

Python for Data Science

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Agenda

1. Pop Quiz
2. Common Python Libraries for Data Science
3. NumPy and Pandas
4. Common NumPy functions
5. Common Pandas functions
6. Merge vs Join in Pandas
7. Example of Join

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pop Quiz

1. What are the data types in Python?
2. What are some of the common Python libraries for Data Science?
3. Can you list some of the common functions in Pandas?
4. What are the applications of the functions like group by, merge, join etc?

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Common Python Libraries for Data Science

Library	Use
NumPy	Handling multi-dimensional arrays
Scipy	Scientific computation package
Matplotlib, Seaborn	Data visualisation
Pandas	Handling tabular data
Scikit-learn	Machine learning

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

- Stands for Numerical Python
- It is one of the fundamental packages for mathematical, logical, and statistical operations with Python
- It contains
 - Powerful N-dimensional array object, called ndarray
 - Large set of functions for creating, manipulating, and transforming ndarrays
- ndarrays can only contain data of a single datatype
- Useful in linear algebra, vector calculus, random number capabilities, etc

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Pandas

- Pandas is one of the fundamental packages for analysis and manipulation of tabular data
- Offers two major data structures - series & dataframe
- We can think of a pandas dataframe like an excel spreadsheet that is storing some data in rows and columns.
- A pandas dataframe is made up of several pandas series
 - Each column of a dataframe is a series.
- Pandas dataframes can contain data of multiple datatypes

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Common NumPy Functions

Function	Description
<code>np.array()</code>	To create an array
<code>np.arange()</code>	Return evenly spaced values within a given interval
<code>np.linspace()</code>	Return evenly spaced numbers over a specified interval
<code>np.zeros()</code>	To create an array of zeros
<code>np.ones()</code>	To create an array of ones
<code>np.transpose()</code>	Permute array dimensions

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Common NumPy Functions

Function	Description
<code>np.random.rand()</code>	To create an array of specified shape filled with random values
<code>np.random.randint()</code>	Return random integers from low (inclusive) to high (exclusive)
<code>np.random.randn()</code>	Return a sample (or samples) from the “standard normal” distribution.
<code>np.concatenate()</code>	Concatenate two arrays
<code>np.save()</code>	Save an array to a binary file in .npy format.
<code>np.savez()</code>	Save several arrays into a single file in uncompressed .npz format.

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Common Pandas Functions

Function	Description
<code>pd.read_csv()</code>	Read a comma-separated values (csv) file into DataFrame
<code>df.loc[]</code>	Access a group of rows and columns by label(s)
<code>df.iloc[]</code>	Purely integer-location based indexing for selection by position
<code>df.drop()</code>	Drop specified labels from rows or columns
<code>pd.concat()</code>	To concatenate two pandas objects
<code>pd.merge()</code>	To merge the pandas dataframes
<code>df.groupby()</code>	To split, apply or combine the data structures

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Common Pandas Functions

Function	Description
df.value_counts()	To get count of some attributes
df.unique()	To get unique values
df.dtypes	To get the data types
df.shape	To get the shape (number of rows and columns)
df.head()	To get the top rows
df.tail()	To get the last rows
df.describe()	To get the quick statistic summary

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Merge vs Join

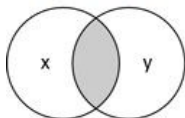
- **Join:** The **join** method works best when we are joining dataframes on their indexes (though you can specify another column to join on for the left dataframe).
- **Merge:** The **merge** method is more versatile and allows us to specify columns besides the index to join on for both dataframes.

Natural join - Intersection

To keep only rows that match from the data frames

how='inner'.

how='inner'



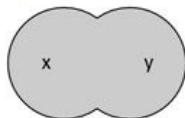
natural join

Full outer join - Union

To keep all rows from both data frames,

how='outer'.

how='outer'

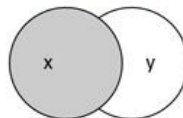


full outer join

Left outer join

To include all the rows of your data frame x and only those from y that match
how = 'left'.

how='left'

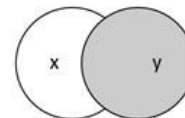


left outer join

Right outer join

To include all the rows of your data frame y and only those from x that match
how='right'.

how='right'



right outer join

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Example of Join

Left Join

Index	Customer_id	Product
0	1	Oven
1	2	Oven
2	3	Oven
3	4	Television
4	5	Television
5	6	Television

Left Table

Index	Customer_id	State
0	2	California
1	4	California
2	6	Texas

Right Table

Index	Customer_id	Product	State
0	1	Oven	nan
1	2	Oven	California
2	3	Oven	nan
3	4	Television	California
4	5	Television	nan
5	6	Television	Texas

After Left Join datasets on Customer_id

Syntax: `merged = pd.merge(left, right, on = 'Customer_id', how = 'left')`

Right Join

Index	Customer_id	Product
0	1	Oven
1	2	Oven
2	3	Oven
3	4	Television
4	5	Television
5	6	Television

Left Table

Index	Customer_id	State
0	2	California
1	4	California
2	6	Texas

Right Table

Index	Customer_id	State	State
0	2	Oven	California
1	4	Television	California
2	6	Television	Texas

After right Join datasets on Customer_id

Syntax: `merged = pd.merge(left, right, on = 'Customer_id', how = 'right')`

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Example of Join

Inner Join

Index	Id	Name	Age
0	1	Alex	25
1	2	Amy	23
2	3	Allen	22
3	4	Alice	21
4	5	Ayoung	24

Left Table

Index	Id	Subject
0	1	sub2
1	2	sub4
2	3	sub3
3	4	sub6
4	5	sub5

Right Table

Index	Id	Name	Age	Subject
0	1	Alex	25	sub2
1	2	Amy	23	sub4
2	3	Allen	22	sub3
3	4	Alice	21	sub6
4	5	Ayoung	24	sub5

After Merging datasets on Id

Syntax: `merged = pd.merge(left, right, on = 'id')`

Outer/Full Join

Index	Customer_id	Product
0	1	Oven
1	2	Oven
2	3	Oven
3	4	Television
4	5	Television
5	6	Television

Left Table

Index	Customer_id	State
0	2	California
1	4	California
2	6	Texas

Right Table

Index	Customer_id	Product	State
0	1	Oven	nan
1	2	Oven	California
2	3	Oven	nan
3	4	Television	California
4	5	Television	nan
5	6	Television	Texas

After Outer Join datasets on Customer_id

Syntax: `merged = pd.merge(left, right, on = 'Customer_id', how = 'outer')`

This file is meant for personal use by brentjones@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Proprietary content. © Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.



Happy Learning !

