Glossary terms from module 4
Terms and definitions from Course 2, Module 4

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Keys: The shared points of reference between different dataframes

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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