Computer programming

The process of giving instructions to computer to perform an action or set of actions

What you'll learn

- Variables
- Data types
- Functions
- Conditional statements
- Loops
- Strings
- Data structures

What you'll learn

- Collaborate on data projects
- Jupyter Notebooks
- Object-oriented programming
- Variables
- Naming conventions
- Data types

Programming Languages

The words and symbols we use to write instructions for computers to follow

Library

A reusable collection of code

Jupyter Notebook

An open-source web application for creating and sharing documents containing live code, mathematical formulas, visualizations, and text

Cells

The modular code input/output fields into which Jupyter Notebook are partitioned

Object-Oriented Programming

A programming system that is based around objects, which can contain both data and code that manipulates that data

Object

An instance of a class; a fundamental building block of Python

Class

A class is an object's data type that bundles data and functionality together

Method

A function that belongs to a class and typically performs an action or operation

Dot Notation

How to access the methods and attributes that belongs to an instance of a class

Core Python Class

- Integers
- Floats
- Strings
- Booleans
- Lists
- Dictionaries

- Tuples
- Sets
- Frozensets
- Functions
- Ranges
- None
- Custom-defined

Attribute

A value associated with an object or class which is referenced by name using dot notation

Data type

An attribute that describes a piece of data based on its values, its programming language, or the operations it can perform

Variable algorithm questions

- What's the variable's name?
- What's the variable's type?
- What's the variable's starting value?

Assignment

The process of storing a value in a variable

Expression

A combination of numbers, symbols, or other variables that produce a result when evaluated

Dynamic typing

Variables can point to objects of any data type

Naming conventions

Consistent guidelines that describe the content, creation date, and version of a file in its name

Naming restrictions

Rules built into syntax of the language itself that must be followed

Keyword

A special word that is reversed for a specific purpose and that can only be used for that purpose

Variable naming conventions

- Don’t use reversed keywords like or, in, if, else, etc.
- Don’t use reversed functions such as print, str, etc.

Naming restrictions for variables

- Only include letters, numbers, and underscores
- Must start with a letter or underscore
- Case-sensitive
- Cannot include parentheses

Example variable names

any_a_variable valid

any_a_variable_2 valid

1_is_a_number invalid

Apples_&_oranges invalid

String

A sequence of characters and punctuation that contains textual information

Immutable data type

A data type in which the values can never be altered or updated

Integer

A data type used to represent whole numbers without fractions

Float

A data type that represent numbers that contains decimals

You can use the type() function to have the computer tell you the data type.

Implicit conversion

Python automatically converts one data type to another without user involvement

Explicit conversion

Users convert the data type of an object to a required data type

Review

- Introduction to Python
- Jupyter Notebook
- Object-oriented programming
- Variables
- Naming conventions
- Data types

What you'll learn

- Functions
- Write clean code
- Commenting
- Operators
- Conditional statements

Function

A body of reusable code for performing specific processes or tasks

Def

A keyword that defines a function at the start of the function block

Return

A reserved keyword in Python that makes a function produce new results, which are saved for later use

Reusability

Defining code once and using it many times without having to rewrite it

Modularity

The ability to write code in separate components that works together and that can be reused for other programs

Refactoring

The process of restructuring code while maintaining its original functionality

Self-documenting code

Code written in a way that is readable and makes its purpose clear

Algorithm

A set of instructions for solving a problem or accomplishing a task

Docstring

A string at the beginning of a function's body that summarizes the function's behavior and explains its arguments and return values

Boolean

A data type that has only two possible values, usually true or false

Comparators

Operators that compare two values and produce boolean values (True/False)

Python comparators	Symbols
Greater than	>
Greater than or equal to	>=
Less than	<
Less than or equal to	<=
Equal to	==
Not equal to	!=

Logical operators

Operators that connect multiple statements together and perform more complex complex comparisons

- and

- or
- not

The and operator needs both expressions to be true to return a True result.

If we use the or operator the expression will be True if either of the expressions is true, and False only when both expressions are false.

The not operator inverts the value of the expression that follows it.

Branching

The ability of a program to alter its execution sequence

if
A reserved keyword that sets up a condition in Python

else
A reversed keyword that executes when preceding conditions evaluate as False

Modulo

An operator that returns the remainder when one number is divided by another

elif
A reversed keyword that executes subsequent conditions when the previous conditions are not true

The elif keyword lets us handle more than two comparison cases.

Uses of branching

- Bin data based on its value
- Backup files
- Restrict login access

Review

- Functions
- Writing clean code
- Commenting
- Operators
- Conditional statements

What you've learned

- Variables
- Data types
- Functions
- Operators
- Write clean code
- Conditional statements

Loop

A block of code used to carry out iterations

Iteration

The repeated execution of a set of statements, where one iteration is the single execution of a block of code

Iterable

An object that's looped, or iterated, over

While loop

A loop that instructs your computer to continuously execute your code based on the value of a condition

The condition used by the while loop needs to evaluate to true or false.

Break

A keyword that lets you escape a loop without triggering any ELSE statement that follows it in the loop

For loop

A piece of code that iterates over a sequence of values

Range()

A Python function that returns a sequence of numbers starting from zero, increments by 1 by default, and stops before the given number

Range function

1. A range of numbers will start with the value 0 by default

2. The list of numbers generated will be ones less than the given value

Use for loops when there's a sequence of elements that you want to iterate over.

Use while loops when you want to repeat an action until a boolean condition changes.

Booleans are a data type that represents one of two possible states: True or False.

Concatenate

To link or join together

Escape character

A character that changes the typical behavior of the characters that follow it

Indexing

A way to refer to the individual items within an iterable by their relative position

index()

A string method that outputs the index number of a character in a string

String slice

A portion of a string, also known as a substring, that can contain more than one character

To check whether or not a substring is contained in a string, use keyword in.

format()

Formats and inserts specific substrings into designed places

Review

- While loops
- For loops
- Strings

Data structures

Collections of data values or objects that contain different data types

What you'll learn

- Lists
- Tuples
- Dictionaries
- Sets
- Arrays

Python libraries

- Numpy
- Pandas

List

A data structure that helps store and manipulate an ordered collection of items

Lists	Strings
1. Data structure	Data type
2. Allow duplicates elements	Allow duplicate elements
3. Allow indexing and slicing	Allow indexing and slicing
4. Sequences of elements	Sequences of characters

Sequence

A positionally ordered collection of items

Mutability

The ability to change the internal state of a data structure

Immutability

A data structure or elements’s values can never be altered or updated

Append()

Method that adds an element to the end of a list

Insert()

Function that takes an index as the first parameter and an element as the second parameter, then inserts the element into a list at the given index

Remove()

A method that removes an element from a list

POP()

A method that extracts an element from a list by removing it at a given index

Tuple

An immutable sequence that can contain elements of any data type

Tuples are expressed with parentheses or the tuple() function.

Function that transforms input into tuples

List comprehension

Formulaic creation of a new list based on the values in an existing list

Dictionary

A data structure that consists of a collection of key-value pairs

dict()

Function used to create a dictionary

Immutable keys

- Integers
- Floats
- Tuples
- Strings

Mutable data types cannot be used as keys

- Lists
- Sets
- Other dictionaries

Keys()

A dictionary method to retrieve only the dictionary’s keys

values()

A dictionary method to retrieve only the dictionary’s values

items()

A dictionary method to retrieve both the dictionary’s keys and values

Set()

A data structure in python that contains only unordered, non-interchangeable elements

Instantiated with set() function or non-empty braces

Set()

A function that takes an iterable as an argument and returns a new set object

To define an empty set, you have to use set().

intersection()

A function that finds the elements that two sets have in common

union()

A function that finds all the elements from booths sets

difference()

A function that finds the elements present in one set, but not the other

symmetric_difference()

A function that finds elements from both sets that are mutually not present in the other

Library (or package)

Broadly refers to a reusable collection of code

matplotlib

A library for creating static, animated, and interactive visualizations in Python

Seaborn

A visualization library based on matplotlib that provides a simpler interface for working with common plots and graphs

Numpy

An essential library that contains multidimensional array and matrix data structures and functions to manipulate them

pandas

A powerful library built on top of Numpy that’s used to manipulate and analyze tabular data

Module

A simple Python file containing a collection of functions and global variables

Global variables

Variables that can be accessed from anywhere in a program or script

Commonly used Python modules:

- Math
- Random

Vectorization

Enables operations to be performed on multiple components of data object at the same time

Import statement

Uses the import keyword to load an external library, package, module, or function into your computing environment

Aliasing

Let’s you assign an alternate name —- or alias —- by which you can refer to something

N-dimensional array (ndarray)

The core data object of NumPy

dtype

A NumPy attribute used to check the data type of the contents of an array

shape

A NumPy attribute used to check the shape of an array

ndim

A NumPy attribute used to check the number of dimensions of an array

reshape()

NumPy method used to change the shape of an array

Tabular data

Data that is in the form of a table, with rows and columns

Core pandas object classes

- DataFrame
- Series

DataFrame

A two-dimensional, labeled data structure with rows and columns

CSV file

Stands for “comma-separated values.” A plaintext file that uses commas to separate distinct values from one another

Series

A one-dimensional, labeled array

NaN

How null values are represented in pandas, which stands for “not a number”

iloc[]

A way to indicate in pandas that you want to select by integer-location-based position

loc[]

Used to select pandas rows and columns by name

Boolean masking

A filtering technique that overlays a Boolean grid onto a dataframe in order to select only the values in the dataframe that align with the True values of the grid

groupby()

A pandas DataFrame method that groups rows of the dataframe together based on their values at one or more columns, which allows further analysis of the groups

agg()

Short for “aggregate”. A pandas groupby method that allows you to apply multiple calculations to groups of data

Pandas functions

- concat()
- merge()

concat()

A pandas function that combines data either by adding it horizontally as new columns for existing rows, or vertically as new rows for existing columns

merge()

A pandas function that joins two dataframes together; it only combines data by extending along axis one horizontally

Keys

The shared points of reference between different dataframes–what to match on

Experiential learning

The idea of understanding through doing

Data tidying

Structuring datasets to facilitate analysis

Tidy dataset

- Easy to manipulate, model, and visualize
- Each variable is a column
- Each observation is a row
- Each type of observation unit is a table

What we covered

- Importance of communication
- Data professional tools
- Python for dataset management

Keep in mind

- Who your audience is and what their goals are
- What they already know
- What they need to know