Sathish_K_Rajendiran_Week5_Async_5.2_Stacking_and_Unstacking_Dat

July 29, 2020

Week 5 Stacking and Unstacking Data

```
Name: Sathish Kumar Rajendiran
Task: 5.2 Data Transformations
Date: 7/29/2020
```

```python
[1]: #import libraries
     import os
     import csv
     import numpy as np
     import pandas as pd

     #verify current directory
     os.getcwd()
```

```
[1]: '/Users/sathishrajendiran/ist652-python'
```

```python
[33]: # define dictionary with State, Postcode, Area and Population info

      state_data = {'State':['Alabama','Alaska','Arizona','Arkansas'],'PostCode':
       →['AL','AK','AZ','AR'],'Area':['52,423','656,424','*','53,182'],'Pop':
       →['4,040,587','550,043','3,665,228','2,350,750']}

      type(state_data) #dict

      #create a dataframe from state_data dictionary
      statedf = pd.DataFrame(state_data, columns=['State','PostCode','Area','Pop'])
      # statedf.head()
```

```
[33]: dict
```

```python
[15]: #set State as the index column
      statedf = statedf.set_index('State')
      statedf.head()

      #replace '* with '0' value
      statedf = statedf.replace('*','0')
```

```
statedf
```

[15]:
```
           PostCode      Area          Pop
State
Alabama          AL    52,423    4,040,587
Alaska           AK   656,424      550,043
Arizona          AZ         0    3,665,228
Arkansas         AR    53,182    2,350,750
```

[16]:
```
#create a function to replace a string
def item_replace(xstr):
    return xstr.replace(',','')
```

[19]:
```
#replace ',' from Area and Population values
statedf['Area'] = statedf['Area'].map(item_replace)
statedf['Pop'] = statedf['Pop'].map(item_replace)
statedf
```

[24]:
```
#convert Area and Population values to float
statedf['Area'] = statedf['Area'].astype(np.float)
statedf['Pop'] = statedf['Pop'].astype(np.float)
```

[29]:
```
#calculate mean area and replace 0 with mean area value
mean_area = statedf['Area'].mean()
statedf['Area']= statedf['Area'].mask(statedf['Area']==0,mean_area)
```

[36]:
```
#display top 2 rows
statedf.head(2)

#display bottom 2 rows
statedf.tail(2)

#display all rows
statedf
```

[36]:
```
           PostCode       Area          Pop
State
Alabama          AL    52423.00    4040587.0
Alaska           AK   656424.00     550043.0
Arizona          AZ   190507.25    3665228.0
Arkansas         AR    53182.00    2350750.0
```

[ ]: