

[paulgb / convert.py](#)

Last active on Mar 28, 2013

Convert the Yelp Academic dataset from JSON to CSV files with Pandas.

[convert.py](#)

```
1  '''
2  Convert Yelp Academic Dataset from JSON to CSV
3
4  Requires Pandas (https://pypi.python.org/pypi/pandas)
5
6  By Paul Butler, No Rights Reserved
7  '''
8
9  import json
10 import pandas as pd
11 from glob import glob
12
13 def convert(x):
14     ''' Convert a json string to a flat python dictionary
15     which can be passed into Pandas. '''
16     ob = json.loads(x)
17     for k, v in ob.items():
18         if isinstance(v, list):
19             ob[k] = ','.join(v)
20         elif isinstance(v, dict):
21             for kk, vv in v.items():
22                 ob['%s_%s' % (k, kk)] = vv
23     del ob[k]
24     return ob
25
26 for json_filename in glob('*.json'):
27     csv_filename = '%s.csv' % json_filename[:-5]
28     print 'Converting %s to %s' % (json_filename, csv_filename)
29     df = pd.DataFrame([convert(line) for line in file(json_filename)])
30     df.to_csv(csv_filename, encoding='utf-8', index=False)
```

[\(/uttara-cmu\)](#)[uttara-cmu \(/uttara-cmu\)](#) commented on Sep 30, 2014

Hi,

This worked great for reviews and businesses. Thanks a lot for the code. But for users it gives me an error

python convert.py

Converting yelp_academic_dataset_user.json to yelp_academic_dataset_user.csv

Traceback (most recent call last):

File "convert.py", line 29, in

df = pd.DataFrame([convert(line) for line in file(json_filename)])

File "convert.py", line 19, in convert

ob[k] = ','.join(v)

TypeError: sequence item 0: expected string, int found

What should I be doing?

[\(/heng0136\)](#)[heng0136 \(/heng0136\)](#) commented on Oct 5, 2014

Hi uttara-cmu,

For converting of the user.json file, you might wanna try changing the following code
(Line 19) `ob[k] = ','.join(v)`

To this:

`ob[k] = ','.join(str(v))`

Hope it helps...



[\(/tiagoovieira\)](#)

tiagoovieira ([/tiagoovieira](#)) commented on Jan 16

It converts to csv, but a lot of columns are blank :\

Does this happened before?



[\(/RDCooter\)](#)

[\(/\)](#)

[Status \(https://status.github.com/\)](https://status.github.com/) [API \(https://developer.github.com/\)](https://developer.github.com/) [Blog \(https://github.com/blog\)](https://github.com/blog) [About \(https://github.com/about\)](https://github.com/about)

© 2015 GitHub, Inc. [Terms \(https://github.com/site/terms\)](https://github.com/site/terms) [Privacy \(https://github.com/site/privacy\)](https://github.com/site/privacy) [Security \(https://github.com/security\)](https://github.com/security) [Contact \(https://github.com/contact\)](https://github.com/contact)