Joyce Woznica  
IST 652, Week 5

Activity 5.2: Data Transformations

Code:

﻿

﻿# Week 5 Activities

# Joyce Woznica

import pandas as pd

import numpy as np

# state date

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']}

# 1. convert to a dataframe

stateDF = pd.DataFrame(state\_data, columns = ['State', 'PostCode', 'Area', 'Pop'])

# 2. show the dataframe

# 3. Index by state (done) - column statement

stateDF

# 4. need to replace the '\*' for Arizona Area with a 0

stateDF = stateDF.replace('\*', '0')

stateDF

# 5. function to eliminate commas

def repl\_comma(df):

return df.replace(',', '')

# 6. replace commas

stateDF['Area'] = stateDF['Area'].map(repl\_comma)

stateDF

stateDF['Pop'] = stateDF['Pop'].map(repl\_comma)

stateDF

Output:

﻿stateDF

Out[95]:

State PostCode Area Pop

0 Alabama AL 52,423 4,040,587

1 Alaska AK 656,424 550,043

2 Arizona AZ \* 3,665,228

3 Arkansas AR 53,182 2,350,750

﻿Out[96]:

State PostCode Area Pop

0 Alabama AL 52423 4040587

1 Alaska AK 656424 550043

2 Arizona AZ 0 3665228

3 Arkansas AR 53182 2350750