Wk3_workthrough

November 22, 2016

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

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

        df = pd.DataFrame([{'Name': 'Chris', 'Item Purchased': 'Sponge', 'Cost': 22
                            {'Name': 'Kevyn', 'Item Purchased': 'Kitty Litter', 'Cos
                            {'Name': 'Filip', 'Item Purchased': 'Spoon', 'Cost': 5.0
                           index=['Store 1', 'Store 1', 'Store 2'])
        df
```

Out[1]:

	Cost	Item Purchased	Name
Store 1	22.5	Sponge	Chris
Store 1	2.5	Kitty Litter	Kevyn
Store 2	5.0	Spoon	Filip

```
In [2]: df['Date'] = ['December 1', 'January 1', 'mid-May']
        df
```

Out[2]:

	Cost	Item Purchased	Name	Date
Store 1	22.5	Sponge	Chris	December 1
Store 1	2.5	Kitty Litter	Kevyn	January 1
Store 2	5.0	Spoon	Filip	mid-May

```
In [3]: df['Delivered'] = True
        df
```

Out[3]:

	Cost	Item Purchased	Name	Date	Delivered
Store 1	22.5	Sponge	Chris	December 1	True
Store 1	2.5	Kitty Litter	Kevyn	January 1	True
Store 2	5.0	Spoon	Filip	mid-May	True

```
In [5]: df['Feedback'] = ['Positive', None, 'Negative']
        df
```

Out[5]:

	Cost	Item Purchased	Name	Date	Delivered	Feedback
Store 1	22.5	Sponge	Chris	December 1	True	Positive
Store 1	2.5	Kitty Litter	Kevyn	January 1	True	None
Store 2	5.0	Spoon	Filip	mid-May	True	Negative

```
In [6]: adf = df.reset_index()
        adf['Date'] = pd.Series({0: 'December 1', 2: 'mid-May'})

In [7]: adf
```

Out[7]:

	index	Cost	Item Purchased	Name	Date	Delivered	Feedback
0	Store 1	22.5	Sponge	Chris	December 1	True	Positive
1	Store 1	2.5	Kitty Litter	Kevyn	NaN	True	None
2	Store 2	5.0	Spoon	Filip	mid-May	True	Negative

```
In [8]: # intersection => Inner Join
        staff_df = pd.DataFrame([{'Name': 'Kelly', 'Role': 'Director of HR'},
                                 {'Name': 'Sally', 'Role': 'Course liasion'},
                                 {'Name': 'James', 'Role': 'Grader'}])
        staff_df = staff_df.set_index('Name')
        student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business'},
                                   {'Name': 'Mike', 'School': 'Law'},
                                   {'Name': 'Sally', 'School': 'Engineering'}])
        student_df = student_df.set_index('Name')
        print(staff_df.head())
        print()
        print(student_df.head())
```

```
                Role
Name
Kelly  Director of HR
Sally  Course liasion
James          Grader

            School
Name
James     Business
Mike           Law
Sally  Engineering
```

```
In [9]: staff_df
```

Out[9]:

Name	Role
Kelly	Director of HR
Sally	Course liasion
James	Grader

```
In [10]: student_df
```

Out[10]:

Name	School
James	Business
Mike	Law
Sally	Engineering

```
In [11]: # outer merge
         pd.merge(staff_df, student_df, how='outer', left_index=True, right_index=T
```

```
Out[11]:                  Role       School
         Name
         James          Grader     Business
         Kelly  Director of HR          NaN
         Mike              NaN          Law
         Sally  Course liasion  Engineering
```

```
In [12]: # inner merge
         pd.merge(staff_df, student_df, how='inner', left_index=True, right_index=T
```

```
Out[12]:                  Role       School
         Name
         James          Grader     Business
         Sally  Course liasion  Engineering
```

```
In [13]: pd.merge(staff_df, student_df, how='left', left_index=True, right_index=Tr
```

```
Out[13]:                  Role       School
         Name
         Kelly  Director of HR          NaN
         Sally  Course liasion  Engineering
         James          Grader     Business
```

```
In [14]: pd.merge(staff_df, student_df, how='right', left_index=True, right_index=T
```

```
Out[14]:                  Role       School
         Name
         James          Grader     Business
         Mike              NaN          Law
         Sally  Course liasion  Engineering
```

```
In [15]: staff_df = staff_df.reset_index()
         student_df = student_df.reset_index()
         pd.merge(staff_df, student_df, how='left', left_on='Name', right_on='Name'
```

```
Out[15]:     Name            Role       School
         0  Kelly  Director of HR          NaN
         1  Sally  Course liasion  Engineering
         2  James          Grader     Business
```

```
In [16]: staff_df = pd.DataFrame([{'Name': 'Kelly', 'Role': 'Director of HR', 'Loca
                                  {'Name': 'Sally', 'Role': 'Course liasion', 'Loca
                                  {'Name': 'James', 'Role': 'Grader', 'Location': '
         student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business', 'Locati
                                    {'Name': 'Mike', 'School': 'Law', 'Location': '
                                    {'Name': 'Sally', 'School': 'Engineering', 'Loc
         pd.merge(staff_df, student_df, how='left', left_on='Name', right_on='Name'
```

Out[16]:

	Location_x	Name	Role	Location_y	Sc
0	State Street	Kelly	Director of HR	NaN	
1	Washington Avenue	Sally	Course liasion	512 Wilson Crescent	Enginee
2	Washington Avenue	James	Grader	1024 Billiard Avenue	Busi

In [17]:
```
staff_df = pd.DataFrame([{'First Name': 'Kelly', 'Last Name': 'Desjardins'
                          {'First Name': 'Sally', 'Last Name': 'Brooks', 'F
                          {'First Name': 'James', 'Last Name': 'Wilde', 'Ro
student_df = pd.DataFrame([{'First Name': 'James', 'Last Name': 'Hammond',
                            {'First Name': 'Mike', 'Last Name': 'Smith', 'S
                            {'First Name': 'Sally', 'Last Name': 'Brooks',
staff_df
student_df
pd.merge(staff_df, student_df, how='inner', left_on=['First Name','Last Na
```

Out[17]:

	First Name	Last Name	Role	School
0	Sally	Brooks	Course liasion	Engineering

In [ ]: