

6.1 Reading and Writing Data in Text Format

pandas features a number of functions for reading tabular data as a DataFrame object. Table 6-1 summarizes some of them, though read_csv and read_table are likely the ones you'll use the most.

Table 6-1. Parsing functions in pandas

Function	Description
read_csv	Load delimited data from a file, URL, or file-like object; use comma as default delimiter
read_table	Load delimited data from a file, URL, or file-like object; use tab ('\t ') as default delimiter
read_fwf	Read data in fixed-width column format (i.e., no delimiters)
read_clipboard	Version of read_table that reads data from the clipboard; useful for converting tables from web pages
read_excel	Read tabular data from an Excel XLS or XLSX file
read_hdf	Read HDF5 files written by pandas
read_html	Read all tables found in the given HTML document
read_json	Read data from a JSON (JavaScript Object Notation) string representation
read_msgpack	Read pandas data encoded using the MessagePack binary format
read_pickle	Read an arbitrary object stored in Python pickle format

Function	Description
read_sas	Read a SAS dataset stored in one of the SAS system's custom storage formats
read_sql	Read the results of a SQL query (using SQLAlchemy) as a pandas DataFrame
read_stata	Read a dataset from Stata file format
read_feather	Read the Feather binary file format

The optional arguments for these functions may fall into a few categories:

Indexing

Can treat one or more columns as the returned DataFrame, and whether to get column names from the file, the user, or not at all.

Type inference and data conversion

This includes the user-defined value conversions and custom list of missing value markers.

Datetime parsing

Includes combining capability, including combining date and time information spread over multiple columns into a single column in the result.

Iterating

Support for iterating over chunks of very large files.

Unclean data issues

Skipping rows or a footer, comments, or other minor things like numeric data with thousands separated by commas.

Because of how messy data in the real world can be, some of the data loading functions (especially `read_csv`) have grown very complex in their options over time. It's normal to feel overwhelmed by the number of different parameters (`read_csv` has over 50 as of this writing). The online pandas documentation has many examples about how each of them works, so if you're struggling to read a particular file, there might be a similar enough example to help you find the right parameters.

In [21]: `!type ex1.csv`

```
a,b,c,d,message
1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo
```

In [10]: `# from os import system
system("cat pydata-book-master/ch06/ex1.csv")`

Since this is comma-delimited, we can use `read_csv` to read it into a DataFrame:

In [24]: `import pandas as pd
df = pd.read_csv('ex1.csv')
df`

Out[24]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

We could also have used `read_table` and specified the delimiter

In [26]: `pd.read_table('ex1.csv', sep=',')`

Out[26]:

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

In [2]: `directory="C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\`

A file will not always have a header row. Consider this file:

```
In [44]: directory+"\\ch06\\ex2.csv"
```

```
Out[44]: 'C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Desktop
\\Practice_Code\\Python_Practice\\Python_For_Data_Analysis\\pydata-book-master\\ch
06\\ex2.csv'
```

```
In [46]: ! type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\De
1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo
```

```
In [49]: pd.read_csv(directory+'\\ch06\\ex2.csv', header=None)
```

```
Out[49]:
```

	0	1	2	3	4
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

```
In [50]: pd.read_csv(directory+'\\ch06\\ex2.csv', names=['a', 'b', 'c', 'd', 'message'])
```

```
Out[50]:
```

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

Suppose you wanted the message column to be the index of the returned DataFrame. You can either indicate you want the column at index 4 or named 'message' using the `index_col` argument:

```
In [51]: names = ['a', 'b', 'c', 'd', 'message']
```

```
In [52]: pd.read_csv(directory+'\\ch06\\ex2.csv', names=names, index_col='message')
```

```
Out[52]:
```

	a	b	c	d
hello	1	2	3	4
world	5	6	7	8
foo	9	10	11	12

In the event that you want to form a hierarchical index from multiple columns, pass a list of column numbers or names:

```
In [53]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
```

```
key1,key2,value1,value2
one,a,1,2
one,b,3,4
one,c,5,6
one,d,7,8
two,a,9,10
two,b,11,12
two,c,13,14
two,d,15,16
```

```
In [54]: parsed = pd.read_csv(directory+'\\ch06\\csv_mindex.csv',index_col=['key1', 'key2'])
```

```
In [55]: parsed
```

```
Out[55]:
```

		value1	value2
one	key1	key2	
	a	1	2
	b	3	4
	c	5	6
	d	7	8
two	a	9	10
	b	11	12
	c	13	14
	d	15	16

In some cases, a table might not have a fixed delimiter, using whitespace or some other pattern to separate fields. Consider a text file that looks like this:

```
In [56]: list(open(directory+'\\ch06\\ex3.txt'))
```

```
Out[56]: ['          A          B          C\n',
'aaa -0.264438 -1.026059 -0.619500\n',
'bbb  0.927272  0.302904 -0.032399\n',
'ccc -0.264273 -0.386314 -0.217601\n',
'ddd -0.871858 -0.348382  1.100491\n']
```

While you could do some munging by hand, the fields here are separated by a variable amount of whitespace. In these cases, you can pass a regular expression as a delimiter for `read_table`. This can be expressed by the regular expression `\s+`, so we have then:

```
In [57]: result = pd.read_table(directory+'\\ch06\\ex3.txt', sep='\\s+')
result
```

```
Out[57]:
```

	A	B	C
aaa	-0.264438	-1.026059	-0.619500
bbb	0.927272	0.302904	-0.032399
ccc	-0.264273	-0.386314	-0.217601
ddd	-0.871858	-0.348382	1.100491

Because there was one fewer column name than the number of data rows, read_table infers that the first column should be the DataFrame's index in this special case.

The parser functions have many additional arguments to help you handle the wide variety of exception file formats that occur (see a partial listing in Table 6-2). For example, you can skip the first, third, and fourth rows of a file with skiprows:

```
In [59]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
# hey!
a,b,c,d,message
# just wanted to make things more difficult for you
# who reads CSV files with computers, anyway?
1,2,3,4,hello
5,6,7,8,world
9,10,11,12,foo
```

```
In [60]: pd.read_csv(directory+'\\ch06\\ex4.csv', skiprows=[0, 2, 3])
```

```
Out[60]:
```

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

Handling missing values is an important and frequently nuanced part of the file parsing process. Missing data is usually either not present (empty string) or marked by some sentinel value. By default, pandas uses a set of commonly occurring sentinels, such as NA and NULL:

```
In [61]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
something,a,b,c,d,message
one,1,2,3,4,NA
two,5,6,,8,world
three,9,10,11,12,foo
```

```
In [62]: result = pd.read_csv(directory+'\\ch06\\ex5.csv')
```

In [63]: `result`

Out[63]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

In [64]: `pd.isnull(result)`

Out[64]:

	something	a	b	c	d	message
0	False	False	False	False	False	True
1	False	False	False	True	False	False
2	False	False	False	False	False	False

The `na_values` option can take either a list or set of strings to consider missing values:

In [65]: `result = pd.read_csv(directory+'\\ch06\\ex5.csv', na_values=['NULL'])`
`result`

Out[65]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

Different NA sentinels can be specified for each column in a dict:

In [66]: `sentinels = {'message': ['foo', 'NA'], 'something': ['two']}`
`pd.read_csv(directory+'\\ch06\\ex5.csv', na_values=sentinels)`

Out[66]:

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	NaN	5	6	NaN	8	world
2	three	9	10	11.0	12	NaN

Table 6-2 lists some frequently used options in `pandas.read_csv` and `pandas.read_table`.

Table 6-2. Some `read_csv/read_table` function arguments

Argument	Description
<code>path</code>	String indicating filesystem location, URL, or file-like object
<code>sep</code> or <code>delimiter</code>	Character sequence or regular expression to use to split fields in each row
<code>header</code>	Row number to use as column names; defaults to 0 (first row), but should be <code>None</code> if there is no header row
<code>index_col</code>	Column numbers or names to use as the row index in the result; can be a single name/number or a list of them for a hierarchical index
<code>names</code>	List of column names for result, combine with <code>header=None</code>

Argument	Description
<code>skiprows</code>	Number of rows at beginning of file to ignore or list of row numbers (starting from 0) to skip.
<code>na_values</code>	Sequence of values to replace with NA.
<code>comment</code>	Character(s) to split comments off the end of lines.
<code>parse_dates</code>	Attempt to parse data to <code>datetime</code> ; <code>False</code> by default. If <code>True</code> , will attempt to parse all columns. Otherwise can specify a list of column numbers or name to parse. If element of list is tuple or list, will combine multiple columns together and parse to date (e.g., if date/time split across two columns).
<code>keep_date_col</code>	If joining columns to parse date, keep the joined columns; <code>False</code> by default.
<code>converters</code>	Dict containing column number or name mapping to functions (e.g., <code>{ 'foo' : f }</code> would apply the function <code>f</code> to all values in the 'foo' column).
<code>dayfirst</code>	When parsing potentially ambiguous dates, treat as international format (e.g., 7/6/2012 -> June 7, 2012); <code>False</code> by default.
<code>date_parser</code>	Function to use to parse dates.
<code>nrows</code>	Number of rows to read from beginning of file.
<code>iterator</code>	Return a <code>TextParser</code> object for reading file piecemeal.
<code>chunksize</code>	For iteration, size of file chunks.
<code>skip_footer</code>	Number of lines to ignore at end of file.
<code>verbose</code>	Print various parser output information, like the number of missing values placed in non-numeric columns.
<code>encoding</code>	Text encoding for Unicode (e.g., 'utf-8' for UTF-8 encoded text).
<code>squeeze</code>	If the parsed data only contains one column, return a <code>Series</code> .
<code>thousands</code>	Separator for thousands (e.g., ',' or '.').

Reading Text Files in Pieces

When processing very large files or figuring out the right set of arguments to correctly process a large file, you may only want to read in a small piece of a file or iterate through smaller chunks of the file.

Before we look at a large file, we make the pandas display settings more compact

```
In [1]: import pandas as pd
pd.options.display.max_rows = 10
```

```
In [4]: result = pd.read_csv(directory+'\\ch06\\ex6.csv')
result
```

```
Out[4]:
```

	one	two	three	four	key
0	0.467976	-0.038649	-0.295344	-1.824726	L
1	-0.358893	1.404453	0.704965	-0.200638	B
2	-0.501840	0.659254	-0.421691	-0.057688	G
3	0.204886	1.074134	1.388361	-0.982404	R
4	0.354628	-0.133116	0.283763	-0.837063	Q
...
9995	2.311896	-0.417070	-1.409599	-0.515821	L
9996	-0.479893	-0.650419	0.745152	-0.646038	E
9997	0.523331	0.787112	0.486066	1.093156	K
9998	-0.362559	0.598894	-1.843201	0.887292	G
9999	-0.096376	-1.012999	-0.657431	-0.573315	0

10000 rows × 5 columns

To read a file in pieces, specify a chunksize as a number of rows:

```
In [5]: chunker = pd.read_csv(directory+'\\ch06\\ex6.csv', chunksize=1000)
chunker
```

```
Out[5]: <pandas.io.parsers.TextFileReader at 0x261fbe28438>
```

The TextParser object returned by read_csv allows you to iterate over the parts of the file according to the chunksize. For example, we can iterate over ex6.csv, aggregating the value counts in the 'key' column like so:

```
In [7]: chunker = pd.read_csv(directory+'\\ch06\\ex6.csv', chunksize=1000)
```

```
In [8]: tot = pd.Series([])
for piece in chunker:
    tot = tot.add(piece['key'].value_counts(), fill_value=0)
tot = tot.sort_values(ascending=False)
```



```
C:\Users\ankit19.gupta\OneDrive - Reliance Corporate IT Park Limited\Desktop\Practice_Code\Python_Practice\Python_For_Data_Analysis\myenv\lib\site-packages\ipykernel_launcher.py:1: DeprecationWarning: The default dtype for empty Series will be 'object' instead of 'float64' in a future version. Specify a dtype explicitly to silence this warning.
"""Entry point for launching an IPython kernel.
```

```
In [9]: tot
```

```
Out[9]: E    368.0
        X    364.0
        L    346.0
        O    343.0
        Q    340.0
        ...
        5    157.0
        2    152.0
        0    151.0
        9    150.0
        1    146.0
        Length: 36, dtype: float64
```

TextParser is also equipped with a `get_chunk` method that enables you to read pieces of an arbitrary size.

Writing Data to Text Format

Data can also be exported to a delimited format. Let's consider one of the CSV files read before:

```
In [10]: data = pd.read_csv(directory+'\\ch06\\ex5.csv')
        data
```

```
Out[10]:
```

	something	a	b	c	d	message
0	one	1	2	3.0	4	NaN
1	two	5	6	NaN	8	world
2	three	9	10	11.0	12	foo

Using `DataFrame`'s `to_csv` method, we can write the data out to a comma-separated file:

```
In [12]: data.to_csv(directory+'\\ch06\\out.csv')
```

```
In [13]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
, something, a, b, c, d, message
0, one, 1, 2, 3.0, 4,
1, two, 5, 6, , 8, world
2, three, 9, 10, 11.0, 12, foo
```

Other delimiters can be used, of course (writing to `sys.stdout` so it prints the text result to the console):

```
In [14]: import sys
         data.to_csv(sys.stdout, sep='|')
```

```
|something|a|b|c|d|message
0|one|1|2|3.0|4|
1|two|5|6|8|world
2|three|9|10|11.0|12|foo
```

Missing values appear as empty strings in the output. You might want to denote them by some other sentinel value:

```
In [15]: data.to_csv(sys.stdout, na_rep='NULL')
```

```
,something,a,b,c,d,message
0,one,1,2,3.0,4,NULL
1,two,5,6,NULL,8,world
2,three,9,10,11.0,12,foo
```

With no other options specified, both the row and column labels are written. Both of these can be disabled:

```
In [16]: data.to_csv(sys.stdout, index=False, header=False)
```

```
one,1,2,3.0,4,
two,5,6,,8,world
three,9,10,11.0,12,foo
```

Series also has a `to_csv` method:

```
In [25]: import numpy as np
         dates = pd.date_range('1/1/2000', periods=7)
         ts = pd.Series(np.arange(7), index=dates)
         ts.to_csv(directory+'\\ch06\\tseries.csv', header=None)
```

```
In [26]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
```

```
2000-01-01,0
2000-01-02,1
2000-01-03,2
2000-01-04,3
2000-01-05,4
2000-01-06,5
2000-01-07,6
```

Working with Delimited Formats

It's possible to load most forms of tabular data from disk using functions like `pd.read_table`. In some cases, however, some manual processing may be necessary. It's not

uncommon to receive a file with one or more malformed lines that trip up `read_table`. To illustrate the basic tools, consider a small CSV file:

```
In [27]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\Des
"a","b","c"
"1","2","3"
"1","2","3","4"
```

For any file with a single-character delimiter, you can use Python's built-in `csv` module. To use it, pass any open file or file-like object to `csv.reader`:

```
In [28]: import csv
f = open(directory+'\\ch06\\ex7.csv')
reader = csv.reader(f)
reader
```

```
Out[28]: <_csv.reader at 0x26188560b40>
```

Iterating through the reader like a file yields tuples of values with any quote characters removed:

```
In [29]: for line in reader:
print(line)
```

```
['a', 'b', 'c']
['1', '2', '3']
['1', '2', '3', '4']
```

```
In [30]: with open(directory+'\\ch06\\ex7.csv') as f:
lines = list(csv.reader(f))
```

```
In [31]: header, values = lines[0], lines[1:]
```

```
In [32]: data_dict = {h: v for h, v in zip(header, zip(*values))}
data_dict
```

```
Out[32]: {'a': ('1', '1'), 'b': ('2', '2'), 'c': ('3', '3')}
```

CSV files come in many different flavors. To define a new format with a different delimiter, string quoting convention, or line terminator, we define a simple subclass of `csv.Dialect`:

```
In [37]: f=open(directory+'\\ch06\\ex7.csv')
class my_dialect(csv.Dialect):
    lineterminator = '\n'
    delimiter = ';'
    quotechar = '"'
    quoting = csv.QUOTE_MINIMAL
reader = csv.reader(f, dialect=my_dialect)
```

We can also give individual CSV dialect parameters as keywords to `csv.reader` without having to define a subclass:

```
In [38]: reader = csv.reader(f, delimiter='|')

In [39]: reader

Out[39]: <_csv.reader at 0x261884332b8>
```

Table 6-3. CSV dialect options

Argument	Description
delimiter	One-character string to separate fields; defaults to ','.
lineterminator	Line terminator for writing; defaults to '\r\n'. Reader ignores this and recognizes cross-platform line terminators.
quotechar	Quote character for fields with special characters (like a delimiter); default is '" '.

Argument	Description
quoting	Quoting convention. Options include <code>csv.QUOTE_ALL</code> (quote all fields), <code>csv.QUOTE_MINIMAL</code> (only fields with special characters like the delimiter), <code>csv.QUOTE_NONNUMERIC</code> , and <code>csv.QUOTE_NONE</code> (no quoting). See Python’s documentation for full details. Defaults to <code>QUOTE_MINIMAL</code> .
skipinitialspace	Ignore whitespace after each delimiter; default is <code>False</code> .
doublequote	How to handle quoting character inside a field; if <code>True</code> , it is doubled (see online documentation for full detail and behavior).
escapechar	String to escape the delimiter if quoting is set to <code>csv.QUOTE_NONE</code> ; disabled by default.

To write delimited files manually, you can use `csv.writer`. It accepts an open, writable file object and the same dialect and format options as `csv.reader`:

```
In [41]: with open('mydata.csv', 'w') as f:
writer = csv.writer(f, dialect=my_dialect)
writer.writerow(('one', 'two', 'three'))
writer.writerow(('1', '2', '3'))
writer.writerow(('4', '5', '6'))
writer.writerow(('7', '8', '9'))
```

JSON Data

JSON (short for JavaScript Object Notation) has become one of the standard formats for sending data by HTTP request between web browsers and other applications. It is a much more free-form data format than a tabular text form like CSV. Here is an example:

```
In [42]: obj = """
{"name": "Wes",
 "places_lived": ["United States", "Spain", "Germany"],
 "pet": null,
```

```

"siblings": [{"name": "Scott", "age": 30, "pets": ["Zeus", "Zuko"]},
{"name": "Katie", "age": 38,
"pets": ["Sixes", "Stache", "Cisco"]}]
}
"""

```

JSON is very nearly valid Python code with the exception of its null value null and some other nuances (such as disallowing trailing commas at the end of lists). The basic types are objects (dicts), arrays (lists), strings, numbers, booleans, and nulls. All of the keys in an object must be strings. There are several Python libraries for reading and writing JSON data. I'll use json here, as it is built into the Python standard library. To convert a JSON string to Python form, use json.loads:

```

In [43]: import json
result = json.loads(obj)
result

```

```

Out[43]: {'name': 'Wes',
'places_lived': ['United States', 'Spain', 'Germany'],
'pet': None,
'siblings': [{'name': 'Scott', 'age': 30, 'pets': ['Zeus', 'Zuko']},
{'name': 'Katie', 'age': 38, 'pets': ['Sixes', 'Stache', 'Cisco']}]}

```

json.dumps, on the other hand, converts a Python object back to JSON:

```

In [44]: asjson = json.dumps(result)

```

```

In [45]: asjson

```

```

Out[45]: '{"name": "Wes", "places_lived": ["United States", "Spain", "Germany"], "pet": null, "siblings": [{"name": "Scott", "age": 30, "pets": ["Zeus", "Zuko"]}, {"name": "Katie", "age": 38, "pets": ["Sixes", "Stache", "Cisco"]}]}

```

How you convert a JSON object or list of objects to a DataFrame or some other data structure for analysis will be up to you. Conveniently, you can pass a list of dicts (which were previously JSON objects) to the DataFrame constructor and select a sub- set of the data fields:

```

In [46]: siblings = pd.DataFrame(result['siblings'], columns=['name', 'age'])
siblings

```

```

Out[46]:
   name  age
0  Scott   30
1  Katie   38

```

The pandas.read_json can automatically convert JSON datasets in specific arrangements into a Series or DataFrame. For example:

```
In [48]: !type "C:\\Users\\ankit19.gupta\\OneDrive - Reliance Corporate IT Park Limited\\De
{
  "fruit": "Apple",
  "size": "Large",
  "color": "Red"
}
```

```
In [51]: # data = pd.read_json(directory+'\\ch06\\sample1.json')
# data
```

If you need to export data from pandas to JSON, one way is to use the `to_json` methods on Series and DataFrame:

```
In [52]: # print(data.to_json())
# print(data.to_json(orient='records'))
```

XML and HTML: Web Scraping

Python has many libraries for reading and writing data in the ubiquitous HTML and XML formats. Examples include `lxml`, `Beautiful Soup`, and `html5lib`. While `lxml` is comparatively much faster in general, the other libraries can better handle malformed HTML or XML files.

`pandas` has a built-in function, `read_html`, which uses libraries like `lxml` and `Beautiful Soup` to automatically parse tables out of HTML files as DataFrame objects. To show how this works, I downloaded an HTML file (used in the `pandas` documentation) from the United States FDIC government agency showing bank failures.¹ First, you must install some additional libraries used by `read_html`:

```
In [54]: # !pip install lxml
# !pip install beautifulsoup4 html5lib
```

The `pandas.read_html` function has a number of options, but by default it searches for and attempts to parse all tabular data contained within

tags. The result is a list of DataFrame objects:

```
In [56]: tables = pd.read_html('table.html')
len(tables)
```

```
Out[56]: 1
```

```
In [57]: failures = tables[0]
```

```
In [58]: failures.head()
```

Out[58]:

	Bank NameBank	CityCity	StateSt	CertCert	Acquiring InstitutionAI	Closing DateClosing	FundFund
0	Citizens Bank	Sac City	IA	8758	Iowa Trust & Savings Bank	November 3, 2023	10545
1	Heartland Tri-State Bank	Elkhart	KS	25851	Dream First Bank, N.A.	July 28, 2023	10544
2	First Republic Bank	San Francisco	CA	59017	JPMorgan Chase Bank, N.A.	May 1, 2023	10543
3	Signature Bank	New York	NY	57053	Flagstar Bank, N.A.	March 12, 2023	10540
4	Silicon Valley Bank	Santa Clara	CA	24735	First-Citizens Bank & Trust Company	March 10, 2023	10539

In [60]: `failures.columns`Out[60]: `Index(['Bank NameBank', 'CityCity', 'StateSt', 'CertCert', 'Acquiring InstitutionAI', 'Closing DateClosing', 'FundFund'], dtype='object')`In [61]: `close_timestamps = pd.to_datetime(failures['Closing DateClosing'])`
`close_timestamps.dt.year.value_counts()`Out[61]: `2023 5`
`2020 4`
`2019 3`
`Name: Closing DateClosing, dtype: int64`

Parsing XML with lxml.objectify

XML (eXtensible Markup Language) is another common structured data format supporting hierarchical, nested data with metadata. The book you are currently reading was actually created from a series of large XML documents.

Earlier, I showed the `pandas.read_html` function, which uses either `lxml` or `Beautiful Soup` under the hood to parse data from HTML. XML and HTML are structurally similar, but XML is more general. Here, I will show an example of how to use `lxml` to parse data from a more general XML format.

In [62]: `from lxml import objectify`
`path = 'Performance_MNR.xml'`
`parsed = objectify.parse(open(path))`
`root = parsed.getroot()`In [63]: `root`

Out[63]: <Element INFO at 0x2618a7ae848>

In [64]: root.INDICATOR

Out[64]: <Element INDICATOR at 0x2618a7a2548>

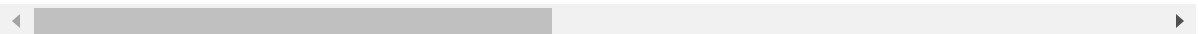
```
In [65]: data = []
skip_fields = ['PARENT_SEQ', 'INDICATOR_SEQ', 'DESIRED_CHANGE', 'DECIMAL_PLACES']
for elt in root.INDICATOR:
    el_data = {}
    for child in elt.getchildren():
        if child.tag in skip_fields:
            continue
        el_data[child.tag] = child.pyval
    data.append(el_data)
```

In [66]: perf = pd.DataFrame(data)

In [67]: perf.head()

Out[67]:

	AGENCY_NAME	INDICATOR_NAME	DESCRIPTION	PERIOD_YEAR	PERIOD_MONTH	CAT
0	Metro-North Railroad	Escalator Availability	Percent of the time that escalators are operat...	2011	12	Inc



XML data can get much more complicated than this example. Each tag can have metadata, too. Consider an HTML link tag, which is also valid XML:

```
In [68]: from io import StringIO
tag = '<a href="http://www.google.com">Google</a>'
root = objectify.parse(StringIO(tag)).getroot()
```

In [69]: root.get('href')

Out[69]: 'http://www.google.com'

In [70]: root.text

Out[70]: 'Google'

6.2 Binary Data Formats

One of the easiest ways to store data (also known as serialization) efficiently in binary format is using Python's built-in pickle serialization. pandas objects all have a `to_pickle` method that writes the data to disk in pickle format:


```
In [71]: frame = pd.read_csv(directory+'\\ch06\\ex1.csv')
         frame
```

```
Out[71]:
```

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

```
In [72]: frame.to_pickle('frame_pickle')
```

You can read any “pickled” object stored in a file by using the built-in pickle directly, or even more conveniently using pandas.read_pickle:

```
In [73]: pd.read_pickle('frame_pickle')
```

```
Out[73]:
```

	a	b	c	d	message
0	1	2	3	4	hello
1	5	6	7	8	world
2	9	10	11	12	foo

pickle is only recommended as a short-term storage format. The problem is that it is hard to guarantee that the format will be stable over time; an object pickled today may not unpickle with a later version of a library. We have tried to maintain backward compatibility when possible, but at some point in the future it may be necessary to “break” the pickle format.

Some other storage formats for pandas or NumPy data include:

bcolz:

A compressable column-oriented binary format based on the Blosc compression library.

Feather:

A cross-language column-oriented file format I designed with the R programming community's Hadley Wickham. Feather uses the Apache Arrow columnar memory format

Using HDF5 Format

HDF5 is a well-regarded file format intended for storing large quantities of scientific array data. It is available as a C library, and it has interfaces available in many other languages, including Java, Julia, MATLAB, and Python. The “HDF” in HDF5 stands for hierarchical data format. Each HDF5 file can store multiple datasets and supporting metadata. Compared with simpler formats, HDF5 supports on-the-fly compression with a variety of compression modes, enabling data with repeated patterns to be stored more efficiently. HDF5 can be a

good choice for working with very large datasets that don't fit into memory, as you can efficiently read and write small sections of much larger arrays.

While it's possible to directly access HDF5 files using either the PyTables or h5py libraries, pandas provides a high-level interface that simplifies storing Series and DataFrame object. The HDFStore class works like a dict and handles the low-level details:

```
In [13]: import pandas as pd
import numpy as np
frame = pd.DataFrame({'a': np.random.randn(100)})
store = pd.HDFStore('mydata.h5')
store['obj1'] = frame
store['obj1_col1'] = frame['a']
store
```

```
Out[13]: <class 'pandas.io.pytables.HDFStore'>
File path: mydata.h5
```

```
In [14]: # !conda install pytables
```

```
In [15]: # !pip install --user tables
```

Objects contained in the HDF5 file can then be retrieved with the same dict-like API:

```
In [16]: store['obj1']
```

```
Out[16]:
```

	a
0	-1.092360
1	-2.203172
2	-1.035297
3	0.046597
4	0.762753
...	...
95	-0.586159
96	0.130564
97	-1.002235
98	-0.667696
99	0.025241

100 rows × 1 columns

HDFStore supports two storage schemas, 'fixed' and 'table'. The latter is generally slower, but it supports query operations using a special syntax:

```
In [17]: store.put('obj2', frame, format='table')
store.select('obj2', where=['index >= 10 and index <= 15'])
```

```
Out[17]:
```

	a
10	-0.074819
11	-1.504490
12	-0.462523
13	-0.656356
14	-0.323957
15	0.314711

```
In [18]: store.close()
```

The put is an explicit version of the store['obj2'] = frame method but allows us to set other options like the storage format.

The pandas.read_hdf function gives you a shortcut to these tools:

```
In [19]: frame.to_hdf('mydata.h5', 'obj3', format='table')
```

```
In [20]: pd.read_hdf('mydata.h5', 'obj3', where=['index < 5'])
```

```
Out[20]:
```

	a
0	-1.092360
1	-2.203172
2	-1.035297
3	0.046597
4	0.762753

If you are processing data that is stored on remote servers, like Amazon S3 or HDFS, using a different binary format designed for distributed storage like Apache Parquet may be more suitable. Python for Parquet and other such storage formats is still developing, so I do not write about them in this book.

If you work with large quantities of data locally, I would encourage you to explore PyTables and h5py to see how they can suit your needs. Since many data analysis problems are I/O-

bound (rather than CPU-bound), using a tool like HDF5 can massively accelerate your applications.

HDF5 is not a database. It is best suited for write-once, read-many datasets. While data can be added to a file at any time, if multiple writers do so simultaneously, the file can become corrupted.

Reading Microsoft Excel Files

pandas also supports reading tabular data stored in Excel 2003 (and higher) files using either the `ExcelFile` class or `pandas.read_excel` function. Internally these tools use the add-on packages `xlrd` and `openpyxl` to read XLS and XLSX files, respectively. You may need to install these manually with `pip` or `conda`.

To use `ExcelFile`, create an instance by passing a path to an `xls` or `xlsx` file:

```
In [4]: import pandas as pd
        xlsx = pd.ExcelFile(directory+'\\ch06\\file_example_XLSX_10.xlsx')
```

```
In [ ]: # !pip install xlrd==1.1.0
```

Data stored in a sheet can then be read into `DataFrame` with `parse`:

```
In [5]: pd.read_excel(xlsx, 'Sheet1')
```

```
Out[5]:
```

	0	First Name	Last Name	Gender	Country	Age	Date	Id
0	1	Dulce	Abril	Female	United States	32	15/10/2017	1562
1	2	Mara	Hashimoto	Female	Great Britain	25	16/08/2016	1582
2	3	Philip	Gent	Male	France	36	21/05/2015	2587
3	4	Kathleen	Hanner	Female	United States	25	15/10/2017	3549
4	5	Nereida	Magwood	Female	United States	58	16/08/2016	2468
5	6	Gaston	Brumm	Male	United States	24	21/05/2015	2554
6	7	Etta	Hurn	Female	Great Britain	56	15/10/2017	3598
7	8	Earlean	Melgar	Female	United States	27	16/08/2016	2456
8	9	Vincenza	Weiland	Female	United States	40	21/05/2015	6548

To write pandas data to Excel format, you must first create an `ExcelWriter`, then write data to it using pandas objects' `to_excel` method:

```
In [8]: writer = pd.ExcelWriter(directory+'\\ch06\\write_excel.xlsx')
```

```
In [9]: # !pip install openpyxl
```

```
In [13]: import numpy as np
frame = pd.DataFrame({'a': np.random.randn(100)})
frame.to_excel(writer, 'Sheet1')
writer.save()
```

You can also pass a file path to `to_excel` and avoid the `ExcelWriter`:

```
In [16]: frame.to_excel(directory+'\\ch06\\write_excel1.xlsx')
```

6.3 Interacting with Web APIs

Many websites have public APIs providing data feeds via JSON or some other format. There are a number of ways to access these APIs from Python; one easy-to-use method that I recommend is the `requests` package.

To find the last 30 GitHub issues for pandas on GitHub, we can make a GET HTTP request using the add-on `requests` library:

```
In [12]: import requests
url = 'https://api.github.com/repos/pandas-dev/pandas/issues'
#url = 'https://www.geeksforgeeks.org/'
resp = requests.get(url)
resp
```

```
Out[12]: <Response [200]>
```

The `Response` object's `json` method will return a dictionary containing JSON parsed into native Python objects:

```
In [13]: data = resp.json()
data[0]['title']
```

```
Out[13]: 'BUG: concat with datetime index returns Series instead of scalar if microsecond=0'
```

```
In [14]: data
```

```

Out[14]: [{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57835',
  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57835/labels{/name}',
  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57835/comments',
  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57835/events',
  'html_url': 'https://github.com/pandas-dev/pandas/issues/57835',
  'id': 2185343261,
  'node_id': 'I_kwDOAA0YD86CQbEd',
  'number': 57835,
  'title': 'BUG: concat with datetime index returns Series instead of scalar if microsecond=0',
  'user': {'login': 'davetapley',
    'id': 48232,
    'node_id': 'MDQ6VXNlcjQ4MjMy',
    'avatar_url': 'https://avatars.githubusercontent.com/u/48232?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/davetapley',
    'html_url': 'https://github.com/davetapley',
    'followers_url': 'https://api.github.com/users/davetapley/followers',
    'following_url': 'https://api.github.com/users/davetapley/following{/other_user}',
    'gists_url': 'https://api.github.com/users/davetapley/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/davetapley/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/davetapley/subscriptions',
    'organizations_url': 'https://api.github.com/users/davetapley/orgs',
    'repos_url': 'https://api.github.com/users/davetapley/repos',
    'events_url': 'https://api.github.com/users/davetapley/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/davetapley/received_events',
    'type': 'User',
    'site_admin': False},
  'labels': [{'id': 76811,
    'node_id': 'MDU6TGFiZWw3NjgxMQ==',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
    'name': 'Bug',
    'color': 'e10c02',
    'default': False,
    'description': None},
    {'id': 1954720290,
    'node_id': 'MDU6TGFiZWwxOTU0NzIwMjkw',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Needs%20Triage',
    'name': 'Needs Triage',
    'color': '0052cc',
    'default': False,
    'description': 'Issue that has not been reviewed by a pandas team member'}],
  'state': 'open',
  'locked': False,
  'assignee': None,
  'assignees': [],
  'milestone': None,
  'comments': 0,
  'created_at': '2024-03-14T03:52:26Z',

```

```

'updated_at': '2024-03-14T03:52:26Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'body': "### Pandas version checks\n\n- [X] I have checked that this issue has not already been reported.\n\n- [X] I have confirmed this bug exists on the [latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\n\n- [X] I have confirmed this bug exists on the [main branch](https://pandas.pydata.org/docs/dev/getting_started/install.html#installing-the-development-version-of-pandas) of pandas.\n\n\n### Reproducible Example\n\n```\npython\nfrom datetime import UTC, datetime\nfrom pandas import DataFrame, concat\n\nnt1 = datetime.now(tz=UTC)\nnt2 = datetime.now(tz=UTC).replace(microsecond=0)\n\nnt1_str = str(nt1)\nnt2_str = str(nt2)\n\nndf1 = DataFrame({'a': [1]}, index=[nt1])\nprint(type(df1.loc[nt1].a))\nprint(type(df1.loc[nt1_str].a))\n\nndf2 = DataFrame({'a': [2]}, index=[nt2])\nprint(type(df2.loc[nt2].a))\nprint(type(df2.loc[nt2_str].a))\n\nndf = concat([df1, df2])\nprint(type(df.loc[nt1].a))\nprint(type(df.loc[nt1_str].a))\nprint(type(df.loc[nt2].a))\nprint(type(df.loc[nt2_str].a))\n\n```\n\n\n### Issue Description\n\n`.a` is correctly returned as `numpy.int64` in all cases, except for the last line when I use `nt2_str` and suddenly it's a one element `Series` containing the `numpy.int64`.\n\nI have no idea what on earth is going on, it took a lot of 🔍 to get a repro.\n\nI found it while writing a unit test where I was passing a timestamp from test data as a string.\n\nIf you remove the `.replace(microsecond=0)` you'll see it works as expected 🤔\n\n\n### Expected Behavior\n\n`.loc` should be consistent before and after a `concat`.\n\n\n### Installed Versions\n\n<details>\n\nINSTALLED VERSIONS\n\n-----\n\ncommit : bdc79c146c2e32f2cab629be240f01658cfb6cc2\npython : 3.11.6.final.0\npython-bits : 64\nOS : Linux\nOS-release : 6.2.0-1019-azure\nVersion : #19~22.04.1-Ubuntu SMP Wed Jan 10 22:57:03 UTC 2024\nmachine : x86_64\nprocessor : x86_64\nbyteorder : little\nLC_ALL : None\nLANG : C.UTF-8\nLOCALE : en_US.UTF-8\n\npandas : 2.2.1\nnumpy : 1.26.1\npytz : 2023.3.post1\ndateutil : 2.8.2\nsetuptools : 68.2.2\npip : 23.3.1\nCython : None\npytest : None\nhypothesis : None\nsphinx : None\nblosc : None\nfeather : None\nxlsxwriter : None\nlxml.etree : None\nhtml5lib : None\npymysql : None\npsycopg2 : None\nninja2 : 3.1.2\nIPython : None\npandas_datareader : None\nadbc-driver-postgresql : None\nadbc-driver-sqlite : None\nbs4 : None\nbottleneck : None\ndataframe-api-compat : None\nfastparquet : 2023.2.0\nfsspec : 2023.10.0\ngeospatial : None\nmatplotlib : None\nnumba : None\nnumexpr : None\nnodfpy : None\nopenpyxl : None\npandas_gbq : None\npyarrow : 14.0.1\npyreadstat : None\npython-calamine : None\npyxlsb : None\ns3fs : None\nscipy : 1.10.1\nsqlalchemy : None\ntables : None\nntables : None\nntabulate : None\nnxarray : None\nxlrd : None\nnzstandard : None\ntzdata : 2023.3\nnqtpy : None\npyqt5 : None\n\n</details>\n",
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5783/5/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,

```

```

    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57835/timeline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57833',
     'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
     'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57833/labels{/name}',
     'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57833/comments',
     'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57833/events',
     'html_url': 'https://github.com/pandas-dev/pandas/pull/57833',
     'id': 2185047630,
     'node_id': 'PR_kwDOAA0YD85pj-Lw',
     'number': 57833,
     'title': 'PERF: RangeIndex.insert maintains RangeIndex when empty',
     'user': {'login': 'mroeschke',
              'id': 10647082,
              'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
              'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
              'gravatar_id': '',
              'url': 'https://api.github.com/users/mroeschke',
              'html_url': 'https://github.com/mroeschke',
              'followers_url': 'https://api.github.com/users/mroeschke/followers',
              'following_url': 'https://api.github.com/users/mroeschke/following{/other_user}',
              'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
              'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{/repo}',
              'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',
              'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
              'repos_url': 'https://api.github.com/users/mroeschke/repos',
              'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
              'received_events_url': 'https://api.github.com/users/mroeschke/received_events',
              'type': 'User',
              'site_admin': False},
     'labels': [{'id': 8935311,
                  'node_id': 'MDU6TGFiZWw4OTM1MzEx',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Performance',
                  'name': 'Performance',
                  'color': 'a10c02',
                  'default': False,
                  'description': 'Memory or execution speed performance'},
                 {'id': 1218227310,
                  'node_id': 'MDU6TGFiZWwzMjE4MjI3MzEw',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Index',
                  'name': 'Index',
                  'color': 'e99695',
                  'default': False,
                  'description': 'Related to the Index class or subclasses'}]},

```



```

'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 0,
'created_at': '2024-03-13T23:02:56Z',
'updated_at': '2024-03-14T00:46:10Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
833',
  'html_url': 'https://github.com/pandas-dev/pandas/pull/57833',
  'diff_url': 'https://github.com/pandas-dev/pandas/pull/57833.diff',
  'patch_url': 'https://github.com/pandas-dev/pandas/pull/57833.patch',
  'merged_at': None},
'body': 'Discovered in https://github.com/pandas-dev/pandas/pull/57441',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5783
3/reactions',
  'total_count': 0,
  '+1': 0,
  '-1': 0,
  'laugh': 0,
  'hooray': 0,
  'confused': 0,
  'heart': 0,
  'rocket': 0,
  'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57833/tim
eline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57831',
 'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
 'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57831/label
s{/name}',
 'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57831/com
ments',
 'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57831/event
s',
 'html_url': 'https://github.com/pandas-dev/pandas/issues/57831',
 'id': 2184582507,
 'node_id': 'I_kwDOAA0YD86CNhVr',
 'number': 57831,
 'title': 'DOC: Remove Dask and Modin sections in `scale.rst` in favor of linking
to ecosystem docs.',
 'user': {'login': 'mroeschke',
  'id': 10647082,
  'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
  'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
  'gravatar_id': '',
  'url': 'https://api.github.com/users/mroeschke',
  'html_url': 'https://github.com/mroeschke',
  'followers_url': 'https://api.github.com/users/mroeschke/followers',

```

```

    'following_url': 'https://api.github.com/users/mroeschke/following{/other_user}',
    'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',
    'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
    'repos_url': 'https://api.github.com/users/mroeschke/repos',
    'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/mroeschke/received_events',
    'type': 'User',
    'site_admin': False,
    'labels': [{ 'id': 134699,
                  'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
                  'name': 'Docs',
                  'color': '3465A4',
                  'default': False,
                  'description': None},
                { 'id': 717120670,
                  'node_id': 'MDU6TGFiZWw3MTcxMjA2NzA=',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/good%20first%20issue',
                  'name': 'good first issue',
                  'color': '0e8a16',
                  'default': True,
                  'description': None}],
    'state': 'open',
    'locked': False,
    'assignee': { 'login': 'yukikitayama',
                  'id': 47092819,
                  'node_id': 'MDQ6VXNlcjQ3MDkyODE5',
                  'avatar_url': 'https://avatars.githubusercontent.com/u/47092819?v=4',
                  'gravatar_id': '',
                  'url': 'https://api.github.com/users/yukikitayama',
                  'html_url': 'https://github.com/yukikitayama',
                  'followers_url': 'https://api.github.com/users/yukikitayama/followers',
                  'following_url': 'https://api.github.com/users/yukikitayama/following{/other_user}'},
    'gists_url': 'https://api.github.com/users/yukikitayama/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/yukikitayama/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/yukikitayama/subscriptions',
    'organizations_url': 'https://api.github.com/users/yukikitayama/orgs',
    'repos_url': 'https://api.github.com/users/yukikitayama/repos',
    'events_url': 'https://api.github.com/users/yukikitayama/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/yukikitayama/received_events',
    'type': 'User',
    'site_admin': False,
    'assignees': [{ 'login': 'yukikitayama',
                    'id': 47092819,
                    'node_id': 'MDQ6VXNlcjQ3MDkyODE5',
                    'avatar_url': 'https://avatars.githubusercontent.com/u/47092819?v=4',
                    'gravatar_id': '',
                    'url': 'https://api.github.com/users/yukikitayama',

```

```

    'html_url': 'https://github.com/yukikitayama',
    'followers_url': 'https://api.github.com/users/yukikitayama/followers',
    'following_url': 'https://api.github.com/users/yukikitayama/following{/other_u
ser}',
    'gists_url': 'https://api.github.com/users/yukikitayama/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/yukikitayama/starred{/owner}/{/rep
o}',
    'subscriptions_url': 'https://api.github.com/users/yukikitayama/subscription
s',
    'organizations_url': 'https://api.github.com/users/yukikitayama/orgs',
    'repos_url': 'https://api.github.com/users/yukikitayama/repos',
    'events_url': 'https://api.github.com/users/yukikitayama/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/yukikitayama/received_eve
nts',
    'type': 'User',
    'site_admin': False}],
    'milestone': None,
    'comments': 1,
    'created_at': '2024-03-13T17:52:22Z',
    'updated_at': '2024-03-13T22:34:59Z',
    'closed_at': None,
    'author_association': 'MEMBER',
    'active_lock_reason': None,
    'body': 'xref https://github.com/pandas-dev/pandas/pull/57586#pullrequestreview-
1922295036\r\n\r\n1. Remove `Use Dask` and `Use Modin` sections in `doc/source/use
r_guide/scale.rst`\r\n2. Add a new section in `doc/source/user_guide/scale.rst`
(`Use Other Libraries`) and link to `Out-of-core` section in `web/pandas/communit
y/ecosystem.md`\r\n3. Remove `dask-expr` in `environment.yml` and `requirements.tx
t`\r\n4. Remove version pinning on `dask` post https://github.com/pandas-dev/panda
s/pull/57830',
    'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5783
1/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57831/tim
eline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57829',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57829/label
s{/name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57829/com
ments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57829/event
s',
    'html_url': 'https://github.com/pandas-dev/pandas/issues/57829',
    'id': 2183819691,
    'node_id': 'I_kwDOAA0YD86CKnGr',

```

```

'number': 57829,
'title': 'Potential regression induced by PR "PERF: Categorical(range).categorie
s returns RangeIndex instead of Index"',
'user': {'login': 'DeaMariaLeon',
'id': 11835246,
'node_id': 'MDQ6VXNlcjExODM1MjQ2',
'avatar_url': 'https://avatars.githubusercontent.com/u/11835246?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/DeaMariaLeon',
'html_url': 'https://github.com/DeaMariaLeon',
'followers_url': 'https://api.github.com/users/DeaMariaLeon/followers',
'following_url': 'https://api.github.com/users/DeaMariaLeon/following{/other_us
er}',
'gists_url': 'https://api.github.com/users/DeaMariaLeon/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/DeaMariaLeon/starred{/owner}/{/rep
o}',
'subscriptions_url': 'https://api.github.com/users/DeaMariaLeon/subscriptions',
'organizations_url': 'https://api.github.com/users/DeaMariaLeon/orgs',
'repos_url': 'https://api.github.com/users/DeaMariaLeon/repos',
'events_url': 'https://api.github.com/users/DeaMariaLeon/events{/privacy}',
'received_events_url': 'https://api.github.com/users/DeaMariaLeon/received_even
ts',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 8935311,
'node_id': 'MDU6TGFiZWw4OTM1MzEx',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Performance',
'name': 'Performance',
'color': 'a10c02',
'default': False,
'description': 'Memory or execution speed performance'},
{ 'id': 71268330,
'node_id': 'MDU6TGFiZWw3MTI2ODMzMzMA==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/MultiIndex',
'name': 'MultiIndex',
'color': '207de5',
'default': False,
'description': None}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-13T11:57:05Z',
'updated_at': '2024-03-13T17:11:09Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'body': 'PR #57787 \r\n\r\nIf this is expected please ignore the issue.\r\n\r\nT
he regressions seem to be here:\r\n\r\n`indexing.MultiIndexing.time_xs_full_key`
(Python) with unique_levels=True\r\n\r\n`indexing.MultiIndexing.time_loc_all_scalars`
(Python) with unique_levels=True\r\n\r\n@mroeschke\r\n\r\n\r\n\r\n\r\n\r\n',
  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5782
9/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57829/tim
eline',
  'performed_via_github_app': None,
  'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57828',
  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57828/label
s{/name}',
  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57828/com
ments',
  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57828/event
s',
  'html_url': 'https://github.com/pandas-dev/pandas/issues/57828',
  'id': 2183235082,
  'node_id': 'I_kwDOAA0YD86CIYYK',
  'number': 57828,
  'title': 'Copy-on-Write Guide - "Previous behavior" output is what would happen
post-migration to CoW',
  'user': {'login': 'gah-bo',
    'id': 64602073,
    'node_id': 'MDQ6VXNlcjY0NjAyMDcz',
    'avatar_url': 'https://avatars.githubusercontent.com/u/64602073?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/gah-bo',
    'html_url': 'https://github.com/gah-bo',
    'followers_url': 'https://api.github.com/users/gah-bo/followers',
    'following_url': 'https://api.github.com/users/gah-bo/following{/other_user}',
    'gists_url': 'https://api.github.com/users/gah-bo/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/gah-bo/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/gah-bo/subscriptions',
    'organizations_url': 'https://api.github.com/users/gah-bo/orgs',
    'repos_url': 'https://api.github.com/users/gah-bo/repos',
    'events_url': 'https://api.github.com/users/gah-bo/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/gah-bo/received_events',
    'type': 'User',
    'site_admin': False},
  'labels': [{'id': 134699,
    'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
    'name': 'Docs',
    'color': '3465A4',
    'default': False,
    'description': None},
    {'id': 2085877452,

```

```

'node_id': 'MDU6TGFiZWwYMDg1ODc3NDUy',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Copy%20/%20view%
20semantics',
'name': 'Copy / view semantics',
'color': '70e5ca',
'default': False,
'description': ''}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-13T06:42:29Z',
'updated_at': '2024-03-13T21:08:11Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'body': '### Pandas version checks\r\n\r\n- [X] I have checked that the issue st
ill exists on the latest versions of the docs on `main` [here](https://pandas.pyda
ta.org/docs/dev/)\r\n\r\n\r\n### Location of the documentation\r\n\r\nhttps://pand
as.pydata.org/docs/dev/user_guide/copy_on_write.html\r\n\r\n\r\n### Documentation prob
lem\r\n\r\n\r\nCopy-on-Write (CoW) section of "Previous behavior" states that the beha
vior before CoW leads to this output:\r\n\r\n\r\n```\r\nIn [1]: df = pd.DataFrame({"fo
o": [1, 2, 3], "bar": [4, 5, 6]})\r\n\r\n\r\nIn [2]: subset = df["foo"]\r\n\r\n\r\nIn [3]:
subset.iloc[0] = 100\r\n\r\n\r\nIn [4]: df\r\nOut[4]: \r\n   foo bar\r\n0    1    4\r
\r\n1    2    5\r\n2    3    6\r\n\r\n```\r\n\r\n\r\nBut that's the behavior that would hap
pen after migrating to CoW\r\n\r\n\r\nNote: this codeblock seems to be used more than
once, each time having this issue\r\n\r\n\r\n### Suggested fix for documentation\r\n\r\n\r
\r\nActually show "Previous behavior" on the codeblock ',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5782
8/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57828/tim
eline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826/label
s{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826/com
ments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826/event
s',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57826',
'id': 2182894169,
'node_id': 'PR_kwDOAA0YD85pc13P',

```

```

'number': 57826,
'title': 'CI: speedup docstring check consecutive runs',
'user': {'login': 'dontgoto',
'id': 6079615,
'node_id': 'MDQ6VXNlcjYwNzk2MTU=',
'avatar_url': 'https://avatars.githubusercontent.com/u/6079615?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/dontgoto',
'html_url': 'https://github.com/dontgoto',
'followers_url': 'https://api.github.com/users/dontgoto/followers',
'following_url': 'https://api.github.com/users/dontgoto/following{/other_user}',
'gists_url': 'https://api.github.com/users/dontgoto/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/dontgoto/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/dontgoto/subscriptions',
'organizations_url': 'https://api.github.com/users/dontgoto/orgs',
'repos_url': 'https://api.github.com/users/dontgoto/repos',
'events_url': 'https://api.github.com/users/dontgoto/events{/privacy}',
'received_events_url': 'https://api.github.com/users/dontgoto/received_events',
'type': 'User',
'site_admin': False},
'labels': [{'id': 134699,
'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
'name': 'Docs',
'color': '3465A4',
'default': False,
'description': None},
{'id': 48070600,
'node_id': 'MDU6TGFiZWw0ODA3MDYwMA==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/CI',
'name': 'CI',
'color': 'a2bca7',
'default': False,
'description': 'Continuous Integration'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 4,
'created_at': '2024-03-13T00:30:04Z',
'updated_at': '2024-03-13T16:27:13Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57826',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57826',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57826.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57826.patch',
'merged_at': None},
'body': "- [x] [Tests added and passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if fixing a bug or adding a new feature\r\n- [x] All [code checks passed](https://pandas.pydata.org/pandas-d

```

ocs/dev/development/contributing_codebase.html#pre-commit).\r\n\r\nThis PR brings the runtime of the docstring check CI down to 2-3 minutes from about 20 minutes.\r\n\r\nCurrently, `check_code.sh docstring` does multiple calls to `validate_docstrings.py` due to various error types that have function exceptions. Each run of `validate_docstrings` takes about 2-3 minutes, leading to the current 20 minutes runtime just for the docstring checks.\r\n\r\nThe runtime for consecutive calls is brought down to that of a single call by changing the argument parsing of `validate_docstrings`, adding `--` to separate parameters that previously were separate function calls. Additionally, a cache is added to reuse the parsing results between runs. The cache size should pose no problem, the Python instance ran by `check_code.sh docstring` only reserves about 95MB of memory on my machine.",

```
{
  'reactions': {
    'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0
  },
  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57826/timeline',
  'performed_via_github_app': None,
  'state_reason': None
}, {
  'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57825',
  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57825/labels{/name}',
  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57825/comments',
  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57825/events',
  'html_url': 'https://github.com/pandas-dev/pandas/pull/57825',
  'id': 2182861259,
  'node_id': 'PR_kwDOAA0YD85pcewM',
  'number': 57825,
  'title': 'PERF: Unary methods on RangeIndex returns RangeIndex',
  'user': {
    'login': 'mroeschke',
    'id': 10647082,
    'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
    'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/mroeschke',
    'html_url': 'https://github.com/mroeschke',
    'followers_url': 'https://api.github.com/users/mroeschke/followers',
    'following_url': 'https://api.github.com/users/mroeschke/following{/other_user}',
    'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{/repo}',
    'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',
    'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
    'repos_url': 'https://api.github.com/users/mroeschke/repos',
    'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/mroeschke/received_event'
  }
}
```



```

s',
  'type': 'User',
  'site_admin': False},
  'labels': [{ 'id': 8935311,
    'node_id': 'MDU6TGFiZWw4OTM1MzEx',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Performance',
    'name': 'Performance',
    'color': 'a10c02',
    'default': False,
    'description': 'Memory or execution speed performance'},
    { 'id': 1218227310,
    'node_id': 'MDU6TGFiZWwzMjE4MjI3MzEw',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Index',
    'name': 'Index',
    'color': 'e99695',
    'default': False,
    'description': 'Related to the Index class or subclasses'}],
  'state': 'open',
  'locked': False,
  'assignee': None,
  'assignees': [],
  'milestone': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/
102',
    'html_url': 'https://github.com/pandas-dev/pandas/milestone/102',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102/la
bels',
    'id': 9021906,
    'node_id': 'MI_kwDOAA0YD84AianS',
    'number': 102,
    'title': '3.0',
    'description': '',
    'creator': { 'login': 'jbrockmendel',
    'id': 8078968,
    'node_id': 'MDQ6VXNlcjgwNzg5Njg=',
    'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/jbrockmendel',
    'html_url': 'https://github.com/jbrockmendel',
    'followers_url': 'https://api.github.com/users/jbrockmendel/followers',
    'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_u
ser}',
    'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}/{rep
o}',
    'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscription
s',
    'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',
    'repos_url': 'https://api.github.com/users/jbrockmendel/repos',
    'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/jbrockmendel/received_eve
nts',
    'type': 'User',
    'site_admin': False},
    'open_issues': 53,
    'closed_issues': 402,
    'state': 'open',

```

```

    'created_at': '2023-02-08T18:16:49Z',
    'updated_at': '2024-03-14T02:02:54Z',
    'due_on': '2024-04-01T07:00:00Z',
    'closed_at': None},
    'comments': 1,
    'created_at': '2024-03-12T23:48:52Z',
    'updated_at': '2024-03-14T02:05:13Z',
    'closed_at': None,
    'author_association': 'MEMBER',
    'active_lock_reason': None,
    'draft': False,
    'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
825',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57825',
    'diff_url': 'https://github.com/pandas-dev/pandas/pull/57825.diff',
    'patch_url': 'https://github.com/pandas-dev/pandas/pull/57825.patch',
    'merged_at': None},
    'body': None,
    'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5782
5/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57825/tim
eline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57820',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57820/label
s{/name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57820/com
ments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57820/event
s',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57820',
    'id': 2181624666,
    'node_id': 'PR_kwDOAA0YD85pYK8w',
    'number': 57820,
    'title': 'CLN: enforce deprecation of `interpolate` with object dtype',
    'user': {'login': 'natmokval',
    'id': 91160475,
    'node_id': 'MDQ6VXNlcjkxMTYwNDc1',
    'avatar_url': 'https://avatars.githubusercontent.com/u/91160475?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/natmokval',
    'html_url': 'https://github.com/natmokval',
    'followers_url': 'https://api.github.com/users/natmokval/followers',
    'following_url': 'https://api.github.com/users/natmokval/following{/other_use
r}'},

```

```

'gists_url': 'https://api.github.com/users/natmokval/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/natmokval/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/natmokval/subscriptions',
'organizations_url': 'https://api.github.com/users/natmokval/orgs',
'repos_url': 'https://api.github.com/users/natmokval/repos',
'events_url': 'https://api.github.com/users/natmokval/events{/privacy}',
'received_events_url': 'https://api.github.com/users/natmokval/received_event
s',
'type': 'User',
'site_admin': False,
'labels': [{ 'id': 2822342,
  'node_id': 'MDU6TGFiZWwYODIyMzQy',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Missing-data',
  'name': 'Missing-data',
  'color': 'd7e102',
  'default': False,
  'description': 'np.nan, pd.NaT, pd.NA, dropna, isnull, interpolate'},
{ 'id': 211029535,
  'node_id': 'MDU6TGFiZWwYMTUwMjk1MzU=',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Clean',
  'name': 'Clean',
  'color': '207de5',
  'default': False,
  'description': None}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-12T13:27:42Z',
'updated_at': '2024-03-13T22:52:14Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'draft': False,
'pull_request': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
820',
  'html_url': 'https://github.com/pandas-dev/pandas/pull/57820',
  'diff_url': 'https://github.com/pandas-dev/pandas/pull/57820.diff',
  'patch_url': 'https://github.com/pandas-dev/pandas/pull/57820.patch',
  'merged_at': None},
'body': 'xref #53638\r\n\r\nenforced deprecation of `interpolate` with object dt
ype\r\n',
'reactions': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5782
0/reactions',
  'total_count': 0,
  '+1': 0,
  '-1': 0,
  'laugh': 0,
  'hooray': 0,
  'confused': 0,
  'heart': 0,
  'rocket': 0,
  'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57820/tim

```

```

eline',
  'performed_via_github_app': None,
  'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817',
 'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
 'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817/labels{/name}',
 'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817/comments',
 'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817/events',
 'html_url': 'https://github.com/pandas-dev/pandas/pull/57817',
 'id': 2180589157,
 'node_id': 'PR_kwDOAA0YD85pUo9E',
 'number': 57817,
 'title': 'DOC: Updated the returns for DataFrame.any/all to return either a Series or scalar',
 'user': {'login': 'sammig6i',
 'id': 115059655,
 'node_id': 'U_kgDOBturxw',
 'avatar_url': 'https://avatars.githubusercontent.com/u/115059655?v=4',
 'gravatar_id': '',
 'url': 'https://api.github.com/users/sammig6i',
 'html_url': 'https://github.com/sammig6i',
 'followers_url': 'https://api.github.com/users/sammig6i/followers',
 'following_url': 'https://api.github.com/users/sammig6i/following{/other_user}',
 'gists_url': 'https://api.github.com/users/sammig6i/gists{/gist_id}',
 'starred_url': 'https://api.github.com/users/sammig6i/starred{/owner}/{/repo}',
 'subscriptions_url': 'https://api.github.com/users/sammig6i/subscriptions',
 'organizations_url': 'https://api.github.com/users/sammig6i/orgs',
 'repos_url': 'https://api.github.com/users/sammig6i/repos',
 'events_url': 'https://api.github.com/users/sammig6i/events{/privacy}',
 'received_events_url': 'https://api.github.com/users/sammig6i/received_events',
 'type': 'User',
 'site_admin': False},
'labels': [{'id': 134699,
 'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
 'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
 'name': 'Docs',
 'color': '3465A4',
 'default': False,
 'description': None}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 2,
'created_at': '2024-03-12T02:21:46Z',
'updated_at': '2024-03-12T12:39:32Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57

```

```

817',
  'html_url': 'https://github.com/pandas-dev/pandas/pull/57817',
  'diff_url': 'https://github.com/pandas-dev/pandas/pull/57817.diff',
  'patch_url': 'https://github.com/pandas-dev/pandas/pull/57817.patch',
  'merged_at': None},
  'body': '#57088 \r\nupdated documentation for reflect the return types for Dataframe.any/all calls.\r\n\r\n',
  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57817/timeline',
  'performed_via_github_app': None,
  'state_reason': None},
  {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57813',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57813/labels{/name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57813/comments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57813/events',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57813',
    'id': 2180108115,
    'node_id': 'PR_kwDOAA0YD85pS-7X',
    'number': 57813,
    'title': 'CI: Fail tests on all builds for FutureWarning/DeprecationWarning from numpy or pyarrow',
    'user': {'login': 'mroeschke',
      'id': 10647082,
      'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
      'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
      'gravatar_id': '',
      'url': 'https://api.github.com/users/mroeschke',
      'html_url': 'https://github.com/mroeschke',
      'followers_url': 'https://api.github.com/users/mroeschke/followers',
      'following_url': 'https://api.github.com/users/mroeschke/following{/other_user}',
      'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
      'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{/repo}',
      'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',
      'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
      'repos_url': 'https://api.github.com/users/mroeschke/repos',
      'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
      'received_events_url': 'https://api.github.com/users/mroeschke/received_events',
      'type': 'User',
      'site_admin': False},

```

```

'labels': [{ 'id': 127685,
  'node_id': 'MDU6TGFiZWwzMjc2ODU=',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Testing',
  'name': 'Testing',
  'color': 'C4A000',
  'default': False,
  'description': 'pandas testing functions or related to the test suite'},
{ 'id': 48070600,
  'node_id': 'MDU6TGFiZWw0ODA3MDYwMA==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/CI',
  'name': 'CI',
  'color': 'a2bca7',
  'default': False,
  'description': 'Continuous Integration'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/117',
  'html_url': 'https://github.com/pandas-dev/pandas/milestone/117',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/117/labels',
  'id': 10539704,
  'node_id': 'MI_kwDOAA0YD84AoNK4',
  'number': 117,
  'title': '2.2.2',
  'description': 'on-merge: backport to 2.2.x',
  'creator': { 'login': 'lithomas1',
    'id': 47963215,
    'node_id': 'MDQ6VXNlcjQ3OTYzMjE1',
    'avatar_url': 'https://avatars.githubusercontent.com/u/47963215?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/lithomas1',
    'html_url': 'https://github.com/lithomas1',
    'followers_url': 'https://api.github.com/users/lithomas1/followers',
    'following_url': 'https://api.github.com/users/lithomas1/following{/other_users}',
    'gists_url': 'https://api.github.com/users/lithomas1/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/lithomas1/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/lithomas1/subscriptions',
    'organizations_url': 'https://api.github.com/users/lithomas1/orgs',
    'repos_url': 'https://api.github.com/users/lithomas1/repos',
    'events_url': 'https://api.github.com/users/lithomas1/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/lithomas1/received_events',
    'type': 'User',
    'site_admin': False},
  'open_issues': 42,
  'closed_issues': 26,
  'state': 'open',
  'created_at': '2024-02-10T23:53:00Z',
  'updated_at': '2024-03-14T00:39:41Z',
  'due_on': None,
  'closed_at': None},

```

```

'comments': 0,
'created_at': '2024-03-11T20:01:39Z',
'updated_at': '2024-03-12T02:49:21Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
813',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57813',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57813.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57813.patch',
'merged_at': None},
'body': "I don't think we should necessarily limit this to just to nightly build
of these libraries",
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5781
3/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57813/tim
eline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57812',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57812/label
s{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57812/com
ments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57812/event
s',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57812',
'id': 2179938698,
'node_id': 'PR_kwDOAA0YD85pSZSj',
'number': 57812,
'title': 'PERF: Avoid np.divmod in RangeIndex._shallow_copy',
'user': {'login': 'mroeschke',
'id': 10647082,
'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/mroeschke',
'html_url': 'https://github.com/mroeschke',
'followers_url': 'https://api.github.com/users/mroeschke/followers',
'following_url': 'https://api.github.com/users/mroeschke/following{/other_use
r}',
'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',

```

```

'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
'repos_url': 'https://api.github.com/users/mroeschke/repos',
'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
'received_events_url': 'https://api.github.com/users/mroeschke/received_events',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 8935311,
  'node_id': 'MDU6TGFiZWw4OTM1MzEx',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Performance',
  'name': 'Performance',
  'color': 'a10c02',
  'default': False,
  'description': 'Memory or execution speed performance'},
{ 'id': 1218227310,
  'node_id': 'MDU6TGFiZWwzMjE4MjI3MzEw',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Index',
  'name': 'Index',
  'color': 'e99695',
  'default': False,
  'description': 'Related to the Index class or subclasses'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102',
  'html_url': 'https://github.com/pandas-dev/pandas/milestone/102',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102/labels',
  'id': 9021906,
  'node_id': 'MI_kwDOAA0YD84AianS',
  'number': 102,
  'title': '3.0',
  'description': '',
  'creator': { 'login': 'jbrockmendel',
    'id': 8078968,
    'node_id': 'MDQ6VXNlcjgwNzg5Njg=',
    'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/jbrockmendel',
    'html_url': 'https://github.com/jbrockmendel',
    'followers_url': 'https://api.github.com/users/jbrockmendel/followers',
    'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_user}',
    'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscriptions',
    'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',
    'repos_url': 'https://api.github.com/users/jbrockmendel/repos',
    'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/jbrockmendel/received_events',
    'type': 'User',

```



```

    'site_admin': False},
    'open_issues': 53,
    'closed_issues': 402,
    'state': 'open',
    'created_at': '2023-02-08T18:16:49Z',
    'updated_at': '2024-03-14T02:02:54Z',
    'due_on': '2024-04-01T07:00:00Z',
    'closed_at': None},
    'comments': 0,
    'created_at': '2024-03-11T18:34:44Z',
    'updated_at': '2024-03-14T01:48:59Z',
    'closed_at': None,
    'author_association': 'MEMBER',
    'active_lock_reason': None,
    'draft': False,
    'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
812',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57812',
    'diff_url': 'https://github.com/pandas-dev/pandas/pull/57812.diff',
    'patch_url': 'https://github.com/pandas-dev/pandas/pull/57812.patch',
    'merged_at': None},
    'body': 'xref https://github.com/pandas-dev/pandas/pull/57534#issuecomment-19578
32841\r\n\r\nMade a `is_range` method (like `is_range_indexer`) that avoids a `np.
divmod` operation\r\n\r\n``python\r\nIn [1]: from pandas import *; import numpy a
s np\r\n    ...: np.random.seed(123)\r\n    ...: size = 1_000_000\r\n    ...: ngroups
= 1000\r\n    ...: data = Series(np.random.randint(0, ngroups, size=size))\r\n+ /op
t/miniconda3/envs/pandas-dev/bin/ninja\r\n[1/1] Generating write_version_file with
a custom command\r\n\r\nIn [2]: %timeit data.groupby(data).groups\r\n14 ms ± 552 µ
s per loop (mean ± std. dev. of 7 runs, 100 loops each) # PR\r\n\r\nIn [3]: %time
it data.groupby(data).groups\r\n17.8 ms ± 84.8 µs per loop (mean ± std. dev. of 7
runs, 100 loops each) # main\r\n```\r\n',
    'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5781
2/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57812/tim
eline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57811',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57811/label
s{name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57811/com
ments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57811/event
s',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57811',
    'id': 2179797779,

```

```

'node_id': 'PR_kwDOAA0YD85pR46a',
'number': 57811,
'title': 'BUG: improve pd.io.json_normalize',
'user': {'login': 'slavanorm',
'id': 48209342,
'node_id': 'MDQ6VXNlcjQ4MjA5MzQy',
'avatar_url': 'https://avatars.githubusercontent.com/u/48209342?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/slavanorm',
'html_url': 'https://github.com/slavanorm',
'followers_url': 'https://api.github.com/users/slavanorm/followers',
'following_url': 'https://api.github.com/users/slavanorm/following{/other_user}',
'gists_url': 'https://api.github.com/users/slavanorm/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/slavanorm/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/slavanorm/subscriptions',
'organizations_url': 'https://api.github.com/users/slavanorm/orgs',
'repos_url': 'https://api.github.com/users/slavanorm/repos',
'events_url': 'https://api.github.com/users/slavanorm/events{/privacy}',
'received_events_url': 'https://api.github.com/users/slavanorm/received_events',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 76811,
'node_id': 'MDU6TGFiZWw3NjgxMQ==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
'name': 'Bug',
'color': 'e10c02',
'default': False,
'description': None},
{ 'id': 49379259,
'node_id': 'MDU6TGFiZWw0OTM3OTI1OQ==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/IO%20JSON',
'name': 'IO JSON',
'color': '207de5',
'default': False,
'description': 'read_json, to_json, json_normalize'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 4,
'created_at': '2024-03-11T17:48:55Z',
'updated_at': '2024-03-14T01:56:17Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57811',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57811',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57811.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57811.patch',
'merged_at': None},
'body': '- [x] closes #57810\r\n- [x] [Tests added and passed](https://pandas.py

```

```

data.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if
fixing a bug or adding a new feature\r\n- [x] All [code checks passed](https://pan
das.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commi
t).\r\n- [x] Added [type annotations](https://pandas.pydata.org/pandas-docs/dev/de
velopment/contributing_codebase.html#type-hints) to new arguments/methods/function
s.\r\n- [x] Added an entry in the latest `doc/source/whatsnew/v2.2.2.rst` file if
fixing a bug or adding a new feature.\r\n',
  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5781
1/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57811/tim
eline',
  'performed_via_github_app': None,
  'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810',
  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810/label
s{/name}',
  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810/com
ments',
  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810/event
s',
  'html_url': 'https://github.com/pandas-dev/pandas/issues/57810',
  'id': 2179589085,
  'node_id': 'I_kwDOAA0YD86B6ePd',
  'number': 57810,
  'title': 'BUG: pd.json_normalize improvement',
  'user': {'login': 'slavanorm',
    'id': 48209342,
    'node_id': 'MDQ6VXNlcjQ4MjA5MzQy',
    'avatar_url': 'https://avatars.githubusercontent.com/u/48209342?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/slavanorm',
    'html_url': 'https://github.com/slavanorm',
    'followers_url': 'https://api.github.com/users/slavanorm/followers',
    'following_url': 'https://api.github.com/users/slavanorm/following{/other_use
r}',
    'gists_url': 'https://api.github.com/users/slavanorm/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/slavanorm/starred{/owner}/{/repo}',
    'subscriptions_url': 'https://api.github.com/users/slavanorm/subscriptions',
    'organizations_url': 'https://api.github.com/users/slavanorm/orgs',
    'repos_url': 'https://api.github.com/users/slavanorm/repos',
    'events_url': 'https://api.github.com/users/slavanorm/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/slavanorm/received_event
s',
    'type': 'User',
    'site_admin': False},
  'labels': [{'id': 76811,

```

```
'node_id': 'MDU6TGFiZWw3NjgxMQ==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
'name': 'Bug',
'color': 'e10c02',
'default': False,
'description': None},
{'id': 1954720290,
'node_id': 'MDU6TGFiZWw3OTU0NzIwMjkx',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Needs%20Triage',
'name': 'Needs Triage',
'color': '0052cc',
'default': False,
'description': 'Issue that has not been reviewed by a pandas team member'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-11T16:36:25Z',
'updated_at': '2024-03-11T20:15:26Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'body': "### Pandas version checks\r\n\r\n- [X] I have checked that this issue h
as not already been reported.\r\n\r\n- [X] I have confirmed this bug exists on the
[latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\r
\r\n- [X] I have confirmed this bug exists on the [main branch](https://pandas.p
ydata.org/docs/dev/getting_started/install.html#installing-the-development-version
-of-pandas) of pandas.\r\n\r\n\r\n### Reproducible Example\r\n\r\npython\r\nimp
ort pandas as pd\r\npd.json_normalize(\r\n    data=dict(x=[1, 2], y=[]),\r\n    re
cord_path='x',\r\n    meta=[[ 'y', 'yy']])\r\n\r\npython-traceback\r\nTypeError: list indices must be integers or slices, not str\r\n\r\n\r\n### Iss
ue Description\r\n\r\nHello. \r\njson_normalize is great function. it has feature
that allows to coerce errors when provided json structure has some missing keys.\r
\r\nI wish to improve the coercion algorithm.\r\n\r\n\r\n### Expected Behavior\r\n\r
nabove code should result with {x:[1,2],y:[pd.nan,pd.nan]}.\r\nnow it just throws
TypeError.\r\n\r\nI have already made a fix for that, but would be happy to discus
some other improvements before proposing a PR.\r\n\r\n\r\n\r\n### Installed Versions
\r\n\r\n<details>\r\n\r\nINSTALLED VERSIONS\r\n-----\r\ncommit
: bdc79c146c2e32f2cab629be240f01658cfb6cc2\r\npython : 3.11.3.fina
l.0\r\npython-bits : 64\r\nOS : Darwin\r\nOS-release
: 21.4.0\r\nVersion : Darwin Kernel Version 21.4.0: Mon Feb 21 20:3
6:53 PST 2022; root:xnu-8020.101.4~2/RELEASE_ARM64_T8010\r\nmachine
: arm64\r\nprocessor : arm\r\nbyteorder : little\r\nLC_ALL
: None\r\nLANG : None\r\nLOCALE : en_US.UTF-8\r\n
\r\npandas : 2.2.1\r\nnumpy : 1.26.3\r\npytz
: 2023.3.post1\r\ndateutil : 2.8.2\r\nsetuptools : 68.2.2
\r\npip : 24.0\r\nCython : None\r\npytest
: None\r\nhypothesis : None\r\nsphinx : None\r\nblosc
: None\r\nfeather : None\r\nxlswriter : None\r\nxlsxwriter
: 5.0.1\r\nhtml5lib : 1.1\r\npymysql : No
ne\r\npsycopg2 : 2.9.9\r\njinja2 : 3.1.2\r\nIPython
: 8.19.0\r\npandas_datareader : None\r\nadbc-driver-postgresql: None\r\nadbc-d
river-sqlite : None\r\nbs4 : 4.12.2\r\nbottleneck
: None\r\ndataframe-api-compat : None\r\nfastparquet : None\r\nfsspec
```

```

: 2023.12.2\r\ngcsfs : None\r\nmatplotlib : 3.8.2\r\nnumba
: 0.58.1\r\nnumexpr : None\r\nnodfpy
: None\r\nopenpyxl : None\r\npandas_gbq : None\r\npyarrow
: None\r\npyreadstat : None\r\npython-calamine : None\r\npyxlsb
: None\r\ns3fs : None\r\nscipy : 1.11.4\r\nsqlalchemy
: None\r\ntables : None\r\nxtabulate :
0.9.0\r\nxarray : None\r\nxlrd : None\r\nzstandard
: 0.22.0\r\nntzdata : 2023.4\r\nnqtpy : None\r\nipyqt
5 : None\r\n\r\n</details>\r\n",
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57810/timeline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57798',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57798/labels{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57798/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57798/events',
'html_url': 'https://github.com/pandas-dev/pandas/issues/57798',
'id': 2177827298,
'node_id': 'I_kwDOAA0YD86BzwHi',
'number': 57798,
'title': 'ENH: DataFrame argument `columns` should accept dict/iterable type',
'user': {'login': '5j9',
'id': 1055747,
'node_id': 'MDQ6VXNlcjEwNTU3NDc=',
'avatar_url': 'https://avatars.githubusercontent.com/u/1055747?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/5j9',
'html_url': 'https://github.com/5j9',
'followers_url': 'https://api.github.com/users/5j9/followers',
'following_url': 'https://api.github.com/users/5j9/following{/other_user}',
'gists_url': 'https://api.github.com/users/5j9/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/5j9/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/5j9/subscriptions',
'organizations_url': 'https://api.github.com/users/5j9/orgs',
'repos_url': 'https://api.github.com/users/5j9/repos',
'events_url': 'https://api.github.com/users/5j9/events{/privacy}',
'received_events_url': 'https://api.github.com/users/5j9/received_events',
'type': 'User',
'site_admin': False},
'labels': [{'id': 76812,

```

[illegible]

```

st[str]" is incompatible with protocol "SequenceNotStr[Unknown]"\r\n  \xa0\xa0\xa0
\xa0\xa0\xa0"index" is an incompatible type\r\n  \xa0\xa0\xa0\xa0\xa0\xa0\xa0
type "(__value: str, __start: SupportsIndex = 0, __stop: SupportsIndex = sys.maxsiz
e, /) -> int" cannot be assigned to type "(value: Any, /, start: int = 0, stop: in
t = ...) -> int"\r\n    ... (reportArgumentType)\r\n1 error, 0 warnings, 0 informa
tions\r\n```\r\n-->\r\n### Additional Context\r\n\r\n\r\nNo response_',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5779
8/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57798/tim
eline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57794',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57794/label
s{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57794/com
ments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57794/event
s',
'html_url': 'https://github.com/pandas-dev/pandas/issues/57794',
'id': 2177595898,
'node_id': 'I_kwDOAA0YD86By3n6',
'number': 57794,
'title': 'DEPR: rename `startingMonth` to `starting_month` (argument in BQuarter
Begin)',
'user': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_us
er}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{/rep
o}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_even
ts',
'type': 'User',
'site_admin': False},

```

```

'labels': [{ 'id': 53181044,
  'node_id': 'MDU6TGFiZWw1MzE4MTA0NA==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Frequency',
  'name': 'Frequency',
  'color': '0052cc',
  'default': False,
  'description': 'DateOffsets'},
{ 'id': 87485152,
  'node_id': 'MDU6TGFiZWw4NzQ4NTE1Mg==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Deprecate',
  'name': 'Deprecate',
  'color': '5319e7',
  'default': False,
  'description': 'Functionality to remove in pandas'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 0,
'created_at': '2024-03-10T08:13:29Z',
'updated_at': '2024-03-10T08:14:21Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'body': "Renamings should be done with care, but this one strikes me as especial
ly odd\r\n\r\n``python\r\npandas.tseries.offsets.QuarterBegin(startingMonth=1)\r
\r\n``\r\nIt looks very odd in Python to have a camelcase argument name...I thought
this was probably a typo in the docs when I saw it, but no, it runs\r\n\r\nOK with
deprecating in favour of `starting_month`?\r\n\r\nI think this is the only place i
n pandas where this happens:\r\n\r\n`console\r\n$ git grep -E ' [a-z]+[A-Z][a-z]+: '
pandas\r\npandas/_libs/tslibs/offsets.pyi:         self, n: int = ..., normalize: b
ool = ..., startingMonth: int | None = ... \r\npandas/_libs/tslibs/offsets.pyi:
startingMonth: int = ..., \r\npandas/_libs/tslibs/offsets.pyi:         startingMont
h: int = ..., \r\n$ git grep -E ' [a-z]+[A-Z][a-z]+ : ' pandas\r\npandas/_libs/tsli
bs/offsets.pyx:         startingMonth : int, default 3\r\npandas/_libs/tslibs/offsets.p
yx:         startingMonth : int, default 3\r\npandas/_libs/tslibs/offsets.pyx:     start
ingMonth : int, default 3\r\npandas/_libs/tslibs/offsets.pyx:         startingMonth : i
nt, default 3\r\npandas/_libs/tslibs/offsets.pyx:         startingMonth : int {1, 2,
... 12}, default 1\r\npandas/_libs/tslibs/offsets.pyx:         startingMonth : int {1,
2, ..., 12}, default 1\r\n\r\n",
'reactions': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5779
4/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57794/tim
eline',
'performed_via_github_app': None,
'state_reason': None},

```



```

{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57792',
 'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
 'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57792/labels{/name}',
 'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57792/comments',
 'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57792/events',
 'html_url': 'https://github.com/pandas-dev/pandas/issues/57792',
 'id': 2177448416,
 'node_id': 'I_kwDOAA0YD86ByTng',
 'number': 57792,
 'title': 'BUG: read_csv inconsistent behavior',
 'user': {'login': 'fgr1986',
 'id': 10627849,
 'node_id': 'MDQ6VXNlcjEwNjI3ODQ5',
 'avatar_url': 'https://avatars.githubusercontent.com/u/10627849?v=4',
 'gravatar_id': '',
 'url': 'https://api.github.com/users/fgr1986',
 'html_url': 'https://github.com/fgr1986',
 'followers_url': 'https://api.github.com/users/fgr1986/followers',
 'following_url': 'https://api.github.com/users/fgr1986/following{/other_user}',
 'gists_url': 'https://api.github.com/users/fgr1986/gists{/gist_id}',
 'starred_url': 'https://api.github.com/users/fgr1986/starred{/owner}/{/repo}',
 'subscriptions_url': 'https://api.github.com/users/fgr1986/subscriptions',
 'organizations_url': 'https://api.github.com/users/fgr1986/orgs',
 'repos_url': 'https://api.github.com/users/fgr1986/repos',
 'events_url': 'https://api.github.com/users/fgr1986/events{/privacy}',
 'received_events_url': 'https://api.github.com/users/fgr1986/received_events',
 'type': 'User',
 'site_admin': False},
 'labels': [{ 'id': 134699,
 'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
 'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
 'name': 'Docs',
 'color': '3465A4',
 'default': False,
 'description': None},
 { 'id': 47229171,
 'node_id': 'MDU6TGFiZWw0NzIyOTE3MQ==',
 'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/IO%20CSV',
 'name': 'IO CSV',
 'color': '5319e7',
 'default': False,
 'description': 'read_csv, to_csv'},
 { 'id': 717120670,
 'node_id': 'MDU6TGFiZWw3MTcxMjA2NzA=',
 'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/good%20first%20issue',
 'name': 'good first issue',
 'color': '0e8a16',
 'default': True,
 'description': None}],
 'state': 'open',
 'locked': False,
 'assignee': None,

```

```

'assignees': [],
'milestone': None,
'comments': 3,
'created_at': '2024-03-09T23:44:19Z',
'updated_at': '2024-03-13T21:32:35Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'body': '### Pandas version checks\n\n- [X] I have checked that this issue has not already been reported.\n\n- [X] I have confirmed this bug exists on the [latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\n\n- [X] I have confirmed this bug exists on the [main branch](https://pandas.pydata.org/docs/dev/getting_started/install.html#installing-the-development-version-of-pandas) of pandas.\n\n\n### Reproducible Example\n\n```\npython\nimport pandas as pd\nfrom io import StringIO\n\ncsv_data = """,1.992422433360555e-06,0,1.992422433374335e-06,0,0,0,0.01867750880504259,0,0.8,0,0.002722413402762714,0,0.002722413278797551\n1\n\n3e-12,1.992418292306889e-06,3e-12,1.992418292319521e-06,3e-12,0,3e-12,0.01867750887016915,3e-12,0.8,3e-12,0.002722413686257528,3e-12,0.002722413562292257\n\n9.000000000000001e-12,1.992416874989637e-06,9.000000000000001e-12,1.992416875007223e-06,9.000000000000001e-12,0,9.000000000000001e-12,0.01867750900743637,9.000000000000001e-12,0.8,9.000000000000001e-12,0.00272241426436429,9.000000000000001e-12,0.002722414140398993"""\n\n# Use StringIO to create a file-like object from the string\n\ncsv_data_io1 = StringIO(csv_data)\n\ncsv_data_io2 = StringIO(csv_data)\n\n# Use pandas.read_csv to read from the file-like object\n\ndf1 = pd.read_csv(csv_data_io1)\n\ndf2 = pd.read_csv(csv_data_io2, header=None)\n\n# assertions\n\nprint(df1.columns.tolist())\n\nprint(df2.iloc[0].tolist())\n\n```\n\n\n### Issue Description\n\n\nColumns are incorrectly read from CSV\n\n\n### Expected Behavior\n\n\ndf1.columns.tolist() being the same as df2.iloc[0].tolist()\n\n\ni.e.: [0.0, 1.992422433360555e-06, 0.0, 1.992422433374335e-06, 0.0, 0.0, 0.0, 0.0186775088050425, 0.0, 0.8, 0.0, 0.0027224134027627, 0.0, 0.0027224132787975]\n\n\nHowever, it is:\n\n['\\0\\', '\\1.992422433360555e-06\\', '\\0.1\\', '\\1.992422433374335e-06\\', '\\0.2\\', '\\0.3\\', '\\0.4\\', '\\0.01867750880504259\\', '\\0.5\\', '\\0.8\\', '\\0.6\\', '\\0.002722413402762714\\', '\\0.7\\', '\\0.002722413278797551\\']\n\n\nwhich is not any row in the file\n\n\n### Installed Versions\n\n<details>\n\n\nINSTALLED VERSIONS\n\n-----\n\n-----\n\ncommit : bdc79c146c2e32f2cab629be240f01658cfb6cc2\n\npython : 3.9.7.final.0\n\npython-bits : 64\n\nOS : Linux\n\nOS-release : 5.15.133.1-microsoft-standard-WSL2\n\nVersion : #1 SMP Thu Oct 5 21:02:42 UTC 2023\n\nmachine : x86_64\n\nprocessor : x86_64\n\nbyteorder : little\n\nLC_ALL : None\n\nLANG : C.UTF-8\n\nLOCALE : en_US.UTF-8\n\npandas : 2.2.1\n\nnumpy : 1.26.4\n\npytz : 2023.3.post1\n\ndateutil : 2.8.2\n\nsetuptools : 68.2.2\n\npip : 23.3.1\n\nCython : None\n\npytest : 7.4.1\n\nhypothesis : None\n\nsphinx : 5.0.2\n\nblosc : None\n\nfeather : None\n\nxlswriter : None\n\nxlml.etree : 4.9.3\n\nhtml5lib : None\n\npymysql : None\n\npsycopg2 : None\n\njinja2 : 3.1.3\n\nIPython : 8.15.0\n\npandas_datareader : None\n\nnadb-driver-postgresql : None\n\nnadb-driver-sqlite : None\n\nnbs4 : 4.12.2\n\nbottleneck : 1.3.7\n\ndataframe-api-compat : None\n\nfastparquet : None\n\nfsspec : None\n\nngcsfs : None\n\nmatplotlib : 3.8.0\n\nnumba : None\n\nnumexpr : 2.8.7\n\nnodfpy : None\n\nnopenpyxl : 3.1.2\n\npandas_gbq : None\n\npyarrow : None\n\npyreadstat : None\n\npython-calamine : None\n\npyxlsb : None\n\nns3fs : None\n\nnscipy : 1.11.4\n\nsqlalchemy : None\n\nntables : None\n\nntabulate :

```

```

None\r\nxarray          : None\r\nxlrd          : None\r\nzstandard
: None\r\nntzdata        : 2023.3\r\nnqtpy          : None\r\nnpq5
: None\r\n\r\n</details>\r\n',
  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5779
2/reactions',
    'total_count': 1,
    '+1': 1,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57792/tim
eline',
    'performed_via_github_app': None,
    'state_reason': None},
  {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790/label
s{name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790/com
ments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790/event
s',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57790',
    'id': 2177354563,
    'node_id': 'PR_kwDOAA0YD85pJsZG',
    'number': 57790,
    'title': 'Updated the pandas.DatetimeIndex.day_name and pandas.DatetimeIndex.mon
th_name docstring',
    'user': {'login': 'pmhatre1',
      'id': 58086649,
      'node_id': 'MDQ6VXNlcjU4MDg2NjQ5',
      'avatar_url': 'https://avatars.githubusercontent.com/u/58086649?v=4',
      'gravatar_id': '',
      'url': 'https://api.github.com/users/pmhatre1',
      'html_url': 'https://github.com/pmhatre1',
      'followers_url': 'https://api.github.com/users/pmhatre1/followers',
      'following_url': 'https://api.github.com/users/pmhatre1/following{other_use
r}',
      'gists_url': 'https://api.github.com/users/pmhatre1/gists{/gist_id}',
      'starred_url': 'https://api.github.com/users/pmhatre1/starred{/owner}/{/repo}',
      'subscriptions_url': 'https://api.github.com/users/pmhatre1/subscriptions',
      'organizations_url': 'https://api.github.com/users/pmhatre1/orgs',
      'repos_url': 'https://api.github.com/users/pmhatre1/repos',
      'events_url': 'https://api.github.com/users/pmhatre1/events{/privacy}',
      'received_events_url': 'https://api.github.com/users/pmhatre1/received_events',
      'type': 'User',
      'site_admin': False},
    'labels': [],
    'state': 'open',
    'locked': False,
    'assignee': None,
    'assignees': [],

```

```

'milestone': None,
'comments': 1,
'created_at': '2024-03-09T18:50:12Z',
'updated_at': '2024-03-10T20:04:39Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57790',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57790',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57790.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57790.patch',
'merged_at': None},
'body': '- closes #57111 partially\r\n- All [code checks passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commit).\r\n\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57790/timeline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788/labels{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788/events',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57788',
'id': 2177028218,
'node_id': 'PR_kwDOAA0YD85pIpxe',
'number': 57788,
'title': 'API: Revert 57042 - MultiIndex.names|codes|levels returns tuples',
'user': {'login': 'rhshadrach',
'id': 45562402,
'node_id': 'MDQ6VXNlcjQ1NTYyNDAY',
'avatar_url': 'https://avatars.githubusercontent.com/u/45562402?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/rhshadrach',
'html_url': 'https://github.com/rhshadrach',
'followers_url': 'https://api.github.com/users/rhshadrach/followers',
'following_url': 'https://api.github.com/users/rhshadrach/following{/other_user}'},
'gists_url': 'https://api.github.com/users/rhshadrach/gists{/gist_id}',

```

```

    'starred_url': 'https://api.github.com/users/rhshadrach/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/rhshadrach/subscriptions',
    'organizations_url': 'https://api.github.com/users/rhshadrach/orgs',
    'repos_url': 'https://api.github.com/users/rhshadrach/repos',
    'events_url': 'https://api.github.com/users/rhshadrach/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/rhshadrach/received_events',
    'type': 'User',
    'site_admin': False,
    'labels': [],
    'state': 'open',
    'locked': False,
    'assignee': None,
    'assignees': [],
    'milestone': None,
    'comments': 3,
    'created_at': '2024-03-09T03:18:34Z',
    'updated_at': '2024-03-12T17:27:52Z',
    'closed_at': None,
    'author_association': 'MEMBER',
    'active_lock_reason': None,
    'draft': True,
    'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57788',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57788',
    'diff_url': 'https://github.com/pandas-dev/pandas/pull/57788.diff',
    'patch_url': 'https://github.com/pandas-dev/pandas/pull/57788.patch',
    'merged_at': None},
    'body': '- [ ] closes #xxxx (Replace xxxx with the GitHub issue number)\r\n- [ ] [Tests added and passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if fixing a bug or adding a new feature\r\n- [ ] All [code checks passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commit).\r\n- [ ] Added [type annotations](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#type-hints) to new arguments/methods/functions.\r\n- [ ] Added an entry in the latest `doc/source/whatsnew/vX.X.X.rst` file if fixing a bug or adding a new feature.\r\n\r\nManual revert of #57042\r\n\r\nCloses #57607\r\n\r\nThe behavior described in that issue seems quite undesirable, especially for a breaking change.\r\n\r\nncc @mroeschke ',
    'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57788/timeline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57786',

```

```

'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57786/labels{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57786/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57786/events',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57786',
'id': 2176910155,
'node_id': 'PR_kwDOAA0YD85pIQm1',
'number': 57786,
'title': 'PERF: Allow ensure_index_from_sequence to return RangeIndex',
'user': {'login': 'mroeschke',
'id': 10647082,
'node_id': 'MDQ6VXNlcjEwNjQ3MDgy',
'avatar_url': 'https://avatars.githubusercontent.com/u/10647082?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/mroeschke',
'html_url': 'https://github.com/mroeschke',
'followers_url': 'https://api.github.com/users/mroeschke/followers',
'following_url': 'https://api.github.com/users/mroeschke/following{/other_user}',
'gists_url': 'https://api.github.com/users/mroeschke/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/mroeschke/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/mroeschke/subscriptions',
'organizations_url': 'https://api.github.com/users/mroeschke/orgs',
'repos_url': 'https://api.github.com/users/mroeschke/repos',
'events_url': 'https://api.github.com/users/mroeschke/events{/privacy}',
'received_events_url': 'https://api.github.com/users/mroeschke/received_events'},
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 8935311,
'node_id': 'MDU6TGFiZWw4OTM1MzEx',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Performance',
'name': 'Performance',
'color': 'a10c02',
'default': False,
'description': 'Memory or execution speed performance'},
{ 'id': 1218227310,
'node_id': 'MDU6TGFiZWwzMjE4MjI3MzEw',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Index',
'name': 'Index',
'color': 'e99695',
'default': False,
'description': 'Related to the Index class or subclasses'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': {'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102',
'html_url': 'https://github.com/pandas-dev/pandas/milestone/102',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102/labels',
'id': 9021906,

```

```

'node_id': 'MI_kwDOAA0YD84AianS',
'number': 102,
'title': '3.0',
'description': '',
'creator': {'login': 'jbrockmendel',
'id': 8078968,
'node_id': 'MDQ6VXNlcjgwNzg5Njg=',
'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/jbrockmendel',
'html_url': 'https://github.com/jbrockmendel',
'followers_url': 'https://api.github.com/users/jbrockmendel/followers',
'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_u
ser}',
'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}/{rep
o}',
'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscription
s',
'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',
'repos_url': 'https://api.github.com/users/jbrockmendel/repos',
'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',
'received_events_url': 'https://api.github.com/users/jbrockmendel/received_eve
nts',
'type': 'User',
'site_admin': False},
'open_issues': 53,
'closed_issues': 402,
'state': 'open',
'created_at': '2023-02-08T18:16:49Z',
'updated_at': '2024-03-14T02:02:54Z',
'due_on': '2024-04-01T07:00:00Z',
'closed_at': None},
'comments': 0,
'created_at': '2024-03-08T23:15:11Z',
'updated_at': '2024-03-13T23:33:25Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
786',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57786',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57786.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57786.patch',
'merged_at': None},
'body': 'Discovered in https://github.com/pandas-dev/pandas/pull/57441\r\n\r\nBu
ilds on https://github.com/pandas-dev/pandas/pull/57752\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5778
6/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,

```

```

    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57786/timeline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781',
     'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
     'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781/labels{/name}',
     'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781/comments',
     'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781/events',
     'html_url': 'https://github.com/pandas-dev/pandas/issues/57781',
     'id': 2176500721,
     'node_id': 'I_kwDOAA0YD86BusPx',
     'number': 57781,
     'title': 'BUG: timedelta.round fails to round to nearest day',
     'user': {'login': 'jbogaardt',
              'id': 19401133,
              'node_id': 'MDQ6VXNlcjE5NDAxMTMz',
              'avatar_url': 'https://avatars.githubusercontent.com/u/19401133?v=4',
              'gravatar_id': '',
              'url': 'https://api.github.com/users/jbogaardt',
              'html_url': 'https://github.com/jbogaardt',
              'followers_url': 'https://api.github.com/users/jbogaardt/followers',
              'following_url': 'https://api.github.com/users/jbogaardt/following{/other_user}',
              'gists_url': 'https://api.github.com/users/jbogaardt/gists{/gist_id}',
              'starred_url': 'https://api.github.com/users/jbogaardt/starred{/owner}/{/repo}',
              'subscriptions_url': 'https://api.github.com/users/jbogaardt/subscriptions',
              'organizations_url': 'https://api.github.com/users/jbogaardt/orgs',
              'repos_url': 'https://api.github.com/users/jbogaardt/repos',
              'events_url': 'https://api.github.com/users/jbogaardt/events{/privacy}',
              'received_events_url': 'https://api.github.com/users/jbogaardt/received_events',
              'type': 'User',
              'site_admin': False},
     'labels': [{'id': 76811,
                  'node_id': 'MDU6TGFiZWw3NjgxMQ==',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
                  'name': 'Bug',
                  'color': 'e10c02',
                  'default': False,
                  'description': None},
                 {'id': 1954720290,
                  'node_id': 'MDU6TGFiZWw3OTU0NzIwMjkx',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Needs%20Triage',
                  'name': 'Needs Triage',
                  'color': '0052cc',
                  'default': False,
                  'description': 'Issue that has not been reviewed by a pandas team member'}],
     'state': 'open',
     'locked': False,

```



```
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781/reactions',
'total_count': 0,
'+1': 0,
```

```

'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57781/timeline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57777',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57777/labels{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57777/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57777/events',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57777',
'id': 2175400926,
'node_id': 'PR_kwDOAA0YD85pDEsx',
'number': 57777,
'title': 'Clarify doc for converting timestamps to epoch',
'user': {'login': 'tqa236',
'id': 25203655,
'node_id': 'MDQ6VXNlcjI1MjAzNjU1',
'avatar_url': 'https://avatars.githubusercontent.com/u/25203655?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/tqa236',
'html_url': 'https://github.com/tqa236',
'followers_url': 'https://api.github.com/users/tqa236/followers',
'following_url': 'https://api.github.com/users/tqa236/following{/other_user}',
'gists_url': 'https://api.github.com/users/tqa236/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/tqa236/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/tqa236/subscriptions',
'organizations_url': 'https://api.github.com/users/tqa236/orgs',
'repos_url': 'https://api.github.com/users/tqa236/repos',
'events_url': 'https://api.github.com/users/tqa236/events{/privacy}',
'received_events_url': 'https://api.github.com/users/tqa236/received_events',
'type': 'User',
'site_admin': False},
'labels': [{'id': 134699,
'node_id': 'MDU6TGFiZWwzMzQ2OTk=',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Docs',
'name': 'Docs',
'color': '3465A4',
'default': False,
'description': None}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 0,
'created_at': '2024-03-08T07:00:50Z',

```

```

'updated_at': '2024-03-08T17:53:23Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/5777',
'html_url': 'https://github.com/pandas-dev/pandas/pull/5777',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/5777.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/5777.patch',
'merged_at': None},
'body': '- [x] closes #56343 (Replace xxxx with the GitHub issue number)\r\n- [ ] [Tests added and passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if fixing a bug or adding a new feature\r\n- [ ] All [code checks passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commit).\r\n- [ ] Added [type annotations](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#type-hints) to new arguments/methods/functions.\r\n- [ ] Added an entry in the latest `doc/source/whatsnew/vX.X.X.rst` file if fixing a bug or adding a new feature.\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5777/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5777/timeline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773/labels/{name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773/events',
'html_url': 'https://github.com/pandas-dev/pandas/issues/57773',
'id': 2174962145,
'node_id': 'I_kwDOAA0YD86Bo0nh',
'number': 57773,
'title': 'BUG: to_parquet on column containing StringArray objects fails',
'user': {'login': 'paavanb',
'id': 1556995,
'node_id': 'MDQ6VXNlcjE1NTY5OTU=',
'avatar_url': 'https://avatars.githubusercontent.com/u/1556995?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/paavanb',
'html_url': 'https://github.com/paavanb',
'followers_url': 'https://api.github.com/users/paavanb/followers',

```

```

'following_url': 'https://api.github.com/users/paavanb/following{/other_user}',
'gists_url': 'https://api.github.com/users/paavanb/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/paavanb/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/paavanb/subscriptions',
'organizations_url': 'https://api.github.com/users/paavanb/orgs',
'repos_url': 'https://api.github.com/users/paavanb/repos',
'events_url': 'https://api.github.com/users/paavanb/events{/privacy}',
'received_events_url': 'https://api.github.com/users/paavanb/received_events',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 76811,
  'node_id': 'MDU6TGFiZWw3NjgxMQ==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
  'name': 'Bug',
  'color': 'e10c02',
  'default': False,
  'description': None},
{ 'id': 1954720290,
  'node_id': 'MDU6TGFiZWw0OTU0NzIwMjkw',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Needs%20Triage',
  'name': 'Needs Triage',
  'color': '0052cc',
  'default': False,
  'description': 'Issue that has not been reviewed by a pandas team member'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-07T22:50:44Z',
'updated_at': '2024-03-08T16:50:45Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'body': '### Pandas version checks\r\n\r\n\r\n- [X] I have checked that this issue h
as not already been reported.\r\n\r\n\r\n- [X] I have confirmed this bug exists on the
[latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\r
\r\n\r\n\r\n- [X] I have confirmed this bug exists on the [main branch](https://pandas.p
ydata.org/docs/dev/getting_started/install.html#installing-the-development-version
-of-pandas) of pandas.\r\n\r\n\r\n\r\n### Reproducible Example\r\n\r\n\r\n\r\n``python\r\nnimp
ort pandas as pd\r\n\r\n\r\nrecords = [\r\n\r\n    {"name": "A", "project": "foo"}, \r\n\r\n
{"name": "A", "project": "bar"}, \r\n\r\n    {"name": "A", "project": pd.NA},\r\n\r\n]\r\n\r
df = pd.DataFrame.from_records(records).astype({"project": "string"})\r\nnagg_df =
df.groupby("name").agg({"project": "unique"})\r\n\r\n\r\nnagg_df.to_parquet("test.parqu
et")\r\n\r\n``\r\n\r\n\r\n\r\n\r\n### Issue Description\r\n\r\n\r\n\r\nIt seems that when aggregating
on a ``"string"`` type column (potentially nullable), the resulting dataframe uses ``
StringArray`` to represent the aggregated value (here ``project``). When attempting t
o write to parquet, this fails with the following error:\r\n\r\n\r\n\r\n``\r\n\r\npyarrow.li
b.ArrowInvalid: ("Could not convert <StringArray>\n['foo', 'bar', <NA>]\n\nLen
gth: 3, dtype: string with type StringArray: did not recognize Python value type w
hen inferring an Arrow data type", 'Conversion failed for column project with typ
e object')\r\n\r\n``\r\n\r\n\r\n\r\nI'm not sure why this is happening, because ``StringArra
y`` clearly implements the ``__arrow_array__`` [protocol](https://arrow.apache.org/do
cs/python/extending_types.html) that should be allowing this conversion. This also
happens if the string is of type ``string[pyarrow]``.\r\n\r\n\r\n\r\nThe workaround I've

```

been using so far is to re-type the column as `object` prior to the `groupby`. This results in values of type `np.ndarray`, which are parquet-able. But this is suboptimal because we use `read_parquet` upstream that types the column as `"string"`, which means that we have to remember to retype it every time we do this aggregation, else it will fail downstream when writing to parquet. It seems like this should work with the column typed as `"string"`.
 Expected Behavior
 It should successfully write the parquet file without failures.
 Installed Versions
 <details>
 INSTALLED VERSIONS

 commit : bdc79c146c2e32f2cab629be240f01658cfb6cc2
 python : 3.9.17.final.0
 python-bits : 64
 OS : Darwin
 OS-release : 23.3.0
 Version : Darwin Kernel Version 23.3.0: Wed Dec 20 21:30:44 PST 2023; root:xnu-10002.81.5~7/RELEASE_ARM64_T6000
 machine : arm64
 processor : arm
 byteorder : little
 LC_ALL : None
 LANG : en_US.UTF-8
 LOCALE : en_US.UTF-8
 pandas : 2.2.1
 numpy : 1.26.4
 npytz : 2024.1
 dateutil : 2.9.0.post0
 setuptools : 58.1.0
 pip : 23.0.1
 Cython : None
 pytest : None
 hypothesis : None
 sphinx : None
 blosc : None
 feather : None
 nxlsxwriter : None
 lxml.etree : None
 html5lib : None
 pymysql : None
 No : No
 npsycopg2 : None
 njinja2 : None
 IPython : None
 8.18.1
 pandas_datareader : None
 nadbc-driver-postgresql : None
 nadbc-driver-sqlite : None
 nbs4 : None
 bottleneck : None
 dataframe-api-compat : None
 fastparquet : None
 fsspec : None
 gcfsfs : None
 matplotlib : None
 numba : None
 numexpr : None
 nodfpy : None
 openpyxl : None
 pandas_gbq : None
 pyarrow : 15.0.0
 pyreadstat : None
 python-calamine : None
 pyxlsb : None
 ns3fs : None
 nscipy : None
 sqlalchemy : None
 ntables : None
 ntabulate : None
 nxarray : None
 nxlrd : None
 nzstandard : None
 ntzdata : None
 2024.1
 nqtpy : None
 npyqt5 : None
 </details>
 'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773/reactions',
 'total_count': 0,
 '+1': 0,
 '-1': 0,
 'laugh': 0,
 'hooray': 0,
 'confused': 0,
 'heart': 0,
 'rocket': 0,
 'eyes': 0},
 'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57773/timeline',
 'performed_via_github_app': None,
 'state_reason': None},
 {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57771',
 'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
 'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57771/labels/{name}',
 'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57771/comments',
 'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57771/events',

```

'html_url': 'https://github.com/pandas-dev/pandas/pull/57771',
'id': 2174815824,
'node_id': 'PR_kwDOAA0YD85pBHBH',
'number': 57771,
'title': 'DOC: Fix PR01 and SA01 errors in pandas.ExcelFile',
'user': {'login': 'iridiium',
'id': 60609879,
'node_id': 'MDQ6VXNlcjYwNjA5ODc5',
'avatar_url': 'https://avatars.githubusercontent.com/u/60609879?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/iridiium',
'html_url': 'https://github.com/iridiium',
'followers_url': 'https://api.github.com/users/iridiium/followers',
'following_url': 'https://api.github.com/users/iridiium/following{/other_use
r}',
'gists_url': 'https://api.github.com/users/iridiium/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/iridiium/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/iridiium/subscriptions',
'organizations_url': 'https://api.github.com/users/iridiium/orgs',
'repos_url': 'https://api.github.com/users/iridiium/repos',
'events_url': 'https://api.github.com/users/iridiium/events{/privacy}',
'received_events_url': 'https://api.github.com/users/iridiium/received_events',
'type': 'User',
'site_admin': False},
'labels': [],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 2,
'created_at': '2024-03-07T21:17:12Z',
'updated_at': '2024-03-13T00:10:24Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
771',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57771',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57771.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57771.patch',
'merged_at': None},
'body': '- [x] xref #57438 \r\n- [x] xref #57417 \r\n- [ ] [Tests added and pass
ed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.ht
ml#writing-tests) if fixing a bug or adding a new feature\r\n- [ ] All [code check
s passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codeb
ase.html#pre-commit).\r\n- [ ] Added [type annotations](https://pandas.pydata.org/
pandas-docs/dev/development/contributing_codebase.html#type-hints) to new argument
s/methods/functions.\r\n- [ ] Added an entry in the latest `doc/source/whatsnew/v
X.X.X.rst` file if fixing a bug or adding a new feature.\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5777
1/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,

```

```

    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57771/timeline',
    'performed_via_github_app': None,
    'state_reason': None},
    {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765',
     'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
     'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765/labels{/name}',
     'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765/comments',
     'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765/events',
     'html_url': 'https://github.com/pandas-dev/pandas/issues/57765',
     'id': 2174333453,
     'node_id': 'I_kwDOAA0YD86BmbIN',
     'number': 57765,
     'title': 'BUG: Setting a numpy array as a column in Pandas uses only the first column of the array. ',
     'user': {'login': 'kgourgou',
              'id': 2594318,
              'node_id': 'MDQ6VXNlcjI1OTQzMTg=',
              'avatar_url': 'https://avatars.githubusercontent.com/u/2594318?v=4',
              'gravatar_id': '',
              'url': 'https://api.github.com/users/kgourgou',
              'html_url': 'https://github.com/kgourgou',
              'followers_url': 'https://api.github.com/users/kgourgou/followers',
              'following_url': 'https://api.github.com/users/kgourgou/following{/other_user}',
              'gists_url': 'https://api.github.com/users/kgourgou/gists{/gist_id}',
              'starred_url': 'https://api.github.com/users/kgourgou/starred{/owner}/{/repo}',
              'subscriptions_url': 'https://api.github.com/users/kgourgou/subscriptions',
              'organizations_url': 'https://api.github.com/users/kgourgou/orgs',
              'repos_url': 'https://api.github.com/users/kgourgou/repos',
              'events_url': 'https://api.github.com/users/kgourgou/events{/privacy}',
              'received_events_url': 'https://api.github.com/users/kgourgou/received_events',
              'type': 'User',
              'site_admin': False},
     'labels': [{'id': 76811,
                  'node_id': 'MDU6TGFiZWw3Njg5MQ==',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
                  'name': 'Bug',
                  'color': 'e10c02',
                  'default': False,
                  'description': None},
                 {'id': 2822098,
                  'node_id': 'MDU6TGFiZWw0ODIyMDk4',
                  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Indexing',
                  'name': 'Indexing',
                  'color': '0b02e1',
                  'default': False,

```

```
'description': 'Related to indexing on series/frames, not to indexes themselves'},
{'id': 219960758,
'node_id': 'MDU6TGFiZWwyMTk5NjA3NTg=',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Needs%20Discussion',
'name': 'Needs Discussion',
'color': '207de5',
'default': False,
'description': 'Requires discussion from core team before further action'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 2,
'created_at': '2024-03-07T17:02:55Z',
'updated_at': '2024-03-08T17:49:08Z',
'closed_at': None,
'author_association': 'NONE',
'active_lock_reason': None,
'body': "### Pandas version checks\r\n\r\n [X] I have checked that this issue has not already been reported.\r\n\r\n [X] I have confirmed this bug exists on the [latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\r\n\r\n [ ] I have confirmed this bug exists on the [main branch](https://pandas.pydata.org/docs/dev/getting_started/install.html#installing-the-development-version-of-pandas) of pandas.\r\n\r\n\r\n### Reproducible Example\r\n\r\n```python\r\nimport pandas \r\nimport numpy as np \r\nnp = np.array([ \r\n    [1,2,3], \r\n    [1,2,3]\r\n])\r\n\r\nndf = pandas.DataFrame(columns=['n', 'p'])\r\nndf['n'] = np.array([0,0])\r\nndf['p'] = p \r\n\r\n\r\nprint(df)\r\n\r\n\r\n n p\r\n0 0 1\r\n1 0 1\r\n\r\n\r\n### Issue Description\r\n\r\n\r\nPassing the columns as separate arrays, but one of the arrays has the wrong dimensions Nx2 instead of Nx1. In that case, the dataframe column 'p' can be assigned that Nx2 array, but only the first column of the array is actually assigned to 'p'. \r\n\r\n\r\nWhile this is hard to happen by accident, it's not impossible. \r\n\r\n\r\n### Expected Behavior\r\n\r\n\r\nEither store each row of the array to the corresponding row of the dataframe or raise a warning/error for trying to store a NxM array as a column. \r\n\r\n\r\nSee for example: \r\n\r\n\r\n```python\r\nimport pandas \r\nimport numpy as np \r\n\r\nnp = \r\n    [1,2], \r\n    [1,2]\r\n\r\n\r\nndf = pandas.DataFrame(columns=['n', 'p'])\r\nndf['n'] = [0,0]\r\nndf['p'] = p \r\n\r\n\r\nprint(df)\r\n\r\n\r\n n p\r\n0 0 [1, 2]\r\n1 0 [1, 2]\r\n\r\n\r\n\r\n### Installed Versions\r\n\r\n<details>\r\n\r\nINSTALLED VERSIONS\r\n-----\r\ncommit : bdc79c146c2e32f2cab629be240f01658cfb6cc2\r\npython : 3.10.12.final.0\r\npython-bits : 64\r\nOS : Linux\r\nOS-release : 6.1.58+\r\nVersion : #1 SMP PREEMPT_DYNAMIC Sat Nov 18 15:31:17 UTC 2023\r\nmachine : x86_64\r\nprocessor : x86_64\r\nbyteorder : little\r\nLC_ALL : en_US.UTF-8\r\nLANG : en_US.UTF-8\r\nLOCALE : en_US.UTF-8\r\npandas : 2.2.1\r\nnumpy : 1.25.2\r\npytz : 2023.4\r\ndateutil : 2.8.2\r\nsetuptools : 67.7.2\r\npip : 23.1.2\r\nCython : 3.0.8\r\npytest : 7.4.4\r\nhypothesis : None\r\nsphinx : 5.0.2\r\nblosc : None\r\nfeather : None\r\nxlswriter : None\r\nlxml.etree : 4.9.4\r\nhtml5lib : 1.1\r\npymysql : No\r\npsycopg2 : 2.9.9\r\njinia2 : 3.1.3\r\nIPython
```



```

: 7.34.0\r\npandas_datareader      : 0.10.0\r\nnadb-driver-postgresql: None\r\nnadb-driver-sqlite      : None\r\nnbs4      : 4.12.3\r\nnbottleneck
: None\r\nndataframe-api-compat : None\r\nnfastparquet      : None\r\nnfsspec
: 2023.6.0\r\nngcsfs      : 2023.6.0\r\nnmatplotlib      : 3.7.1\r\n\nnumba      : 0.58.1\r\nnnumexpr      : 2.9.0\r\nnodfpy
: None\r\nnopenpyxl      : 3.1.2\r\nnpandas_gbq      : 0.19.2\r\nnpyarrow
: 14.0.2\r\nnpyreadstat      : None\r\nnpython-calamine
: None\r\nnpyxlsb      : None\r\nns3fs      : None\r\nnscipy
: 1.11.4\r\nnsqlalchemy      : 2.0.28\r\nntables      : 3.8.0\r\nntabulate
: 0.9.0\r\nnxarray      : 2023.7.0\r\nnxlrd
: 2.0.1\r\nnzstandard      : None\r\nntzdata      : 2024.1\r\nnqtpy
: None\r\nnpyqt5      : None\r\nn</details>\r\n",
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57765/timeline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764/labels{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764/comments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764/events',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57764',
'id': 2173990499,
'node_id': 'PR_kwDOAA0YD85o-RR0',
'number': 57764,
'title': 'BUG: PyArrow dtypes were not supported in the interchange protocol',
'user': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',

```

```

    'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events',
    'type': 'User',
    'site_admin': False},
  'labels': [{ 'id': 4344388253,
    'node_id': 'LA_kwDOAA0YD88AAAABAvIanQ',
    'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Interchange',
    'name': 'Interchange',
    'color': '130443',
    'default': False,
    'description': 'Dataframe Interchange Protocol'}],
  'state': 'open',
  'locked': False,
  'assignee': None,
  'assignees': [],
  'milestone': None,
  'comments': 0,
  'created_at': '2024-03-07T14:26:05Z',
  'updated_at': '2024-03-14T02:48:15Z',
  'closed_at': None,
  'author_association': 'MEMBER',
  'active_lock_reason': None,
  'draft': False,
  'pull_request': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57764',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57764',
    'diff_url': 'https://github.com/pandas-dev/pandas/pull/57764.diff',
    'patch_url': 'https://github.com/pandas-dev/pandas/pull/57764.patch',
    'merged_at': None},
  'body': 'fixes a few things:\r\n\r\n- closes #57762\r\n- closes #57761\r\n- closes #57664\r\n',
  'reactions': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57764/timeline',
  'performed_via_github_app': None,
  'state_reason': None},
  { 'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57762',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57762/labels{/name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57762/comments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57762/events',
    'html_url': 'https://github.com/pandas-dev/pandas/issues/57762',

```

```

'id': 2173463666,
'node_id': 'I_kwDOAA0YD86BjGxy',
'number': 57762,
'title': "BUG: interchange buffer show bytemask (instead of bitmask) validity fo
r `string[pyarrow]`",
'user': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_us
er}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{/rep
o}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_even
ts',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 76811,
'node_id': 'MDU6TGFiZWw3NjgxmQ==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
'name': 'Bug',
'color': 'e10c02',
'default': False,
'description': None},
{ 'id': 4344388253,
'node_id': 'LA_kwDOAA0YD88AAAABAvIanQ',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Interchange',
'name': 'Interchange',
'color': '130443',
'default': False,
'description': 'Dataframe Interchange Protocol'}],
'state': 'open',
'locked': False,
'assignee': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_us
er}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{/rep
o}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',

```

```
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events',
'type': 'User',
'site_admin': False},
'assignees': [{'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNmY',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events',
'type': 'User',
'site_admin': False}],
'milestone': None,
'comments': 0,
'created_at': '2024-03-07T09:59:13Z',
'updated_at': '2024-03-07T11:07:43Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'body': '### Pandas version checks\r\n\r\n [X] I have checked that this issue has not already been reported.\r\n\r\n [X] I have confirmed this bug exists on the [latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\r\n\r\n [X] I have confirmed this bug exists on the [main branch](https://pandas.pydata.org/docs/dev/getting_started/install.html#installing-the-development-version-of-pandas) of pandas.\r\n\r\n#### Reproducible Example\r\n\r\n```python\r\ndf = pd.DataFrame({"a": ["x", None]}, dtype="large_string[pyarrow]")\r\nprint(df.__dataframe__.get_column_by_name('a').get_buffers())\r\nprint(df['a'].array._pa_array.chunks[0].buffers())\r\nprint(pai.from_dataframe(df))\r\n```\r\n\r\nIssue Description\r\n\r\nThis shows\r\n\r\n`df['data']`: (PandasBuffer({'bufsize': 1, 'ptr': 140019260284416, 'device': 'CPU'}), (<DtypeKind.UINT: 1>, 8, 'C', '=')), `validity`: (PandasBuffer({'bufsize': 2, 'ptr': 94883712583296, 'device': 'CPU'}), (<DtypeKind.BOOL: 20>, 8, 'b', '=')), `offsets`: (PandasBuffer({'bufsize': 24, 'ptr': 94883723016784, 'device': 'CPU'}), (<DtypeKind.INT: 0>, 64, 'l', '=')))\r\n\r\n[<pyarrow.Buffer address=0x7f58c8608040 size=1 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7f58c8608000 size=24 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7f58c8608080 size=1 is_cpu=True is_mutable=True>]\r\n\r\nExpected Behavior\r\n\r\nI think\r\n\r\n`df['data']`: (PandasBuffer({'bufsize': 1, 'ptr': 140019260284416, 'device': 'CPU'}), (<DtypeKind.UINT: 1>, 8, 'C', '=')), `validity`: (PandasBuffer
```

```

r({'\bufsize\': 1, '\ptr\': 94883712583296, '\device\': '\CPU\'}), (<DtypeKind.BOOL: 20>, 1, '\b\ ', '\=')), '\offsets\': (PandasBuffer({'\bufsize\': 24, '\ptr\': 94883723016784, '\device\': '\CPU\'}), (<DtypeKind.INT: 0>, 64, '\l\ ', '\=')))\r\n
[<pyarrow.Buffer address=0x7f58c8608040 size=2 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7f58c8608000 size=24 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7f58c8608080 size=1 is_cpu=True is_mutable=True>]\r\n```\r\n\r\n###
# Installed Versions\r\n\r\n<details>\r\n\r\nINSTALLED VERSIONS\r\n\r\n-----
--\r\nncommit : b033ca94e7ae6e1320c9d65a8163bd0a6049f40a\r\npython
: 3.10.12.final.0\r\npython-bits : 64\r\nOS : Linux\r\n
\r\nOS-release : 5.15.146.1-microsoft-standard-WSL2\r\nVersion
: #1 SMP Thu Jan 11 04:09:03 UTC 2024\r\nnmachine : x86_64\r\nnprocessor
or : x86_64\r\nnbyteorder : little\r\nnLC_ALL
: None\r\nnLANG : C.UTF-8\r\nnLOCALE : en_US.UTF-8\r\n
\r\n\r\nnpandas : 2.1.0.dev0+761.gb033ca94e7\r\nnnumpy
: 1.26.4\r\nnpytz : 2024.1\r\nndateutil : 2.8.2\r\nnset
uptools : 69.1.1\r\nnpip : 24.0\r\nnCython
: 3.0.8\r\nnpytest : 8.0.2\r\nnhypothesis : 6.98.15\r\nnsph
inx : 7.2.6\r\nnblosc : None\r\nnfeather
: None\r\nnxlsxwriter : 3.2.0\r\nnlxml.etree : 5.1.0\r\nnhtml5l
ib : 1.1\r\nnpymysql : 1.4.6\r\nnpsycopg2
: 2.9.9\r\nnjinja2 : 3.1.3\r\nnIPython : 8.22.1\r\nnpandas
_datareader : None\r\nnadb-driver-postgresql: None\r\nnadb-driver-sqlite :
None\r\nnbs4 : 4.12.3\r\nnbottleneck : 1.3.8\r\nnfastpar
quet : 2024.2.0\r\nnfsspec : 2024.2.0\r\nngcsfs
: 2024.2.0\r\nnmatplotlib : 3.8.3\r\nnumba : 0.59.0\r\nnn
umexpr : 2.9.0\r\nnodfpy : None\r\nnopenpyxl
: 3.1.2\r\nnpyarrow : 15.0.0\r\nnpyreadstat : 1.2.6\r\nnpyth
on-calamine : None\r\nnpyxlsb : 1.0.10\r\nns3fs
: 2024.2.0\r\nnscipy : 1.12.0\r\nnsqlalchemy : 2.0.27\r\nn
tables : 3.9.2\r\nntabulate : 0.9.0\r\nnxarray
: 2024.2.0\r\nnxlrd : 2.0.1\r\nnzstandard : 0.22.0\r\nnt
zdata : 2024.1\r\nnqtpy : 2.4.1\r\nnpyqt5
: None\r\n\r\n\r\n</details>\r\n\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5776
2/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57762/tim
eline',
'performed_via_github_app': None,
'state_reason': None},
{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57761',
'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57761/label
s{/name}',
'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57761/com
ments',
'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57761/event
s',

```

```

'html_url': 'https://github.com/pandas-dev/pandas/issues/57761',
'id': 2173441344,
'node_id': 'I_kwDOAA0YD86BjBVA',
'number': 57761,
'title': "BUG: validity of interchange column is incorrectly set to Some for 'large-string[pyarrow]'",
'user': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events'},
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 76811,
'node_id': 'MDU6TGFiZWw3Njg5MQ==',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Bug',
'name': 'Bug',
'color': 'e10c02',
'default': False,
'description': None},
{ 'id': 4344388253,
'node_id': 'LA_kwDOAA0YD88AAAABAvIanQ',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Interchange',
'name': 'Interchange',
'color': '130443',
'default': False,
'description': 'Dataframe Interchange Protocol'}],
'state': 'open',
'locked': False,
'assignee': {'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}'},

```

```

'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events',
'type': 'User',
'site_admin': False},
'assignees': [{'login': 'MarcoGorelli',
'id': 33491632,
'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/MarcoGorelli',
'html_url': 'https://github.com/MarcoGorelli',
'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}',
'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events',
'type': 'User',
'site_admin': False}],
'milestone': None,
'comments': 0,
'created_at': '2024-03-07T09:49:14Z',
'updated_at': '2024-03-07T09:49:14Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'body': '### Pandas version checks\n\n- [X] I have checked that this issue has not already been reported.\n\n- [X] I have confirmed this bug exists on the [latest version](https://pandas.pydata.org/docs/whatsnew/index.html) of pandas.\n\n- [X] I have confirmed this bug exists on the [main branch](https://pandas.pydata.org/docs/dev/getting_started/install.html#installing-the-development-version-of-pandas) of pandas.\n\n\n### Reproducible Example\n\n```\npython\nndf = pd.DataFrame({"a": ["x"]}, dtype="large_string[pyarrow]")\nprint(df.__dataframe__.get_column_by_name("a").get_buffers())\nprint(df["a"].array._pa_array.chunks[0].buffers())\n\n```\n\n### Issue Description\n\nthis outputs\n\n```\n{"data": (PandasBuffer({'bufsize': 1, 'ptr': 140384569941504, 'device': 'CPU'}), (<DtypeKind.UINT: 1>, 8, 'C', '=')), 'validity': (PandasBuffer({'bufsize': 1, 'ptr': 94849242700288, 'device': 'CPU'}), (<DtypeKind.BOOL: 20>, 8, 'b', '=')), 'offsets': (PandasBuffer({'bufsize': 16, 'ptr': 94849235835952, 'device': 'CPU'}), (<DtypeKind.INT: 0>, 64, 'l', '='))}\n\n[None, <pyarrow.Buffer address=0x7fadd6808000 size=16 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7fadd6808080 size=1 is_cpu=True is_mutable=True>]\n\n```\n\n### Expected Behavior\n\nthis outputs\n\n```\n{"data": (PandasBuffer({'bufsize': 1, 'ptr': 140384569941504, 'device': 'CPU'}), (<DtypeKind.UINT: 1>, 8, 'C', '=')), 'validity': None, 'offsets': (PandasBuffer({'bufsize': 16, 'ptr': 94849235835952, 'device':

```

```

{'CPU\'}}, (<DtypeKind.INT: 0>, 64, \l', \=''))}\r\n[None, <pyarrow.Buffer address=0x7fadd6808000 size=16 is_cpu=True is_mutable=True>, <pyarrow.Buffer address=0x7fadd6808080 size=1 is_cpu=True is_mutable=True>]\r\n```\n\n### Installed Versions\n\n<details>\r\n\r\nINSTALLED VERSIONS\r\n-----\r\ncommit
: b033ca94e7ae6e1320c9d65a8163bd0a6049f40a\r\npython
: 3.10.12.fina
1.0\r\npython-bits
: 64\r\nOS
: Linux\r\nOS-release
: 5.15.146.1-microsoft-standard-WSL2\r\nVersion
: #1 SMP Thu Jan 11
04:09:03 UTC 2024\r\nmachine
: x86_64\r\nprocessor
: x86
_64\r\nbyteorder
: little\r\nLC_ALL
: None\r\nLANG
: C.UTF-8\r\nLOCALE
: en_US.UTF-8\r\n\r\npandas
: 2.
1.0.dev0+761.gb033ca94e7\r\nnumpy
: 1.26.4\r\npytz
: 2024.1\r\ndateutil
: 2.8.2\r\nsetuptools
: 69.1.1\r\npip
: 24.0\r\nCython
: 3.0.8\r\npytest
: 8.0.2\r\nhypothesis
: 6.98.15\r\nsphinx
: 7.2.6\r\nblosc
: None\r\nfeather
: None\r\nxlsxwriter
: 3.2.0\r\nlxml.et
ree
: 5.1.0\r\nhtml5lib
: 1.1\r\npymysql
:
1.4.6\r\npsycopg2
: 2.9.9\r\nninja2
: 3.1.3\r\nIPython
: 8.22.1\r\npandas_datareader
: None\r\nadbc-driver-postgresql: None\r\nadbc-d
river-sqlite
: None\r\nbs4
: 4.12.3\r\nbottleneck
: 1.3.8\r\nfastparquet
: 2024.2.0\r\nfsspec
: 2024.2.0\r\n
ngcsfs
: 2024.2.0\r\nmatplotlib
: 3.8.3\r\nnumba
: 0.59.0\r\nnumexpr
: 2.9.0\r\nnodfpy
: None\r\nopenp
yxl
: 3.1.2\r\npyarrow
: 15.0.0\r\npyreadstat
: 1.2.6\r\npython-calamine
: None\r\npyxlsb
: 1.0.10\r\ns3fs
: 2024.2.0\r\nscipy
: 1.12.0\r\nsqlalchemy
: 2.0.27\r\n
tables
: 3.9.2\r\nntabulate
: 0.9.0\r\nnxarray
: 2024.2.0\r\nxlrd
: 2.0.1\r\nnzstandard
: 0.22.0\r\nnt
zdata
: 2024.1\r\nnqtpy
: 2.4.1\r\npyqt5
: None\r\n</details>\r\n',
  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/5776
1/reactions',
    'total_count': 0,
    '+1': 0,
    '-1': 0,
    'laugh': 0,
    'hooray': 0,
    'confused': 0,
    'heart': 0,
    'rocket': 0,
    'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57761/tim
eline',
    'performed_via_github_app': None,
    'state_reason': None},
  {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760',
    'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760/label
s{/name}',
    'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760/com
ments',
    'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760/event
s',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57760',
    'id': 2173431071,
    'node_id': 'PR_kwDOAA0YD85o8V0e',
    'number': 57760,

```



```

'title': 'CLN: enforce deprecation of the `method` keyword on `df.fillna`',
'user': {'login': 'natmokval',
'id': 91160475,
'node_id': 'MDQ6VXNlcjkxMTYwNDc1',
'avatar_url': 'https://avatars.githubusercontent.com/u/91160475?v=4',
'gravatar_id': '',
'url': 'https://api.github.com/users/natmokval',
'html_url': 'https://github.com/natmokval',
'followers_url': 'https://api.github.com/users/natmokval/followers',
'following_url': 'https://api.github.com/users/natmokval/following{/other_user}',
'gists_url': 'https://api.github.com/users/natmokval/gists{/gist_id}',
'starred_url': 'https://api.github.com/users/natmokval/starred{/owner}/{/repo}',
'subscriptions_url': 'https://api.github.com/users/natmokval/subscriptions',
'organizations_url': 'https://api.github.com/users/natmokval/orgs',
'repos_url': 'https://api.github.com/users/natmokval/repos',
'events_url': 'https://api.github.com/users/natmokval/events{/privacy}',
'received_events_url': 'https://api.github.com/users/natmokval/received_events',
's',
'type': 'User',
'site_admin': False},
'labels': [{ 'id': 211029535,
'node_id': 'MDU6TGFiZWwMTewMjk1MzU=',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Clean',
'name': 'Clean',
'color': '207de5',
'default': False,
'description': None},
{ 'id': 5432589519,
'node_id': 'LA_kwDOAA0YD88AAAABQ868zw',
'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Transformation
s',
'name': 'Transformations',
'color': '5319e7',
'default': False,
'description': 'e.g. cumsum, diff, rank'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': None,
'comments': 1,
'created_at': '2024-03-07T09:45:25Z',
'updated_at': '2024-03-14T00:50:38Z',
'closed_at': None,
'author_association': 'CONTRIBUTOR',
'active_lock_reason': None,
'draft': False,
'pull_request': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
760',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57760',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57760.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57760.patch',
'merged_at': None},
'body': 'xref #53496\r\n\r\nenforced deprecation of the `method` keyword on `df.
fillna/ser.fillna`',

```

```

    'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760/reactions',
                  'total_count': 0,
                  '+1': 0,
                  '-1': 0,
                  'laugh': 0,
                  'hooray': 0,
                  'confused': 0,
                  'heart': 0,
                  'rocket': 0,
                  'eyes': 0},
    'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57760/timeline',
    'performed_via_github_app': None,
    'state_reason': None},
  {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758',
   'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',
   'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758/labels{/name}',
   'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758/comments',
   'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758/events',
   'html_url': 'https://github.com/pandas-dev/pandas/pull/57758',
   'id': 2173293118,
   'node_id': 'PR_kwDOAA0YD85o73Rg',
   'number': 57758,
   'title': 'BUG: DataFrame Interchange Protocol errors on Boolean columns',
   'user': {'login': 'MarcoGorelli',
            'id': 33491632,
            'node_id': 'MDQ6VXNlcjMzNDkxNjMy',
            'avatar_url': 'https://avatars.githubusercontent.com/u/33491632?v=4',
            'gravatar_id': '',
            'url': 'https://api.github.com/users/MarcoGorelli',
            'html_url': 'https://github.com/MarcoGorelli',
            'followers_url': 'https://api.github.com/users/MarcoGorelli/followers',
            'following_url': 'https://api.github.com/users/MarcoGorelli/following{/other_user}',
            'gists_url': 'https://api.github.com/users/MarcoGorelli/gists{/gist_id}',
            'starred_url': 'https://api.github.com/users/MarcoGorelli/starred{/owner}/{repo}',
            'subscriptions_url': 'https://api.github.com/users/MarcoGorelli/subscriptions',
            'organizations_url': 'https://api.github.com/users/MarcoGorelli/orgs',
            'repos_url': 'https://api.github.com/users/MarcoGorelli/repos',
            'events_url': 'https://api.github.com/users/MarcoGorelli/events{/privacy}',
            'received_events_url': 'https://api.github.com/users/MarcoGorelli/received_events'},
   'type': 'User',
   'site_admin': False},
  {'labels': [{'id': 4344388253,
               'node_id': 'LA_kwDOAA0YD88AAAABAvIanQ',
               'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Interchange',
               'name': 'Interchange',
               'color': '130443',
               'default': False,
               'description': 'Dataframe Interchange Protocol'}]},

```

```

    'state': 'open',
    'locked': False,
    'assignee': None,
    'assignees': [],
    'milestone': {'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/
117',
    'html_url': 'https://github.com/pandas-dev/pandas/milestone/117',
    'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/117/la
bels',
    'id': 10539704,
    'node_id': 'MI_kwDOAA0YD84AoNK4',
    'number': 117,
    'title': '2.2.2',
    'description': 'on-merge: backport to 2.2.x',
    'creator': {'login': 'lithomas1',
    'id': 47963215,
    'node_id': 'MDQ6VXNlcjQ3OTYzMjE1',
    'avatar_url': 'https://avatars.githubusercontent.com/u/47963215?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/lithomas1',
    'html_url': 'https://github.com/lithomas1',
    'followers_url': 'https://api.github.com/users/lithomas1/followers',
    'following_url': 'https://api.github.com/users/lithomas1/following{/other_use
r}',
    'gists_url': 'https://api.github.com/users/lithomas1/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/lithomas1/starred{/owner}/{/rep
o}',
    'subscriptions_url': 'https://api.github.com/users/lithomas1/subscriptions',
    'organizations_url': 'https://api.github.com/users/lithomas1/orgs',
    'repos_url': 'https://api.github.com/users/lithomas1/repos',
    'events_url': 'https://api.github.com/users/lithomas1/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/lithomas1/received_event
s',
    'type': 'User',
    'site_admin': False},
    'open_issues': 42,
    'closed_issues': 26,
    'state': 'open',
    'created_at': '2024-02-10T23:53:00Z',
    'updated_at': '2024-03-14T00:39:41Z',
    'due_on': None,
    'closed_at': None},
    'comments': 0,
    'created_at': '2024-03-07T08:38:04Z',
    'updated_at': '2024-03-08T07:47:14Z',
    'closed_at': None,
    'author_association': 'MEMBER',
    'active_lock_reason': None,
    'draft': True,
    'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57
758',
    'html_url': 'https://github.com/pandas-dev/pandas/pull/57758',
    'diff_url': 'https://github.com/pandas-dev/pandas/pull/57758.diff',
    'patch_url': 'https://github.com/pandas-dev/pandas/pull/57758.patch',
    'merged_at': None},
    'body': '- [ ] closes #xxxx (Replace xxxx with the GitHub issue number)\r\n- [ ]

```

```
[Tests added and passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if fixing a bug or adding a new feature\r\n- [ ] All [code checks passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commit).\r\n- [ ] Added [type annotations](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#type-hints) to new arguments/methods/functions.\r\n- [ ] Added an entry in the latest `doc/source/whatsnew/vX.X.X.rst` file if fixing a bug or adding a new feature.\r\n\r\n\ncloses #55332\r\n\r\nneeds rebasing onto #57764',\n  'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758/reactions',\n    'total_count': 0,\n    '+1': 0,\n    '-1': 0,\n    'laugh': 0,\n    'hooray': 0,\n    'confused': 0,\n    'heart': 0,\n    'rocket': 0,\n    'eyes': 0},\n  'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57758/timeline',\n  'performed_via_github_app': None,\n  'state_reason': None},\n{'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754',\n  'repository_url': 'https://api.github.com/repos/pandas-dev/pandas',\n  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754/labels{/name}',\n  'comments_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754/comments',\n  'events_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754/events',\n  'html_url': 'https://github.com/pandas-dev/pandas/pull/57754',\n  'id': 2172784612,\n  'node_id': 'PR_kwDOAA0YD85o6JD9',\n  'number': 57754,\n  'title': 'DEPR: make_block',\n  'user': {'login': 'jbrockmendel',\n    'id': 8078968,\n    'node_id': 'MDQ6VXNlcjgwNzg5Njg=',\n    'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',\n    'gravatar_id': '',\n    'url': 'https://api.github.com/users/jbrockmendel',\n    'html_url': 'https://github.com/jbrockmendel',\n    'followers_url': 'https://api.github.com/users/jbrockmendel/followers',\n    'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_user}',\n    'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',\n    'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}/{repo}',\n    'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscriptions',\n    'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',\n    'repos_url': 'https://api.github.com/users/jbrockmendel/repos',\n    'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',\n    'received_events_url': 'https://api.github.com/users/jbrockmendel/received_events',\n    'type': 'User',
```

```

'site_admin': False},
'labels': [{ 'id': 49094459,
  'node_id': 'MDU6TGFiZWw0OTA5NDQ1OQ==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Internals',
  'name': 'Internals',
  'color': 'fbca04',
  'default': False,
  'description': 'Related to non-user accessible pandas implementation'},
{ 'id': 87485152,
  'node_id': 'MDU6TGFiZWw4NzQ4NTE1Mg==',
  'url': 'https://api.github.com/repos/pandas-dev/pandas/labels/Deprecate',
  'name': 'Deprecate',
  'color': '5319e7',
  'default': False,
  'description': 'Functionality to remove in pandas'}],
'state': 'open',
'locked': False,
'assignee': None,
'assignees': [],
'milestone': { 'url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102',
  'html_url': 'https://github.com/pandas-dev/pandas/milestone/102',
  'labels_url': 'https://api.github.com/repos/pandas-dev/pandas/milestones/102/labels',
  'id': 9021906,
  'node_id': 'MI_kwDOAA0YD84AianS',
  'number': 102,
  'title': '3.0',
  'description': '',
  'creator': { 'login': 'jbrockmendel',
    'id': 8078968,
    'node_id': 'MDQ6VXNlcjgwNzg5Njg=',
    'avatar_url': 'https://avatars.githubusercontent.com/u/8078968?v=4',
    'gravatar_id': '',
    'url': 'https://api.github.com/users/jbrockmendel',
    'html_url': 'https://github.com/jbrockmendel',
    'followers_url': 'https://api.github.com/users/jbrockmendel/followers',
    'following_url': 'https://api.github.com/users/jbrockmendel/following{/other_user}',
    'gists_url': 'https://api.github.com/users/jbrockmendel/gists{/gist_id}',
    'starred_url': 'https://api.github.com/users/jbrockmendel/starred{/owner}/{repo}',
    'subscriptions_url': 'https://api.github.com/users/jbrockmendel/subscriptions',
    'organizations_url': 'https://api.github.com/users/jbrockmendel/orgs',
    'repos_url': 'https://api.github.com/users/jbrockmendel/repos',
    'events_url': 'https://api.github.com/users/jbrockmendel/events{/privacy}',
    'received_events_url': 'https://api.github.com/users/jbrockmendel/received_events',
    'type': 'User',
    'site_admin': False},
  'open_issues': 53,
  'closed_issues': 402,
  'state': 'open',
  'created_at': '2023-02-08T18:16:49Z',
  'updated_at': '2024-03-14T02:02:54Z',

```

```

'due_on': '2024-04-01T07:00:00Z',
'closed_at': None},
'comments': 0,
'created_at': '2024-03-07T01:55:30Z',
'updated_at': '2024-03-07T19:52:05Z',
'closed_at': None,
'author_association': 'MEMBER',
'active_lock_reason': None,
'draft': False,
'pull_request': {'url': 'https://api.github.com/repos/pandas-dev/pandas/pulls/57754',
'html_url': 'https://github.com/pandas-dev/pandas/pull/57754',
'diff_url': 'https://github.com/pandas-dev/pandas/pull/57754.diff',
'patch_url': 'https://github.com/pandas-dev/pandas/pull/57754.patch',
'merged_at': None},
'body': '- [x] closes #56815 (Replace xxxx with the GitHub issue number)\r\n- [ ] [Tests added and passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#writing-tests) if fixing a bug or adding a new feature\r\n- [ ] All [code checks passed](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#pre-commit).\r\n- [ ] Added [type annotations](https://pandas.pydata.org/pandas-docs/dev/development/contributing_codebase.html#type-hints) to new arguments/methods/functions.\r\n- [x] Added an entry in the latest `doc/source/whatsnew/vX.X.X.rst` file if fixing a bug or adding a new feature.\r\n',
'reactions': {'url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754/reactions',
'total_count': 0,
'+1': 0,
'-1': 0,
'laugh': 0,
'hooray': 0,
'confused': 0,
'heart': 0,
'rocket': 0,
'eyes': 0},
'timeline_url': 'https://api.github.com/repos/pandas-dev/pandas/issues/57754/timeline',
'performed_via_github_app': None,
'state_reason': None}]

```

Each element in data is a dictionary containing all of the data found on a GitHub issue page (except for the comments). We can pass data directly to DataFrame and extract fields of interest:

```
In [20]: issues = pd.DataFrame(data, columns=['number', 'title', 'labels', 'state'])
issues
```

Out[20]:

	number	title	labels	state
0	56630	QST: stack overflow	[{'id': 34444536, 'node_id': 'MDU6TGFZWwzNDQ0...}	open
1	56629	BUG: inconsistent behaviour of 'to_timedelta' ...	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open
2	56628	TYP: more simple return types from ruff		open
3	56627	added negative value support in pct_changes of...		open
4	56626	TST: Remove arraymanager markers	[{'id': 127685, 'node_id': 'MDU6TGFZWwxMjc2OD...}	open
5	56625	TST/CLN: Remove more seldom used fixtures	[{'id': 127685, 'node_id': 'MDU6TGFZWwxMjc2OD...}	open
6	56622	DOC: "Copy On Write: Patterns to avoid" confusion	[{'id': 134699, 'node_id': 'MDU6TGFZWwxMzQ2OT...}	open
7	56619	ENH/PERF: dispatch is_monotonic_increasing / d...	[{'id': 8935311, 'node_id': 'MDU6TGFZWw4OTM1M...}	open
8	56618	ENH: Add option to use absolute value in pct_c...	[{'id': 76812, 'node_id': 'MDU6TGFZWw3NjgxMg=...}	open
9	56616	BUG: Add limit_area to EA ffill/bfill	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open
10	56614	DEPR: Remove SettingWithCopyWarning		open
11	56613	BUG: Added raising when merging datetime colum...		open
12	56611	BUG: bar a line plots are not aligned on the x...		open
13	56608	CLN/TST: Resample fixtures	[{'id': 127685, 'node_id': 'MDU6TGFZWwxMjc2OD...}	open
14	56607	BUG: Series.diff does no arg validation unlike...	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open
15	56606	BUG: Series.apply with arrow integral types pa...	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open
16	56605	ENH: Add alternative html for Styler tooltips	[{'id': 76812, 'node_id': 'MDU6TGFZWw3NjgxMg=...}	open
17	56600	BUG: incompatible dtype when creating bool col...	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open
18	56599	BUG: DataFrame.replace with regex on `StringDt...	[{'id': 76811, 'node_id': 'MDU6TGFZWw3NjgxMQ=...}	open

	number	title	labels	state
19	56598	BUG: Fix inconsistency when constructing a Ser...		open
20	56596	DEPR: List of deprecations to be removed in 4.0		open
21	56594	DEPR: is_anchored()		open
22	56592	BUG: Docs won't build (S3 bucket does not exist)	[{'id': 134699, 'node_id': 'MDU6TGFiZWwxMzQ2OT...}	open
23	56591	BUG: read_html has unexpected behavior parsing...	[{'id': 76811, 'node_id': 'MDU6TGFiZWw3NjgxMQ=...}	open
24	56589	ENH: Update CFF with publication reference, Ze...		open
25	56588	QST: Roadmap for deprecations of Period types	[{'id': 34444536, 'node_id': 'MDU6TGFiZWwzNDQ0...}	open
26	56587	ENH: support the Arrow PyCapsule Interface on ...	[{'id': 3303158446, 'node_id': 'MDU6TGFiZWwzMz...}	open
27	56586	BUG: dt64[non_nano] + some_offsets incorrectly...	[{'id': 76811, 'node_id': 'MDU6TGFiZWw3NjgxMQ=...}	open
28	56583	TST/CLN: Reuse top level fixtures instead of p...	[{'id': 127685, 'node_id': 'MDU6TGFiZWwxMjc2OD...}	open
29	56576	BUG: PytzUsageWarning	[{'id': 76811, 'node_id': 'MDU6TGFiZWw3NjgxMQ=...}	open

With a bit of elbow grease, you can create some higher-level interfaces to common web APIs that return DataFrame objects for easy analysis.

6.4 Interacting with Databases

In a business setting, most data may not be stored in text or Excel files. SQL-based relational databases (such as SQL Server, PostgreSQL, and MySQL) are in wide use, and many alternative databases have become quite popular. The choice of database is usually dependent on the performance, data integrity, and scalability needs of an application. Loading data from SQL into a DataFrame is fairly straightforward, and pandas has some functions to simplify the process. As an example, I'll create a SQLite database using Python's built-in sqlite3 driver:

```
In [22]: import sqlite3
query = """CREATE TABLE test (a VARCHAR(20), b VARCHAR(20), c REAL, d INTEGER );"""
con = sqlite3.connect('mydata.sqlite')
con.execute(query)
```


Out[22]: <sqlite3.Cursor at 0x236374ab810>

In [23]: `con.commit()`

Then, insert a few rows of data:

In [24]: `data = [('Atlanta', 'Georgia', 1.25, 6), ('Tallahassee', 'Florida', 2.6, 3), ('Sacramento', 'California', 1.7, 5)]`
`stmt = "INSERT INTO test VALUES(?, ?, ?, ?)"`
`con.executemany(stmt, data)`
`con.commit()`

Most Python SQL drivers (PyODBC, psycopg2, MySQLdb, pymssql, etc.) return a list of tuples when selecting data from a table:

In [25]: `cursor = con.execute('select * from test')`
`rows = cursor.fetchall()`
`rows`

Out[25]: `[('Atlanta', 'Georgia', 1.25, 6), ('Tallahassee', 'Florida', 2.6, 3), ('Sacramento', 'California', 1.7, 5)]`

You can pass the list of tuples to the DataFrame constructor, but you also need the column names, contained in the cursor's description attribute:

In [26]: `cursor.description`

Out[26]: `((('a', None, None, None, None, None, None), ('b', None, None, None, None, None, None), ('c', None, None, None, None, None, None), ('d', None, None, None, None, None, None)))`

In [27]: `pd.DataFrame(rows, columns=[x[0] for x in cursor.description])`

Out[27]:

	a	b	c	d
0	Atlanta	Georgia	1.25	6
1	Tallahassee	Florida	2.60	3
2	Sacramento	California	1.70	5

The SQLAlchemy project is a popular Python SQL toolkit that abstracts away many of the common differences between SQL databases. pandas has a `read_sql` function that enables you to read data easily from a general SQLAlchemy connection. Here, we'll connect to the same SQLite database with SQLAlchemy and read data from the table created before:

In [30]: `import sqlalchemy as sqla`
`db = sqla.create_engine('sqlite:///mydata.sqlite')`
`pd.read_sql('select * from test', db)`

Out[30]:

	a	b	c	d
0	Atlanta	Georgia	1.25	6
1	Tallahassee	Florida	2.60	3
2	Sacramento	California	1.70	5

In [31]: `# !pip install sqlalchemy`

In []: