

Assignment - 1: EE4708 ('exercise_pandas.ipynb')

Raj Jain | CH17B066

Pandas Exercises

Assignments for Pandas skills

Import pandas and read in the banklist.csv file into a dataframe called banks.

In [1]:

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

In [2]:

```
data = pd.read_csv("banklist.csv")
```

Show the head of the dataframe

In [3]:

```
# CODE HERE
```

In [4]:

```
data.head()
```

Out[4]:

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date
0	Fayette County Bank	Saint Elmo	IL	1802	United Fidelity Bank, fsb	26-May-17	1-Jun-17
1	Guaranty Bank, (d/b/a BestBank in Georgia & Mi...	Milwaukee	WI	30003	First-Citizens Bank & Trust Company	5-May-17	1-Jun-17
2	First NBC Bank	New Orleans	LA	58302	Whitney Bank	28-Apr-17	23-May-17
3	Proficio Bank	Cottonwood Heights	UT	35495	Cache Valley Bank	3-Mar-17	18-May-17
4	Seaway Bank and Trust Company	Chicago	IL	19328	State Bank of Texas	27-Jan-17	18-May-17

What are the column names?

In [5]:

```
# CODE HERE
```

In [6]:

```
print(data.columns)
```

```
Index(['Bank Name', 'City', 'ST', 'CERT', 'Acquiring Institution',  
      'Closing Date', 'Updated Date'],  
      dtype='object')
```

How many States (ST) are represented in this data set?

In [7]:

```
# CODE HERE
```

In [8]:

```
print("The number of states represented in this dataset is",data.ST.nunique(),".")
```

The number of states represented in this dataset is 44 .

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

In [9]:

```
# CODE HERE
```

In [10]:

```
print("The list of all the states in the data set:\n\n",data.ST.unique())
```

The list of all the states in the data set:

```
['IL' 'WI' 'LA' 'UT' 'NJ' 'AR' 'GA' 'PA' 'TN' 'WA' 'CO' 'PR' 'FL' 'MN'  
'CA' 'MD' 'OK' 'OH' 'SC' 'VA' 'ID' 'TX' 'CT' 'AZ' 'NV' 'NC' 'KY' 'MO'  
'KS' 'AL' 'MI' 'IN' 'IA' 'NE' 'MS' 'NM' 'OR' 'NY' 'MA' 'SD' 'WY' 'WV'  
'NH' 'HI']
```

What are the top 5 states with the most failed banks?

In [11]:

```
# CODE HERE
```

In [13]:

```
Top5FailStates = data['ST'].value_counts()[0:5]  
print("Top 5 states with the most failed banks: (Adjacent number to each state represents the number of failed banks in that state)\n", Top5FailStates)
```

Top 5 states with the most failed banks: (Adjacent number to each state represents the number of failed banks in that state)

```
GA      93  
FL      75  
IL      67  
CA      41  
MN      23
```

Name: ST, dtype: int64

What are the top 5 acquiring institutions?

In [14]:

```
# CODE HERE
```

In [15]:

```
Top5Acquirers = data['Acquiring Institution'].value_counts()[0:5]
print("Top 5 acquiring institutions:\n",Top5Acquirers)
```

```
Top 5 acquiring institutions:
No Acquirer                      31
State Bank and Trust Company      12
First-Citizens Bank & Trust Company 11
Ameris Bank                      10
U.S. Bank N.A.                   9
Name: Acquiring Institution, dtype: int64
```

How many banks has the State Bank of Texas acquired? How many of them were actually in Texas?

In [16]:

```
# CODE HERE
```

In [18]:

```
data[data['Acquiring Institution']=='State Bank of Texas']
```

Out[18]:

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date
4	Seaway Bank and Trust Company	Chicago	IL	19328	State Bank of Texas	27-Jan-17	18-May-17
21	The National Republic Bank of Chicago	Chicago	IL	916	State Bank of Texas	24-Oct-14	6-Jan-16
450	Millennium State Bank of Texas	Dallas	TX	57667	State Bank of Texas	2-Jul-09	26-Oct-12

There are three banks acquired by the State Bank of Texas, of which only one bank is actually in the state of Texas (Millennium State Bank of Texas).

What is the most common city in California for a bank to fail in?

In [19]:

```
# CODE HERE
```

In [20]:

```
MostCommonCityCA = data[data['ST']=='CA']['City'].value_counts().index[0]  
print(MostCommonCityCA, "is the city with the most bank failures in California")
```

Los Angeles is the city with the most bank failures in California

How many failed banks don't have the word "Bank" in their name?

In [21]:

```
# CODE HERE
```

In [22]:

```
NumWithoutBank = data[~data["Bank Name"].str.contains("Bank")].shape[0]
print("The number of banks which do not contain the word 'Bank' is", NumWithoutBank)
data[~data["Bank Name"].str.contains("Bank")]
```

The number of banks which do not contain the word 'Bank' is 14

Out[22]:

	Bank Name	City	ST	CERT	Acquiring Institution	Closing Date	Updated Date
22	NBRS Financial	Rising Sun	MD	4862	Howard Bank	17-Oct-14	26-Mar-15
74	Second Federal Savings and Loan Association of...	Chicago	IL	27986	Hinsdale Bank & Trust Company	20-Jul-12	14-Jan-13
101	Home Savings of America	Little Falls	MN	29178	No Acquirer	24-Feb-12	17-Dec-12
222	First Arizona Savings, A FSB	Scottsdale	AZ	32582	No Acquirer	22-Oct-10	20-Aug-12
248	Imperial Savings and Loan Association	Martinsville	VA	31623	River Community Bank, N.A.	20-Aug-10	24-Aug-12
300	CF Bancorp	Port Huron	MI	30005	First Michigan Bank	30-Apr-10	13-Apr-16
301	Westernbank Puerto Rico	Mayaguez	PR	31027	Banco Popular de Puerto Rico	30-Apr-10	21-Mar-14
303	Eurobank	San Juan	PR	27150	Oriental Bank and Trust	30-Apr-10	21-Mar-14
423	ebank	Atlanta	GA	34682	Stearns Bank, N.A.	21-Aug-09	21-Aug-12
428	Dwelling House Savings and Loan Association	Pittsburgh	PA	31559	PNC Bank, N.A.	14-Aug-09	15-Jan-13
496	Suburban FSB	Crofton	MD	30763	Bank of Essex	30-Jan-09	4-Feb-16
505	Downey Savings & Loan	Newport Beach	CA	30968	U.S. Bank, N.A.	21-Nov-08	4-Jan-13
514	Ameribank	Northfork	WV	6782	The Citizens Savings Bank, Pioneer Community B...	19-Sep-08	21-Sep-15
523	ANB Financial, NA	Bentonville	AR	33901	Pulaski Bank and Trust Company	9-May-08	28-Aug-12

How many bank names start with the letter 's' ?

In [23]:

```
# CODE HERE
```

In [24]:

```
NumBankStartS = data[data["Bank Name"].str.startswith("S")].shape[0]  
print("The number of bank names starting with the letter 'S' is", NumBankStartS)
```

The number of bank names starting with the letter 'S' is 53

How many CERT values are above 20000 ?

In [25]:

```
# CODE HERE
```

In [26]:

```
NumCERT2000 = data[data["CERT"]>20000].shape[0]  
print("The number of CERT values above 20000 is", NumCERT2000)
```

The number of CERT values above 20000 is 417

How many bank names consist of just two words? (e.g. "First Bank" , "Bank Georgia")

In [27]:

```
# CODE HERE
```

In [28]:

```
NumBanksTwoWords = data[data['Bank Name'].str.split().str.len()==2].shape[0]  
print("The number of banks which consist of just two words is:", NumBanksTwoWords)
```

The number of banks which consist of just two words is: 114

Bonus: How many banks closed in the year 2008? (this is hard because we technically haven't learned about time series with pandas yet! Feel free to skip this one!)

In [29]:

```
# CODE HERE
```

In [30]:

```
NumBanksClosedIn2008 = data[pd.DatetimeIndex(data['Closing Date']).year == 2008].shape[0]  
print("The number of banks closed in the year 2008 is:", NumBanksClosedIn2008)
```

The number of banks closed in the year 2008 is: 25