Module Summary

Introduction to the GroupBy function

Calculating aggregate values for groups

Filter and transform using GroupBy

Grouping multi-index data

The GroupBy Function

GroupBy Functiom

GroupBy function allows you to split a dataset into different groups or categories based on some dataset feature or column
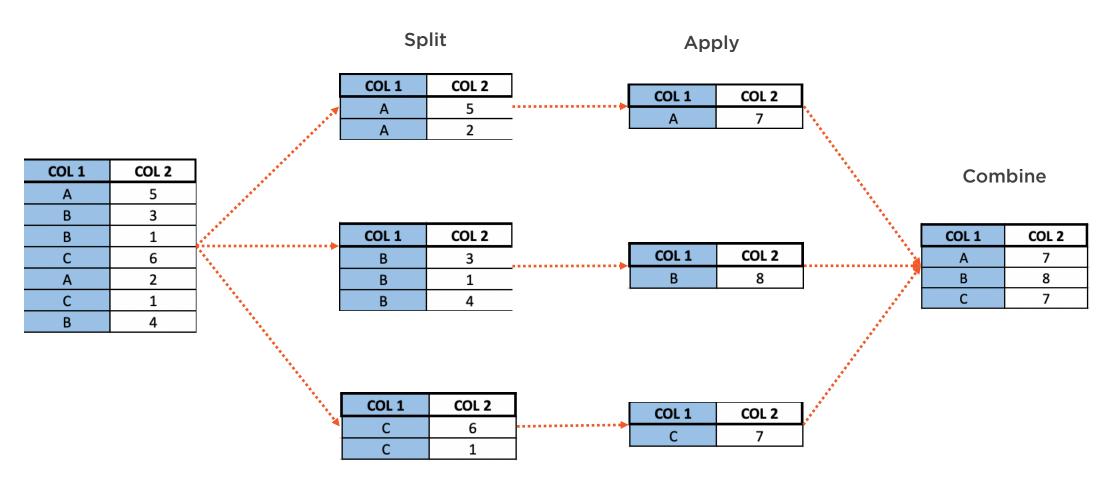
GroupBy Operating Principle

Split

Apply

Combine

Split-Apply-Combine

Split

COL 1	COL 2
A	5
A	2

COL 1	COL 2
B	3
B	1
B	4

COL 1	COL 2
C	6
C	1

Apply

COL 1	COL 2
A	7

COL 1	COL 2
B	8

COL 1	COL 2
C	7

Combine

COL 1	COL 2
A	7
B	8
C	7

COL 1	COL 2
A	5
B	3
B	1
C	6
A	2
C	1
B	4

Sorting Behavior

Default sorting behavior

	age	dim	dim.1
animal			
alligator	16	27	11
cat	42	52	26
hamster	23	45	30
snake	21	22	14

df.groupby(sort=False)

	age	dim	dim.1
animal			
hamster	23	45	30
alligator	16	27	11
cat	42	52	26
snake	21	22	14
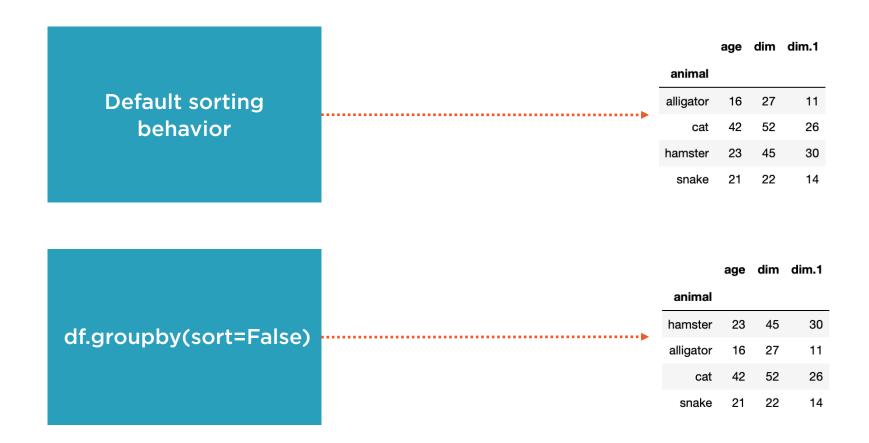
Demo

Calculating aggregate values for groups

- Why and how to use the groupby function

Filter and Transform with GroupBy

Filtration

Process of discarding groups based on some criteria

Argument to filter must be a function or lambda that will take a group and return True or False

This will determine whether that group should be in the output

Transformation

Returns a transformed version of the full data on a group level

New output has the same shape as the input data

Demo

Grouping multi-index data