Nullable Data

In [9]:

```
ort pandas as pd
hase_1 = pd.Series({'Name' : 'Chris',
                    'Item Purchased' : 'Dog Food',
                    'Cost' : 22.50})
hase_2 = pd.Series({'Name' : 'Kevyn',
                    'Item Purchased' : 'Kitty Litter',
                    'Cost' : 2.50})
hase_3 = pd.Series({'Name' : 'Vinod',
                    'Item Purchased' : 'Bird Seed',
                    'Cost' : 5.00})
 pd.DataFrame([purchase_1, purchase_2, purchase_3], index = ['Store 1', 'Store 1', 'Store 2'])
ead()
```

Out[9]:

	Name	Item Purchased	Cost
Store 1	Chris	Dog Food	22.5
Store 1	Kevyn	Kitty Litter	2.5
Store 2	Vinod	Bird Seed	5.0

In [10]:

```
df['Date'] = ['Dec. 1', 'January 1', 'mid-May']
df
```

Out[10]:

	Name	Item Purchased	Cost	Date
Store 1	Chris	Dog Food	22.5	Dec. 1
Store 1	Kevyn	Kitty Litter	2.5	January 1
Store 2	Vinod	Bird Seed	5.0	mid-May

In [11]:

```
df['Delivered'] = True
df
```

Out[11]:

	Name	Item Purchased	Cost	Date	Delivered
Store 1	Chris	Dog Food	22.5	Dec. 1	True
Store 1	Kevyn	Kitty Litter	2.5	January 1	True
Store 2	Vinod	Bird Seed	5.0	mid-May	True

In [12]:

```python
df['Feedback'] = ['Positive', None, 'Negative']
df
```

Out[12]:

	Name	Item Purchased	Cost	Date	Delivered	Feedback
Store 1	Chris	Dog Food	22.5	Dec. 1	True	Positive
Store 1	Kevyn	Kitty Litter	2.5	January 1	True	None
Store 2	Vinod	Bird Seed	5.0	mid-May	True	Negative

In [13]:

```python
df['Date'] = pd.Series({'Store 1' : 'May 5'})
df
```

Out[13]:

	Name	Item Purchased	Cost	Date	Delivered	Feedback
Store 1	Chris	Dog Food	22.5	May 5	True	Positive
Store 1	Kevyn	Kitty Litter	2.5	May 5	True	None
Store 2	Vinod	Bird Seed	5.0	NaN	True	Negative

Merging Dataframes

In [14]:

```python
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')
staff_df
```

Out[14]:

	Role
Name	
Kelly	Director of HR
Sally	Course liasion
James	Grader

In [15]:

```python
student_df = pd.DataFrame([{'Name' : 'Mike', 'School' : 'Business'},
                           {'Name' : 'Sally', 'School' : 'Law'},
                           {'Name' : 'James', 'School' : 'Computer'}])
student_df.set_index('Name', inplace = True)    #inplace는 메소드를 사용했을때 바로 고쳐줌
student_df
```

Out[15]:

	School
Name	
Mike	Business
Sally	Law
James	Computer

In [16]:

```python
pd.merge(staff_df, student_df, how = 'outer', left_index = True, right_index = True)
```

Out[16]:

	Role	School
Name		
James	Grader	Computer
Kelly	Director of HR	NaN
Mike	NaN	Business
Sally	Course liasion	Law

In [17]:

```python
pd.merge(staff_df, student_df, how = 'left', left_index = True, right_index = True)
```

Out[17]:

	Role	School
Name		
Kelly	Director of HR	NaN
Sally	Course liasion	Law
James	Grader	Computer

In [18]:

```
pd.merge(staff_df, student_df, how = 'right', left_index = True, right_index = True)
```

Out[18]:

Name	Role	School
Mike	NaN	Business
Sally	Course liasion	Law
James	Grader	Computer

In [19]:

```
pd.merge(staff_df, student_df, how = 'inner', left_index = True, right_index = True)
```

Out[19]:

Name	Role	School
Sally	Course liasion	Law
James	Grader	Computer

In [20]:

```
staff_df.reset_index(inplace = True)
staff_df
```

Out[20]:

	Name	Role
0	Kelly	Director of HR
1	Sally	Course liasion
2	James	Grader

In [21]:

```python
student_df = student_df.reset_index()
student_df
```

Out[21]:

	Name	School
0	Mike	Business
1	Sally	Law
2	James	Computer

In [22]:

```python
pd.merge(staff_df, student_df, how = 'inner', left_on = 'Name', right_on = 'Name')
```

Out[22]:

	Name	Role	School
0	Sally	Course liasion	Law
1	James	Grader	Computer

In [23]:

```python
staff_df['Location'] = ['101', '102', 's101']
student_df['Location'] = ['401', '402', 's101']
print(staff_df)
print()
print(student_df)
```

```
    Name           Role Location
0  Kelly  Director of HR      101
1  Sally  Course liasion      102
2  James          Grader     s101

    Name     School Location
0   Mike   Business      401
1  Sally        Law      402
2  James   Computer     s101
```

In [24]:

```python
pd.merge(staff_df, student_df, how = 'left', left_on = 'Name', right_on = 'Name')
```

Out[24]:

	Name	Role	Location_x	School	Location_y
0	Kelly	Director of HR	101	NaN	NaN
1	Sally	Course liasion	102	Law	402
2	James	Grader	s101	Computer	s101

In [25]:

```
, student_df, how = 'inner', left_on = ['Name', 'Location'], right_on = ['Name', 'Location'])
```

Out[25]:

	Name	Role	Location	School
0	James	Grader	s101	Computer

Group By

In [26]:

```
import numpy as np
df = pd.read_csv('census.csv')
df.head()
```

Out[26]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	EST
0	40	3	6	1	0	Alabama	Alabama	4779736	
1	50	3	6	1	1	Alabama	Autauga County	54571	
2	50	3	6	1	3	Alabama	Baldwin County	182265	
3	50	3	6	1	5	Alabama	Barbour County	27457	
4	50	3	6	1	7	Alabama	Bibb County	22915	

5 rows × 100 columns

In [27]:

```
df['STNAME'].unique()
```

Out[27]:

```
array(['Alabama', 'Alaska', 'Arizona', 'Arkansas', 'California',
       'Colorado', 'Connecticut', 'Delaware', 'District of Columbia',
       'Florida', 'Georgia', 'Hawaii', 'Idaho', 'Illinois', 'Indiana',
       'Iowa', 'Kansas', 'Kentucky', 'Louisiana', 'Maine', 'Maryland',
       'Massachusetts', 'Michigan', 'Minnesota', 'Mississippi',
       'Missouri', 'Montana', 'Nebraska', 'Nevada', 'New Hampshire',
       'New Jersey', 'New Mexico', 'New York', 'North Carolina',
       'North Dakota', 'Ohio', 'Oklahoma', 'Oregon', 'Pennsylvania',
       'Rhode Island', 'South Carolina', 'South Dakota', 'Tennessee',
       'Texas', 'Utah', 'Vermont', 'Virginia', 'Washington',
       'West Virginia', 'Wisconsin', 'Wyoming'], dtype=object)
```

In [30]:

```
%%timeit -r 1
for state in df['STNAME'].unique():
    avg = np.average(df.where(df['STNAME'] == state).dropna()['CENSUS2010POP'])
    print('Counties in state ' + state + 'have an average population of' + str(avg))
```

```
Counties in state Alabamahave an average population of140580.4705882353
Counties in state Alaskahave an average population of47348.73333333333
Counties in state Arizonahave an average population of799002.125
Counties in state Arkansashave an average population of76734.68421052632
Counties in state Californiahave an average population of1262845.9661016949
Counties in state Coloradohave an average population of154744.4923076923
Counties in state Connecticuthave an average population of794243.7777777778
Counties in state Delawarehave an average population of448967.0
Counties in state District of Columbiahave an average population of601723.0
Counties in state Floridahave an average population of552979.7058823529
Counties in state Georgiahave an average population of121095.6625
Counties in state Hawaiihave an average population of453433.6666666667
Counties in state Idahohave an average population of69670.3111111111
Counties in state Illinoishave an average population of249138.4854368932
Counties in state Indianahave an average population of139436.60215053763
Counties in state Iowahave an average population of60927.1
Counties in state Kansashave an average population of53832.41509433962
Counties in state Kentuckyhave an average population of71725.07438016529
Counties in state Louisianahave an average population of139488.36923076923
Counties in state Mainehave an average population of156277.76470588235
Counties in state Marylandhave an average population of461884.16
Counties in state Massachusettshave an average population of873017.2
Counties in state Michiganhave an average population of235324.7619047619
Counties in state Minnesotahave an average population of120543.75
Counties in state Mississippihave an average population of71501.13253012048
Counties in state Missourihave an average population of103257.36206896552
Counties in state Montanahave an average population of34716.31578947369
Counties in state Nebraskahave an average population of38858.31914893617
Counties in state Nevadahave an average population of300061.22222222225
Counties in state New Hampshirehave an average population of239358.18181818182
Counties in state New Jerseyhave an average population of799263.0909090909
Counties in state New Mexicohave an average population of121128.17647058824
Counties in state New Yorkhave an average population of615177.8412698413
Counties in state North Carolinahave an average population of188821.44554455444
Counties in state North Dakotahave an average population of24910.777777777777
Counties in state Ohiohave an average population of259247.2808988764
Counties in state Oklahomahave an average population of96188.48717948717
Counties in state Oregonhave an average population of207085.0810810811
Counties in state Pennsylvaniahave an average population of373599.3823529412
Counties in state Rhode Islandhave an average population of350855.6666666667
Counties in state South Carolinahave an average population of196824.0
Counties in state South Dakotahave an average population of24303.880597014926
Counties in state Tennesseehave an average population of132210.52083333334
Counties in state Texashave an average population of197220.0862745098
Counties in state Utahhave an average population of184259.0
Counties in state Vermonthave an average population of83432.13333333333
Counties in state Virginiahave an average population of119371.83582089552
Counties in state Washingtonhave an average population of336227.0
Counties in state West Virginiahave an average population of66178.35714285714
Counties in state Wisconsinhave an average population of155807.83561643836
Counties in state Wyominghave an average population of46968.833333333336
Counties in state Alabamahave an average population of140580.4705882353
```

```
Counties in state Alaskahave an average population of47348.73333333333
Counties in state Arizonahave an average population of799002.125
Counties in state Arkansashave an average population of76734.68421052632
Counties in state Californiahave an average population of1262845.9661016949
Counties in state Coloradohave an average population of154744.4923076923
Counties in state Connecticuthave an average population of794243.7777777778
Counties in state Delawarehave an average population of448967.0
Counties in state District of Columbiahave an average population of601723.0
Counties in state Floridahave an average population of552979.7058823529
Counties in state Georgiahave an average population of121095.6625
Counties in state Hawaiihave an average population of453433.6666666667
Counties in state Idahohave an average population of69670.3111111111
Counties in state Illinoishave an average population of249138.4854368932
Counties in state Indianahave an average population of139436.60215053763
Counties in state Iowahave an average population of60927.1
Counties in state Kansashave an average population of53832.41509433962
Counties in state Kentuckyhave an average population of71725.07438016529
Counties in state Louisianahave an average population of139488.36923076923
Counties in state Mainehave an average population of156277.76470588235
Counties in state Marylandhave an average population of461884.16
Counties in state Massachusettshave an average population of873017.2
Counties in state Michiganhave an average population of235324.7619047619
Counties in state Minnesotahave an average population of120543.75
Counties in state Mississippihave an average population of71501.13253012048
Counties in state Missourihave an average population of103257.36206896552
Counties in state Montanahave an average population of34716.31578947369
Counties in state Nebraskahave an average population of38858.31914893617
Counties in state Nevadahave an average population of300061.22222222225
Counties in state New Hampshirehave an average population of239358.18181818182
Counties in state New Jerseyhave an average population of799263.0909090909
Counties in state New Mexicohave an average population of121128.17647058824
Counties in state New Yorkhave an average population of615177.8412698413
Counties in state North Carolinahave an average population of188821.44554455444
Counties in state North Dakotahave an average population of24910.777777777777
Counties in state Ohiohave an average population of259247.2808988764
Counties in state Oklahomahave an average population of96188.48717948717
Counties in state Oregonhave an average population of207085.0810810811
Counties in state Pennsylvaniahave an average population of373599.3823529412
Counties in state Rhode Islandhave an average population of350855.6666666667
Counties in state South Carolinahave an average population of196824.0
Counties in state South Dakotahave an average population of24303.880597014926
Counties in state Tennesseehave an average population of132210.52083333334
Counties in state Texashave an average population of197220.0862745098
Counties in state Utahhave an average population of184259.0
Counties in state Vermonthave an average population of83432.13333333333
Counties in state Virginiahave an average population of119371.83582089552
Counties in state Washingtonhave an average population of336227.0
Counties in state West Virginiahave an average population of66178.35714285714
Counties in state Wisconsinhave an average population of155807.83561643836
Counties in state Wyominghave an average population of46968.833333333336
2.08 s ± 0 ns per loop (mean ± std. dev. of 1 run, 1 loop each)
```

In [31]:

```
df.groupby('STNAME')
```

Out[31]:

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000027CAB299BE0>

In [32]:

```python
df.groupby('STNAME').mean()
```

Out[32]:

STNAME	SUMLEV	REGION	DIVISION	STATE	COUNTY	CENSUS2010POP	ESTIMATE
Alabama	49.852941	3.0	6.0	1.0	66.014706	1.405805e+05	1
Alaska	49.666667	4.0	9.0	2.0	144.666667	4.734873e+04	4
Arizona	49.375000	4.0	8.0	4.0	13.000000	7.990021e+05	7
Arkansas	49.868421	3.0	7.0	5.0	74.013158	7.673468e+04	7
California	49.830508	4.0	9.0	6.0	57.016949	1.262846e+06	1
Colorado	49.846154	4.0	8.0	8.0	61.276923	1.547445e+05	1
Connecticut	48.888889	1.0	1.0	9.0	7.111111	7.942438e+05	7
Delaware	47.500000	3.0	5.0	10.0	2.250000	4.489670e+05	4
District of Columbia	45.000000	3.0	5.0	11.0	0.500000	6.017230e+05	6
Florida	49.852941	3.0	5.0	12.0	66.911765	5.529797e+05	5
Georgia	49.937500	3.0	5.0	13.0	160.481250	1.210957e+05	1
Hawaii	48.333333	4.0	9.0	15.0	4.166667	4.534337e+05	4
Idaho	49.777778	4.0	8.0	16.0	43.022222	6.967031e+04	6
Illinois	49.902913	2.0	3.0	17.0	101.009709	2.491385e+05	2
Indiana	49.892473	2.0	3.0	18.0	91.010753	1.394366e+05	1
Iowa	49.900000	2.0	4.0	19.0	98.010000	6.092710e+04	6
Kansas	49.905660	2.0	4.0	20.0	104.009434	5.383242e+04	5
Kentucky	49.917355	3.0	6.0	21.0	119.008264	7.172507e+04	7
Louisiana	49.846154	3.0	7.0	22.0	63.015385	1.394884e+05	1
Maine	49.411765	1.0	1.0	23.0	15.058824	1.562778e+05	1
Maryland	49.600000	3.0	5.0	24.0	43.160000	4.618842e+05	4
Massachusetts	49.333333	1.0	1.0	25.0	13.066667	8.730172e+05	8
Michigan	49.880952	2.0	3.0	26.0	82.011905	2.353248e+05	2
Minnesota	49.886364	2.0	4.0	27.0	86.011364	1.205438e+05	1
Mississippi	49.879518	3.0	6.0	28.0	81.012048	7.150113e+04	7
Missouri	49.913793	2.0	4.0	29.0	116.698276	1.032574e+05	1
Montana	49.824561	4.0	8.0	30.0	55.017544	3.471632e+04	3
Nebraska	49.893617	2.0	4.0	31.0	92.010638	3.885832e+04	3
Nevada	49.444444	4.0	8.0	32.0	43.000000	3.000612e+05	3
New Hampshire	49.090909	1.0	1.0	33.0	9.090909	2.393582e+05	2
New Jersey	49.545455	1.0	2.0	34.0	20.045455	7.992631e+05	7
New Mexico	49.705882	4.0	8.0	35.0	29.264706	1.211282e+05	1

STNAME	SUMLEV	REGION	DIVISION	STATE	COUNTY	CENSUS2010POP	ESTIMATE
New York	49.841270	1.0	2.0	36.0	61.015873	6.151778e+05	6
North Carolina	49.900990	3.0	5.0	37.0	99.009901	1.888214e+05	1
North Dakota	49.814815	2.0	4.0	38.0	52.018519	2.491078e+04	2
Ohio	49.887640	2.0	3.0	39.0	87.011236	2.592473e+05	2
Oklahoma	49.871795	3.0	7.0	40.0	76.012821	9.618849e+04	9
Oregon	49.729730	4.0	9.0	41.0	35.027027	2.070851e+05	2
Pennsylvania	49.852941	1.0	2.0	42.0	66.014706	3.735994e+05	3
Rhode Island	48.333333	1.0	1.0	44.0	4.166667	3.508557e+05	3
South Carolina	49.787234	3.0	5.0	45.0	45.021277	1.968240e+05	1
South Dakota	49.850746	2.0	4.0	46.0	66.940299	2.430388e+04	2
Tennessee	49.895833	3.0	6.0	47.0	94.010417	1.322105e+05	1
Texas	49.960784	3.0	7.0	48.0	253.003922	1.972201e+05	1
Utah	49.666667	4.0	8.0	49.0	28.033333	1.842590e+05	1
Vermont	49.333333	1.0	1.0	50.0	13.066667	8.343213e+04	8
Virginia	49.925373	3.0	5.0	51.0	262.022388	1.193718e+05	1
Washington	49.750000	4.0	9.0	53.0	38.025000	3.362270e+05	3
West Virginia	49.821429	3.0	5.0	54.0	54.017857	6.617836e+04	6
Wisconsin	49.863014	2.0	3.0	55.0	70.123288	1.558078e+05	1
Wyoming	49.583333	4.0	8.0	56.0	22.041667	4.696883e+04	4

51 rows × 98 columns

In [33]:

```
%%timeit -r 1
for group, frame in df.groupby('STNAME') :
    avg = np.average(frame['CENSUS2010POP'])
    print('Counties in state ' + group + 'have an average population of' + str(avg))
```

```
Counties in state Washingtonhave an average population of338227.0
Counties in state West Virginiahave an average population of66178.35714285714
Counties in state Wisconsinhave an average population of155807.83561643836
Counties in state Wyominghave an average population of46968.833333333336
Counties in state Alabamahave an average population of140580.4705882353
Counties in state Alaskahave an average population of47348.73333333333
Counties in state Arizonahave an average population of799002.125
Counties in state Arkansashave an average population of76734.68421052632
Counties in state Californiahave an average population of1262845.9661016949
Counties in state Coloradohave an average population of154744.4923076923
Counties in state Connecticuthave an average population of794243.7777777778
Counties in state Delawarehave an average population of448967.0
Counties in state District of Columbiahave an average population of601723.0
Counties in state Floridahave an average population of552979.7058823529
Counties in state Georgiahave an average population of121095.6625
Counties in state Hawaiihave an average population of453433.6666666667
Counties in state Idahohave an average population of69670.3111111111
Counties in state Illinoishave an average population of249138.4854368932
Counties in state Indianahave an average population of139436.60215053763
Counties in state Iowahave an average population of60927.1
```

In [34]:

```
df = df.set_index('STNAME')
df
```

Out[34]:

STNAME	SUMLEV	REGION	DIVISION	STATE	COUNTY	CTYNAME	CENSUS2010POP	ESTIM
Alabama	40	3	6	1	0	Alabama	4779736	
Alabama	50	3	6	1	1	Autauga County	54571	
Alabama	50	3	6	1	3	Baldwin County	182265	
Alabama	50	3	6	1	5	Barbour County	27457	
Alabama	50	3	6	1	7	Bibb County	22915	
...	...	...	...	...	...	...	...	
Wyoming	50	4	8	56	37	Sweetwater County	43806	
Wyoming	50	4	8	56	39	Teton County	21294	
Wyoming	50	4	8	56	41	Uinta County	21118	
Wyoming	50	4	8	56	43	Washakie County	8533	
Wyoming	50	4	8	56	45	Weston County	7208	

3193 rows × 99 columns

In [37]:

```
distribute(item) :
if item[0] < 'M' :
    return 0
if item[0] < 'Q' :
    return 1
return 2

group, frame in df.groupby(distribute) :
print('There are ' + str(len(frame)) + ' records in group' + str(group) + ' for processing.')
```

```
There are 1196 records in group0 for processing.
There are 1154 records in group1 for processing.
There are 843 records in group2 for processing.
```