Data in Motion Pandas Challenge Week 9

import

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

0. Create Entries

In [2]:
```
raw_data_1 = { 'subject_id': ['1', '2', '3', '4', '5'], 'first_name': ['Alex', 'Amy', 'A

raw_data_2 = { 'subject_id': ['4', '5', '6', '7', '8'], 'first_name': ['Billy', 'Brian',

raw_data_3 = { 'subject_id': ['1', '2', '3', '4', '5', '7', '8', '9', '10', '11'], 'test
```

1. Assign each to a variable called data1, data2, data3.

In [3]:
```
data1=pd.DataFrame.from_dict(raw_data_1)
data1.head()
```

Out[3]:

	subject_id	first_name	last_name
0	1	Alex	Anderson
1	2	Amy	Ackerman
2	3	Allen	Ali
3	4	Alice	Aoni
4	5	Ayoung	Atiches

In [4]:
```
data2=pd.DataFrame.from_dict(raw_data_2)
data2.head()
```

Out[4]:

	subject_id	first_name	last_name
0	4	Billy	Bonder
1	5	Brian	Black
2	6	Bran	Balwner
3	7	Bryce	Brice
4	8	Betty	Btisan

In [5]:
```
data3=pd.DataFrame.from_dict(raw_data_3)
data3.head()
```

Out[5]:

	subject_id	test_id
0	1	51
1	2	15
2	3	15
3	4	61
4	5	16

2. Join data1 and data2 along rows and assign all_data.

In [6]:
```
all_data=pd.concat([data1,data2],ignore_index=True)
all_data.head(10)
```

Out[6]:

	subject_id	first_name	last_name
0	1	Alex	Anderson
1	2	Amy	Ackerman
2	3	Allen	Ali
3	4	Alice	Aoni
4	5	Ayoung	Atiches
5	4	Billy	Bonder
6	5	Brian	Black
7	6	Bran	Balwner
8	7	Bryce	Brice
9	8	Betty	Btisan

3. Join the two dataframes along columns and assign to all_data_col.

In [7]:
```
all_data_col=pd.concat([data1,data2],axis=1)
all_data_col.head(10)
```

Out[7]:

	subject_id	first_name	last_name	subject_id	first_name	last_name
0	1	Alex	Anderson	4	Billy	Bonder
1	2	Amy	Ackerman	5	Brian	Black
2	3	Allen	Ali	6	Bran	Balwner
3	4	Alice	Aoni	7	Bryce	Brice
4	5	Ayoung	Atiches	8	Betty	Btisan

4. Print data3.

In [8]:
```
display(data3)
```

	subject_id	test_id
0	1	51
1	2	15
2	3	15
3	4	61
4	5	16
5	7	14
6	8	15
7	9	1
8	10	61
9	11	16

5. Merge all_data and data3 along the subject_id value.

In [9]:
```
allMerged=all_data.merge(data3,on='subject_id')
allMerged.head()
```

Out[9]:

	subject_id	first_name	last_name	test_id
0	1	Alex	Anderson	51
1	2	Amy	Ackerman	15
2	3	Allen	Ali	15
3	4	Alice	Aoni	61
4	4	Billy	Bonder	61

6. Merge only the data that has the same 'subject_id' on both data1 and data2.

In [10]:
```
DfInnerId=data1.merge(data2,on='subject_id')
DfInnerId.head()
```

Out[10]:

	subject_id	first_name_x	last_name_x	first_name_y	last_name_y
0	4	Alice	Aoni	Billy	Bonder
1	5	Ayoung	Atiches	Brian	Black

7. Merge all values in data1 and data2, with matching records from both sides where available.

In [11]:
```
DfInner=data1.merge(data2)
DfInner.head()
```

Out[11]:

	subject_id	first_name	last_name