08-Pandas-Exercises-Solutions

October 19, 2022

---

1 Pandas Exercises

Time to test your new pandas skills! Use the population_by_county.csv file in the Data folder to complete the tasks in bold below!

NOTE: ALL TASKS CAN BE DONE IN ONE LINE OF PANDAS CODE. GET STUCK? NO PROBLEM! CHECK OUT THE SOLUTIONS LECTURE!

IMPORTANT NOTE! Make sure you don't run the cells directly above the example output shown, otherwise you will end up writing over the example output!

1. Import pandas and read in the population_by_county.csv file into a dataframe called pop.

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

[2]:
```
pop = pd.read_csv('../Data/population_by_county.csv')
```

2. Show the head of the dataframe

[ ]:
```
# CODE HERE
```

[3]:
```
# DON'T WRITE HERE
pop.head()
```

[3]:
```
             County           State  2010Census  2017PopEstimate
0  Abbeville County  South Carolina       25417            24722
1     Acadia Parish       Louisiana       61773            62590
2   Accomack County        Virginia       33164            32545
3        Ada County           Idaho      392365           456849
4      Adair County            Iowa        7682             7054
```

3. What are the column names?

[ ]:

```

```

[4]:

```
# DON'T WRITE HERE
pop.columns
```

[4]:

```
Index(['County', 'State', '2010Census', '2017PopEstimate'], dtype='object')
```

4. How many States are represented in this data set? Note: the data includes the District of Columbia

[ ]:

```

```

[5]:

```
# DON'T WRITE HERE
pop['State'].nunique()
```

[5]:

```
51
```

5. Get a list or array of all the states in the data set.

[ ]:

```

```

[6]:

```
# DON'T WRITE HERE
pop['State'].unique()
```

[6]:

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

6. What are the five most common County names in the U.S.?

[ ]:

```

```

[7]:

```
# DON'T WRITE HERE
pop['County'].value_counts().head()
```

[7]:

```
Washington County    30
Jefferson County     25
Franklin County      24
```

```
Jackson County       23
Lincoln County       23
Name: County, dtype: int64
```

7. What are the top 5 most populated Counties according to the 2010 Census?

```
[ ]:
```

```
[8]: # DON'T WRITE HERE
     pop.sort_values('2010Census', ascending=False).head()
```

```
[8]:                   County       State  2010Census  2017PopEstimate
     1713  Los Angeles County  California     9818605         10163507
     628          Cook County    Illinois     5194675          5211263
     1209       Harris County       Texas     4092459          4652980
     1784     Maricopa County     Arizona     3817117          4307033
     2501    San Diego County  California     3095313          3337685
```

8. What are the top 5 most populated States according to the 2010 Census?

```
[ ]:
```

```
[9]: # DON'T WRITE HERE
     pop.groupby('State').sum().sort_values('2010Census', ascending=False).head()
```

```
[9]:             2010Census  2017PopEstimate
     State
     California    37253956         39536653
     Texas         25145561         28304596
     New York      19378102         19849399
     Florida       18801310         20984400
     Illinois      12830632         12802023
```

9. How many Counties have 2010 populations greater than 1 million?

```
[ ]:
```

```
[10]: # DON'T WRITE HERE
      # pop['2010Census'].apply(lambda qty: qty>1000000).value_counts()
      sum(pop['2010Census']>1000000)
```

```
[10]: 39
```

10. How many Counties don't have the word "County" in their name?

```
[ ]:
```

[11]:

```
# DON'T WRITE HERE
# pop['County'].apply(lambda name: 'County' not in name).value_counts()
sum(pop['County'].apply(lambda name: 'County' not in name))
```

[11]:

```
135
```

11. Add a column that calculates the percent change between the 2010 Census and the 2017 Population Estimate

[ ]:

```
# CODE HERE

# USE THIS TO SHOW THE RESULT
pop.head()
```

[12]:

```
# DON'T WRITE HERE
pop['PercentChange'] = 100*(pop['2017PopEstimate']-pop['2010Census'])/
 ↪pop['2010Census']
pop.head()
```

[12]:

```
             County           State  2010Census  2017PopEstimate  \
0  Abbeville County  South Carolina       25417            24722
1     Acadia Parish       Louisiana       61773            62590
2   Accomack County        Virginia       33164            32545
3        Ada County           Idaho      392365           456849
4      Adair County            Iowa        7682             7054

   PercentChange
0      -2.734390
1       1.322584
2      -1.866482
3      16.434697
4      -8.174954
```

Bonus: What States have the highest estimated percent change between the 2010 Census and the 2017 Population Estimate?This will take several lines of code, as it requires a recalculation of PercentChange.

[ ]:

```
# CODE HERE
```

[13]:

```
# DON'T WRITE HERE
pop2 = pd.DataFrame(pop.groupby('State').sum())
pop2['PercentChange'] = 100*(pop2['2017PopEstimate']-pop2['2010Census'])/
 ↪pop2['2010Census']
pop2.sort_values('PercentChange', ascending=False).head()
```

[13]:

```
                      2010Census  2017PopEstimate  PercentChange
State
```

```
District of Columbia      601723          693972      15.330808
Texas                   25145561        28304596      12.562993
North Dakota              672591          755393      12.310899
Utah                     2763885         3101833      12.227282
Florida                 18801310        20984400      11.611372
```

2 GREAT JOB!