2022

ICQCM

Introduction to Python

ASYNCHRONOUS/SYNCHRONUS INSTRUCTION
ALFRED D. HULL, MBA, CIMP: BUSINESS ANALYTICS
CIMP MASTER DATA MANAGEMENT & DATA QUALITY
SPRINGBOARD CERTIFIED DATA SCIENTIST
SPRINGBOARD CERTIFIED DATA ENGINEER

Connect with me here:

INSTITUTE IN CRITICAL QUANTITATIVE, COMPUTATIONAL, & MIXED METHODOLOGIES

Roger

Setting up your Data Science Lab

1. Anaconda installation documentation:

As a Data Scientist, you will need an integrated development environment (IDE) to create code. Please follow the instructions listed in the anaconda documentation provided below.

Anaconda is a free and open-source distribution of Python and R. The distribution comes with the Python interpreter and various packages related to machine learning and data science. Anaconda's idea is that it is easy and free for individuals in the field of Data Science to work in with a single installation.

Link: https://docs.anaconda.com/anaconda/install/

2. JupyterLab documentation:

JupyterLab is an open-source web application primarily designed to provide a user interface based on Jupyter Notebook. The features are like that of the latter, such as text editor, web browser support, and many more, except that it offers improved support for third-party extensions. The installation can be done using simple Python code for conda and pip software package. The packages are available for Windows, Mac, and Linux operating systems(OS) and is necessary to run JupyterLab. One needs to have the latest version of Jupyter Notebook installed (version 5.3 or above) [Sharma 2018].

Link: https://jupyterlab.readthedocs.io/en/stable/getting_started/overview.html

3. BONUS/OPTIONAL Intro to GIT:

By far, the most widely used modern version control system in the world today is Git. Git is a mature, actively maintained open-source project initially developed in 2005 by Linus Torvalds, the Linux operating system kernel's famous creator. A staggering number of software projects rely on Git for version control, including commercial projects as well as open-source. Developers who have worked with Git are well represented in the pool of available software development talent. It works well on a wide range of operating systems and IDEs (Integrated Development Environments) [Atlassian 2019].

Link: https://www.youtube.com/watch?v=dnDHVbO7ycM

Data Structures

In this subunit, we'll cover data structures in Python. As you learn to be a data scientist, how you structure your data is of principal importance. It is essential to know which data structure is the right one to use facing various scenarios.

1. Intro to Data Structures

This video will explain what a data structure is. Data structures store data as objects designed to make manipulating their underlying data simple. Data structures can be linear, like arrays and linked lists, or non-linear, like trees and graphs. They organize and prioritize information into a dataset that can compress easily to save storage space. Since they are structured logically, retrieving information is straightforward.

Video 1. (10 -20 minutes)
What are Data Structures? & Why we need them? DS Real World Example
https://www.youtube.com/watch?v=NzUVn8Y9deQ

2. Abstract Data Types

This video will dive into abstract data types (ADTs). ADTs are logical descriptions of how data is viewed and the operations allowed when working with it. They are not an implementation of those operations but a representation of what is possible. This is a bit of an ABSTRACT concept, so if you don't totally understand ADTs, this video will clear them up.

Video 2. (10 -20 minutes)
What is Abstract Data Types(ADT) in Data Structures ? | with Example
https://www.youtube.com/watch?v=n0e27Cpc88E

Lab Exercise: Getting started writing Python Code

In this subunit, we will look at a .pynb file that has a series of exercises to work through that will build you up to getting use to Anaconda, JupyterLab, and Python.

1. Lab exercise

The lab exercise will cover a series of topics below:

1. Topics: Quick Overview
- Data types
- Expressions
- Statements
- Variables

2. Topics: Quick overview
- data types
- Boolean examples
- Number examples
- Identity, using id() function
- Importing -- Math module: sqrt() function
- String examples
- Input from users

3. Topics: Quick overview
- Basic usage of the str.zfill() method
- Basic usage of the str.format() method
- keyword arguments
- Formatting Strings

4. Topics: Quick overview
- Numbers
- Strings
- Lists
- Dictionaries
- Tuples

5. Topics: Quick overview
- Redefining/redirecting standard out (stdout) to print to a file
- Redirecting standard out (stdout) back to printing to the screen

6. Topics: Quick overview
- Getting input from the user
- If-statement // if-else statement // if-else-if statement

Reference:

Sharma, A. (2018, March 5). JupyterLab - What is it? - Analytics India Magazine. Analytics India Magazine. https://analyticsindiamag.com/jupyterlab-what-is-it/

PROGIT. (n.d.). Git - About Version Control. Www.git-Scm.com. Retrieved February 3, 2021, from https://www.git-scm.com/book/en/v2/Getting-Started-About-Version-Control

Atlassian. (2019). What is Git: become a pro at Git with this guide | Atlassian Git Tutorial. Atlassian. https://www.atlassian.com/git/tutorials/what-is-git

Vanderplas, J. T. (2017). Python data science handbook: essential tools for working with data. O'Reilly, Cop.

Adams, C. (2014). Learning Python data visualization: master how to build dynamic HTML5-ready SVG charts using Python and the pygal library. Packt Publishing Ltd.

Data Science Project Management. (n.d.). CRISP-DM. Data Science Project Management. https://www.datascience-pm.com/crisp-dm-2/