

KÖNIGSWEG

Piter  Py #4

Efficient Data Mangling with Pandas Indexes



Alexander C. S. Hendorf

- Senior Consultant Information Technology
- Program Chair EuroPython, PyConDE & PyData Karlsruhe,
PSF Managing Member
- Program Committee EuroSciPy, Percona Live
- MongoDB Masters / MongoDB Certified DBA
- Speaker Europe & USA MongoDB World New York / San José,
PyCon Italy, CEBIT, BI Forum, IT-Tage FFM, PyData, PyParis,
PyCon UK.



ah@koenigsweg.com
@hendorf

Königsweg Strategy Consulting

INNOVATION STRATEGY

We expose your innovation potential through in-depth assessments and convert it through on-track strategies and full cycle implementation.

DIGITAL TRANSFORMATION

We guide SMEs and institutions along the transformational agendas of digitalization and Industry 4.0.

BUSINESS INTELLIGENCE

We introduce state-of-the-art data science to your business, allowing for more accurate and reliable forecasts, and decision-making

CLUSTER OPERATION

We connect high-tech start-ups with key actors, influencers and industry partners in their respective field.

START-UP FINANCING

We support start-ups and innovative high tech companies in obtaining financing through equity, debt and quasi-equity financing.

Effective Data Analysis with Pandas Indexes

Pandas

- Accessing data via the Index
- **Index Types**
- **MultilIndex**
- **DateTimeIndex and Resampling**
- **Categorical**

Demo

The image shows a Jupyter Notebook interface with two overlapping windows. The background window displays a page titled "data2day 2018" with the subtitle "Interaktive Datenanalyse n". It includes a section "3 minutes Markup Crash-Course" and a list of heading levels (h1, h2, h3, h4, h5) with corresponding examples of how to create them in Markdown. The foreground window shows a Jupyter Notebook with the title "data2day2018" and subtitle "Interaktive Datenanalyse mit Jupyter und Pandas". It contains six input cells with Python code for data manipulation and visualization, followed by a hexbin plot.

data2day 2018
Interaktive Datenanalyse n

3 minutes Markup Crash-Course

Dies ist eine kleine Demo über die Möglichk unterschiedlicher Gewichtung können mit #

h1

h2

h3

h4

h5

Mit > können wir ein Zitat erzeugen:

```
> Hello World!
```

Mit

```
ESC+m
```

kann man eine Codezelle in ine Markdown

```
SHIFT+RETURN
```

führen wir die Zelle aus, es wird automatisch

Auch stehen Stilmittel wie *kursiv* oder **fett** zu

data2day2018
Interaktive Datenanalyse mit Jupyter und Pandas

```
In [1]: 1 import pandas as pd
        2 import numpy as np
        3 %matplotlib inline

In [2]: 1 %config InlineBackend.figure_format = 'retina'

In [3]: 1 df = pd.DataFrame(np.random.randn(1000, 2), columns=['a', 'b'])

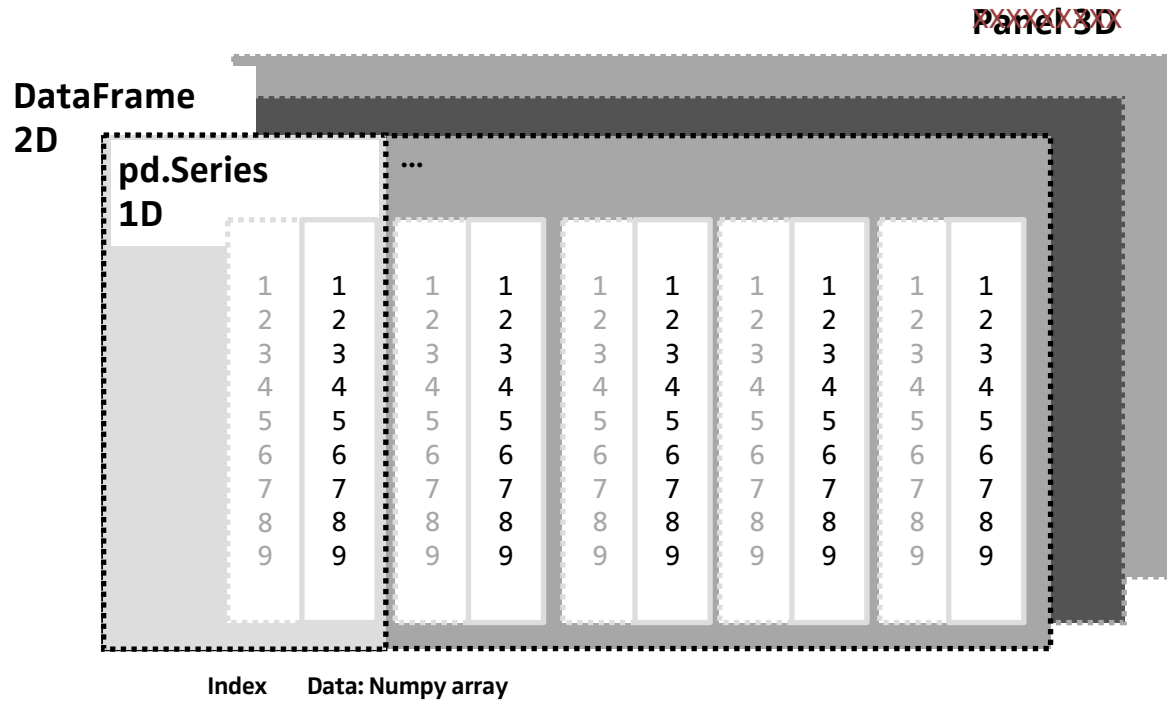
In [4]: 1 df['b'] = df['b'] + np.arange(1000)

In [5]: 1 df['z'] = np.random.uniform(0, 3, 1000)

In [6]: 1 df.plot.hexbin(x='a', y='b', C='z', reduce_C_function=np.max,
        2               gridsize=25);
```

The plot is a hexbin plot with 'a' on the x-axis and 'b' on the y-axis, both ranging from 0 to 1000. The color of each hexagon represents the value of 'z', with a color scale from 0.5 (light green) to 2.5 (dark green). The points are distributed across the plot area, with a higher density in the center.

Structure



The Index

- Label of a DataSeries
- Immutable but replaceable
- One or more Dimensions
- Labels are not necessarily *unique*

Index Types

- Index
- MultiIndex
- DateTimeIndex
- TimeDelta
- IntervalIndex
- CategoricalIndex

Thank you!

Q & A



koenigsweg.com

ah@koenigsweg.com

 @hendorf

