

Project

September 2, 2019

```
In [1]: import pandas as pd
```

```
def JSON_to_dataframe(file_name):
    # read the entire file into a python array
    with open(file_name, 'r') as f:
        data = f.readlines()

    # remove the trailing "\n" from each line
    data = map(lambda x: x.rstrip(), data)

    # parse json into string format
    data_json_str = "[" + ','.join(data) + "]"

    # now, load it into pandas
    data_df = pd.read_json(data_json_str)

    return data_df
```

```
In [2]: review_df = JSON_to_dataframe("review.json")
        tip_df = JSON_to_dataframe("tip.json")
        checkin_df = JSON_to_dataframe("checkin.json")
        business_df = JSON_to_dataframe("business.json")
```

```
In [9]: frames = [review_df, tip_df, checkin_df, business_df]
```

```
        cumulative_df = pd.concat(frames, axis=1, keys=['review', 'tip', 'checkin', 'business'])
```

```
In [18]: cumulative_df.shape
```

```
Out[18]: (6685900, 30)
```

```
In [15]: cumulative_df.head()
```

```
Out[15]:
```

	review				
	business_id	cool	date	funny	
0	ujmEBvifdJM6h6RLv4wQIg	0	2013-05-07 04:34:36	1	
1	NZnhc2sEQy3RmzKTZnqtwQ	0	2017-01-14 21:30:33	0	
2	WTqjgwHlXbSFevF32_DJVw	0	2016-11-09 20:09:03	0	

3	ikCg8xy5JIg_NGPx-MSIDA	0	2018-01-09 20:56:38	0
4	b1b1eb3uo-w561D0ZfCEiQ	0	2018-01-30 23:07:38	0

	review_id	stars
0	Q1sbwvVQXV2734tPgoKj4Q	1
1	GJXCdrto3ASJOqKeVWPi6Q	5
2	2TzJjDVDEuAW6MR5Vuc1ug	5
3	yi0R0Ugj_xUx_Nek0-_Qig	5
4	11a8sVPMUftaC7_ABRkmtw	1

	text	useful
0	Total bill for this horrible service? Over \$8G...	6
1	I *adore* Travis at the Hard Rock's new Kelly ...	0
2	I have to say that this office really has it t...	3
3	Went in for a lunch. Steak sandwich was delici...	0
4	Today was my second out of three sessions I ha...	7

	user_id	business_id	tip	business	city
0	hG7b0MtEbXx5QzbzE6C_VA	VaKXUpmWTTWdKbpJ3aQdMw	...	Phoenix	Phoenix
1	yXQM5uF2jS6es16SJzNHfg	OPiPeoJiv92rENwbq76orA	...	Mississauga	Mississauga
2	n6-Gk65cPZL6Uz8qRm3NYw	5KheTjYPu1HcQzQFtm4_vw	...	Charlotte	Charlotte
3	dacAlZ6fTM6mqwW5uxkskg	TkoyGi8J7YFjA6SbaRzrxg	...	Goodyear	Goodyear
4	ssoyf2_x0EQMed6fgHeMyQ	AkL60us6A1atZejfZXn1Bg	...	Charlotte	Charlotte

	hours	is_open	latitude
0	None	0.0	33.522143
1	{'Monday': '9:0-0:0', 'Tuesday': '9:0-0:0', 'W...	1.0	43.605499
2	{'Monday': '17:30-21:30', 'Wednesday': '17:30-...	1.0	35.092564
3	{'Monday': '8:0-17:0', 'Tuesday': '8:0-17:0', ...	1.0	33.455613
4	{'Monday': '7:0-23:0', 'Tuesday': '7:0-23:0', ...	1.0	35.190012

	longitude	name	postal_code	review_count	stars
0	-112.018481	Arizona Biltmore Golf Club	85016	5.0	3.0
1	-79.652289	Emerald Chinese Restaurant	L5R 3E7	128.0	2.5
2	-80.859132	Musashi Japanese Restaurant	28210	170.0	4.0
3	-112.395596	Farmers Insurance - Paul Lorenz	85338	3.0	5.0
4	-80.887223	Queen City Plumbing	28217	4.0	4.0

	state
0	AZ
1	ON
2	NC

3 AZ
4 NC

[5 rows x 30 columns]