Mini-Project: JSON

Import Modules and Files

In [98]:

```python
import pandas as pd
import json
from pandas.io.json import json_normalize
import numpy as np
```

In [99]:

```python
world_bank_df = pd.read_json('world_bank_projects.json')
world_bank_df
```

Out[99]:

	_id	approvalfy	board_approval_month	boardapprovaldate	
0	{'$oid': '52b213b38594d8a2be17c780'}	1999	November	2013-11-12T00:00:00Z	DE RE
1	{'$oid': '52b213b38594d8a2be17c781'}	2015	November	2013-11-04T00:00:00Z	GOVER
2	{'$oid': '52b213b38594d8a2be17c782'}	2014	November	2013-11-01T00:00:00Z	M FIN ECONC
3	{'$oid': '52b213b38594d8a2be17c783'}	2014	October	2013-10-31T00:00:00Z	MIN. OF CO
4	{'$oid': '52b213b38594d8a2be17c784'}	2014	October	2013-10-31T00:00:00Z	M
5	{'$oid': '52b213b38594d8a2be17c785'}	2014	October	2013-10-31T00:00:00Z	RE
6	{'$oid': '52b213b38594d8a2be17c786'}	2014	October	2013-10-29T00:00:00Z	GOVER

7	{'$oid': '52b213b38594d8a2be17c787'}	2014	October	2013-10-29T00:00:00Z	RE
8	{'$oid': '52b213b38594d8a2be17c788'}	2014	October	2013-10-29T00:00:00Z	THE GO
9	{'$oid': '52b213b38594d8a2be17c789'}	2014	October	2013-10-29T00:00:00Z	THE KI
10	{'$oid': '52b213b38594d8a2be17c78a'}	2014	October	2013-10-25T00:00:00Z	GOVEF SOL
11	{'$oid': '52b213b38594d8a2be17c78b'}	2014	October	2013-10-25T00:00:00Z	
12	{'$oid': '52b213b38594d8a2be17c78c'}	2014	October	2013-10-24T00:00:00Z	GOVEF
13	{'$oid': '52b213b38594d8a2be17c78d'}	2014	October	2013-10-22T00:00:00Z	GOVEF TI
14	{'$oid': '52b213b38594d8a2be17c78e'}	2014	October	2013-10-22T00:00:00Z	GOVEF
15	{'$oid': '52b213b38594d8a2be17c78f'}	2014	October	2013-10-17T00:00:00Z	M
16	{'$oid': '52b213b38594d8a2be17c790'}	2014	October	2013-10-17T00:00:00Z	M
17	{'$oid': '52b213b38594d8a2be17c791'}	2014	October	2013-10-16T00:00:00Z	M FIN BU
18	{'$oid': '52b213b38594d8a2be17c792'}	2014	October	2013-10-16T00:00:00Z	GOVEF
19	{'$oid': '52b213b38594d8a2be17c793'}	2014	October	2013-10-10T00:00:00Z	M
20	{'$oid': '52b213b38594d8a2be17c794'}	2014	October	2013-10-09T00:00:00Z	DEV NETW
21	{'$oid': '52b213b38594d8a2be17c795'}	2014	October	2013-10-07T00:00:00Z	

22	{'$oid': '52b213b38594d8a2be17c796'}	2014	October	2013-10-07T00:00:00Z	M PLA INTERN
23	{'$oid': '52b213b38594d8a2be17c797'}	2014	October	2013-10-03T00:00:00Z	RE
24	{'$oid': '52b213b38594d8a2be17c798'}	2014	September	2013-09-30T00:00:00Z	RE A
25	{'$oid': '52b213b38594d8a2be17c799'}	2014	September	2013-09-30T00:00:00Z	UNI QL
26	{'$oid': '52b213b38594d8a2be17c79a'}	2014	September	2013-09-30T00:00:00Z	LA DF
27	{'$oid': '52b213b38594d8a2be17c79b'}	2014	September	2013-09-30T00:00:00Z	PACIF SECUF
28	{'$oid': '52b213b38594d8a2be17c79c'}	2014	September	2013-09-30T00:00:00Z	GO
29	{'$oid': '52b213b38594d8a2be17c79d'}	2014	September	2013-09-27T00:00:00Z	GOVEF MC
...	...	...	...	...	
470	{'$oid': '52b213b38594d8a2be17c956'}	2013	September	2012-09-20T00:00:00Z	GOVEF
471	{'$oid': '52b213b38594d8a2be17c957'}	2013	September	2012-09-20T00:00:00Z	GOVEF BA
472	{'$oid': '52b213b38594d8a2be17c958'}	2013	September	2012-09-13T00:00:00Z	M PI
473	{'$oid': '52b213b38594d8a2be17c959'}	2013	September	2012-09-13T00:00:00Z	M ENV FORE\$
474	{'$oid': '52b213b38594d8a2be17c95a'}	2013	September	2012-09-11T00:00:00Z	GOVEF
475	{'$oid': '52b213b38594d8a2be17c95b'}	2013	September	2012-09-11T00:00:00Z	GOVEF

476	{'$oid': '52b213b38594d8a2be17c95c'}	2013	September	2012-09-11T00:00:00Z	RE
477	{'$oid': '52b213b38594d8a2be17c95d'}	2013	September	2012-09-11T00:00:00Z	RE U
478	{'$oid': '52b213b38594d8a2be17c95e'}	2013	September	2012-09-11T00:00:00Z	ISLAMIC O
479	{'$oid': '52b213b38594d8a2be17c95f'}	2013	September	2012-09-10T00:00:00Z	
480	{'$oid': '52b213b38594d8a2be17c960'}	2013	September	2012-09-10T00:00:00Z	GOVEF
481	{'$oid': '52b213b38594d8a2be17c961'}	2013	September	2012-09-06T00:00:00Z	GOVEF
482	{'$oid': '52b213b38594d8a2be17c962'}	2013	September	2012-09-06T00:00:00Z	GOVEF
483	{'$oid': '52b213b38594d8a2be17c963'}	2013	September	2012-09-06T00:00:00Z	GOVEF
484	{'$oid': '52b213b38594d8a2be17c964'}	2013	September	2012-09-06T00:00:00Z	ILO, C
485	{'$oid': '52b213b38594d8a2be17c965'}	2013	September	2012-09-05T00:00:00Z	GOVEF
486	{'$oid': '52b213b38594d8a2be17c966'}	2013	September	2012-09-01T00:00:00Z	GOVEF
487	{'$oid': '52b213b38594d8a2be17c967'}	2013	August	2012-08-30T00:00:00Z	UNITE
488	{'$oid': '52b213b38594d8a2be17c968'}	2013	August	2012-08-30T00:00:00Z	THE ST
489	{'$oid': '52b213b38594d8a2be17c969'}	2013	August	2012-08-30T00:00:00Z	U

490	{'$oid': '52b213b38594d8a2be17c96a'}	2013	August	2012-08-29T00:00:00Z	GOVEF
491	{'$oid': '52b213b38594d8a2be17c96b'}	2013	August	2012-08-24T00:00:00Z	P,
492	{'$oid': '52b213b38594d8a2be17c96c'}	2013	August	2012-08-21T00:00:00Z	GOVEF
493	{'$oid': '52b213b38594d8a2be17c96d'}	2013	August	2012-08-21T00:00:00Z	GOVEF BA
494	{'$oid': '52b213b38594d8a2be17c96e'}	2013	August	2012-08-17T00:00:00Z	M I
495	{'$oid': '52b213b38594d8a2be17c96f'}	2013	August	2012-08-10T00:00:00Z	COMPE
496	{'$oid': '52b213b38594d8a2be17c970'}	2013	August	2012-08-09T00:00:00Z	LA DI
497	{'$oid': '52b213b38594d8a2be17c971'}	2013	August	2012-08-03T00:00:00Z	GOVEF THE RE
498	{'$oid': '52b213b38594d8a2be17c972'}	2013	August	2012-08-02T00:00:00Z	RE
499	{'$oid': '52b213b38594d8a2be17c973'}	2013	August	2012-08-02T00:00:00Z	GOVI

500 rows × 50 columns

Problem 1: Fin Top 10 Countries with Most Projects

```python
# Find 10 countries with most projects
projects_per_country = world_bank_df.countryname.value_counts()
projects_per_country.head(10)
```

```
Republic of Indonesia              19
People's Republic of China         19
Socialist Republic of Vietnam      17
Republic of India                  16
Republic of Yemen                  13
Nepal                              12
Kingdom of Morocco                 12
People's Republic of Bangladesh    12
Africa                             11
Republic of Mozambique             11
Name: countryname, dtype: int64
```

Problem 2: Find Top 10 Major Project Themes

```
In [101]:
```

```python
# Load JSON file into string and normalize
world_bank_projects = json.load((open('world_bank_projects.json')))
new_df = json_normalize(world_bank_projects,'mjtheme_namecode')
new_df.head(15)
```

```
Out[101]:
```

	code	name
0	8	Human development
1	11	
2	1	Economic management
3	6	Social protection and risk management
4	5	Trade and integration
5	2	Public sector governance
6	11	Environment and natural resources management
7	6	Social protection and risk management
8	7	Social dev/gender/inclusion
9	7	Social dev/gender/inclusion
10	5	Trade and integration
11	4	Financial and private sector development
12	6	Social protection and risk management
13	6	
14	2	Public sector governance

```
# List Top 10 projects by count per code
project_theme = new_df.code.value_counts()
project_theme.head(10)
```

Out[102]:

```
11     250
10     216
8      210
2      199
6      168
4      146
7      130
5       77
9       50
1       38
Name: code, dtype: int64
```

Problem 3: Find and Replace Missing Values

In [103]:

```
# Create data frame sorted by 'code' and 'name'
new_df = new_df.sort_values(['code','name'])
new_df.head(10)
```

Out[103]:

	code	name
212	1	
363	1	
1024	1	
1114	1	
1437	1	
2	1	Economic management
88	1	Economic management
175	1	Economic management
204	1	Economic management
205	1	Economic management

```
In [104]:
```

```python
# Replace missing values with Nan using numpy
new_df.name[new_df['name'] == ''] = np.nan
new_df.head(10)
```

Out[104]:

	code	name
212	1	NaN
363	1	NaN
1024	1	NaN
1114	1	NaN
1437	1	NaN
2	1	Economic management
88	1	Economic management
175	1	Economic management
204	1	Economic management
205	1	Economic management

```
In [105]:
```

```python
# Backfill all values listed as NaN
new_df = new_df.fillna(method='bfill')
new_df.head(10)
```

Out[105]:

	code	name
212	1	Economic management
363	1	Economic management
1024	1	Economic management
1114	1	Economic management
1437	1	Economic management
2	1	Economic management
88	1	Economic management
175	1	Economic management
204	1	Economic management
205	1	Economic management

Country	Number of Projects
Republic of Indonesia	19
People's Republic of China	19
Socialist Republic of Vietnam	17
Republic of India	16
Republic of Yemen	13
Nepal	12
Kingdom of Morocco	12
People's Republic of Bangladesh	12
Africa	12
Republic of Mozambique	11

Final Results

Project Code	Number of Projects
11	250
10	216
8	210
2	199
6	168
4	146
7	130
5	77
9	50
1	38