provide auto-suggest. For example, press Tab after the `'.'` to see a list of members callable on the `g` variable that was just declared.

```
# Intentionally incomplete for purposes of auto-suggest demo, rather than running unmodified.  
g.
```

Function signature help is also available. For example, press Tab in the empty parentheses below.

```
str()  
  
"
```

Note that help in Python relies on the interpreter being able to resolve the type of the expression that you are invoking help on.

If you have not yet executed code, you may be able to invoke help directly on the class or method you're interested in, rather than the variable itself. Try this.

```
import datetime  
  
datetime.datetime?
```

Python Libraries

Datalab includes the standard Python library and a set of libraries that you can easily import. Most of the libraries were installed using pip, the Python package manager, or pip3 for Python 3.

```
%%bash
```

```
pip list --format=columns
```

Package	Version

appdirs	1.4.2
attrs	16.3.0
Automat	0.5.0
avro	1.8.1
backports-abc	0.5
backports.shutil-get-terminal-size	1.0.0
beautifulsoup4	4.5.3
bleach	1.5.0
brewer2mpl	1.4.1
bs4	0.0.1
certifi	2017.1.23
cffi	0.8.6
chardet	2.3.0
cloudml	0.1.9.1a0
colorama	0.3.2
configparser	3.5.0
constantly	15.1.0
crcmod	1.7
cryptography	0.6.1
cssselect	1.0.1
cycler	0.10.0
datalab	0.1.1703070322
decorator	4.0.11
dill	0.2.6
entrypoints	0.2.2
enum34	1.1.6
funcsigs	1.0.2
functools32	3.2.3.post2
future	0.16.0
futures	3.0.5
gapi-google-cloud-logging-v2	0.91.3
gapi-google-logging-v2	0.9.3
gapi-google-pubsub-v1	0.9.3
ggplot	0.6.8
google-api-python-client	1.5.1
google-apitools	0.5.8
google-auth	0.8.0
google-auth-httpplib2	0.0.2
google-cloud	0.19.0
google-cloud-core	0.23.1
google-cloud-dataflow	0.5.5
google-cloud-logging	0.23.1
google-gax	0.13.0
googleapis-common-protos	1.5.2
googledatastore	6.4.1
grpc-google-logging-v2	0.9.3
grpc-google-pubsub-v1	0.9.3

grpcio	1.1.3
html5lib	0.9999999
httplib2	0.9.2
incremental	16.10.1
ipykernel	4.4.1
ipython	5.3.0
ipython-genutils	0.1.0
ipywidgets	5.2.2
Jinja2	2.8
jsonschema	2.6.0
jupyter-client	5.0.0
jupyter-core	4.3.0
lxml	3.7.3
MarkupSafe	0.23
matplotlib	1.5.3
mistune	0.7.3
mltoolbox-datalab-classification-and-regression	1.0.0
mltoolbox-datalab-image-classification	0.1
mock	2.0.0
nb2kg	0.0.1.dev0
nbconvert	5.1.1
nbformat	4.3.0
ndg-httpsclient	0.3.2
nltk	3.2.1
notebook	4.2.3
numpy	1.11.2
oauth2client	2.2.0
packaging	16.8
pandas	0.19.1
pandas-profiling	1.4.0
pandocfilters	1.4.1
parsel	1.1.0
pathlib2	2.2.1
patsy	0.4.1
pbr	2.0.0
pexpect	4.2.1
pickleshare	0.7.4
Pillow	3.4.1
pip	9.0.1
plotly	1.12.5
ply	3.8
prompt-toolkit	1.0.13
proto-google-cloud-logging-v2	0.91.3
proto-google-datastore-v1	1.3.1
protobuf	3.1.0
protorpc	0.11.1
psutil	4.3.0
ptyprocess	0.5.1
pyasn1	0.2.3
pyasn1-modules	0.0.8

pycparser	2.10
PyDispatcher	2.0.5
Pygments	2.1.3
pyOpenSSL	0.14
pyparsing	2.2.0
python-dateutil	2.5.0
python-gflags	3.1.1
pytz	2016.7
PyYAML	3.11
pyzmq	16.0.2
queuelib	1.4.2
requests	2.9.1
rsa	3.4.2
scandir	1.5
scikit-learn	0.17.1
scipy	0.18.0
Scrapy	1.3.2
seaborn	0.7.0
service-identity	16.0.0
setuptools	34.3.1
simplegeneric	0.8.1
simplejson	3.10.0
singledispatch	3.4.0.3
six	1.10.0
statsmodels	0.6.1
sympy	0.7.6.1
tensorflow	1.0.0
terminado	0.6
testpath	0.3
tornado	4.4.2
traitlets	4.3.2
Twisted	17.1.0
uritemplate	0.6
urllib3	1.9.1
w3lib	1.17.0
wcwidth	0.1.7
wheel	0.29.0
widetsnbextension	2.0.0
zope.interface	4.3.3

If you have suggestions for additional packages to include, please submit feedback proposing the inclusion of the packages in a future version.

Installing a Python Library

You can use pip to install your own Python 2 libraries, or pip3 to install Python 3 libraries.

Keep in mind that this will install the library within the virtual machine instance being used for Datalab, and the library will become available to all notebooks and all users sharing the same instance.

The library installation is temporary. If the virtual machine instance is recreated, you will need to reinstall the library.

The example, below, installs scrapy (<http://scrapy.org/>), a library that helps in scraping web content.

```
%%bash
apt-get update -y
apt-get install -y -q python-dev python-pip libxml2-dev libxslt1-dev zlib1g-dev libffi-dev libssl-dev
pip install -q scrapy
```

```
Hit http://security.debian.org jessie/updates InRelease
Hit http://ftp.us.debian.org testing InRelease
Ign http://deb.debian.org jessie InRelease
Hit http://deb.debian.org jessie-updates InRelease
Get:1 http://security.debian.org jessie/updates/main amd64 Packages [444 kB]
Hit http://deb.debian.org jessie Release.gpg
Hit http://deb.debian.org jessie Release
Get:2 http://ftp.us.debian.org testing/main Sources [8936 kB]
Get:3 http://deb.debian.org jessie-updates/main amd64 Packages [17.6 kB]
Get:4 http://deb.debian.org jessie/main amd64 Packages [9049 kB]
Fetched 18.4 MB in 14s (1279 kB/s)
Reading package lists...
Reading package lists...
Building dependency tree...
Reading state information...
libffi-dev is already the newest version.
libxml2-dev is already the newest version.
libxslt1-dev is already the newest version.
python-dev is already the newest version.
python-pip is already the newest version.
zlib1g-dev is already the newest version.
libssl-dev is already the newest version.
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
```

Inspecting the Python environment by running pip list, we should now see that Scrapy is installed and ready to use.


```
%%bash
```

```
pip list --format=columns
```

Package	Version

appdirs	1.4.2
attrs	16.3.0
Automat	0.5.0
avro	1.8.1
backports-abc	0.5
backports.shutil-get-terminal-size	1.0.0
beautifulsoup4	4.5.3
bleach	1.5.0
brewer2mpl	1.4.1
bs4	0.0.1
certifi	2017.1.23
cffi	0.8.6
chardet	2.3.0
cloudml	0.1.9.1a0
colorama	0.3.2
configparser	3.5.0
constantly	15.1.0
crcmod	1.7
cryptography	0.6.1
cssselect	1.0.1
cycler	0.10.0
datalab	0.1.1703070322
decorator	4.0.11
dill	0.2.6
entrypoints	0.2.2
enum34	1.1.6
funcsigs	1.0.2
functools32	3.2.3.post2
future	0.16.0
futures	3.0.5
gapi-google-cloud-logging-v2	0.91.3
gapi-google-logging-v2	0.9.3
gapi-google-pubsub-v1	0.9.3
ggplot	0.6.8
google-api-python-client	1.5.1
google-apitools	0.5.8
google-auth	0.8.0
google-auth-httpplib2	0.0.2
google-cloud	0.19.0
google-cloud-core	0.23.1
google-cloud-dataflow	0.5.5
google-cloud-logging	0.23.1
google-gax	0.13.0
googleapis-common-protos	1.5.2
googledatastore	6.4.1
grpc-google-logging-v2	0.9.3
grpc-google-pubsub-v1	0.9.3

grpcio	1.1.3
html5lib	0.9999999
httplib2	0.9.2
incremental	16.10.1
ipykernel	4.4.1
ipython	5.3.0
ipython-genutils	0.1.0
ipywidgets	5.2.2
Jinja2	2.8
jsonschema	2.6.0
jupyter-client	5.0.0
jupyter-core	4.3.0
lxml	3.7.3
MarkupSafe	0.23
matplotlib	1.5.3
mistune	0.7.3
mltoolbox-datalab-classification-and-regression	1.0.0
mltoolbox-datalab-image-classification	0.1
mock	2.0.0
nb2kg	0.0.1.dev0
nbconvert	5.1.1
nbformat	4.3.0
ndg-httpsclient	0.3.2
nlTK	3.2.1
notebook	4.2.3
numpy	1.11.2
oauth2client	2.2.0
packaging	16.8
pandas	0.19.1
pandas-profiling	1.4.0
pandocfilters	1.4.1
parsel	1.1.0
pathlib2	2.2.1
patsy	0.4.1
pbr	2.0.0
pexpect	4.2.1
pickleshare	0.7.4
Pillow	3.4.1
pip	9.0.1
plotly	1.12.5
ply	3.8
prompt-toolkit	1.0.13
proto-google-cloud-logging-v2	0.91.3
proto-google-datastore-v1	1.3.1
protobuf	3.1.0
protorpc	0.11.1
psutil	4.3.0
ptyprocess	0.5.1
pyasn1	0.2.3
pyasn1-modules	0.0.8

pycparser	2.10
PyDispatcher	2.0.5
Pygments	2.1.3
pyOpenSSL	0.14
pyparsing	2.2.0
python-dateutil	2.5.0
python-gflags	3.1.1
pytz	2016.7
PyYAML	3.11
pyzmq	16.0.2
queuelib	1.4.2
requests	2.9.1
rsa	3.4.2
scandir	1.5
scikit-learn	0.17.1
scipy	0.18.0
Scrapy	1.3.2
seaborn	0.7.0
service-identity	16.0.0
setuptools	34.3.1
simplegeneric	0.8.1
simplejson	3.10.0
singledispatch	3.4.0.3
six	1.10.0
statsmodels	0.6.1
sympy	0.7.6.1
tensorflow	1.0.0
terminado	0.6
testpath	0.3
tornado	4.4.2
traitlets	4.3.2
Twisted	17.1.0
uritemplate	0.6
urllib3	1.9.1
w3lib	1.17.0
wcwidth	0.1.7
wheel	0.29.0
widetsnbextension	2.0.0
zope.interface	4.3.3