Pandas Exercises

Assigments for Pandas skills

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

In [2]:

```
#importing pandas
import pandas as pd
```

In [3]:

```
#reading "banklist.csv"
banks = pd.read_csv("banklist.csv")
```

Show the head of the dataframe

In [4]:

```
# Showing head of the dataframe - first 5 entries
banks.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

In [37]:

Out[37]:

	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]:

```
# Column names of the dataframe
banks.columns
```

Out[5]:

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

In [29]:

Out[29]:

```
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 [6]:

```
# printing the number of unique states (ST) in the dataframe
print(len(pd.unique(banks["ST"])))
```

44

In [33]:

Out[33]:

44

44 states are represented in this dataset

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

In [7]:

```
# displaying an array of all the state codes
pd.unique(banks["ST"])
```

Out[7]:

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

In [32]:

Out[32]:

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

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

In [8]:

```
# displaying the top 5 states with the most failed banks
banks.groupby("ST").count()["Bank Name"].sort_values(ascending = False)[:5]
```

Out[8]:

```
ST
GA    93
FL    75
IL    67
CA    41
MN    23
Name: Bank Name, dtype: int64
```

In [35]:

Out[35]:

```
ST
GA    93
FL    75
IL    67
CA    41
MN    23
Name: Bank Name, dtype: int64
```

What are the top 5 acquiring institutions?

In [9]:

```
# displaying the top 5 acquiring institutions
banks["Acquiring Institution"].value_counts()[:5]
```

Out[9]:

```
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
```

In [14]:

Out[14]:

```
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 [10]:

```python
#displaying the banks that State Bank of Texas acquired
banks[banks["Acquiring Institution"]=="State Bank of Texas"]
```

Out[10]:

	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

In [57]:

Out[57]:

	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

In [12]:

```python
#determining how many of the banks acquired by State Bank of Texas are in Texas
sum(banks[banks["Acquiring Institution"]=="State Bank of Texas"]["ST"]=="TX")
```

Out[12]:

1

Therefore, three banks were aqcuired by State Bank of Texas, out of which one is in Texas

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

In [83]:

```
# Displaying the most common city in California for a bank to fail in, and the number o
f fails
banks[banks["ST"]=="CA"].groupby("City",sort = True).count().sort_values("Bank Name",as
cending = False).head(1)
```

Out[83]:

City	Bank Name	ST	CERT	Acquiring Institution	Closing Date	Updated Date
Los Angeles	4	4	4	4	4	4

In [84]:

Out[84]:

City	Bank Name	ST	CERT	Acquiring Institution	Closing Date	Updated Date
Los Angeles	4	4	4	4	4	4

Hence, Los Angeles is the most common city in California for a bank to fail in, with 4 banks to failing

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

In [82]:

```
# Printing the number of banks that do not have "Bank" in their name
c = 0
for i in range(len(banks["Bank Name"])):
    if "Bank" not in banks["Bank Name"][i]:
        c+=1
print(c)
```

14

In [55]:

Out[55]:

14

14 banks do not have "Bank" in their name

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

In [86]:

```python
# printing number of banks that start with S
c = 0
for i in range(len(banks["Bank Name"])):
    if banks["Bank Name"][i][0]=="S":
        c+=1
print(c)
```

53

In [58]:

Out[58]:

53

53 banks start with "S"

How many CERT values are above 20000 ?

In [86]:

```python
# displaying number of entries whose CERT values are above 20000
import numpy as np
np.sum(banks["CERT"]>20000)
```

Out[86]:

417

In [64]:

Out[64]:

417

417 CERT values are above 20000

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

In [91]:

```python
# CODE HERE
c = 0
for i in range(len(banks["Bank Name"])):
    if len(banks["Bank Name"][i].split())==2:
        c+=1
print(c)
```

114

In [67]:

Out[67]:

114

114 bank names have only two words

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 [85]:

```python
# printing the number of banks closed in the year 2008

#importing required package
import datetime as dt
#From 1st Jan 2008 to 31st Dec 2008
start = dt.datetime.strptime("01-Jan-08","%d-%b-%y")
end = dt.datetime.strptime("31-Dec-08","%d-%b-%y")

#Counting the number of banks whose closing date lies within the start and end date
c = 0
for date in banks["Closing Date"]:
    d = dt.datetime.strptime(date,"%d-%b-%y")
    if d >= start and d <= end:
        c+=1
print(c)
```

25

In [54]:

Out[54]:

25

25 banks were closed in the year 2008

In [ ]: