Computing for Data Analytics
CPSC 4800

Python Basics, Jupyter Notebook

CPSC 4800, Nasim Taba

Langara.
THE COLLEGE OF HIGHER LEARNING

Learning Objective

- What is programming?
- Why Python?
- Open and explore Jupyter notebook
- Write your first Python code

What is programming?

At a basic level, a computer program is a recipe of instructions that tells your computer what to do.

The recipe is written in a code called programming language.

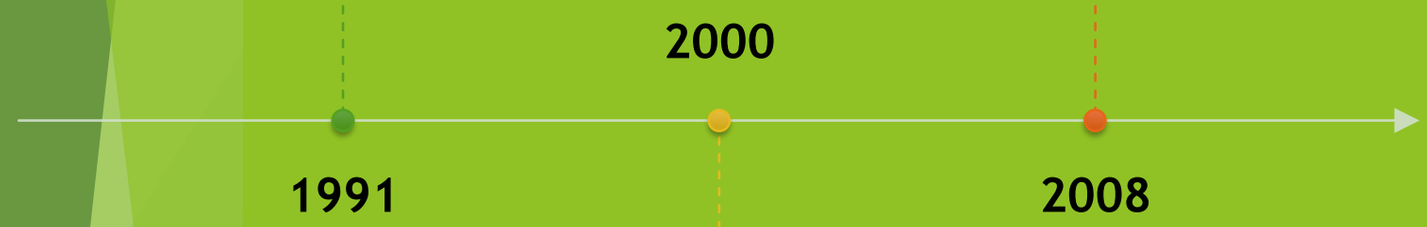
Programming languages are actually similar to humans spoken languages since they have a syntax and semantics.

when you write a program, it's important to be super clear about what you want the computer to do.

Python History

Python isn't new. Its first version was released by Guido van Rossum back in 1991.

In 2008, python 3 was released.



In 2000, python 2 was released

Python

The world’s fastest growing programming language.

It has simple, clean and beginner-friendly syntax.

Popularity of Python

Python language became so popular because:

- The language has become more powerful
- There's more tools available in Python for growing range of applications.
- You can use python to calculate statistics run your e-commerce site, process images, interact with web services
- And bunch of other tasks

Why Python is so popular?

Open Source! Free

Python packages, also for data science

You can almost build anything in python

Python - Language

- Python has well-known data analysis library called Pandas which is designed for data analysis and manipulation
- The Python language design is distinguished by its emphasis on readability, simplicity, and explicitness

source

Syntax

- When writing code, using correct syntax is super important. Even a small typo, like a missing parentheses or an extra comma, can cause a syntax error and the code won't execute at all.
- If your syntax is correct, but the script has unexpected behavior or output, this may be due to a semantic problem.

Python VS SQL

- SQL is designed to query and extract data from tables within a database. SQL is good at allowing you as a developer, to seamlessly join (or merge) several data together. Combining data from multiple tables is a key strength.
- SQL isn't designed for manipulating or transforming data into other formats.
- Higher types of data manipulation such as machine learning, statistical analysis, regression tests, and time series data is very difficult to achieve using SQL exclusively.

source

Keywords

- Reserved words that are used to construct instructions
- If
- While
- for

Set up and Installation

- Anaconda is a distribution of python
- This means it includes not only python, but many libraries that we use in the course.
- Its an “all-in-one” install that is extremely popular in data science and machine learning.

Set up and Installation

Jupyter is a development environment where we can write code, display images, and write down markdown notes.

It is the most popular IDE in data science for exploring and analyzing data

It is also a great learning tool

Jupyter notebook

- One of the major components of the Jupyter project is the *notebook*, a type of interactive document for code, text (with or without markup), data visualizations, and other output.

Python libraries

- **Numpy:** NumPy, short for Numerical Python, has long been a cornerstone of numerical computing in Python.
- **Pandas:** pandas provides high-level data structures and functions designed to make working with structured or tabular data fast, easy, and expressive
- **Matplotlib**: matplotlib is the most popular Python library for producing plots and other two-dimensional data visualizations. It was originally created by John D. Hunter and is now maintained by a large team of developers.
- **scikit-learn:** Since the project's inception in 2010, scikit-learn has become the premier general-purpose machine learning toolkit for Python programmers