Course 4

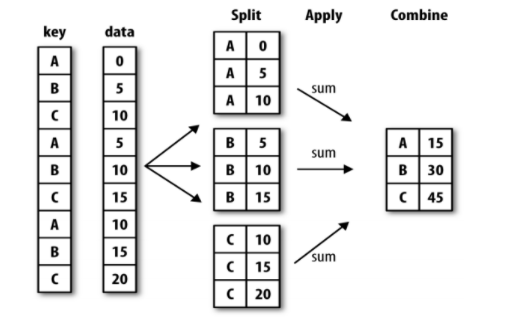
**Q.1 How does the groupby() method works in Pandas?**

Answer: In the first stage of the process, data contained in a pandas object, whether a Series, DataFrame, or otherwise, is split into groups based on one or more keys that we provide.

The splitting is performed on a particular axis of an object. For example, a DataFrame can be grouped on its rows (axis=0) or its columns (axis=1).

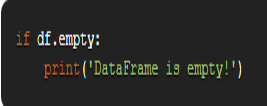
Once this is done, a function is applied to each group, producing a new value. Finally, the results of all those function applications are combined into a result object. The form of the resulting object will usually depend on what's being done to the data.

In the figure below, this process is illustrated for a simple group aggregation.



**Q.2 How to check whether a Pandas DataFrame is empty?**

Answer: You can use the attribute df.empty to check whether it's empty or not:



**Q.3 What is Negative Index in Python?**

Answer: Negative numbers mean that you count from the right instead of the left. So, list[-1] refers to the last element, list[-2] is the second-last, and so on.

**Q.4 How to create new columns derived from existing columns in Pandas?**

Answer: We create a new column by assigning the output to the DataFrame with a new column name in between the [].

Let's say we want to create a new column 'C' whose values are the multiplication of column 'B' with column 'A'. The operation will be easy to implement and will be element-wise, so there's no need to loop over rows.

**Q.5 How are iloc() and loc() different?**

Answer: DataFrame.iloc is a method used to retrieve data from a Data frame, and it is an integer position-based locator (from 0 to length-1 of the axis), but may also be used with a boolean array. It takes input as integer, arrays of integers, a slice object, boolean array and functions. DataFrame.loc gets rows (and/or columns) with particular labels. It takes input as a single label, list of arrays and slice objects with labels.

**Q.6 When to use a tuple vs list vs dictionary in Python?**

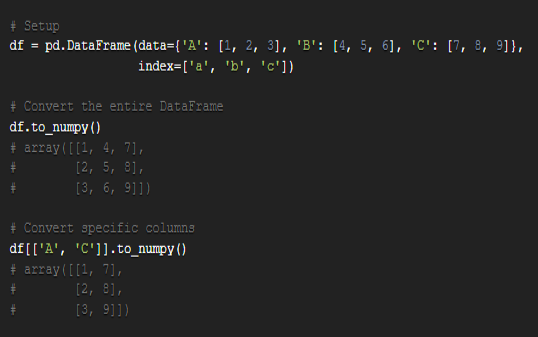
Answer: Use a tuple to store a sequence of items that will not change.

Use a list to store a sequence of items that may change.

Use a dictionary when you want to associate pairs of two items.

**Q.7 How would you convert a Pandas DataFrame into a NumPy array?**

Answer: Use to\_numpy(), which is defined on Index, Series, and DataFrame objects.



**Q.8 What is the difference between ndarray and array in NumPy?**

Answer: numpy.array is just a convenience function to create an ndarray; it is not a class itself.

You can also create an array using numpy.ndarray, but it is not the recommended way.

If you want to create an array from ndarray class you can do it in 2 ways as quoted:

Using array(), zeros() or empty() methods: Arrays should be constructed using array, zeros or empty (refer to the See Also section below). The parameters are given here refer to a low-level method (ndarray(…)) for instantiating an array.

From ndarray class directly: There are two modes of creating an array using \_\_new\_\_: If buffer is None, then only shape, dtype, and order are used. If buffer is an object exposing the buffer interface, then all keywords are interpreted.

**Q.9 What are the benefits of using Python?**

Answer:

* Easy to use– Python is a high-level programming language that is easy to use, read, write and learn.
* Interpreted language– Since python is interpreted language, it executes the code line by line and stops if an error occurs in any line.
* Dynamically typed– the developer does not assign data types to variables at the time of coding. It automatically gets assigned during execution.
* Free and open-source– Python is free to use and distribute. It is open source.
* Extensive support for libraries– Python has vast libraries that contain almost any function needed. It also further provides the facility to import other packages using Python Package Manager(pip).
* Portable– Python programs can run on any platform without requiring any change.
* The data structures used in python are user-friendly.
* It provides more functionality with less coding.