Data In Motion Week 11 Pandas Challenge

By Joshua Ati

Challenge Questions

1. Read in the JSON file
2. Flatten the nested list from JSON object
3. Create a dataframe that includes the flatten list and all original columns

In [1]:
```
# necessary imports
import pandas as pd
```

In [2]:
```
# reading the data
url = "https://raw.githubusercontent.com/BindiChen/machine-learning/main/data-analysis/0
df = pd.read_json(url)
```

In [3]:
```
# return first rows
df.head()
```

Out[3]:

	school_name	class	students
0	ABC primary school	Year 1	{'id': 'A001', 'name': 'Tom', 'math': 60, 'phy...
1	ABC primary school	Year 1	{'id': 'A002', 'name': 'James', 'math': 89, 'p...
2	ABC primary school	Year 1	{'id': 'A003', 'name': 'Jenny', 'math': 79, 'p...

There are two pandas methods that can be used to flatten a JSON object. The pd.Series and pd.json_normalize. I would show an example of the two methods

In [4]:
```
# method 1
df['students'].apply(pd.Series)
```

Out[4]:

	id	name	math	physics	chemistry
0	A001	Tom	60	66	61
1	A002	James	89	76	51
2	A003	Jenny	79	90	78

In [5]:
```
# method 2
pd.json_normalize(df['students'])
```

Out[5]:

	id	name	math	physics	chemistry
0	A001	Tom	60	66	61
1	A002	James	89	76	51
2	A003	Jenny	79	90	78

In [6]:
```
# combining the data
df = pd.concat([pd.json_normalize(df['students']), df[['school_name', 'class']]], axis=1
df.head()
```

Out[6]:

	id	name	math	physics	chemistry	school_name	class
0	A001	Tom	60	66	61	ABC primary school	Year 1
1	A002	James	89	76	51	ABC primary school	Year 1
2	A003	Jenny	79	90	78	ABC primary school	Year 1

In [7]:
```
# checking datatypes
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3 entries, 0 to 2
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   id           3 non-null      object
 1   name         3 non-null      object
 2   math         3 non-null      int64 
 3   physics      3 non-null      int64 
 4   chemistry    3 non-null      int64 
 5   school_name  3 non-null      object
 6   class        3 non-null      object
dtypes: int64(3), object(4)
memory usage: 296.0+ bytes
```

In [2]:
```
#!jupyter nbconvert --to webpdf --allow-chromium-download week_11_pandas.ipynb
```