IE 0015: Introduction to Data Analytics

Lecture 5: Introduction to Python

Announcements

- Assignments
  - Lab 2 is due by midnight this Friday
  - Homework 1 will be posted this week (due 2/19)

Announcements

- Today’s lecture
  - Re-Course schedule and motivation
  - Intro to python
  - A lot of conceptual questions, repeating concepts
  - Some live-coded examples after concepts

Schedule

- *What were we here to learn?*

Mild intro, 5/10

- *Databases*
- *SQL*

Mild intro, more research required

- *Data Science / Machine Learning*

Schedule

- *Non-negotiable:*

Must understand, >=7/10

- *Programming*
- *Python*
- *Data within python*

Unfortunately, we can’t avoid learning this at this point…

Putting Python into context ***(data / IE)***

- Recall the first few steps in data science:
  - Define problem of interest
  - Collect data
  - Store and then retrieve data from a database
- Now we need to clean and analyze the data
  - This is where Python comes in
- Python is a programming language

Putting Python into context ***(data / IE)***

- How does Python compare with other programming languages?
- Three types of programming language:
  - General purpose
    - Java, C++, Python, …
  - Scientific computing
    - Matlab, Mathematica, Python
  - Data science
    - R and Python
    - And sometimes Matlab

Putting Python into context ***(CS)***

*What is a **programming language?***

Putting CS into context ***(science)***

*What is a **language?***

Actual definitions

A **programming language** is a system of notation for writing computer programs.[1] Programming languages are described in terms of their syntax (form) and semantics (meaning), usually defined by a formal language.

language, a system of conventional spoken, manual (signed), or written symbols by means of which human beings, as members of a social group and participants in its culture, express themselves. The functions of language

Compiled vs interpreted ***(pre-survey)***

Is python interpreted or compiled?

Compiled vs interpreted ***(pre-survey)***

Is python interpreted or compiled?

Lets see an example…

Compiled vs interpreted ***(pre-survey)***

What is a compiler?

Compiled vs interpreted ***(pre-survey)***

What is a compiler?

a program that reads in a whole file, and translates it to assembly code

Compiled vs interpreted ***(pre-survey)***

What is an interpreter?

Compiled vs interpreted ***(pre-survey)***

What is an interpreter?

a program that reads in a file line by line, running each line…

Compiled vs interpreted ***(pre-survey)***

What is an interpreter?

a program that reads in a file line by line,
running each line…

Compiled vs interpreted ***(pre-survey)***

What is an interpreter?

a program that reads in a file line by line,
running each line…

Summary – language-agnostic concepts

What is a compiler?

a program that reads in a whole file, and translates it to assembly code…

What is an interpreter?

a program that reads in a file line by line,

running each line…

How does this affect my languages?

Python is almost always interpreted, so compilation and execution are hidden (by the interpreter), C and C++ are compiled.

Types & variables ***(agnostic)***

What is a variable?

Types & variables ***(agnostic)***

What is a variable?

A symbolic name for data, with an associated storage location in memory

Types & variables ***(agnostic)***

What is a variable?

A symbolic name for data, with an associated storage location in memory

Name + storage

Types & variables ***(agnostic)***

What is a variable?

A symbolic name for data, with an associated storage location in memory

Name + storage

Types & variables ***(agnostic)***

What is a type?

Types & variables ***(agnostic)***

What is a type?

Grouping or categorization of data values

Types & variables ***(agnostic)***

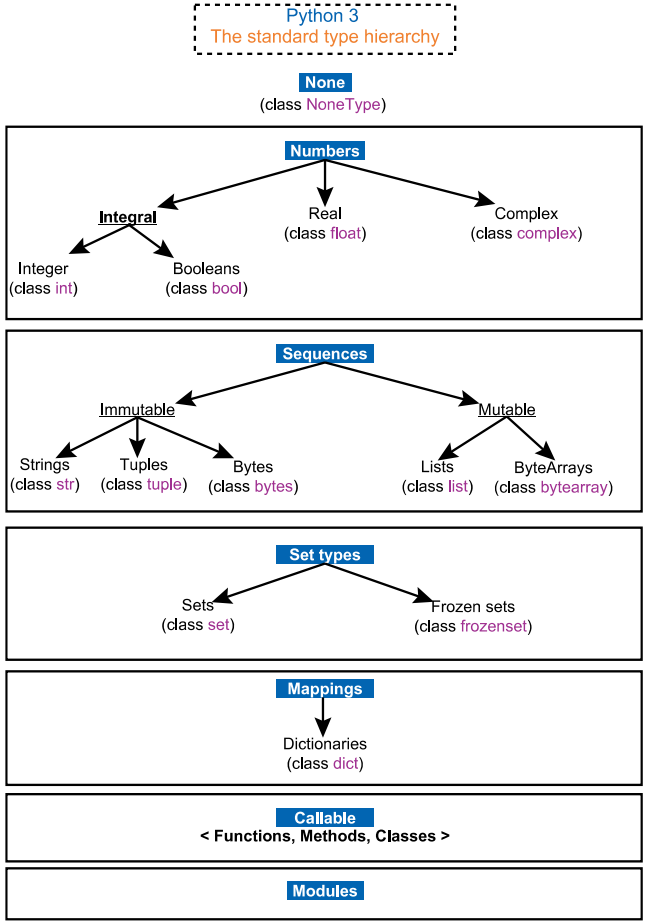
What is a type?

Grouping or categorization of data values

https://en.wikipedia.org/wiki/Data_type
https://en.wikipedia.org/wiki/Type_system

Types & variables ***(Python)***

- Python is dynamically-typed
- You don’t really know what this means
- The alternative is statically-typed



On to live coding

Do we need to install anything…

Following along

- We have not talked about IDE
- I would recommend, prefer, that you try 5-6 in the next two weeks
- Have ***VSCode*** as a starting point

More learning notes for coding

- Optimal learning is probably around 80-90% ***doing*** (coding or thinking about how to solve a problem with code) and 10-20% reading your Python textbook

- You need to code things that you are not qualified to code

Remember that python is general purpose, so you can combine data science with crazy ideas

More learning notes for coding

- Optimal learning is probably around 80-90% ***doing*** (coding or thinking about how to solve a problem with code) and 10-20% reading your Python textbook

- You need to code things that you are not qualified to code

Don’t wait – you have the internet, you can email me questions….

More learning notes for coding

- In summary – if you are a good student and want to invest the extra 5-10 hours in the next 2 weeks

- You should read chapters 1-2 of PD4A (quite quickly, spend 1-2 hrs max) and build 1-2 small projects in python (5-8 hrs)

Reach out to me if you need help or ideas

Python

- Let’s go install VSCode and take a look at it

Looking ahead

- Next week: python control flow and more interesting types

- Lab this week
  - Practice Python basics with twitter data