*You are currently looking at **version 1.0** of this notebook. To download notebooks and datafiles, as well as get help on Jupyter notebooks in the Coursera platform, visit the [Jupyter Notebook FAQ] course resource.*

# Merging Dataframes

In [1]:

```
import pandas as pd

df = pd.DataFrame([{'Name': 'Chris', 'Item Purchased': 'Sponge', 'Cost': 22.50},
                   {'Name': 'Kevyn', 'Item Purchased': 'Kitty Litter', 'Cost': 2.50},
                   {'Name': 'Filip', 'Item Purchased': 'Spoon', 'Cost': 5.00}],
                  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 [4]:

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

Out[4]:

| | Cost | Item Purchased | Name | Date | Delivered | Feedback |
|---|---|---|---|---|---|---|

| | 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'})
adf
```

Out[6]:

| | 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 [7]:

```
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 [8]:

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

Out[8]:

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

In [9]:

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

Out[9]:

| Name | Role | School |
|---|---|---|
| Sally | Course liasion | Engineering |
| James | Grader | Business |

In [10]:

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

Out[10]:

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

In [11]:

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

Out[11]:

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

In [12]:

```
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[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': 'Washin
gton Avenue'},
                         {'Name': 'James', 'Role': 'Grader', 'Location': 'Washington Ave
nue'}])
student_df = pd.DataFrame([{'Name': 'James', 'School': 'Business', 'Location': '1024 Bil
liard Avenue'},
                           {'Name': 'Mike', 'School': 'Law', 'Location': 'Fraternity Hou
se #22'},
                           {'Name': 'Sally', 'School': 'Engineering', 'Location': '512 W
ilson Crescent'}])
pd.merge(staff_df, student_df, how='left', left_on='Name', right_on='Name')
```

Out[13]:

| | Location_x | Name | Role | Location_y | School |
|---|---|---|---|---|---|

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

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

Out[14]:

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

# Idiomatic Pandas: Making Code Pandorable

In [20]:

```
import pandas as pd
df = pd.read_csv('census.csv')
df
```

Out[20]:

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPE |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 40 | 3 | 6 | 1 | 0 | Alabama | Alabama | 4779736 | 4780127 | |
| 1 | 50 | 3 | 6 | 1 | 1 | Alabama | Autauga County | 54571 | 54571 | |
| 2 | 50 | 3 | 6 | 1 | 3 | Alabama | Baldwin County | 182265 | 182265 | |
| 3 | 50 | 3 | 6 | 1 | 5 | Alabama | Barbour County | 27457 | 27457 | |
| 4 | 50 | 3 | 6 | 1 | 7 | Alabama | Bibb County | 22915 | 22919 | |
| 5 | 50 | 3 | 6 | 1 | 9 | Alabama | Blount County | 57322 | 57322 | |
| 6 | 50 | 3 | 6 | 1 | 11 | Alabama | Bullock County | 10914 | 10915 | |
| 7 | 50 | 3 | 6 | 1 | 13 | Alabama | Butler County | 20947 | 20946 | |
| 8 | 50 | 3 | 6 | 1 | 15 | Alabama | Calhoun County | 118572 | 118586 | |
| 9 | 50 | 3 | 6 | 1 | 17 | Alabama | Chambers County | 34215 | 34170 | |
| 10 | 50 | 3 | 6 | 1 | 19 | Alabama | Cherokee County | 25989 | 25986 | |

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPE |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 50 | 3 | 6 | 1 | 21 | Alabama | Chilton County | 43643 | 43631 | |
| 12 | 50 | 3 | 6 | 1 | 23 | Alabama | Choctaw County | 13859 | 13858 | |
| 13 | 50 | 3 | 6 | 1 | 25 | Alabama | Clarke County | 25833 | 25840 | |
| 14 | 50 | 3 | 6 | 1 | 27 | Alabama | Clay County | 13932 | 13932 | |
| 15 | 50 | 3 | 6 | 1 | 29 | Alabama | Cleburne County | 14972 | 14972 | |
| 16 | 50 | 3 | 6 | 1 | 31 | Alabama | Coffee County | 49948 | 49948 | |
| 17 | 50 | 3 | 6 | 1 | 33 | Alabama | Colbert County | 54428 | 54428 | |
| 18 | 50 | 3 | 6 | 1 | 35 | Alabama | Conecuh County | 13228 | 13228 | |
| 19 | 50 | 3 | 6 | 1 | 37 | Alabama | Coosa County | 11539 | 11758 | |
| 20 | 50 | 3 | 6 | 1 | 39 | Alabama | Covington County | 37765 | 37765 | |
| 21 | 50 | 3 | 6 | 1 | 41 | Alabama | Crenshaw County | 13906 | 13906 | |
| 22 | 50 | 3 | 6 | 1 | 43 | Alabama | Cullman County | 80406 | 80410 | |
| 23 | 50 | 3 | 6 | 1 | 45 | Alabama | Dale County | 50251 | 50251 | |
| 24 | 50 | 3 | 6 | 1 | 47 | Alabama | Dallas County | 43820 | 43820 | |
| 25 | 50 | 3 | 6 | 1 | 49 | Alabama | DeKalb County | 71109 | 71115 | |
| 26 | 50 | 3 | 6 | 1 | 51 | Alabama | Elmore County | 79303 | 79296 | |
| 27 | 50 | 3 | 6 | 1 | 53 | Alabama | Escambia County | 38319 | 38319 | |
| 28 | 50 | 3 | 6 | 1 | 55 | Alabama | Etowah County | 104430 | 104427 | |
| 29 | 50 | 3 | 6 | 1 | 57 | Alabama | Fayette County | 17241 | 17241 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 3163 | 50 | 2 | 3 | 55 | 131 | Wisconsin | Washington County | 131887 | 131885 | |
| 3164 | 50 | 2 | 3 | 55 | 133 | Wisconsin | Waukesha County | 389891 | 389938 | |
| 3165 | 50 | 2 | 3 | 55 | 135 | Wisconsin | Waupaca County | 52410 | 52410 | |
| 3166 | 50 | 2 | 3 | 55 | 137 | Wisconsin | Waushara County | 24496 | 24496 | |
| 3167 | 50 | 2 | 3 | 55 | 139 | Wisconsin | Winnebago County | 166994 | 166994 | |
| 3168 | 50 | 2 | 3 | 55 | 141 | Wisconsin | Wood County | 74749 | 74749 | |
| 3169 | 40 | 4 | 8 | 56 | 0 | Wyoming | Wyoming | 563626 | 563767 | |
| 3170 | 50 | 4 | 8 | 56 | 1 | Wyoming | Albany County | 36299 | 36299 | |
| 3171 | 50 | 4 | 8 | 56 | 3 | Wyoming | Big Horn County | 11668 | 11668 | |

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPE |
|---|---|---|---|---|---|---|---|---|---|---|
| 3172 | 50 | 4 | 8 | 56 | 5 | Wyoming | Campbell County | 46133 | 46133 | |
| 3173 | 50 | 4 | 8 | 56 | 7 | Wyoming | Carbon County | 15885 | 15885 | |
| 3174 | 50 | 4 | 8 | 56 | 9 | Wyoming | Converse County | 13833 | 13833 | |
| 3175 | 50 | 4 | 8 | 56 | 11 | Wyoming | Crook County | 7083 | 7083 | |
| 3176 | 50 | 4 | 8 | 56 | 13 | Wyoming | Fremont County | 40123 | 40123 | |
| 3177 | 50 | 4 | 8 | 56 | 15 | Wyoming | Goshen County | 13249 | 13247 | |
| 3178 | 50 | 4 | 8 | 56 | 17 | Wyoming | Hot Springs County | 4812 | 4812 | |
| 3179 | 50 | 4 | 8 | 56 | 19 | Wyoming | Johnson County | 8569 | 8569 | |
| 3180 | 50 | 4 | 8 | 56 | 21 | Wyoming | Laramie County | 91738 | 91881 | |
| 3181 | 50 | 4 | 8 | 56 | 23 | Wyoming | Lincoln County | 18106 | 18106 | |
| 3182 | 50 | 4 | 8 | 56 | 25 | Wyoming | Natrona County | 75450 | 75450 | |
| 3183 | 50 | 4 | 8 | 56 | 27 | Wyoming | Niobrara County | 2484 | 2484 | |
| 3184 | 50 | 4 | 8 | 56 | 29 | Wyoming | Park County | 28205 | 28205 | |
| 3185 | 50 | 4 | 8 | 56 | 31 | Wyoming | Platte County | 8667 | 8667 | |
| 3186 | 50 | 4 | 8 | 56 | 33 | Wyoming | Sheridan County | 29116 | 29116 | |
| 3187 | 50 | 4 | 8 | 56 | 35 | Wyoming | Sublette County | 10247 | 10247 | |
| 3188 | 50 | 4 | 8 | 56 | 37 | Wyoming | Sweetwater County | 43806 | 43806 | |
| 3189 | 50 | 4 | 8 | 56 | 39 | Wyoming | Teton County | 21294 | 21294 | |
| 3190 | 50 | 4 | 8 | 56 | 41 | Wyoming | Uinta County | 21118 | 21118 | |
| 3191 | 50 | 4 | 8 | 56 | 43 | Wyoming | Washakie County | 8533 | 8533 | |
| 3192 | 50 | 4 | 8 | 56 | 45 | Wyoming | Weston County | 7208 | 7208 | |

3193 rows × 100 columns

In [16]:

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

Out[16]:

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | Estimates Base 2010 | POPESTIMATE2010 | PO |
|---|---|---|---|---|---|---|---|---|---|
| STNAME | CTYNAME | | | | | | | | |

| STNAME | CTYNAME | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | Estimates Base 2010 | POPESTIMATE2010 | PO |
|---|---|---|---|---|---|---|---|---|---|---|
| Alabama | Autauga County | 50.0 | 3.0 | 6.0 | 1.0 | 1.0 | 54571.0 | 54571.0 | 54660.0 | |
| | Baldwin County | 50.0 | 3.0 | 6.0 | 1.0 | 3.0 | 182265.0 | 182265.0 | 183193.0 | |
| | Barbour County | 50.0 | 3.0 | 6.0 | 1.0 | 5.0 | 27457.0 | 27457.0 | 27341.0 | |
| | Bibb County | 50.0 | 3.0 | 6.0 | 1.0 | 7.0 | 22915.0 | 22919.0 | 22861.0 | |
| | Blount County | 50.0 | 3.0 | 6.0 | 1.0 | 9.0 | 57322.0 | 57322.0 | 57373.0 | |
| | Bullock County | 50.0 | 3.0 | 6.0 | 1.0 | 11.0 | 10914.0 | 10915.0 | 10887.0 | |
| | Butler County | 50.0 | 3.0 | 6.0 | 1.0 | 13.0 | 20947.0 | 20946.0 | 20944.0 | |
| | Calhoun County | 50.0 | 3.0 | 6.0 | 1.0 | 15.0 | 118572.0 | 118586.0 | 118437.0 | |
| | Chambers County | 50.0 | 3.0 | 6.0 | 1.0 | 17.0 | 34215.0 | 34170.0 | 34098.0 | |
| | Cherokee County | 50.0 | 3.0 | 6.0 | 1.0 | 19.0 | 25989.0 | 25986.0 | 25976.0 | |
| | Chilton County | 50.0 | 3.0 | 6.0 | 1.0 | 21.0 | 43643.0 | 43631.0 | 43665.0 | |
| | Choctaw County | 50.0 | 3.0 | 6.0 | 1.0 | 23.0 | 13859.0 | 13858.0 | 13841.0 | |
| | Clarke County | 50.0 | 3.0 | 6.0 | 1.0 | 25.0 | 25833.0 | 25840.0 | 25767.0 | |
| | Clay County | 50.0 | 3.0 | 6.0 | 1.0 | 27.0 | 13932.0 | 13932.0 | 13880.0 | |
| | Cleburne County | 50.0 | 3.0 | 6.0 | 1.0 | 29.0 | 14972.0 | 14972.0 | 14973.0 | |
| | Coffee County | 50.0 | 3.0 | 6.0 | 1.0 | 31.0 | 49948.0 | 49948.0 | 50177.0 | |
| | Colbert County | 50.0 | 3.0 | 6.0 | 1.0 | 33.0 | 54428.0 | 54428.0 | 54514.0 | |
| | Conecuh County | 50.0 | 3.0 | 6.0 | 1.0 | 35.0 | 13228.0 | 13228.0 | 13208.0 | |
| | Coosa County | 50.0 | 3.0 | 6.0 | 1.0 | 37.0 | 11539.0 | 11758.0 | 11758.0 | |
| | Covington County | 50.0 | 3.0 | 6.0 | 1.0 | 39.0 | 37765.0 | 37765.0 | 37796.0 | |
| | Crenshaw County | 50.0 | 3.0 | 6.0 | 1.0 | 41.0 | 13906.0 | 13906.0 | 13853.0 | |
| | Cullman County | 50.0 | 3.0 | 6.0 | 1.0 | 43.0 | 80406.0 | 80410.0 | 80473.0 | |
| | Dale County | 50.0 | 3.0 | 6.0 | 1.0 | 45.0 | 50251.0 | 50251.0 | 50358.0 | |
| | Dallas County | 50.0 | 3.0 | 6.0 | 1.0 | 47.0 | 43820.0 | 43820.0 | 43803.0 | |
| | DeKalb County | 50.0 | 3.0 | 6.0 | 1.0 | 49.0 | 71109.0 | 71115.0 | 71142.0 | |
| | Elmore County | 50.0 | 3.0 | 6.0 | 1.0 | 51.0 | 79303.0 | 79296.0 | 79465.0 | |
| | Escambia County | 50.0 | 3.0 | 6.0 | 1.0 | 53.0 | 38319.0 | 38319.0 | 38309.0 | |
| | Etowah County | 50.0 | 3.0 | 6.0 | 1.0 | 55.0 | 104430.0 | 104427.0 | 104442.0 | |
| | Fayette | | | | | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | County | 50.0 | 3.0 | 6.0 | 1.0 | 57.0 | 17241.0 | 17241.0 | 17231.0 | |
| | Franklin County | SUMLEV 50.0 | REGION 3.0 | DIVISION 6.0 | STATE 1.0 | COUNTY 59.0 | CENSUS2010POP 31704.0 | Estimates Base 2010 31709.0 | POPESTIMATE2010 31734.0 | PO |
| STNAME ... | CTYNAME ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| Wisconsin | Washburn County | 50.0 | 2.0 | 3.0 | 55.0 | 129.0 | 15911.0 | 15911.0 | 15930.0 | |
| | Washington County | 50.0 | 2.0 | 3.0 | 55.0 | 131.0 | 131887.0 | 131885.0 | 131967.0 | |
| | Waukesha County | 50.0 | 2.0 | 3.0 | 55.0 | 133.0 | 389891.0 | 389938.0 | 390076.0 | |
| | Waupaca County | 50.0 | 2.0 | 3.0 | 55.0 | 135.0 | 52410.0 | 52410.0 | 52422.0 | |
| | Waushara County | 50.0 | 2.0 | 3.0 | 55.0 | 137.0 | 24496.0 | 24496.0 | 24506.0 | |
| | Winnebago County | 50.0 | 2.0 | 3.0 | 55.0 | 139.0 | 166994.0 | 166994.0 | 167059.0 | |
| | Wood County | 50.0 | 2.0 | 3.0 | 55.0 | 141.0 | 74749.0 | 74749.0 | 74807.0 | |
| Wyoming | Albany County | 50.0 | 4.0 | 8.0 | 56.0 | 1.0 | 36299.0 | 36299.0 | 36428.0 | |
| | Big Horn County | 50.0 | 4.0 | 8.0 | 56.0 | 3.0 | 11668.0 | 11668.0 | 11672.0 | |
| | Campbell County | 50.0 | 4.0 | 8.0 | 56.0 | 5.0 | 46133.0 | 46133.0 | 46244.0 | |
| | Carbon County | 50.0 | 4.0 | 8.0 | 56.0 | 7.0 | 15885.0 | 15885.0 | 15837.0 | |
| | Converse County | 50.0 | 4.0 | 8.0 | 56.0 | 9.0 | 13833.0 | 13833.0 | 13826.0 | |
| | Crook County | 50.0 | 4.0 | 8.0 | 56.0 | 11.0 | 7083.0 | 7083.0 | 7114.0 | |
| | Fremont County | 50.0 | 4.0 | 8.0 | 56.0 | 13.0 | 40123.0 | 40123.0 | 40222.0 | |
| | Goshen County | 50.0 | 4.0 | 8.0 | 56.0 | 15.0 | 13249.0 | 13247.0 | 13408.0 | |
| | Hot Springs County | 50.0 | 4.0 | 8.0 | 56.0 | 17.0 | 4812.0 | 4812.0 | 4813.0 | |
| | Johnson County | 50.0 | 4.0 | 8.0 | 56.0 | 19.0 | 8569.0 | 8569.0 | 8581.0 | |
| | Laramie County | 50.0 | 4.0 | 8.0 | 56.0 | 21.0 | 91738.0 | 91881.0 | 92271.0 | |
| | Lincoln County | 50.0 | 4.0 | 8.0 | 56.0 | 23.0 | 18106.0 | 18106.0 | 18091.0 | |
| | Natrona County | 50.0 | 4.0 | 8.0 | 56.0 | 25.0 | 75450.0 | 75450.0 | 75472.0 | |
| | Niobrara County | 50.0 | 4.0 | 8.0 | 56.0 | 27.0 | 2484.0 | 2484.0 | 2492.0 | |
| | Park County | 50.0 | 4.0 | 8.0 | 56.0 | 29.0 | 28205.0 | 28205.0 | 28259.0 | |
| | Platte County | 50.0 | 4.0 | 8.0 | 56.0 | 31.0 | 8667.0 | 8667.0 | 8678.0 | |
| | Sheridan County | 50.0 | 4.0 | 8.0 | 56.0 | 33.0 | 29116.0 | 29116.0 | 29146.0 | |
| | Sublette County | 50.0 | 4.0 | 8.0 | 56.0 | 35.0 | 10247.0 | 10247.0 | 10244.0 | |
| | Sweetwater County | 50.0 | 4.0 | 8.0 | 56.0 | 37.0 | 43806.0 | 43806.0 | 43593.0 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Teton County | 50.0 | 4.0 | 8.0 | 56.0 | 39.0 | 21294.0 | 21294.0 | 21297.0 |
| | Uinta County | 50.0 | 4.0 | 8.0 | 56.0 | 41.0 | 21118.0 | 21118.0 | 21102.0 |
| | Washakie County | 50.0 | 4.0 | 8.0 | 56.0 | 43.0 | 8533.0 | 8533.0 | 8545.0 |
| | Weston County | 50.0 | 4.0 | 8.0 | 56.0 | 45.0 | 7208.0 | 7208.0 | 7181.0 |

3142 rows × 98 columns

In [17]:

```
#print(df.drop(df[df['Quantity'] == 0].index).rename(columns={'Weight': 'Weight (oz.)'}))
```

In [21]:

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

Out[21]:

| | | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | Estimates Base 2010 | POPESTIMATE2010 | PO |
|---|---|---|---|---|---|---|---|---|---|---|
| STNAME | CTYNAME | | | | | | | | | |
| Alabama | Autauga County | 50 | 3 | 6 | 1 | 1 | 54571 | 54571 | 54660 | |
| | Baldwin County | 50 | 3 | 6 | 1 | 3 | 182265 | 182265 | 183193 | |
| | Barbour County | 50 | 3 | 6 | 1 | 5 | 27457 | 27457 | 27341 | |
| | Bibb County | 50 | 3 | 6 | 1 | 7 | 22915 | 22919 | 22861 | |
| | Blount County | 50 | 3 | 6 | 1 | 9 | 57322 | 57322 | 57373 | |
| | Bullock County | 50 | 3 | 6 | 1 | 11 | 10914 | 10915 | 10887 | |
| | Butler County | 50 | 3 | 6 | 1 | 13 | 20947 | 20946 | 20944 | |
| | Calhoun County | 50 | 3 | 6 | 1 | 15 | 118572 | 118586 | 118437 | |
| | Chambers County | 50 | 3 | 6 | 1 | 17 | 34215 | 34170 | 34098 | |
| | Cherokee County | 50 | 3 | 6 | 1 | 19 | 25989 | 25986 | 25976 | |
| | Chilton County | 50 | 3 | 6 | 1 | 21 | 43643 | 43631 | 43665 | |
| | Choctaw County | 50 | 3 | 6 | 1 | 23 | 13859 | 13858 | 13841 | |
| | Clarke County | 50 | 3 | 6 | 1 | 25 | 25833 | 25840 | 25767 | |
| | Clay County | 50 | 3 | 6 | 1 | 27 | 13932 | 13932 | 13880 | |
| | Cleburne County | 50 | 3 | 6 | 1 | 29 | 14972 | 14972 | 14973 | |
| | Coffee County | 50 | 3 | 6 | 1 | 31 | 49948 | 49948 | 50177 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Colbert County | 50 | 3 | 6 | 1 | 33 | 54428 | Estimates 54429 | | 54514 |
| STNAME | CTYNAME | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | Base 2010 | POPESTIMATE2010 | PO |
| | Conecuh County | 50 | 3 | 6 | 1 | 35 | 13228 | 13228 | 13208 | |
| | Coosa County | 50 | 3 | 6 | 1 | 37 | 11539 | 11758 | 11758 | |
| | Covington County | 50 | 3 | 6 | 1 | 39 | 37765 | 37765 | 37796 | |
| | Crenshaw County | 50 | 3 | 6 | 1 | 41 | 13906 | 13906 | 13853 | |
| | Cullman County | 50 | 3 | 6 | 1 | 43 | 80406 | 80410 | 80473 | |
| | Dale County | 50 | 3 | 6 | 1 | 45 | 50251 | 50251 | 50358 | |
| | Dallas County | 50 | 3 | 6 | 1 | 47 | 43820 | 43820 | 43803 | |
| | DeKalb County | 50 | 3 | 6 | 1 | 49 | 71109 | 71115 | 71142 | |
| | Elmore County | 50 | 3 | 6 | 1 | 51 | 79303 | 79296 | 79465 | |
| | Escambia County | 50 | 3 | 6 | 1 | 53 | 38319 | 38319 | 38309 | |
| | Etowah County | 50 | 3 | 6 | 1 | 55 | 104430 | 104427 | 104442 | |
| | Fayette County | 50 | 3 | 6 | 1 | 57 | 17241 | 17241 | 17231 | |
| | Franklin County | 50 | 3 | 6 | 1 | 59 | 31704 | 31709 | 31734 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| Wisconsin | Washburn County | 50 | 2 | 3 | 55 | 129 | 15911 | 15911 | 15930 | |
| | Washington County | 50 | 2 | 3 | 55 | 131 | 131887 | 131885 | 131967 | |
| | Waukesha County | 50 | 2 | 3 | 55 | 133 | 389891 | 389938 | 390076 | |
| | Waupaca County | 50 | 2 | 3 | 55 | 135 | 52410 | 52410 | 52422 | |
| | Waushara County | 50 | 2 | 3 | 55 | 137 | 24496 | 24496 | 24506 | |
| | Winnebago County | 50 | 2 | 3 | 55 | 139 | 166994 | 166994 | 167059 | |
| | Wood County | 50 | 2 | 3 | 55 | 141 | 74749 | 74749 | 74807 | |
| Wyoming | Albany County | 50 | 4 | 8 | 56 | 1 | 36299 | 36299 | 36428 | |
| | Big Horn County | 50 | 4 | 8 | 56 | 3 | 11668 | 11668 | 11672 | |
| | Campbell County | 50 | 4 | 8 | 56 | 5 | 46133 | 46133 | 46244 | |
| | Carbon County | 50 | 4 | 8 | 56 | 7 | 15885 | 15885 | 15837 | |
| | Converse County | 50 | 4 | 8 | 56 | 9 | 13833 | 13833 | 13826 | |
| | Crook County | 50 | 4 | 8 | 56 | 11 | 7083 | 7083 | 7114 | |
| | Fremont County | 50 | 4 | 8 | 56 | 13 | 40123 | 40123 | 40222 | |

| | | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | Estimates Base 2010 | POPESTIMATE2010 | PO |
|---|---|---|---|---|---|---|---|---|---|---|
| | Goshen County | 50 | 4 | 8 | 56 | 15 | 13249 | 13247 | 13408 | |
| STNAME | Hot Springs CTYNAME County | 50 | 4 | 8 | 56 | 17 | 4812 | 4812 | 4813 | |
| | Johnson County | 50 | 4 | 8 | 56 | 19 | 8569 | 8569 | 8581 | |
| | Laramie County | 50 | 4 | 8 | 56 | 21 | 91738 | 91881 | 92271 | |
| | Lincoln County | 50 | 4 | 8 | 56 | 23 | 18106 | 18106 | 18091 | |
| | Natrona County | 50 | 4 | 8 | 56 | 25 | 75450 | 75450 | 75472 | |
| | Niobrara County | 50 | 4 | 8 | 56 | 27 | 2484 | 2484 | 2492 | |
| | Park County | 50 | 4 | 8 | 56 | 29 | 28205 | 28205 | 28259 | |
| | Platte County | 50 | 4 | 8 | 56 | 31 | 8667 | 8667 | 8678 | |
| | Sheridan County | 50 | 4 | 8 | 56 | 33 | 29116 | 29116 | 29146 | |
| | Sublette County | 50 | 4 | 8 | 56 | 35 | 10247 | 10247 | 10244 | |
| | Sweetwater County | 50 | 4 | 8 | 56 | 37 | 43806 | 43806 | 43593 | |
| | Teton County | 50 | 4 | 8 | 56 | 39 | 21294 | 21294 | 21297 | |
| | Uinta County | 50 | 4 | 8 | 56 | 41 | 21118 | 21118 | 21102 | |
| | Washakie County | 50 | 4 | 8 | 56 | 43 | 8533 | 8533 | 8545 | |
| | Weston County | 50 | 4 | 8 | 56 | 45 | 7208 | 7208 | 7181 | |

3142 rows × 98 columns

In [22]:

```
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 [23]:

```
df.apply(min_max, axis=1)
```

Out[23]:

| | | 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 |

| STNAME | CTYNAME | min | max |
|---|---|---|---|
| | Bibb County | 22512.0 | 22861.0 |
| | Blount County | 57373.0 | 57776.0 |
| | Bullock County | 10606.0 | 10887.0 |
| | Butler County | 20154.0 | 20944.0 |
| | Calhoun County | 115620.0 | 118437.0 |
| | Chambers County | 33993.0 | 34153.0 |
| | Cherokee County | 25859.0 | 26084.0 |
| | Chilton County | 43665.0 | 43943.0 |
| | Choctaw County | 13170.0 | 13841.0 |
| | Clarke County | 24675.0 | 25767.0 |
| | Clay County | 13456.0 | 13880.0 |
| | Cleburne County | 14921.0 | 15072.0 |
| | Coffee County | 50177.0 | 51211.0 |
| | Colbert County | 54354.0 | 54514.0 |
| | Conecuh County | 12662.0 | 13208.0 |
| | Coosa County | 10724.0 | 11758.0 |
| | Covington County | 37796.0 | 38060.0 |
| | Crenshaw County | 13853.0 | 13963.0 |
| | Cullman County | 80374.0 | 82005.0 |
| | Dale County | 49501.0 | 50358.0 |
| | Dallas County | 41131.0 | 43803.0 |
| | DeKalb County | 70869.0 | 71387.0 |
| | Elmore County | 79465.0 | 81468.0 |
| | Escambia County | 37784.0 | 38309.0 |
| | Etowah County | 103057.0 | 104442.0 |
| | Fayette County | 16759.0 | 17231.0 |
| | Franklin County | 31507.0 | 31734.0 |
| ... | ... | ... | ... |
| Wisconsin | Washburn County | 15552.0 | 15930.0 |
| | Washington County | 131967.0 | 133674.0 |
| | Waukesha County | 390076.0 | 396488.0 |
| | Waupaca County | 51945.0 | 52422.0 |
| | Waushara County | 24033.0 | 24581.0 |
| | Winnebago County | 167059.0 | 169639.0 |
| | Wood County | 73435.0 | 74807.0 |
| Wyoming | Albany County | 36428.0 | 37956.0 |
| | Big Horn County | 11672.0 | 12022.0 |
| | Campbell County | 46244.0 | 49220.0 |
| | Carbon County | 15559.0 | 15856.0 |
| | Converse County | 13728.0 | 14343.0 |
| | Crook County | 7114.0 | 7444.0 |
| | Fremont County | 40222.0 | 41129.0 |
| | Goshen County | 13383.0 | 13666.0 |
| | Hot Springs County | 4741.0 | 4846.0 |

| | | min | max |
|---|---|---|---|
| | Johnson County | 8552.0 | 8636.0 |
| STNAME | CTYNAME | | |
| | Laramie County | 92271.0 | 97121.0 |
| | Lincoln County | 17943.0 | 18722.0 |
| | Natrona County | 75472.0 | 82178.0 |
| | Niobrara County | 2475.0 | 2548.0 |
| | Park County | 28259.0 | 29237.0 |
| | Platte County | 8678.0 | 8812.0 |
| | Sheridan County | 29146.0 | 30020.0 |
| | Sublette County | 9899.0 | 10418.0 |
| | 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 [24]:

```
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[24]:

| | | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | ESTIMATESBASE2010 | POPESTIMA |
|---|---|---|---|---|---|---|---|---|---|
| STNAME | CTYNAME | | | | | | | | |
| 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 | 1 |
| | 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 | |
| | Bullock County | 50.0 | 3.0 | 6.0 | 1.0 | 11.0 | 10914.0 | 10915.0 | |
| | Butler County | 50.0 | 3.0 | 6.0 | 1.0 | 13.0 | 20947.0 | 20946.0 | |
| | Calhoun County | 50.0 | 3.0 | 6.0 | 1.0 | 15.0 | 118572.0 | 118586.0 | 1 |
| | Chambers County | 50.0 | 3.0 | 6.0 | 1.0 | 17.0 | 34215.0 | 34170.0 | |
| | Cherokee County | 50.0 | 3.0 | 6.0 | 1.0 | 19.0 | 25989.0 | 25986.0 | |

| STNAME | CTYNAME | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | ESTIMATESBASE2010 | POPESTIMA |
|---|---|---|---|---|---|---|---|---|---|
| | Chilton County | 50.0 | 3.0 | 6.0 | 1.0 | 21.0 | 43643.0 | 43651.0 | |
| | Choctaw County | 50.0 | 3.0 | 6.0 | 1.0 | 23.0 | 13859.0 | 13858.0 | |
| | Clarke County | 50.0 | 3.0 | 6.0 | 1.0 | 25.0 | 25833.0 | 25840.0 | |
| | Clay County | 50.0 | 3.0 | 6.0 | 1.0 | 27.0 | 13932.0 | 13932.0 | |
| | Cleburne County | 50.0 | 3.0 | 6.0 | 1.0 | 29.0 | 14972.0 | 14972.0 | |
| | Coffee County | 50.0 | 3.0 | 6.0 | 1.0 | 31.0 | 49948.0 | 49948.0 | |
| | Colbert County | 50.0 | 3.0 | 6.0 | 1.0 | 33.0 | 54428.0 | 54428.0 | |
| | Conecuh County | 50.0 | 3.0 | 6.0 | 1.0 | 35.0 | 13228.0 | 13228.0 | |
| | Coosa County | 50.0 | 3.0 | 6.0 | 1.0 | 37.0 | 11539.0 | 11758.0 | |
| | Covington County | 50.0 | 3.0 | 6.0 | 1.0 | 39.0 | 37765.0 | 37765.0 | |
| | Crenshaw County | 50.0 | 3.0 | 6.0 | 1.0 | 41.0 | 13906.0 | 13906.0 | |
| | Cullman County | 50.0 | 3.0 | 6.0 | 1.0 | 43.0 | 80406.0 | 80410.0 | |
| | Dale County | 50.0 | 3.0 | 6.0 | 1.0 | 45.0 | 50251.0 | 50251.0 | |
| | Dallas County | 50.0 | 3.0 | 6.0 | 1.0 | 47.0 | 43820.0 | 43820.0 | |
| | DeKalb County | 50.0 | 3.0 | 6.0 | 1.0 | 49.0 | 71109.0 | 71115.0 | |
| | Elmore County | 50.0 | 3.0 | 6.0 | 1.0 | 51.0 | 79303.0 | 79296.0 | |
| | Escambia County | 50.0 | 3.0 | 6.0 | 1.0 | 53.0 | 38319.0 | 38319.0 | |
| | Etowah County | 50.0 | 3.0 | 6.0 | 1.0 | 55.0 | 104430.0 | 104427.0 | 1 |
| | Fayette County | 50.0 | 3.0 | 6.0 | 1.0 | 57.0 | 17241.0 | 17241.0 | |
| | Franklin County | 50.0 | 3.0 | 6.0 | 1.0 | 59.0 | 31704.0 | 31709.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| Wisconsin | Washburn County | 50.0 | 2.0 | 3.0 | 55.0 | 129.0 | 15911.0 | 15911.0 | |
| | Washington County | 50.0 | 2.0 | 3.0 | 55.0 | 131.0 | 131887.0 | 131885.0 | 1 |
| | Waukesha County | 50.0 | 2.0 | 3.0 | 55.0 | 133.0 | 389891.0 | 389938.0 | 3 |
| | Waupaca County | 50.0 | 2.0 | 3.0 | 55.0 | 135.0 | 52410.0 | 52410.0 | |
| | Waushara County | 50.0 | 2.0 | 3.0 | 55.0 | 137.0 | 24496.0 | 24496.0 | |
| | Winnebago County | 50.0 | 2.0 | 3.0 | 55.0 | 139.0 | 166994.0 | 166994.0 | 1 |
| | Wood County | 50.0 | 2.0 | 3.0 | 55.0 | 141.0 | 74749.0 | 74749.0 | |
| Wyoming | Albany County | 50.0 | 4.0 | 8.0 | 56.0 | 1.0 | 36299.0 | 36299.0 | |

| STNAME | CTYNAME | SUMLEV | REGION | DIVISION | STATE | COUNTY | CENSUS2010POP | ESTIMATESBASE2010 | POPESTIMA |
|---|---|---|---|---|---|---|---|---|---|
| | Big Horn County | 50.0 | 4.0 | 8.0 | 56.0 | 3.0 | 11668.0 | 11668.0 | |
| | Campbell County | 50.0 | 4.0 | 8.0 | 56.0 | 5.0 | 46133.0 | 46133.0 | |
| | Carbon County | 50.0 | 4.0 | 8.0 | 56.0 | 7.0 | 15885.0 | 15885.0 | |
| | Converse County | 50.0 | 4.0 | 8.0 | 56.0 | 9.0 | 13833.0 | 13833.0 | |
| | Crook County | 50.0 | 4.0 | 8.0 | 56.0 | 11.0 | 7083.0 | 7083.0 | |
| | Fremont County | 50.0 | 4.0 | 8.0 | 56.0 | 13.0 | 40123.0 | 40123.0 | |
| | Goshen County | 50.0 | 4.0 | 8.0 | 56.0 | 15.0 | 13249.0 | 13247.0 | |
| | Hot Springs County | 50.0 | 4.0 | 8.0 | 56.0 | 17.0 | 4812.0 | 4812.0 | |
| | Johnson County | 50.0 | 4.0 | 8.0 | 56.0 | 19.0 | 8569.0 | 8569.0 | |
| | Laramie County | 50.0 | 4.0 | 8.0 | 56.0 | 21.0 | 91738.0 | 91881.0 | |
| | Lincoln County | 50.0 | 4.0 | 8.0 | 56.0 | 23.0 | 18106.0 | 18106.0 | |
| | Natrona County | 50.0 | 4.0 | 8.0 | 56.0 | 25.0 | 75450.0 | 75450.0 | |
| | Niobrara County | 50.0 | 4.0 | 8.0 | 56.0 | 27.0 | 2484.0 | 2484.0 | |
| | Park County | 50.0 | 4.0 | 8.0 | 56.0 | 29.0 | 28205.0 | 28205.0 | |
| | Platte County | 50.0 | 4.0 | 8.0 | 56.0 | 31.0 | 8667.0 | 8667.0 | |
| | Sheridan County | 50.0 | 4.0 | 8.0 | 56.0 | 33.0 | 29116.0 | 29116.0 | |
| | Sublette County | 50.0 | 4.0 | 8.0 | 56.0 | 35.0 | 10247.0 | 10247.0 | |
| | 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 × 100 columns

In [25]:

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

Out[25]:

STNAME CTYNAME

```
STNAME      CTYNAME
Alabama     Autauga County          55347.0
            Baldwin County         203709.0
            Barbour County          27341.0
            Bibb County             22861.0
            Blount County           57776.0
            Bullock County          10887.0
            Butler County           20944.0
            Calhoun County         118437.0
            Chambers County         34153.0
            Cherokee County         26084.0
            Chilton County          43943.0
            Choctaw County          13841.0
            Clarke County           25767.0
            Clay County             13880.0
            Cleburne County         15072.0
            Coffee County           51211.0
            Colbert County          54514.0
            Conecuh County          13208.0
            Coosa County            11758.0
            Covington County        38060.0
            Crenshaw County         13963.0
            Cullman County          82005.0
            Dale County             50358.0
            Dallas County           43803.0
            DeKalb County           71387.0
            Elmore County           81468.0
            Escambia County         38309.0
            Etowah County          104442.0
            Fayette County          17231.0
            Franklin County         31734.0
                                      ...
Wisconsin   Washburn County         15930.0
            Washington County      133674.0
            Waukesha County        396488.0
            Waupaca County          52422.0
            Waushara County         24581.0
            Winnebago County       169639.0
            Wood County             74807.0
Wyoming     Albany County           37956.0
            Big Horn County         12022.0
            Campbell County         49220.0
            Carbon County           15856.0
            Converse County         14343.0
            Crook County             7444.0
            Fremont County          41129.0
            Goshen County           13666.0
            Hot Springs County       4846.0
            Johnson County           8636.0
            Laramie County          97121.0
            Lincoln County          18722.0
            Natrona County          82178.0
            Niobrara County          2548.0
            Park County             29237.0
            Platte County            8812.0
            Sheridan County         30020.0
            Sublette County         10418.0
            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 [26]:

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('census.csv')
df = df[df['SUMLEV']==50]
df
```

Out[26]:

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPE |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 50 | 3 | 6 | 1 | 1 | Alabama | Autauga County | 54571 | 54571 | |
| 2 | 50 | 3 | 6 | 1 | 3 | Alabama | Baldwin County | 182265 | 182265 | |
| 3 | 50 | 3 | 6 | 1 | 5 | Alabama | Barbour County | 27457 | 27457 | |
| 4 | 50 | 3 | 6 | 1 | 7 | Alabama | Bibb County | 22915 | 22919 | |
| 5 | 50 | 3 | 6 | 1 | 9 | Alabama | Blount County | 57322 | 57322 | |
| 6 | 50 | 3 | 6 | 1 | 11 | Alabama | Bullock County | 10914 | 10915 | |
| 7 | 50 | 3 | 6 | 1 | 13 | Alabama | Butler County | 20947 | 20946 | |
| 8 | 50 | 3 | 6 | 1 | 15 | Alabama | Calhoun County | 118572 | 118586 | |
| 9 | 50 | 3 | 6 | 1 | 17 | Alabama | Chambers County | 34215 | 34170 | |
| 10 | 50 | 3 | 6 | 1 | 19 | Alabama | Cherokee County | 25989 | 25986 | |
| 11 | 50 | 3 | 6 | 1 | 21 | Alabama | Chilton County | 43643 | 43631 | |
| 12 | 50 | 3 | 6 | 1 | 23 | Alabama | Choctaw County | 13859 | 13858 | |
| 13 | 50 | 3 | 6 | 1 | 25 | Alabama | Clarke County | 25833 | 25840 | |
| 14 | 50 | 3 | 6 | 1 | 27 | Alabama | Clay County | 13932 | 13932 | |
| 15 | 50 | 3 | 6 | 1 | 29 | Alabama | Cleburne County | 14972 | 14972 | |
| 16 | 50 | 3 | 6 | 1 | 31 | Alabama | Coffee County | 49948 | 49948 | |
| 17 | 50 | 3 | 6 | 1 | 33 | Alabama | Colbert County | 54428 | 54428 | |
| 18 | 50 | 3 | 6 | 1 | 35 | Alabama | Conecuh County | 13228 | 13228 | |
| 19 | 50 | 3 | 6 | 1 | 37 | Alabama | Coosa County | 11539 | 11758 | |
| 20 | 50 | 3 | 6 | 1 | 39 | Alabama | Covington County | 37765 | 37765 | |
| 21 | 50 | 3 | 6 | 1 | 41 | Alabama | Crenshaw County | 13906 | 13906 | |
| 22 | 50 | 3 | 6 | 1 | 43 | Alabama | Cullman County | 80406 | 80410 | |
| 23 | 50 | 3 | 6 | 1 | 45 | Alabama | Dale County | 50251 | 50251 | |
| 24 | 50 | 3 | 6 | 1 | 47 | Alabama | Dallas County | 43820 | 43820 | |
| 25 | 50 | 3 | 6 | 1 | 49 | Alabama | DeKalb County | 71109 | 71115 | |
| | | | | | | | Elmore | | | |

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPE |
|---|---|---|---|---|---|---|---|---|---|---|
| 26 | 50 | 3 | 6 | 1 | 51 | Alabama | Elmore County | 79303 | 79296 | |
| 27 | 50 | 3 | 6 | 1 | 53 | Alabama | Escambia County | 38319 | 38319 | |
| 28 | 50 | 3 | 6 | 1 | 55 | Alabama | Etowah County | 104430 | 104427 | |
| 29 | 50 | 3 | 6 | 1 | 57 | Alabama | Fayette County | 17241 | 17241 | |
| 30 | 50 | 3 | 6 | 1 | 59 | Alabama | Franklin County | 31704 | 31709 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 3162 | 50 | 2 | 3 | 55 | 129 | Wisconsin | Washburn County | 15911 | 15911 | |
| 3163 | 50 | 2 | 3 | 55 | 131 | Wisconsin | Washington County | 131887 | 131885 | |
| 3164 | 50 | 2 | 3 | 55 | 133 | Wisconsin | Waukesha County | 389891 | 389938 | |
| 3165 | 50 | 2 | 3 | 55 | 135 | Wisconsin | Waupaca County | 52410 | 52410 | |
| 3166 | 50 | 2 | 3 | 55 | 137 | Wisconsin | Waushara County | 24496 | 24496 | |
| 3167 | 50 | 2 | 3 | 55 | 139 | Wisconsin | Winnebago County | 166994 | 166994 | |
| 3168 | 50 | 2 | 3 | 55 | 141 | Wisconsin | Wood County | 74749 | 74749 | |
| 3170 | 50 | 4 | 8 | 56 | 1 | Wyoming | Albany County | 36299 | 36299 | |
| 3171 | 50 | 4 | 8 | 56 | 3 | Wyoming | Big Horn County | 11668 | 11668 | |
| 3172 | 50 | 4 | 8 | 56 | 5 | Wyoming | Campbell County | 46133 | 46133 | |
| 3173 | 50 | 4 | 8 | 56 | 7 | Wyoming | Carbon County | 15885 | 15885 | |
| 3174 | 50 | 4 | 8 | 56 | 9 | Wyoming | Converse County | 13833 | 13833 | |
| 3175 | 50 | 4 | 8 | 56 | 11 | Wyoming | Crook County | 7083 | 7083 | |
| 3176 | 50 | 4 | 8 | 56 | 13 | Wyoming | Fremont County | 40123 | 40123 | |
| 3177 | 50 | 4 | 8 | 56 | 15 | Wyoming | Goshen County | 13249 | 13247 | |
| 3178 | 50 | 4 | 8 | 56 | 17 | Wyoming | Hot Springs County | 4812 | 4812 | |
| 3179 | 50 | 4 | 8 | 56 | 19 | Wyoming | Johnson County | 8569 | 8569 | |
| 3180 | 50 | 4 | 8 | 56 | 21 | Wyoming | Laramie County | 91738 | 91881 | |
| 3181 | 50 | 4 | 8 | 56 | 23 | Wyoming | Lincoln County | 18106 | 18106 | |
| 3182 | 50 | 4 | 8 | 56 | 25 | Wyoming | Natrona County | 75450 | 75450 | |
| 3183 | 50 | 4 | 8 | 56 | 27 | Wyoming | Niobrara County | 2484 | 2484 | |
| 3184 | 50 | 4 | 8 | 56 | 29 | Wyoming | Park County | 28205 | 28205 | |
| 3185 | 50 | 4 | 8 | 56 | 31 | Wyoming | Platte County | 8667 | 8667 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3186 | 50 | 4 | 8 | 56 | 33 | Wyoming | Sheridan County | 29116 | 29116 | POPE |
| 3187 | 50 | 4 | 8 | 56 | 35 | Wyoming | Sublette County | 10247 | 10247 | |
| 3188 | 50 | 4 | 8 | 56 | 37 | Wyoming | Sweetwater County | 43806 | 43806 | |
| 3189 | 50 | 4 | 8 | 56 | 39 | Wyoming | Teton County | 21294 | 21294 | |
| 3190 | 50 | 4 | 8 | 56 | 41 | Wyoming | Uinta County | 21118 | 21118 | |
| 3191 | 50 | 4 | 8 | 56 | 43 | Wyoming | Washakie County | 8533 | 8533 | |
| 3192 | 50 | 4 | 8 | 56 | 45 | Wyoming | Weston County | 7208 | 7208 | |

**3142 rows × 100 columns**

In [ ]:

```
%%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(avg))
```

In [27]:

```
#%%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(avg))
```

```
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 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 100551.39130434782
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 [28]:

```
df.head()
```

Out[28]:

| | SUMLEV | REGION | DIVISION | STATE | COUNTY | STNAME | CTYNAME | CENSUS2010POP | ESTIMATESBASE2010 | POPESTIMA |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 50 | 3 | 6 | 1 | 1 | Alabama | Autauga County | 54571 | 54571 | |
| 2 | 50 | 3 | 6 | 1 | 3 | Alabama | Baldwin County | 182265 | 182265 | |
| 3 | 50 | 3 | 6 | 1 | 5 | Alabama | Barbour County | 27457 | 27457 | |
| 4 | 50 | 3 | 6 | 1 | 7 | Alabama | Bibb County | 22915 | 22919 | |
| 5 | 50 | 3 | 6 | 1 | 9 | Alabama | Blount County | 57322 | 57322 | |

5 rows × 100 columns

In [29]:

```
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) + ' for pro
cessing.')
```

```
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 [30]:

```
df = pd.read_csv('census.csv')
df = df[df['SUMLEV']==50]
```

In [31]:

```
df.groupby('STNAME').agg({'CENSUS2010POP': np.average})
```

Out[31]:

| | CENSUS2010POP |
|---|---|

| STNAME | CENSUS2010POP |
|---|---|
| STNAME | |
| 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 |
| Rhode Island | 210513.400000 |
| South Carolina | 100551.391304 |
| South Dakota | 12336.060606 |
| Tennessee | 66801.105263 |
| Texas | 98998.271654 |

| | CENSUS2010POP |
|---|---|
| Utah | 95306.379310 |
| STNAME | |
| Vermont | 44695.785714 |
| Virginia | 60111.293233 |
| Washington | 172424.102564 |
| West Virginia | 33690.800000 |
| Wisconsin | 78985.916667 |
| Wyoming | 24505.478261 |

In [32]:

```
#print(df.groupby('Category').apply(lambda df,a,b: sum(df[a] * df[b]), 'Weight (oz.)', 'Q
uantity'))


# Or alternatively without using a lambda:
# def totalweight(df, w, q):
#         return sum(df[w] * df[q])
#
# print(df.groupby('Category').apply(totalweight, 'Weight (oz.)', 'Quantity'))
```

In [33]:

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

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

In [34]:

```
(df.set_index('STNAME').groupby(level=0)['CENSUS2010POP']
    .agg({'avg': np.average, 'sum': np.sum}))
```

```
C:\Users\Parth\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: FutureWarning: using
a dict on a Series for aggregation
is deprecated and will be removed in a future version
```

Out[34]:

| | avg | sum |
|---|---|---|
| STNAME | | |
| Alabama | 71339.343284 | 4779736 |
| Alaska | 24490.724138 | 710231 |
| Arizona | 426134.466667 | 6392017 |
| Arkansas | 38878.906667 | 2915918 |
| California | 642309.586207 | 37253956 |
| Colorado | 78581.187500 | 5029196 |
| Connecticut | 446762.125000 | 3574097 |
| Delaware | 299311.333333 | 897934 |
| District of Columbia | 601723.000000 | 601723 |
| Florida | 280616.567164 | 18801310 |
| Georgia | 60928.635220 | 9687653 |
| Hawaii | 272060.200000 | 1360301 |
| Idaho | 35626.863636 | 1567582 |
| Illinois | 125790.509804 | 12830632 |
| Indiana | 70476.108696 | 6483802 |

| | | |
|---|---|---|
| Iowa | 99771.262626 | 9946355 |
| [illegible] | 27172.552381 | 2853118 |
| Kentucky | 36161.391667 | 4339367 |
| Louisiana | 70833.937500 | 4533372 |
| Maine | 83022.562500 | 1328361 |
| Maryland | 240564.666667 | 5773552 |
| Massachusetts | 467687.785714 | 6547629 |
| Michigan | 119080.000000 | 9883640 |
| Minnesota | 60964.655172 | 5303925 |
| Mississippi | 36186.548780 | 2967297 |
| Missouri | 52077.626087 | 5988927 |
| Montana | 17668.125000 | 989415 |
| Nebraska | 19638.075269 | 1826341 |
| Nevada | 158855.941176 | 2700551 |
| New Hampshire | 131647.000000 | 1316470 |
| New Jersey | 418661.619048 | 8791894 |
| New Mexico | 62399.363636 | 2059179 |
| New York | 312550.032258 | 19378102 |
| North Carolina | 95354.830000 | 9535483 |
| North Dakota | 12690.396226 | 672591 |
| Ohio | 131096.636364 | 11536504 |
| Oklahoma | 48718.844156 | 3751351 |
| Oregon | 106418.722222 | 3831074 |
| Pennsylvania | 189587.746269 | 12702379 |
| Rhode Island | 210513.400000 | 1052567 |
| South Carolina | 100551.391304 | 4625364 |
| South Dakota | 12336.060606 | 814180 |
| Tennessee | 66801.105263 | 6346105 |
| Texas | 98998.271654 | 25145561 |
| Utah | 95306.379310 | 2763885 |
| Vermont | 44695.785714 | 625741 |
| Virginia | 60111.293233 | 7994802 |
| Washington | 172424.102564 | 6724540 |
| West Virginia | 33690.800000 | 1852994 |
| Wisconsin | 78985.916667 | 5686986 |
| Wyoming | 24505.478261 | 563626 |

In [35]:

```
(df.set_index('STNAME').groupby(level=0)['POPESTIMATE2010','POPESTIMATE2011']
    .agg({'avg': np.average, 'sum': np.sum}))
```

```
C:\Users\Parth\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py:4656: FutureWar
ning: using a dict with renaming is deprecated and will be removed in a future version
  return super(DataFrameGroupBy, self).aggregate(arg, *args, **kwargs)
```

Out[35]:

| | avg | | sum | |
|---|---|---|---|---|
| | POPESTIMATE2010 | POPESTIMATE2011 | POPESTIMATE2010 | POPESTIMATE2011 |

| STNAME | avg | | sum | |
|---|---|---|---|---|
| | POPESTIMATE2010 | POPESTIMATE2011 | POPESTIMATE2010 | POPESTIMATE2011 |
| Alabama | 71420.313433 | 71658.328358 | 4785161 | 4801108 |
| STNAME | | | | |
| Alaska | 24621.413793 | 24921.379310 | 714021 | 722720 |
| Arizona | 427213.866667 | 431248.800000 | 6408208 | 6468732 |
| Arkansas | 38965.253333 | 39180.506667 | 2922394 | 2938538 |
| California | 643691.017241 | 650000.586207 | 37334079 | 37700034 |
| Colorado | 78878.968750 | 79991.875000 | 5048254 | 5119480 |
| Connecticut | 447464.625000 | 448719.875000 | 3579717 | 3589759 |
| Delaware | 299930.333333 | 302638.666667 | 899791 | 907916 |
| District of Columbia | 605126.000000 | 620472.000000 | 605126 | 620472 |
| Florida | 281341.641791 | 285157.208955 | 18849890 | 19105533 |
| Georgia | 61090.905660 | 61712.452830 | 9713454 | 9812280 |
| Hawaii | 272796.000000 | 275645.400000 | 1363980 | 1378227 |
| Idaho | 35704.227273 | 36003.045455 | 1570986 | 1584134 |
| Illinois | 125894.598039 | 126096.882353 | 12841249 | 12861882 |
| Indiana | 70549.891304 | 70835.271739 | 6490590 | 6516845 |
| Iowa | 30815.090909 | 30963.525253 | 3050694 | 3065389 |
| Kansas | 27226.895238 | 27332.542857 | 2858824 | 2869917 |
| Kentucky | 36232.808333 | 36399.016667 | 4347937 | 4367882 |
| Louisiana | 71014.859375 | 71490.328125 | 4544951 | 4575381 |
| Maine | 82980.937500 | 83016.062500 | 1327695 | 1328257 |
| Maryland | 241183.708333 | 243507.125000 | 5788409 | 5844171 |
| Massachusetts | 468931.142857 | 472271.214286 | 6565036 | 6611797 |
| Michigan | 119004.445783 | 118995.048193 | 9877369 | 9876589 |
| Minnesota | 61044.862069 | 61472.632184 | 5310903 | 5348119 |
| Mississippi | 36223.365854 | 36317.060976 | 2970316 | 2977999 |
| Missouri | 52139.582609 | 52265.973913 | 5996052 | 6010587 |
| Montana | 17690.053571 | 17816.892857 | 990643 | 997746 |
| Nebraska | 19677.688172 | 19810.569892 | 1830025 | 1842383 |
| Nevada | 159025.882353 | 159930.529412 | 2703440 | 2718819 |
| New Hampshire | 131670.800000 | 131834.400000 | 1316708 | 1318344 |
| New Jersey | 419232.428571 | 421092.095238 | 8803881 | 8842934 |
| New Mexico | 62567.909091 | 62976.545455 | 2064741 | 2078226 |
| New York | 312950.322581 | 314890.354839 | 19402920 | 19523202 |
| North Carolina | 95589.790000 | 96510.250000 | 9558979 | 9651025 |
| North Dakota | 12726.981132 | 12930.679245 | 674530 | 685326 |
| Ohio | 131145.068182 | 131198.204545 | 11540766 | 11545442 |
| Oklahoma | 48825.922078 | 49176.961039 | 3759596 | 3786626 |
| Oregon | 106610.333333 | 107458.583333 | 3837972 | 3868509 |
| Pennsylvania | 189731.552239 | 190226.895522 | 12712014 | 12745202 |
| Rhode Island | 210643.800000 | 210371.200000 | 1053219 | 1051856 |
| South Carolina | 100780.304348 | 101581.152174 | 4635894 | 4672733 |
| South Dakota | 12368.166667 | 12489.227273 | 816299 | 824289 |
| Tennessee | 66911.421053 | 67351.663158 | 6356585 | 6398408 |
| Texas | 99387.255906 | 101001.826772 | 25244363 | 25654464 |

| | avg | | sum | |
|---|---|---|---|---|
| Utah | 95704.344828 | 97118.620690 | 2775426 | 2816440 |
| | POPESTIMATE2010 | POPESTIMATE2011 | POPESTIMATE2010 | POPESTIMATE2011 |
| Vermont | 44713.142857 | 44763.357143 | 625984 | 626687 |
| STNAME | | | | |
| ~~Virginia~~ | ~~60344.263158~~ | ~~60983.330827~~ | ~~8025787~~ | ~~8110783~~ |
| Washington | 172898.974359 | 174954.589744 | 6743060 | 6823229 |
| West Virginia | 33713.181818 | 33726.327273 | 1854225 | 1854948 |
| Wisconsin | 79030.611111 | 79301.666667 | 5690204 | 5709720 |
| Wyoming | 24544.173913 | 24685.565217 | 564516 | 567768 |

In [36]:

```
(df.set_index('STNAME').groupby(level=0)['POPESTIMATE2010','POPESTIMATE2011']
    .agg({'POPESTIMATE2010': np.average, 'POPESTIMATE2011': np.sum}))
```

Out[36]:

| | POPESTIMATE2010 | POPESTIMATE2011 |
|---|---|---|
| STNAME | | |
| Alabama | 71420.313433 | 4801108 |
| Alaska | 24621.413793 | 722720 |
| Arizona | 427213.866667 | 6468732 |
| Arkansas | 38965.253333 | 2938538 |
| California | 643691.017241 | 37700034 |
| Colorado | 78878.968750 | 5119480 |
| Connecticut | 447464.625000 | 3589759 |
| Delaware | 299930.333333 | 907916 |
| District of Columbia | 605126.000000 | 620472 |
| Florida | 281341.641791 | 19105533 |
| Georgia | 61090.905660 | 9812280 |
| Hawaii | 272796.000000 | 1378227 |
| Idaho | 35704.227273 | 1584134 |
| Illinois | 125894.598039 | 12861882 |
| Indiana | 70549.891304 | 6516845 |
| Iowa | 30815.090909 | 3065389 |
| Kansas | 27226.895238 | 2869917 |
| Kentucky | 36232.808333 | 4367882 |
| Louisiana | 71014.859375 | 4575381 |
| Maine | 82980.937500 | 1328257 |
| Maryland | 241183.708333 | 5844171 |
| Massachusetts | 468931.142857 | 6611797 |
| Michigan | 119004.445783 | 9876589 |
| Minnesota | 61044.862069 | 5348119 |
| Mississippi | 36223.365854 | 2977999 |
| Missouri | 52139.582609 | 6010587 |
| Montana | 17690.053571 | 997746 |
| Nebraska | 19677.688172 | 1842383 |
| Nevada | 159025.882353 | 2718819 |
| New Hampshire | 131670.800000 | 1318344 |

| | POPESTIMATE2010 | POPESTIMATE2011 |
|---|---|---|
| New Jersey | 419232.428571 | 8842934 |
| New Mexico | 62567.909091 | 2078226 |
| STNAME | | |
| New York | 312950.322581 | 19523202 |
| North Carolina | 95589.790000 | 9651025 |
| North Dakota | 12726.981132 | 685326 |
| Ohio | 131145.068182 | 11545442 |
| Oklahoma | 48825.922078 | 3786626 |
| Oregon | 106610.333333 | 3868509 |
| Pennsylvania | 189731.552239 | 12745202 |
| Rhode Island | 210643.800000 | 1051856 |
| South Carolina | 100780.304348 | 4672733 |
| South Dakota | 12368.166667 | 824289 |
| Tennessee | 66911.421053 | 6398408 |
| Texas | 99387.255906 | 25654464 |
| Utah | 95704.344828 | 2816440 |
| Vermont | 44713.142857 | 626687 |
| Virginia | 60344.263158 | 8110783 |
| Washington | 172898.974359 | 6823229 |
| West Virginia | 33713.181818 | 1854948 |
| Wisconsin | 79030.611111 | 5709720 |
| Wyoming | 24544.173913 | 567768 |

# Scales

In [37]:

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

Out[37]:

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

In [38]:

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

Out[38]:

```
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+]
```

In [39]:

```
grades = df['Grades'].astype('category',
                             categories=['D', 'D+', 'C-', 'C', 'C+', 'B-', 'B', 'B+', 'A
-', 'A', 'A+'],
                             ordered=True)
grades.head()
```

```
C:\Users\Parth\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: FutureWarning: specif
ying 'categories' or 'ordered' in .astype() is deprecated; pass a CategoricalDtype instea
d
  This is separate from the ipykernel package so we can avoid doing imports until
```

Out[39]:

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

In [40]:

```
grades > 'C'
```

Out[40]:

```
excellent     True
excellent     True
excellent     True
good          True
good          True
good          True
ok            True
ok           False
ok           False
poor         False
poor         False
Name: Grades, dtype: bool
```

In [41]:

```
df = pd.read_csv('census.csv')
df = df[df['SUMLEV']==50]
df = df.set_index('STNAME').groupby(level=0)['CENSUS2010POP'].agg({'avg': np.average})
pd.cut(df['avg'],10)
```

```
C:\Users\Parth\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: FutureWarning: using
a dict on a Series for aggregation
is deprecated and will be removed in a future version
  This is separate from the ipykernel package so we can avoid doing imports until
```

Out[41]:

```
STNAME
Alabama                   (11706.087, 75333.413]
Alaska                    (11706.087, 75333.413]
Arizona                  (390320.176, 453317.529]
Arkansas                  (11706.087, 75333.413]
California               (579312.234, 642309.586]
Colorado                  (75333.413, 138330.766]
Connecticut              (390320.176, 453317.529]
```

```
Connecticut           (390320.176, 453317.529]
Delaware              (264325.471, 327322.823]
District of Columbia  (579312.234, 642309.586]
Florida               (264325.471, 327322.823]
Georgia                 (11706.087, 75333.413]
Hawaii                (264325.471, 327322.823]
Idaho                   (11706.087, 75333.413]
Illinois               (75333.413, 138330.766]
Indiana                 (11706.087, 75333.413]
Iowa                    (11706.087, 75333.413]
Kansas                  (11706.087, 75333.413]
Kentucky                (11706.087, 75333.413]
Louisiana               (11706.087, 75333.413]
Maine                  (75333.413, 138330.766]
Maryland              (201328.118, 264325.471]
Massachusetts         (453317.529, 516314.881]
Michigan               (75333.413, 138330.766]
Minnesota               (11706.087, 75333.413]
Mississippi             (11706.087, 75333.413]
Missouri                (11706.087, 75333.413]
Montana                 (11706.087, 75333.413]
Nebraska                (11706.087, 75333.413]
Nevada                (138330.766, 201328.118]
New Hampshire          (75333.413, 138330.766]
New Jersey            (390320.176, 453317.529]
New Mexico              (11706.087, 75333.413]
New York              (264325.471, 327322.823]
North Carolina         (75333.413, 138330.766]
North Dakota            (11706.087, 75333.413]
Ohio                   (75333.413, 138330.766]
Oklahoma                (11706.087, 75333.413]
Oregon                 (75333.413, 138330.766]
Pennsylvania          (138330.766, 201328.118]
Rhode Island          (201328.118, 264325.471]
South Carolina         (75333.413, 138330.766]
South Dakota            (11706.087, 75333.413]
Tennessee               (11706.087, 75333.413]
Texas                  (75333.413, 138330.766]
Utah                   (75333.413, 138330.766]
Vermont                 (11706.087, 75333.413]
Virginia                (11706.087, 75333.413]
Washington            (138330.766, 201328.118]
West Virginia           (11706.087, 75333.413]
Wisconsin              (75333.413, 138330.766]
Wyoming                 (11706.087, 75333.413]
Name: avg, dtype: category
Categories (10, interval[float64]): [(11706.087, 75333.413] < (75333.413, 138330.766] < (
138330.766, 201328.118] < (201328.118, 264325.471] ... (390320.176, 453317.529] < (453317
.529, 516314.881] < (516314.881, 579312.234] < (579312.234, 642309.586]]
```

# Pivot Tables

In [42]:

```
#http://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64
df = pd.read_csv('cars.csv')
```

In [43]:

```
df.head()
```

Out[43]:

| | YEAR | Make | Model | Size | (kW) | Unnamed: 5 | TYPE | CITY (kWh/100 km) | HWY (kWh/100 km) | COMB (kWh/100 km) | CITY (Le/100 km) | HWY (Le/100 km) | ( |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2012 | MITSUBISHI | i-MiEV | SUBCOMPACT | 49 | A1 | B | 16.9 | 21.4 | 18.7 | 1.9 | 2.4 | |
| 1 | 2012 | NISSAN | LEAF | MID-SIZE | 80 | A1 | B | 19.3 | 23.0 | 21.1 | 2.2 | 2.6 | |

| | YEAR | Make | Model | Size | (kW) | Unnamed: 5 | TYPE | (kWh/100 km) | (kWh/100 km) | (kWh/100 km) | (Le/100 km) | (Le/100 km) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 2013 | FORD | FOCUS ELECTRIC | COMPACT | 107 | A1 | B | 19.7 | 21.1 | 20.3 | 2.1 | 2.2 |
| 3 | 2013 | MITSUBISHI | i-MiEV | SUBCOMPACT | 49 | A1 | B | 16.9 | 21.4 | 18.7 | 1.9 | 2.4 |
| 4 | 2013 | NISSAN | LEAF | MID-SIZE | 80 | A1 | B | 19.3 | 23.0 | 21.1 | 2.2 | 2.6 |

In [44]:

```
df.pivot_table(values='(kW)', index='YEAR', columns='Make', aggfunc=np.mean)
```

Out[44]:

| 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 [45]:

```
df.pivot_table(values='(kW)', index='YEAR', columns='Make', aggfunc=[np.mean,np.min], ma
rgins=True)
```

Out[45]:

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

# Date Functionality in Pandas

In [2]:

```
import pandas as pd
import numpy as np
```

## Timestamp

In [3]:

```
pd.Timestamp('9/1/2016 10:05AM')
```

Out[3]:

```
Timestamp('2016-09-01 10:05:00')
```

## Period

In [4]:

```
pd.Period('1/2016')
```

Out[4]:

```
Period('2016-01', 'M')
```

In [5]:

```
pd.Period('3/5/2016')
```

Out[5]:

```
Period('2016-03-05', 'D')
```

## DatetimeIndex

In [6]:

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

Out[6]:

```
2016-09-01    a
2016-09-02    b
2016-09-03    c
dtype: object
```

In [7]:

```
type(t1.index)
```

Out[7]:

```
pandas.tseries.index.DatetimeIndex
```

## PeriodIndex

In [8]:

```
t2 = pd.Series(list('def'), [pd.Period('2016-09'), pd.Period('2016-10'), pd.Period('2016
-11')])
t2
```

Out[8]:

```
2016-09    d
2016-10    e
2016-11    f
Freq: M, dtype: object
```

In [9]:

```
type(t2.index)
```

Out[9]:

```
pandas.tseries.period.PeriodIndex
```

## Converting to Datetime

In [10]:

```
d1 = ['2 June 2013', '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[10]:

| | a | b |
|---|---|---|
| **2 June 2013** | 16 | 46 |
| **Aug 29, 2014** | 14 | 66 |
| **2015-06-26** | 59 | 99 |
| **7/12/16** | 27 | 17 |

In [11]:

```
ts3.index = pd.to_datetime(ts3.index)
ts3
```

Out[11]:

| | a | b |
|---|---|---|
| **2013-06-02** | 16 | 46 |
| **2014-08-29** | 14 | 66 |
| **2015-06-26** | 59 | 99 |
| **2016-07-12** | 27 | 17 |

In [12]:

```
pd.to_datetime('4.7.12', dayfirst=True)
```

Out[12]:

```
Timestamp('2012-07-04 00:00:00')
```

## Timedeltas

In [13]:

```
pd.Timestamp('9/3/2016')-pd.Timestamp('9/1/2016')
```

Out[13]:

```
Timedelta('2 days 00:00:00')
```

In [14]:

```
pd.Timestamp('9/2/2016 8:10AM') + pd.Timedelta('12D 3H')
```

Out[14]:

```
Timestamp('2016-09-14 11:10:00')
```

## Working with Dates in a Dataframe

In [15]:

```
dates = pd.date_range('10-01-2016', periods=9, freq='2W-SUN')
dates
```

Out[15]:

```
DatetimeIndex(['2016-10-02', '2016-10-16', '2016-10-30', '2016-11-13',
               '2016-11-27', '2016-12-11', '2016-12-25', '2017-01-08',
               '2017-01-22'],
              dtype='datetime64[ns]', freq='2W-SUN')
```

In [16]:

```
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[16]:

| | Count 1 | Count 2 |
|---|---|---|
| **2016-10-02** | 104 | 125 |
| **2016-10-16** | 109 | 122 |
| **2016-10-30** | 111 | 127 |
| **2016-11-13** | 117 | 126 |
| **2016-11-27** | 114 | 126 |
| **2016-12-11** | 109 | 121 |
| **2016-12-25** | 105 | 126 |
| **2017-01-08** | 105 | 125 |
| **2017-01-22** | 101 | 123 |

In [17]:

```
df.index.weekday_name
```

Out[17]:

```
array(['Sunday', 'Sunday', 'Sunday', 'Sunday', 'Sunday', 'Sunday',
       'Sunday', 'Sunday', 'Sunday'], dtype=object)
```

In [18]:

```
df.diff()
```

Out[18]:

| | Count 1 | Count 2 |
|---|---|---|
| **2016-10-02** | NaN | NaN |
| **2016-10-16** | 5.0 | -3.0 |
| **2016-10-30** | 2.0 | 5.0 |
| **2016-11-13** | 6.0 | -1.0 |
| **2016-11-27** | -3.0 | 0.0 |
| **2016-12-11** | -5.0 | -5.0 |
| **2016-12-25** | -4.0 | 5.0 |
| **2017-01-08** | 0.0 | -1.0 |
| **2017-01-22** | -4.0 | -2.0 |

In [19]:

```
df.resample('M').mean()
```

Out[19]:

| | Count 1 | Count 2 |
|---|---|---|
| **2016-10-31** | 108.0 | 124.666667 |
| **2016-11-30** | 115.5 | 126.000000 |
| **2016-12-31** | 107.0 | 123.500000 |
| **2017-01-31** | 103.0 | 124.000000 |

In [20]:

```
df['2017']
```

Out[20]:

| | Count 1 | Count 2 |
|---|---|---|
| **2017-01-08** | 105 | 125 |
| **2017-01-22** | 101 | 123 |

In [21]:

```
df['2016-12']
```

Out[21]:

| | Count 1 | Count 2 |
|---|---|---|
| **2016-12-11** | 109 | 121 |
| **2016-12-25** | 105 | 126 |

In [22]:

```
df['2016-12':]
```

Out[22]:

| | Count 1 | Count 2 |
|---|---|---|
| **2016-12-11** | 109 | 121 |
| **2016-12-25** | 105 | 126 |
| **2017-01-08** | 105 | 125 |
| **2017-01-22** | 101 | 123 |

In [ ]:

```
df.asfreq('W', method='ffill')
```

In [ ]:

```
import matplotlib.pyplot as plt
%matplotlib inline

df.plot()
```