SciPy + Pandas

Data Science

June 8th, 2021

This Lecture

Scipy: Don't reinvent the wheel. Pandas: making data manageable.

Before we start...

Before we start...

1. Quizzes

SciPy

SciPy is one of the best things to come out of the Python ecosystem

SciPy

SciPy is one of the best things to come out of the Python ecosystem

1. Lots of mathematical functions, implemented over man standard python/numpy types.

SciPy

SciPy is one of the best things to come out of the Python ecosystem

1. Lots of mathematical functions, implemented over man standard python/numpy types.
2. Numerical integration: (scipy.integrate)

SciPy

SciPy is one of the best things to come out of the Python ecosystem

1. Lots of mathematical functions, implemented over man standard python/numpy types.

2. Numerical integration: (scipy.integrate)

3. Solving optimization problems (scipy.optimize)

SciPy

SciPy is one of the best things to come out of the Python ecosystem

1. Lots of mathematical functions, implemented over man standard python/numpy types.
2. Numerical integration: (scipy.integrate)
3. Solving optimization problems (scipy.optimize)
4. Lots of linear algebra

SciPy

SciPy is one of the best things to come out of the Python ecosystem

1. Lots of mathematical functions, implemented over man standard python/numpy types.
2. Numerical integration: (scipy.integrate)
3. Solving optimization problems (scipy.optimize)
4. Lots of linear algebra
5. and much more!

What you need to know about SciPy

If other library authors have done their job:

What you need to know about SciPy

If other library authors have done their job: not much

What you need to know about SciPy

It's good to know that SciPy functions are there, for when you need them. Things that might come up for you during this class:

What you need to know about SciPy

It's good to know that SciPy functions are there, for when you need them. Things that might come up for you during this class:

1. Need a statistical function: scipy.stats

What you need to know about SciPy

It's good to know that SciPy functions are there, for when you need them. Things that might come up for you during this class:

1. Need a statistical function: scipy.stats
2. Need to process an image: scipy.ndimage

What you need to know about SciPy

It's good to know that SciPy functions are there, for when you need them. Things that might come up for you during this class:

1. Need a statistical function: scipy.stats
2. Need to process an image: scipy.ndimage
3. if these things come up: scipy.io often has the functionality for getting various libraries to 'talk' to each other.

Pandas

Tables are important!

Pandas

Tables are important!

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

Pandas

Pandas is for working with and processing tabular data.

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Selection (rows)

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Selection (rows)
2. Slicing (columns)

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Reduce

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Reduce
2. Aggregate

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Map

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. Map
2. may increase or decrease number of columns!

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Sunday	0C	8C	Sunny
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

Pandas

Pandas is for working with and processing tabular data.

Day	Min Temp	Max Temp	Sky
Sunday	0C	8C	Sunny
Monday	0C	5C	Cloudy
Tuesday	1C	3C	Overcast
Wednesday	3C	8C	Sunny

1. GroupBy

High-level view

Many of the same operations you'd see in SQL (but be warned!).

High-level view

Many of the same operations you'd see in SQL (but be warned!).

1. Join/Merge

High-level view

Many of the same operations you'd see in SQL (but be warned!).

1. Join/Merge
2. Intersection/Union/Concat

High-level view

Many of the same operations you'd see in SQL (but be warned!).

1. Join/Merge
2. Intersection/Union/Concat
3. 'index' in pandas is like 'key' in SQL

To the Notebook

What the title says

Thanks for your time!