Glossary

Advanced Data Analytics

Terms and definitions from Course 2

A

agg(): A pandas groupby method that allows the user to apply multiple calculations to groups of data

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

Aliasing: A process that allows the user to assign an alternate name—or alias—to something

append(): A method that adds an element to the end of a list

Argument: Information given to a function in its parentheses

Assignment: The process of storing a value in a variable

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

B

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

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

Branching: The ability of a program to alter its execution sequence

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

C

Cells: The modular code input and output fields into which Jupyter Notebooks are partitioned

Class: An object's data type that bundles data and functionality together

Comparator: An operator that compares two values and produces Boolean values (True/False)

Computer programming: The process of giving instructions to a computer to perform an action or set of actions

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

Concatenate: To link or join together

CSV file: A plaintext file that uses commas to separate distinct values from one another; Stands for "comma-separated values"

D

Data structure: A collection of data values or objects that contain different data types

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

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

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

dict(): A function used to create a dictionary

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

difference(): A function that finds the elements present in one set but not the other

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

Dot notation: How to access the methods and attributes that belong to an instance of a class

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

Dynamic typing: Variables that can point to objects of any data type

E

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

else: A reserved keyword that executes when preceding conditions evaluate as False

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

Explicit conversion: The process of converting a data type of an object to a required data type

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

F

Float: A data type that represents numbers that contain decimals

For loop: A piece of code that iterates over a sequence of values

format(): A string method that formats and inserts specific substrings into designated places within a larger string

Function: A body of reusable code for performing specific processes or tasks

G

Global variable: A variable that can be accessed from anywhere in a program or script

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

I

if: A reserved keyword that sets up a condition in Python

iloc[]: A type of notation in pandas that indicates when the user wants to select by integer-location-based position

Immutability: The concept that a data structure or element's values can never be altered or updated

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

Implicit conversion: The process Python uses to automatically convert one data type to another without user involvement

Import statement: A statement that uses the import keyword to load an external library, package, module, or function into the computing environment

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

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

Inner join: A way of combining data such that only the keys that are in both dataframes get included in the merge

insert(): A 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

Integer: A data type used to represent whole numbers without fractions

intersection(): A function that finds the elements that two sets have in common

items(): A dictionary method to retrieve both the dictionary's keys and values

Iterable: An object that's looped, or iterated, over

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

J

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

K

Keys: The shared points of reference between different dataframes

keys(): A dictionary method to retrieve only the dictionary's keys

Keyword: A special word in a programming language that is reserved for a specific purpose and that can only be used for that purpose

L

Left join: A way of combining data such that all of the keys in the left dataframe are included, even if they aren't in the right dataframe

Library: A reusable collection of code; also referred to as a "package"

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

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

loc[]: Notation that is used to select pandas rows and columns by name

Logical operator: An operator that connects multiple statements together and performs complex comparisons

Loop: A block of code used to carry out iterations

M

Markdown: A markup language that lets the user write formatted text in a coding environment or plain-text editor

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

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

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

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

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

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

Mutability: The ability to change the internal state of a data structure

N

N-dimensional array: The core data object of NumPy; also referred to as “ndarray”

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

Naming restrictions: Rules built into the syntax of a programming language

NaN: How null values are represented in pandas; stands for “not a number”

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

Nested loop: A loop inside of another loop

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

O

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

Object-oriented programming: A programming system that is based around objects which can contain both data and code that manipulates that data

Outer join: A way of combining data such that all of the keys from both dataframes get included in the merge

P

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

pop(): A method that extracts an element from a list by removing it at a given index

Programming languages: The words and symbols used to write instructions for computers to follow

R

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

Refactoring: The process of restructuring code while maintaining its original functionality

remove(): A method that removes an element from a list

reshape(): A NumPy method used to change the shape of an array

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

Reusability: The capability to define code once and using it many times without having to rewrite it

Right join: A way of combining data such that all the keys in the right dataframe are included—even if they aren't in the left dataframe

S

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

Self-documenting code: Code written in a way that is readable and makes its purpose clear

Sequence: A positionally ordered collection of items

Series: A one-dimensional, labeled array where the data type must be the same for all the data in a given series

Set: A data structure in Python that contains only unordered, non-interchangeable elements

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

shape: A NumPy attribute used to check the shape of an array

String: A sequence of characters and punctuation that contains textual information

String slice: A portion of a string that can contain more than one character; also referred to as a substring

symmetric_difference(): A function that finds elements from both sets that are mutually not present in the other

Syntax: The structure of code words, symbols, placement, and punctuation

T

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

Tuple: An immutable sequence that can contain elements of any data type

tuple(): A function that transforms input into tuples

type(): A function used to identify the type of data in a list

Typecasting: Converting data from one type to another (see **explicit conversion**)

U

union(): A function that finds all the elements from both sets

V

values(): A dictionary method to retrieve only the dictionary's values

Variable: A named container which stores values in a reserved location in the computer's memory

Vectorization: A process that enables operations to be performed on multiple components of a data object at the same time

W

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