Merging Dataframes

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

df = pd.DataFrame([{'Name': 'Jayvant', 'Item Purchased': 'Sponge', 'Cost': 22.50},
                   {'Name': 'Ram', 'Item Purchased': 'Kitty Litter', 'Cost': 2.50},
                   {'Name': 'Laxman', 'Item Purchased': 'Spoon', 'Cost': 5.00}],
                  index=['Store 1', 'Store 1', 'Store 2'])
df
```

Out[1]:

	Name	Item Purchased	Cost
Store 1	Jayvant	Sponge	22.5
Store 1	Ram	Kitty Litter	2.5
Store 2	Laxman	Spoon	5.0

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

Out[2]:

	Name	Item Purchased	Cost	Date
Store 1	Jayvant	Sponge	22.5	December 1
Store 1	Ram	Kitty Litter	2.5	January 1
Store 2	Laxman	Spoon	5.0	Mid-May

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

Out[3]:

	Name	Item Purchased	Cost	Date	Delivered
Store 1	Jayvant	Sponge	22.5	December 1	True
Store 1	Ram	Kitty Litter	2.5	January 1	True
Store 2	Laxman	Spoon	5.0	Mid-May	True

In [4]:
```
df['Feedback']=['Possitve',None,'Negative']
df
```

Out[4]:

	Name	Item Purchased	Cost	Date	Delivered	Feedback
Store 1	Jayvant	Sponge	22.5	December 1	True	Possitve
Store 1	Ram	Kitty Litter	2.5	January 1	True	None
Store 2	Laxman	Spoon	5.0	Mid-May	True	Negative

In [5]:
```
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 [6]:
```
pd.merge(staff_df, student_df, how='outer', left_index=True, right_index=True ) #ou
ter join
```

Out[6]:

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

In [7]:
```
pd.merge(staff_df,student_df,how='inner', left_index=True, right_index=True) #inner
join
```

Out[7]:

Name	Role	School
Sally	Course liasion	Engineering
James	Grader	Business

In [8]:
```
pd.merge(staff_df,student_df,how='left', left_index=True, right_index=True) #left o
uter join
```

Out[8]:

Name	Role	School
Kelly	Director of HR	NaN
Sally	Course liasion	Engineering
James	Grader	Business

In [9]:
```
pd.merge(staff_df,student_df,how='right', left_index=True, right_index=True) #right outer join
```

Out[9]:

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

In [10]:
```
staff_df = staff_df.reset_index()
```

In [11]:
```
staff_df
```

Out[11]:

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

In [12]:
```
student_df = student_df.reset_index()
pd.merge(staff_df, student_df, how='left', left_on='Name', right_on='Name')
```

Out[12]:

	Name	Role	School
0	Kelly	Director of HR	NaN
1	Sally	Course liasion	Engineering
2	James	Grader	Business

In [13]:
```
staff_df = pd.DataFrame([{'Name': 'Kelly', 'Role': 'Director of HR', 'Location': 'State Street'},
                         {'Name': 'Sally', 'Role': 'Course liasion', 'Location': 'Washington Avenue'},
                         {'Name': 'James', 'Role': 'Grader', 'Location': 'Washington Avenue'}])
student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business', 'Location': '1024 Billiard Avenue'},
                           {'Name': 'Mike', 'School': 'Law', 'Location': 'Fraternity House #22'},
                           {'Name': 'Sally', 'School': 'Engineering', 'Location': '512 Wilson Crescent'}])
pd.merge(staff_df, student_df, how='left', left_on='Name', right_on='Name')
```

Out[13]:

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

In [14]:
```
staff_df = pd.DataFrame([{'First Name': 'Kelly', 'Last Name': 'Desjardins', 'Role': 'Director of HR'},
                         {'First Name': 'Sally', 'Last Name': 'Brooks', 'Role': 'Course liasion'},
                         {'First Name': 'James', 'Last Name': 'Wilde', 'Role': 'Grader'}])
student_df = pd.DataFrame([{'First Name': 'James', 'Last Name': 'Hammond', 'School': 'Business'},
                           {'First Name': 'Mike', 'Last Name': 'Smith', 'School': 'Law'},
                           {'First Name': 'Sally', 'Last Name': 'Brooks', 'School': 'Engineering'}])
```

In [15]:
```
staff_df
```

Out[15]:

	First Name	Last Name	Role
0	Kelly	Desjardins	Director of HR
1	Sally	Brooks	Course liasion
2	James	Wilde	Grader

In [16]:
```
student_df
```

Out[16]:

	First Name	Last Name	School
0	James	Hammond	Business
1	Mike	Smith	Law
2	Sally	Brooks	Engineering

In [19]:
```
pd.merge(staff_df,student_df,how='inner',left_on=['First Name','Last Name'], right_on=['First Name','Last Name'])
```

Out[19]:

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

Idiomatic Pandas: Making Code Pandorable

In [1]:
```
import pandas as pd
df = pd.read_csv('f:/Python_Programs/census.csv')
df
```

Out[1]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTIMATESBAS
0	40	3	6	1	0	Alabama	Alabama	4779736	47
1	50	3	6	1	1	Alabama	Autauga County	54571	
2	50	3	6	1	3	Alabama	Baldwin County	182265	1
3	50	3	6	1	5	Alabama	Barbour County	27457	
4	50	3	6	1	7	Alabama	Bibb County	22915	
...	...	...	...	...	...	...	...	...	
3188	50	4	8	56	37	Wyoming	Sweetwater County	43806	
3189	50	4	8	56	39	Wyoming	Teton County	21294	
3190	50	4	8	56	41	Wyoming	Uinta County	21118	
3191	50	4	8	56	43	Wyoming	Washakie County	8533	
3192	50	4	8	56	45	Wyoming	Weston County	7208	

3193 rows × 100 columns

In [2]:
```
(df.where(df['SUMLEV']==50)    #chaining
    .dropna()
    .set_index(['STNAME','CTYNAME'])
    .rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'}))
```

Out[2]:

STNAME	CTYNAME	SUMLEV	REGION	DIVISION	STATE	COUNTY	CENSUS2010POP	Estimates Base 2010	POPESTI
Alabama	Autauga County	50.0	3.0	6.0	1.0	1.0	54571.0	54571.0	
	Baldwin County	50.0	3.0	6.0	1.0	3.0	182265.0	182265.0	
	Barbour County	50.0	3.0	6.0	1.0	5.0	27457.0	27457.0	
	Bibb County	50.0	3.0	6.0	1.0	7.0	22915.0	22919.0	
	Blount County	50.0	3.0	6.0	1.0	9.0	57322.0	57322.0	
...	...	...	...	...	...	...	...	...	
Wyoming	Sweetwater County	50.0	4.0	8.0	56.0	37.0	43806.0	43806.0	
	Teton County	50.0	4.0	8.0	56.0	39.0	21294.0	21294.0	
	Uinta County	50.0	4.0	8.0	56.0	41.0	21118.0	21118.0	
	Washakie County	50.0	4.0	8.0	56.0	43.0	8533.0	8533.0	
	Weston County	50.0	4.0	8.0	56.0	45.0	7208.0	7208.0	

3142 rows × 98 columns

In [3]:
```
df = df[df['SUMLEV']==50]
df.set_index(['STNAME','CTYNAME'], inplace=True)
df.rename(columns={'ESTIMATESBASE2010': 'Estimates Base 2010'})
```

Out[3]:

		SUMLEV	REGION	DIVISION	STATE	COUNTY	CENSUS2010POP	Estimates Base 2010	POPESTII
STNAME	**CTYNAME**								
Alabama	**Autauga County**	50	3	6	1	1	54571	54571	
	Baldwin County	50	3	6	1	3	182265	182265	
	Barbour County	50	3	6	1	5	27457	27457	
	Bibb County	50	3	6	1	7	22915	22919	
	Blount County	50	3	6	1	9	57322	57322	
...	**...**	...	...	...	...	...	...	...	
Wyoming	**Sweetwater County**	50	4	8	56	37	43806	43806	
	Teton County	50	4	8	56	39	21294	21294	
	Uinta County	50	4	8	56	41	21118	21118	
	Washakie County	50	4	8	56	43	8533	8533	
	Weston County	50	4	8	56	45	7208	7208	

3142 rows × 98 columns

In [4]:
```
import numpy as np
def min_max(row):
    data = row[['POPESTIMATE2010',
                'POPESTIMATE2011',
                'POPESTIMATE2012',
                'POPESTIMATE2013',
                'POPESTIMATE2014',
                'POPESTIMATE2015']]
    return pd.Series({'min': np.min(data), 'max': np.max(data)})
```

In [6]:
```
df.apply(min_max, axis=1)
```

Out[6]:

		min	max
STNAME	**CTYNAME**		
Alabama	**Autauga County**	54660.0	55347.0
	Baldwin County	183193.0	203709.0
	Barbour County	26489.0	27341.0
	Bibb County	22512.0	22861.0
	Blount County	57373.0	57776.0
...	**...**	...	...
Wyoming	**Sweetwater County**	43593.0	45162.0
	Teton County	21297.0	23125.0
	Uinta County	20822.0	21102.0
	Washakie County	8316.0	8545.0
	Weston County	7065.0	7234.0

3142 rows × 2 columns

In [7]:
```
import numpy as np
def min_max(row):
    data = row[['POPESTIMATE2010',
                'POPESTIMATE2011',
                'POPESTIMATE2012',
                'POPESTIMATE2013',
                'POPESTIMATE2014',
                'POPESTIMATE2015']]
    row['max'] = np.max(data)
    row['min'] = np.min(data)
    return row
df.apply(min_max, axis=1)
```

Out[7]:

		SUMLEV	REGION	DIVISION	STATE	COUNTY	CENSUS2010POP	ESTIMATESBASE201(
STNAME	**CTYNAME**							
Alabama	**Autauga County**	50.0	3.0	6.0	1.0	1.0	54571.0	54571.(
	Baldwin County	50.0	3.0	6.0	1.0	3.0	182265.0	182265.(
	Barbour County	50.0	3.0	6.0	1.0	5.0	27457.0	27457.(
	Bibb County	50.0	3.0	6.0	1.0	7.0	22915.0	22919.(
	Blount County	50.0	3.0	6.0	1.0	9.0	57322.0	57322.(
...	**...**	...	...	...	...	...	...	..
Wyoming	**Sweetwater County**	50.0	4.0	8.0	56.0	37.0	43806.0	43806.(
	Teton County	50.0	4.0	8.0	56.0	39.0	21294.0	21294.(
	Uinta County	50.0	4.0	8.0	56.0	41.0	21118.0	21118.(
	Washakie County	50.0	4.0	8.0	56.0	43.0	8533.0	8533.(
	Weston County	50.0	4.0	8.0	56.0	45.0	7208.0	7208.(

3142 rows × 100 columns

In [8]:
```
rows = ['POPESTIMATE2010',
        'POPESTIMATE2011',
        'POPESTIMATE2012',
        'POPESTIMATE2013',
        'POPESTIMATE2014',
        'POPESTIMATE2015']
df.apply(lambda x: np.max(x[rows]), axis=1)
```

Out[8]:
```
STNAME    CTYNAME
Alabama   Autauga County        55347.0
          Baldwin County       203709.0
          Barbour County        27341.0
          Bibb County           22861.0
          Blount County         57776.0
                                 ...
Wyoming   Sweetwater County      45162.0
          Teton County          23125.0
          Uinta County          21102.0
          Washakie County        8545.0
          Weston County          7234.0
Length: 3142, dtype: float64
```

Group by

In [9]:
```
import pandas as pd
import numpy as np
df = pd.read_csv('f:/Python_Programs/census.csv')
df = df[df['SUMLEV']==50]
df
```

Out[9]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTIMATESBAS
1	50	3	6	1	1	Alabama	Autauga County	54571	
2	50	3	6	1	3	Alabama	Baldwin County	182265	1
3	50	3	6	1	5	Alabama	Barbour County	27457	
4	50	3	6	1	7	Alabama	Bibb County	22915	
5	50	3	6	1	9	Alabama	Blount County	57322	
...	...	...	...	...	...	...	...	...	
3188	50	4	8	56	37	Wyoming	Sweetwater County	43806	
3189	50	4	8	56	39	Wyoming	Teton County	21294	
3190	50	4	8	56	41	Wyoming	Uinta County	21118	
3191	50	4	8	56	43	Wyoming	Washakie County	8533	
3192	50	4	8	56	45	Wyoming	Weston County	7208	

3142 rows × 100 columns

In [10]:
```
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(av
g))
```

```
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of 100551.3913043478
2
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
```

In [11]:
```
%%timeit -n 10
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(av
g))
```

```
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of 100551.3913043478
2
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
```

In [12]:
```
%%timeit -n 10
for group, frame in df.groupby('STNAME'):
    avg = np.average(frame['CENSUS2010POP'])
    print('Counties in state ' + group + ' have an average population of ' + str(av
g))
```

```
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
Counties in state Idaho have an average population of 35626.86363636364
Counties in state Illinois have an average population of 125790.50980392157
Counties in state Indiana have an average population of 70476.10869565218
Counties in state Iowa have an average population of 30771.262626262625
Counties in state Kansas have an average population of 27172.55238095238
Counties in state Kentucky have an average population of 36161.39166666667
Counties in state Louisiana have an average population of 70833.9375
Counties in state Maine have an average population of 83022.5625
Counties in state Maryland have an average population of 240564.66666666666
Counties in state Massachusetts have an average population of 467687.78571428574
Counties in state Michigan have an average population of 119080.0
Counties in state Minnesota have an average population of 60964.65517241379
Counties in state Mississippi have an average population of 36186.54878048781
Counties in state Missouri have an average population of 52077.62608695652
Counties in state Montana have an average population of 17668.125
Counties in state Nebraska have an average population of 19638.075268817203
Counties in state Nevada have an average population of 158855.9411764706
Counties in state New Hampshire have an average population of 131647.0
Counties in state New Jersey have an average population of 418661.61904761905
Counties in state New Mexico have an average population of 62399.36363636364
Counties in state New York have an average population of 312550.03225806454
Counties in state North Carolina have an average population of 95354.83
Counties in state North Dakota have an average population of 12690.396226415094
Counties in state Ohio have an average population of 131096.63636363635
Counties in state Oklahoma have an average population of 48718.844155844155
Counties in state Oregon have an average population of 106418.72222222222
Counties in state Pennsylvania have an average population of 189587.74626865672
Counties in state Rhode Island have an average population of 210513.4
Counties in state South Carolina have an average population of 100551.3913043478
2
Counties in state South Dakota have an average population of 12336.060606060606
Counties in state Tennessee have an average population of 66801.1052631579
Counties in state Texas have an average population of 98998.27165354331
Counties in state Utah have an average population of 95306.37931034483
Counties in state Vermont have an average population of 44695.78571428572
Counties in state Virginia have an average population of 60111.29323308271
Counties in state Washington have an average population of 172424.10256410256
Counties in state West Virginia have an average population of 33690.8
Counties in state Wisconsin have an average population of 78985.91666666667
Counties in state Wyoming have an average population of 24505.478260869564
Counties in state Alabama have an average population of 71339.34328358209
Counties in state Alaska have an average population of 24490.724137931036
Counties in state Arizona have an average population of 426134.4666666667
Counties in state Arkansas have an average population of 38878.90666666667
Counties in state California have an average population of 642309.5862068966
Counties in state Colorado have an average population of 78581.1875
Counties in state Connecticut have an average population of 446762.125
Counties in state Delaware have an average population of 299311.3333333333
Counties in state District of Columbia have an average population of 601723.0
Counties in state Florida have an average population of 280616.5671641791
Counties in state Georgia have an average population of 60928.63522012578
Counties in state Hawaii have an average population of 272060.2
```

In [13]:
```
df.head()
```

Out[13]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTIMATESBASE20
1	50	3	6	1	1	Alabama	Autauga County	54571	545
2	50	3	6	1	3	Alabama	Baldwin County	182265	1822
3	50	3	6	1	5	Alabama	Barbour County	27457	274
4	50	3	6	1	7	Alabama	Bibb County	22915	229
5	50	3	6	1	9	Alabama	Blount County	57322	573

5 rows × 100 columns

In [14]:
```
df = df.set_index('STNAME')

def fun(item):
    if item[0]<'M':
        return 0
    if item[0]<'Q':
        return 1
    return 2

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

```
There are 1177 records in group 0 for processing.
There are 1134 records in group 1 for processing.
There are 831 records in group 2 for processing.
```

In [17]:
```
df = pd.read_csv('f:/Python_Programs/census.csv')
df = df[df['SUMLEV']==50]
```

In [20]:
```
df.head()
```

Out[20]:

	SUMLEV	REGION	DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTIMATESBASE20
1	50	3	6	1	1	Alabama	Autauga County	54571	545
2	50	3	6	1	3	Alabama	Baldwin County	182265	1822
3	50	3	6	1	5	Alabama	Barbour County	27457	274
4	50	3	6	1	7	Alabama	Bibb County	22915	229
5	50	3	6	1	9	Alabama	Blount County	57322	573

5 rows × 100 columns

In [21]:
```
df.groupby('STNAME').agg({'CENSUS2010POP': np.average})
```

```
C:\Users\jayvant\anaconda3\lib\site-packages\numpy\lib\function_base.py:393: Run
timeWarning: Mean of empty slice.
  avg = a.mean(axis)
C:\Users\jayvant\anaconda3\lib\site-packages\numpy\core\_methods.py:161: Runtime
Warning: invalid value encountered in double_scalars
  ret = ret.dtype.type(ret / rcount)
```

Out[21]:

STNAME	CENSUS2010POP
Alabama	71339.343284
Alaska	24490.724138
Arizona	426134.466667
Arkansas	38878.906667
California	642309.586207
Colorado	78581.187500
Connecticut	446762.125000
Delaware	299311.333333
District of Columbia	601723.000000
Florida	280616.567164
Georgia	60928.635220
Hawaii	272060.200000
Idaho	35626.863636
Illinois	125790.509804
Indiana	70476.108696
Iowa	30771.262626
Kansas	27172.552381
Kentucky	36161.391667
Louisiana	70833.937500
Maine	83022.562500
Maryland	240564.666667
Massachusetts	467687.785714
Michigan	119080.000000
Minnesota	60964.655172
Mississippi	36186.548780
Missouri	52077.626087
Montana	17668.125000
Nebraska	19638.075269
Nevada	158855.941176
New Hampshire	131647.000000
New Jersey	418661.619048
New Mexico	62399.363636
New York	312550.032258
North Carolina	95354.830000
North Dakota	12690.396226
Ohio	131096.636364
Oklahoma	48718.844156
Oregon	106418.722222
Pennsylvania	189587.746269

In [22]:
```
print(type(df.groupby(level=0)['POPESTIMATE2010','POPESTIMATE2011']))
print(type(df.groupby(level=0)['POPESTIMATE2010']))
```

```
<class 'pandas.core.groupby.generic.DataFrameGroupBy'>
<class 'pandas.core.groupby.generic.SeriesGroupBy'>

C:\Users\jayvant\anaconda3\lib\site-packages\ipykernel_launcher.py:1: FutureWarn
ing: Indexing with multiple keys (implicitly converted to a tuple of keys) will
be deprecated, use a list instead.
  """Entry point for launching an IPython kernel.
```

Scales

In [26]:
```
df = pd.DataFrame(['A+', 'A', 'A-', 'B+', 'B', 'B-', 'C+', 'C', 'C-', 'D+', 'D'],
                  index=['excellent', 'excellent', 'excellent', 'good', 'good', 'go
od', 'ok', 'ok', 'ok', 'poor', 'poor'])
df.rename(columns={0: 'Grades'}, inplace=True)
df
```

Out[26]:

	Grades
excellent	A+
excellent	A
excellent	A-
good	B+
good	B
good	B-
ok	C+
ok	C
ok	C-
poor	D+
poor	D

In [27]:
```
df['Grades'].astype('category').head()
```

Out[27]:
```
excellent    A+
excellent     A
excellent    A-
good         B+
good          B
Name: Grades, dtype: category
Categories (11, object): [A, A+, A-, B, ..., C+, C-, D, D+]
```

Pivot Tables

In [35]:
```
df = pd.read_csv('f:/Python_Programs/cars.csv')
```

In [36]:
```
df.head()
```

Out[36]:

	YEAR	Make	Model	Size	(kW)	Unnamed: 5	TYPE	CITY (kWh/100 km)	HWY (kWh/100 km)	COMB (kWh/100 km)	C (Le/
0	2012	MITSUBISHI	i-MiEV	SUBCOMPACT	49	A1	B	16.9	21.4	18.7	
1	2012	NISSAN	LEAF	MID-SIZE	80	A1	B	19.3	23.0	21.1	
2	2013	FORD	FOCUS ELECTRIC	COMPACT	107	A1	B	19.0	21.1	20.0	
3	2013	MITSUBISHI	i-MiEV	SUBCOMPACT	49	A1	B	16.9	21.4	18.7	
4	2013	NISSAN	LEAF	MID-SIZE	80	A1	B	19.3	23.0	21.1	

In [37]:
```
df.pivot_table(values='(kW)', index='YEAR', columns='Make', aggfunc=np.mean)
```

Out[37]:

Make	BMW	CHEVROLET	FORD	KIA	MITSUBISHI	NISSAN	SMART	TESLA
YEAR								
2012	NaN	NaN	NaN	NaN	49.0	80.0	NaN	NaN
2013	NaN	NaN	107.0	NaN	49.0	80.0	35.0	280.000000
2014	NaN	104.0	107.0	NaN	49.0	80.0	35.0	268.333333
2015	125.0	104.0	107.0	81.0	49.0	80.0	35.0	320.666667
2016	125.0	104.0	107.0	81.0	49.0	80.0	35.0	409.700000

In [38]:
```
df.pivot_table(values='(kW)', index='YEAR', columns='Make', aggfunc=[np.mean,np.mi
n], margins=True)
```

Out[38]:

	mean									amin	
Make	**BMW**	**CHEVROLET**	**FORD**	**KIA**	**MITSUBISHI**	**NISSAN**	**SMART**	**TESLA**	**All**	**BMW**	**CHEVR**
YEAR											
2012	NaN	NaN	NaN	NaN	49.0	80.0	NaN	NaN	64.500000	NaN	
2013	NaN	NaN	107.0	NaN	49.0	80.0	35.0	280.000000	158.444444	NaN	
2014	NaN	104.0	107.0	NaN	49.0	80.0	35.0	268.333333	135.000000	NaN	
2015	125.0	104.0	107.0	81.0	49.0	80.0	35.0	320.666667	181.428571	125.0	
2016	125.0	104.0	107.0	81.0	49.0	80.0	35.0	409.700000	252.263158	125.0	
All	125.0	104.0	107.0	81.0	49.0	80.0	35.0	345.478261	190.622642	125.0	

Date Functionality in Pandas

In [39]:
```
import pandas as pd
import numpy as np
```

Timestamp

In [40]:
```
pd.Timestamp('9/1/2016 10:05AM')
```

Out[40]:
```
Timestamp('2016-09-01 10:05:00')
```

Period

In [41]:
```
pd.Period('1/2016')
```

Out[41]:
```
Period('2016-01', 'M')
```

In [42]:
```
pd.Period('3/5/2016')
```

Out[42]:
```
Period('2016-03-05', 'D')
```

In [43]:
```
pd.Period('2020')
```

Out[43]:
```
Period('2020', 'A-DEC')
```

In [44]:
```
pd.Period('2020/01')
```

Out[44]:
```
Period('2020-01', 'M')
```

In [45]:
```
pd.Period('2020/Jan')
```

Out[45]:
```
Period('2020-01', 'M')
```

DatetimeIndex

In [46]:
```
t1 = pd.Series(list('abc'), [pd.Timestamp('2016-09-01'), pd.Timestamp('2016-09-02
'), pd.Timestamp('2016-09-03')])
t1
```

Out[46]:
```
2016-09-01    a
2016-09-02    b
2016-09-03    c
dtype: object
```

In [47]:
```
type(t1.index)
```

Out[47]:
```
pandas.core.indexes.datetimes.DatetimeIndex
```

Converting to Datetime

In [62]:
```
d1 = ['2 June 2020', 'Aug 29, 2014', '2015-06-26', '7/12/16']
ts3 = pd.DataFrame(np.random.randint(10, 100, (4,2)), index=d1, columns=list('ab'))
ts3
```

Out[62]:

	a	b
2 June 2020	67	36
Aug 29, 2014	74	45
2015-06-26	91	36
7/12/16	78	96

In [63]:
```
ts3.index = pd.to_datetime(ts3.index)
ts3
```

Out[63]:

	a	b
2020-06-02	67	36
2014-08-29	74	45
2015-06-26	91	36
2016-07-12	78	96

In [64]:
```
pd.to_datetime('4.7.12', dayfirst=True)
```

Out[64]:
```
Timestamp('2012-07-04 00:00:00')
```

Timedeltas

In [65]:
```
pd.Timestamp('9/3/2020')-pd.Timestamp('9/1/2020')
```

Out[65]:
```
Timedelta('2 days 00:00:00')
```

In [66]:
```
pd.Timestamp('9/2/2020 8:10AM') + pd.Timedelta('12D 3H')
```

Out[66]:
```
Timestamp('2020-09-14 11:10:00')
```

Working with Dates in a Dataframe

In [67]:
```
dates = pd.date_range('10-01-2020', periods=9, freq='2W-SUN')
dates
```

Out[67]:
```
DatetimeIndex(['2020-10-04', '2020-10-18', '2020-11-01', '2020-11-15',
               '2020-11-29', '2020-12-13', '2020-12-27', '2021-01-10',
               '2021-01-24'],
              dtype='datetime64[ns]', freq='2W-SUN')
```

In [68]:
```
df = pd.DataFrame({'Count 1': 100 + np.random.randint(-5, 10, 9).cumsum(),
                   'Count 2': 120 + np.random.randint(-5, 10, 9)}, index=dates)
df
```

Out[68]:

	Count 1	Count 2
2020-10-04	98	129
2020-10-18	96	125
2020-11-01	96	117
2020-11-15	91	116
2020-11-29	93	117
2020-12-13	96	128
2020-12-27	101	120
2021-01-10	100	127
2021-01-24	99	122

In [69]:
```
df.diff()
```

Out[69]:

	Count 1	Count 2
2020-10-04	NaN	NaN
2020-10-18	-2.0	-4.0
2020-11-01	0.0	-8.0
2020-11-15	-5.0	-1.0
2020-11-29	2.0	1.0
2020-12-13	3.0	11.0
2020-12-27	5.0	-8.0
2021-01-10	-1.0	7.0
2021-01-24	-1.0	-5.0

In [70]:
```
df.resample('M').mean()
```

Out[70]:

	Count 1	Count 2
2020-10-31	97.000000	127.000000
2020-11-30	93.333333	116.666667
2020-12-31	98.500000	124.000000
2021-01-31	99.500000	124.500000

In [71]:
```
df['2020']
```

Out[71]:

	Count 1	Count 2
2020-10-04	98	129
2020-10-18	96	125
2020-11-01	96	117
2020-11-15	91	116
2020-11-29	93	117
2020-12-13	96	128
2020-12-27	101	120

In [72]:
```
df['2020-12']
```

Out[72]:

	Count 1	Count 2
2020-12-13	96	128
2020-12-27	101	120

In [73]:
```
df['2020-12':]
```

Out[73]:

	Count 1	Count 2
2020-12-13	96	128
2020-12-27	101	120
2021-01-10	100	127
2021-01-24	99	122

In [74]:
```
df.asfreq('W', method='ffill')
```

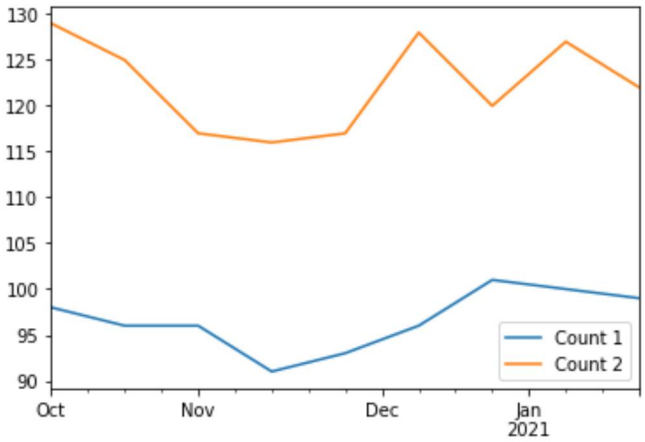
Out[74]:

	Count 1	Count 2
2020-10-04	98	129
2020-10-11	98	129
2020-10-18	96	125
2020-10-25	96	125
2020-11-01	96	117
2020-11-08	96	117
2020-11-15	91	116
2020-11-22	91	116
2020-11-29	93	117
2020-12-06	93	117
2020-12-13	96	128
2020-12-20	96	128
2020-12-27	101	120
2021-01-03	101	120
2021-01-10	100	127
2021-01-17	100	127
2021-01-24	99	122

In [75]:
```
import matplotlib.pyplot as plt
%matplotlib inline

df.plot()
```

Out[75]:
```
<matplotlib.axes._subplots.AxesSubplot at 0xa052fb0>
```



In [ ]:
```

```