Journey to Pandas

2018/01/03

Left for blank

WHAT’S NEW

v0.20.1 (May 5, 2017)

New features

agg API for DataFrame/Series

Now the Series & DataFrame have been enhanced to support the aggregation API, similar to groupby, window operations, and resampling.

dtype keyword for data IO

Now dtype keyword argument is accepted for specifying the types of specific columns.

.to\_datetime() has gained an origin parameter

The new parameter is origin, which is to define a reference date from where to compute the resulting timestamps when parsing numerical values with a specific unit specified.

Groupby Enhancements

Strings passed to DataFrame.groupby() as the by parameter may now reference either column names or index level names.

Better support for compressed URLs in read\_csv

The compression code was refactored so that it now supports additional compression methods: xz, bz2, and zip.

Pickle file I/O now supports compression

read\_pickle(), DataFrame.to\_pickle() and Series.to\_pickle() can now read from and write to compressed pickle files. Compression methods can be an explicit parameter or be inferred from the file extension.

UInt64 Support Improved

A new numerical index, UInt64Index, has been created.

GroupBy on Categoricals

Bug fixed.

Table Schema Output

New argument orient, table, for DataFrame.to\_json() will generate a Table Schema compatible string representation of the data.

SciPy sparse matrix from/to SparseDataFrame

Now Pandas support creating sparse dataframes from scipy.sparse.spmatrix instances.

Excel output for styled DataFrames

Experimental support has been added to export DataFrame.style formats to Excel using the openpyxl engine.

pip install openpyxl

IntervalIndex

Now Pandas has an IntervalIndex with its own dtype.

Besides, an IntervalIndex can also be used in Series and DataFrame as the index.

Other Enhancements

Backwards incompatible API changes

...

Reorganization of the library: Privacy Changes

...

Deprecations

Deprecate .ix

Now the recommend indexing methods are

* .loc if you want to *label* index
* .iloc if you want to *positionally* index.

Deprecate Panel

The recommended way to represent 3-D data are with a MultiIndex on a DataFrame via the to\_frame() or with the **xarray package**.

Deprecate groupby.agg() with a dictionary when renaming

Now you can’t pass a dict to agg() function instead you need to use rename() after that.

Deprecate .plotting

All the public plotting functions from previous pandas.tools.plotting are now available from pandas.plotting.

Also, the top-level pandas.scatter\_matrix and pandas.plot\_params are deprecated and are inside the module pandas.plotting.

Other Deprecations

...

Removal of prior version deprecations/changes

...

Performance Improvements

...

Bug Fixes

...

v0.19.2 (December 24, 2016)

...

INSTALLATION

Python version support

Officially Python 2.7, 3.4, 3.5, and 3.6

Installing pandas

Dependencies

* Setuptools: Needed for installation
* NumPy: 1.7.1 or higher
* python-dateutil: 1.5 or higher
* pytz: Needed for time zone support

CONTRIBUTING TO PANDAS

PACKAGE OVERVIEW

Data structures at a glance

1. One dimension: Series, homogeneously-typed array
2. Two dimensions: DataFrame, size-mutable tabular structure with potentially heterogeneously-typed columns
3. Three dimensions: Panel, size-mutable array

Why with more than 1 data structure?

DataFrame is a container for Series, and Panel is a container for DataFrame objects. And under these structures we can remove objects from their containers in a **dictionary-like** fashion.

And for the semantic reasons, making it much easier to write the code.

Mutability and copying of data

All the data structures of pandas are **value-mutable**, that is the values they contain can be altered, yet not always size-mutable.

Getting Support

...

10 MINUTES TO PANDAS

Object Creation

* Series: pass a list of values with default integer index
* DataFrame: pass a numpy array, with a labeled indexes and columns; or a dict of objects that can be converted to series-like

Viewing Data