



PANDAS BRAIN TEASERS

EXERCISE YOUR MIND

```
1 from io import StringIO
2 import pandas as pd
3
4 csv_data = '''\
5 day,hits
6 2020-01-01,400
7 2020-02-02,800
8 2020-02-03,600
9 '''
10
11 df = pd.read_csv(StringIO(csv_data))
12 print(df['day'].dt.month.unique())
```

WILL THIS CODE RUN? WHAT WILL IT PRINT?

25 MIND BENDING TEASERS & SOLUTIONS

MIKI TEBEKA

Copyright

Copyright © 2020, 353solutions LLC.

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior consent of the publisher.

Late Addition

archer.py

```
1 import pandas as pd
2
3 df = pd.DataFrame([
4     ['Sterling', 83.4],
5     ['Cheryl', 97.2],
6     ['Lana', 13.2],
7 ], columns=['name', 'sum'])
8 df.late_fee = 3.5
9 print(df)
```



Try to guess what the output is before moving to the next page.

This code will print:

```
      name  sum
0  Sterling 83.4
1   Cheryl 97.2
2     Lana 13.2
```

Where did the `late_fee` column go?

Python's objects are very dynamic, you can add attributes to most of them as you please.

```
In [1]: class Point:
...:     def __init__(self, x, y):
...:         self.x, self.y = x, y
In [2]: p = Point(1, 2)
In [3]: p.x, p.y
Out[3]: (1, 2)
In [4]: p.z = 3
In [5]: p.z
Out[5]: 3
```

Pandas lets you access columns both by square brackets access (e.g. `df[name]`) and by attribute access (e.g. `df.name`). I recommend using square brackets at all times. One reason is what we saw - when you add an attribute to a DataFrame, it does not register as a new column. The second is that column names in CSV, JSON and other formats can contain spaces or other characters that are not valid Python identifiers - meaning you won't be able to access them with attribute access - `df.product id` will fail while `df['product id']` will work.

And the last reason is that it's confusing:

```
In [6]: df.sum
Out[6]:
<bound method DataFrame.sum of          name  sum
0  Sterling  83.4
1   Cheryl  97.2
2     Lana  13.2>
```

You get the DataFrame `sum` method and not the `sum` column. Also

```
In [7]: df.late_fee
Out[7]: 3.5
```

You probably expected `late_fee` to be a Series like the other columns.

Sometimes you'd like to add metadata to a DataFrame, say the name of the file the data was read from.

Instead of adding a new attribute, e.g. `df.Originating_file = '/path/to/sales.db'`, there's an experimental attribute called `attrs` for storing metadata in a DataFrame.

```
In [8]: df.attrs['originating_file'] = '/path/to/sales.db'
In [9]: df.attrs
Out[9]: {'originating_file': '/path/to/sales.db'}
```

Further Reading

- [Indexing Basics](#) on the Pandas documentation
- [Identifiers and keywords](#) in the Python documentation
- [DataFrame.attrs](#) documentation