



WORKSHOPS

Week 1 -Introduction



PYTHON

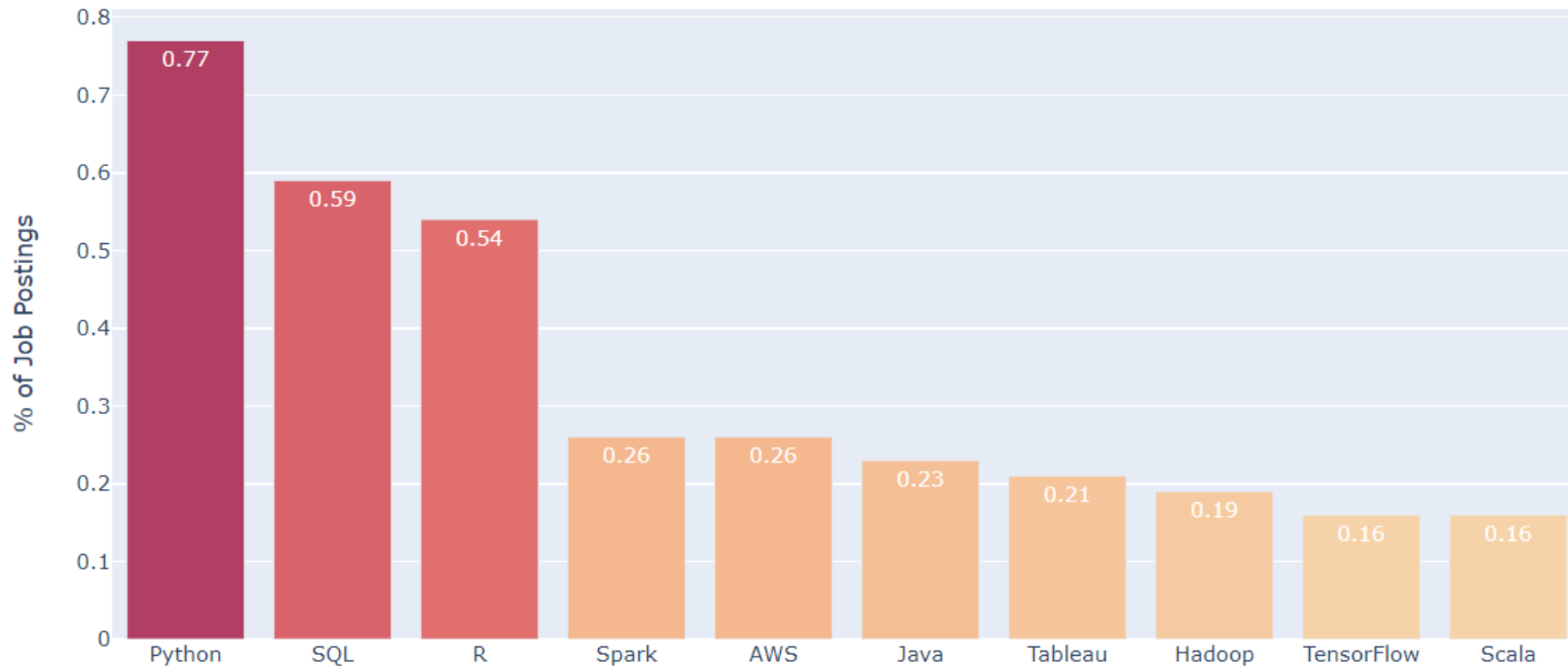
WHAT IS PYTHON?

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed [1].

THE MOST IN-DEMAND SKILLS FOR DATA SCIENTISTS IN 2021

RESULTS FROM WEBSCRAPING OVER 15,000 DATA SCIENTIST JOB POSTINGS

10 Most In-Demand Data Science Skills in 2021



WHY PYTHON?

- It is free and open-source software.
- It is well-documented and runs on all platforms.
- It has a large and constantly growing user-base which includes scientists.
- It is easier for novices to pick up than most other languages.

WHY PYTHON FOR DATA ANALYSIS?

Python has developed a large and active scientific computing and data analysis community. In the last 10 years, Python has become one of the most important languages for data science, machine learning, and general software development in academia and industry. In recent years, Python's improved support for libraries (such as pandas and scikit-learn) has made it a popular choice for data analysis tasks. Combined with Python's overall strength for general-purpose software engineering, it is an excellent option as a primary language for building data applications [2].

ANACONDA



WHAT IS ANACONDA?

Anaconda is a distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS [3].



Anaconda at a glance 🧐

25M

Anaconda users

235

Countries & regions with
Anaconda users

2.4

Billion package
downloads in 2019

24,684

New packages added to
anaconda.org in 2019

192

Pounds of jellybeans
fueled our team last year

3,000+

Lego minifigs given
away at our events

JUPYTER NOTEBOOK



WHAT IS THE JUPYTER NOTEBOOK APP?

The Jupyter Notebook App is a server-client application that allows editing and running notebook documents via a web browser. The Jupyter Notebook App can be executed on a local desktop requiring no internet access (as described in this document) or can be installed on a remote server and accessed through the internet [4].

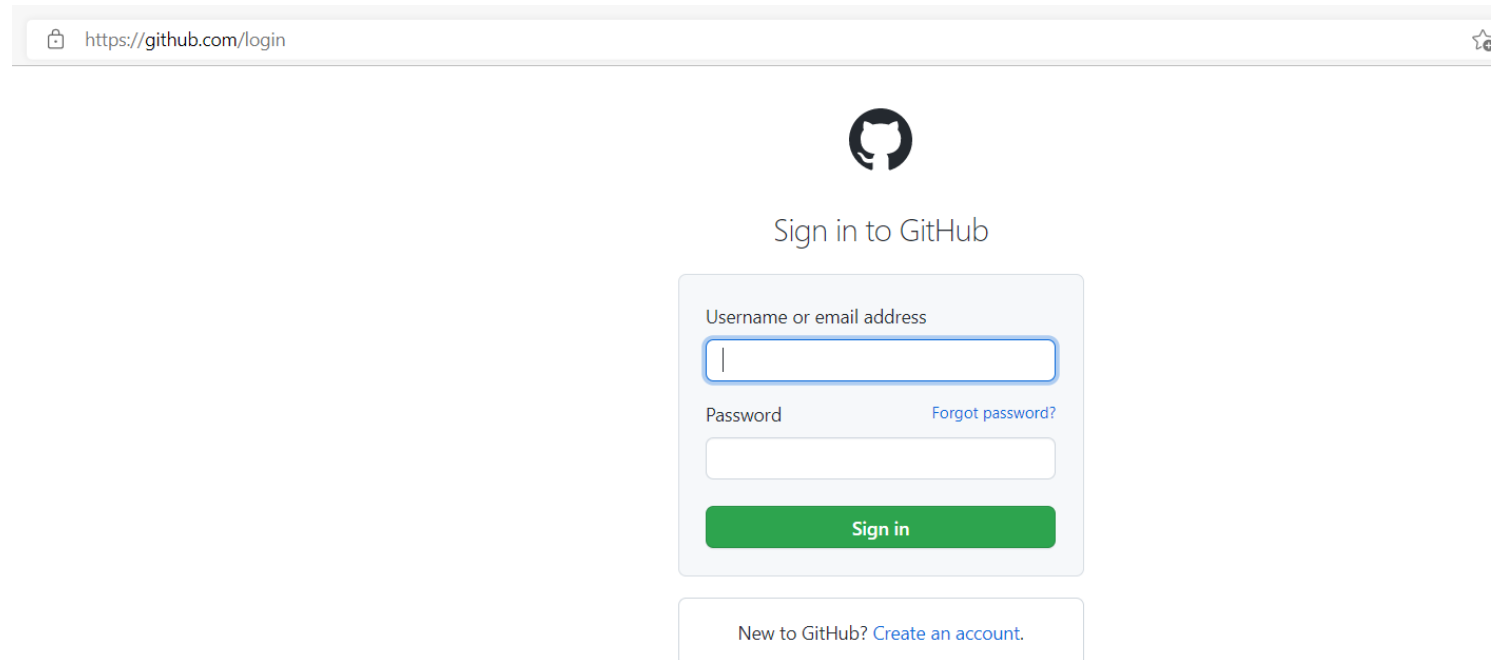
WHAT IS NOTEABLE?

The Noteable service is a cloud-based application providing access to Jupyter notebooks online. Noteable provides a central storage space to store and run Jupyter notebooks in a variety of languages.

The purpose of Noteable is to allow students and staff to access Jupyter notebooks at any time without the need for pre-installation which can be cumbersome and difficult for programming novices. Noteable is integrated with Learn to allow for a central launch point into a pre-set environment without the need for a separate login. Find more: <https://www.ed.ac.uk/information-services/learning-technology/noteable/about>

SET UP

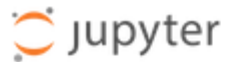
Step 1: Log in to your github account. If you don't have an account, please go to <https://github.com/> and sign up for a free account. It will be best if you used your personal email so you retain access to the material independent of your University account.



The screenshot shows the GitHub login interface. At the top, the browser address bar displays 'https://github.com/login'. Below this is the GitHub logo (Octocat) and the text 'Sign in to GitHub'. The login form consists of two input fields: 'Username or email address' and 'Password'. The 'Username or email address' field is currently empty. To the right of the password field is a link that says 'Forgot password?'. Below the input fields is a green 'Sign in' button. At the bottom of the form, there is a link that says 'New to GitHub? Create an account.'

SET UP

Step 2: Log in to Noteable
(<https://noteable.edina.ac.uk/launch>) using your University of Edinburgh student credentials.



Files

Running

Assignments

Select items to perform actions on them.

+GitRepo

Disk

Empty Trash

Upload

New ▾

↺

☐ 0 ▾

📁 /

Name ▾

Last Modified

File size

The notebook list is empty.

SET UP

Step 3: Click on the +GitRepo icon as shown bellow and add the github repository with the Jupyter Notebooks and data files necessary for the workshops.



FilesRunningAssignments

Select items to perform actions on them.

☐ 0

/

Enter the details of the Git Repository to clone:

Git Repository URL:

https://github.com/GSA/dat

Branch*:

<default>

Username*:

Password*:

* Optional

Clone

Cancel

+GitRepo

Disk

Empty Trash

Upload

New

Pull down a Git repository

Last Modified

File size

SET UP

Step 4:

Git Repository URL: <https://github.com/dsmanufacturing/Workshops.git>

Branch: as is

Username: your github username

Password: your github password



Files

Running

Assignments

Select items to perform actions on them.

0

/

Enter the details of the Git Repository to clone:

Git Repository URL:

https://github.com/GSA/data

Branch*:

<default>

Username*:

Password*:

* Optional

Clone

Cancel

+GitRepo

Disk

Empty Trash

Upload

New

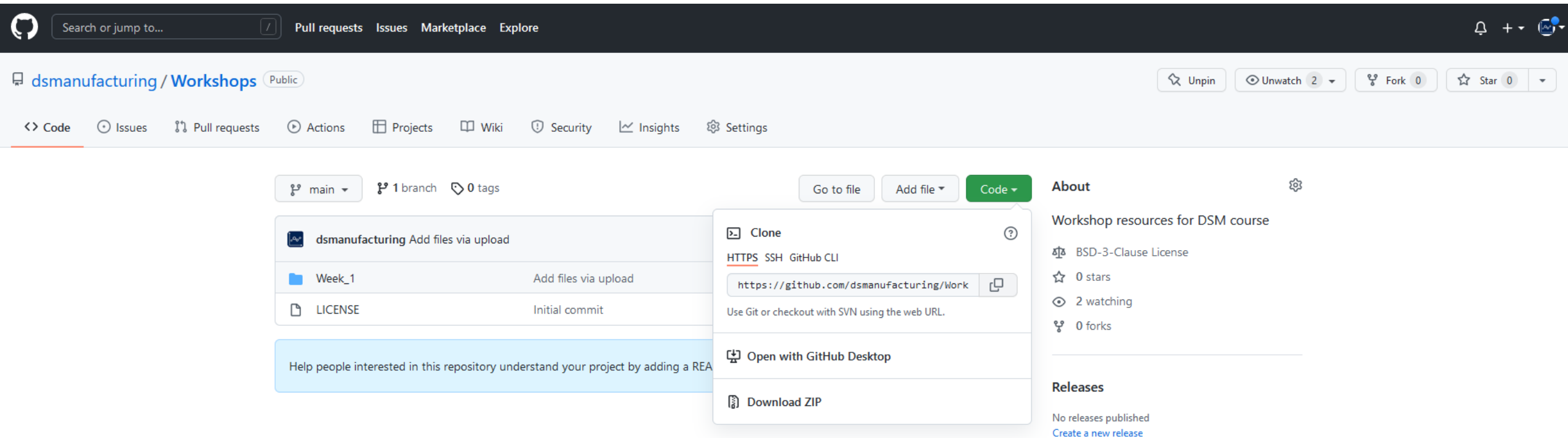
Pull down a Git repository

Last Modified

File size

SET UP

You can get the repository URL by clicking on the Code button on the github repository



The screenshot shows the GitHub interface for the repository `dsmanufacturing / Workshops`. The repository is public and has 1 branch (main) and 0 tags. The 'Code' button is highlighted, and a dropdown menu is open, showing options to clone the repository using HTTPS, SSH, or GitHub CLI, or to open it with GitHub Desktop or download a ZIP file. The repository contains files `Week_1` and `LICENSE`. The 'About' section shows the repository is licensed under BSD-3-Clause License, has 0 stars, 2 watchers, and 0 forks. The 'Releases' section shows no releases published.



SET UP

Step 5: Your Noteable page should look like this



Files Running Assignments

Select items to perform actions on them.

+GitRepo Disk Empty Trash Upload New ↕

<input type="checkbox"/> 0	/ Workshops.git	Name ↓	Last Modified	File size
<input type="checkbox"/>	..		seconds ago	
<input type="checkbox"/>	Week_1		seconds ago	
<input type="checkbox"/>	LICENSE		seconds ago	1.52 kB

REFERENCES

1. Python.org. 2022. *What is Python? Executive Summary*. [online] Available at: <<https://www.python.org/doc/essays/blurb/>>
2. McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.
3. En.wikipedia.org. 2022. *Anaconda (Python distribution)* - Wikipedia. [online] Available at: [https://en.wikipedia.org/wiki/Anaconda_\(Python_distribution\)](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)).



RESOURCES

1. <https://docs.anaconda.com/anacondaorg/faq/#what-is-anaconda-inc>
2. <https://docs.anaconda.com/anacondaorg/glossary/#cloud-glossary-cloud>

