**Glossary**

**variable –** a container for storing data values; a variable is created in Python when you first assign a value to it and can be referenced later to retrieve the stored data:

**>>> x = 'hello world'**

**>>> print(x)**

**hello world**

**data type –** theformat that data stored in a variable takes; some examples are int, float, str, and list. To print the type of data in variable x you can use the **type** function:

**>>> print(type(x))**

**<class 'str'>**

**int –** integer data type in Python; whole number; examples are 12, -3, 175.

**float** – floating decimal data type in Python; examples are 6.667, 12.0, -11.9.

**str** – string data type in Python; examples are “hello”, “Nashville”, “175”.

**boolean** – a binary data type that is either **True** or **False**; typically seen as the output to a comparison.

**dataframe** – tabular format commonly used for storing and working with data in Python and R. A dataframe has rows (observations) and columns (variables).

**series** – a pandas data type; a one dimensional array of data that contains values for a single variable. All the values are of the same type. A column or row in a dataframe is a series.

**csv** – short for **comma separated values**, this is a common text format for sharing data.

**docstring** – text specified in source code that documents what a function or module does.

**function** – reusable code that takes zero, one or more inputs, does something with them, and returns an output. In python a function starts with the **def** keyword. The function below takes a name as input and returns a new string by concatenating the string “Hello “ with the name.

**>>>def say\_hello(name):**

**. . . return "Hello " + name**

**>>>say\_hello('Mary')**

**'Hello Mary'**

**method** – a function that can be applied to a specific object

**attribute** – a characteristic of a specific object: df.shape

**sql** – structured query language; used for creating, retrieving, updating, or deleting data from a database (look up CRUD acronym). Data analysts will only retrieve data.

**list** – a python data type; a 1 dimensional array of values. Individual values can be of any data type. [1, ‘2’, 3.0]. Also a function that can turn something into a list: list()

**dictionary** – a python data type that contains key, value pairs. Typically acts as a lookup table, where the value is returned when searching for a key. Keys are all unique and are typically strings, ints, or floats. Values can be any data type, including other dictionaries.

**tuple** – a python data type similar to a list; a 1 dimensional array of values. Can contain different data types. Also a function that can turn something into a tuple: tuple(). Differs from a list in that tuples cannot be modified once they are created (they are immutable)

**function chaining** – applying multiple functions in a row without saving intermediate steps. The functions are read left to right and the output of one function is directly fed into the next one. df.value\_counts().reset\_index().rename()

**trace back** – occurs when a piece of code produces an error. Shows the line of code that produced the error, the steps the code went through (if calling other functions), and the kind of error that was encountered (such as valueError).

Python Keywords (these should not be used as variable names):

|  |  |  |
| --- | --- | --- |
| FALSE | if | def |
| TRUE | elif | return |
| import | else | for |
| from | not | lambda |
| as | or |  |
| in | None |  |
| and | is |  |

Python Built-In Function (these should not be used as variable names):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| abs() | help() | max() | reversed() | tuple() |
| bool() | int() | min() | round() | type() |
| dict() | len() | pow() | set() | zip() |
| enumerate() | list() | print() | str() |  |
| float() | map() | range() | sum() |  |