Pre-Lectorials Directed Tasks

6G7V0026 Principles of Data Science

Luciano Gerber

Block 1, Week 1, 2021/22 (20-Sep-21)

Suggested Pre-Lectorials Directed Tasks

Familiarisation with Google Colaboratory and Python Notebooks

- Our main data science environment (for all the practical work) is Google Colaboratory (or Colab) and it would be ideal if you could now set your account up, if you haven't done so already. It is accessible with a free Google account.
- We will build **Python Notebooks** for implementing and documenting our data science solutions.
- For an overview of what Colab is, what notebooks look like, and what you can do with them, please go through the following online resources:
  - Introduction to Google Colab
  - Overview of Colaboratory Features
  - Markdown Guide
- Optionally, if you would like to know more about importing data into your Colab area, please visit External Data: Drive, Sheets, and Cloud Storage.

Familiarisation with the Programming Language Python

- You would have received an invitation from **DataCamp** to join for free. Please use your **MMU email address** (i.e., that to which the notification was sent) for completing your registration.
- Please complete the sections *Python Basics*, *Python Lists*, and *Functions and Packages* of the DataCamp's Intro to Python for Data Science
  - please feel free to complete the `numpy` section in your independent study time. This is an essential package for most of the day-to-day of data scientists and

machine learning engineers working with Python, and is also used as the basis and inspiration for many other useful packages.

- Have already some experience with programming or would like to look further? Please go through the following sections of the book A Whirlwind Tour of Python (notebooks accompanying the book are available here):
  - Introduction
  - How to Run Python Code
  - A Quick Tour Of Python Language Syntax
  - Basic Python Semantics: Variables and Objects
  - Basic Python Semantics: Operators
  - Built-In Types: Simple Values
  - Built-In Data Structures
  - Control Flow

Optional: Getting Familiarised with The `bash` Linux Shell (Optional)

- Interface to the operating system for navigating a file system, managing and manipulating files and directories, running programs and managing processes, among others.
- Desirable transferable skill; makes your workflow much more efficient. It is not a requirement for completing this unit successfully.
- Attempt the sections *Manipulating files and directories* and *Manipulating data* of the DataCamp's Introduction to Shell for Data Science.
  - feel free to complete the other sections in your independent study time.

Optional: Custom Python Data Science Track at DataCamp (Optional)

- I have created a custom learning track at DataCamp for your independent study time. It comprises short courses on power tools for Data Science (e.g., `bash` shell, `git`), as well as three courses on introductory Python Data Science.