

In [1]:

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

In [2]:

```
df = pd.read_csv("Lending_Data.csv")
print("Shape of lending df: ", df.shape)
df.head()
```

Shape of lending df: (97, 14)

Out[2]:

| | Name of Lending Institution | HQ State | Rank | TA Ratio1 | TBL Ratio1 | Amount (\$1,000) | Number | Lender Asset Size | Amount (\$1,000).1 | Num |
|---|-----------------------------|----------|------|-----------|------------|------------------|-----------|-------------------|--------------------|--------|
| 0 | NaN | NaN | (1) | (2) | (3) | (4) | (5) | (6) | (7) | |
| 1 | American Express Co. | NY | 1 | 0.238 | 1.000 | 1,63,50,038 | 36,03,226 | >\$50B | 1,63,18,089 | 36,03, |
| 2 | First Citizens Bankshares | NC | 2 | 0.187 | 0.491 | 39,14,507 | 1,17,702 | 10B–50B | 3,78,394 | 1,05, |
| 3 | Wintrust Financial Corp. | IL | 3 | 0.169 | 0.384 | 28,42,793 | 1,36,801 | 10B–50B | 8,00,526 | 1,26, |
| 4 | Zions Bancorp | UT | 4 | 0.118 | 0.261 | 62,31,770 | 52,612 | >\$50B | 5,26,206 | 32, |

In [3]:

```
df.columns
```

Out[3]:

```
Index(['Name of Lending Institution', 'HQ State', 'Rank', 'TA Ratio1',
      'TBL Ratio1', 'Amount ($1,000)', 'Number ', 'Lender Asset Size',
      'Amount ($1,000).1', 'Number .1', 'Unnamed: 10', 'Amount ($1,000).2',
      'Number .2', 'CC Amount/TA1'],
      dtype='object')
```

In [4]:

```
df.rename(columns = {'Amount ($1,000)': 'Amount_in_1000$'}, inplace = True)
df.rename(columns = {'Amount ($1,000).1': 'Amount_in_1000$_1'}, inplace = True)
df.rename(columns = {'Amount ($1,000).2': 'Amount_in_1000$_2'}, inplace = True)

df.rename(columns = {'Number ': 'Number'}, inplace = True)
df.rename(columns = {'Number .1': 'Number_1'}, inplace = True)
df.rename(columns = {'Number .2': 'Number_2'}, inplace = True)

df.rename(columns = {'Lender Asset Size': 'Lender_Asset_Size'}, inplace = True)

df.rename(columns = {'CC Amount/TA1': 'CC_Amount_BY_TA1'}, inplace = True)
```

In [5]:

```
df['Amount_in_1000$'] = df['Amount_in_1000$'] .str.replace(',', '')
df['Amount_in_1000$_1'] = df['Amount_in_1000$_1'] .str.replace(',', '')
df['Amount_in_1000$_2'] = df['Amount_in_1000$_2'] .str.replace(',', '')

df['Number'] = df['Number'] .str.replace(',', '')
df['Number_1'] = df['Number_1'] .str.replace(',', '')
df['Number_2'] = df['Number_2'] .str.replace(',', '')
```

In [6]:

```
df=df.drop(df.index[0])
```

In [7]:

```
print("Shape of lending df: ",df.shape)
df.head()
```

Shape of lending df: (96, 14)

Out[7]:

| | Name of Lending Institution | HQ State | Rank | TA Ratio1 | TBL Ratio1 | Amount_in_1000\$ | Number | Lender_Asset_Size | Amc |
|---|-----------------------------|----------|------|-----------|------------|------------------|---------|-------------------|-----|
| 1 | American Express Co. | NY | 1 | 0.238 | 1.000 | 16350038 | 3603226 | >\$50B | |
| 2 | First Citizens Bancshares | NC | 2 | 0.187 | 0.491 | 3914507 | 117702 | 10B–50B | |
| 3 | Wintrust Financial Corp. | IL | 3 | 0.169 | 0.384 | 2842793 | 136801 | 10B–50B | |
| 4 | Zions Bancorp | UT | 4 | 0.118 | 0.261 | 6231770 | 52612 | >\$50B | |
| 5 | Synovus Financial Corp. | GA | 5 | 0.162 | 0.381 | 4227168 | 22675 | 10B–50B | |

In [8]:

df.columns

Out[8]:

```
Index(['Name of Lending Institution', 'HQ State', 'Rank', 'TA Ratio1',
      'TBL Ratio1', 'Amount_in_1000$', 'Number', 'Lender_Asset_Size',
      'Amount_in_1000$_1', 'Number_1', 'Unnamed: 10', 'Amount_in_1000$_2',
      'Number_2', 'CC_Amount_BY_TA1'],
      dtype='object')
```

In [9]:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 96 entries, 1 to 96
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Name of Lending Institution           96 non-null    object
1   HQ State                             93 non-null    object
2   Rank                                 93 non-null    object
3   TA Ratio1                            93 non-null    object
4   TBL Ratio1                           93 non-null    object
5   Amount_in_1000$                      93 non-null    object
6   Number                               93 non-null    object
7   Lender_Asset_Size                   93 non-null    object
8   Amount_in_1000$_1                   93 non-null    object
9   Number_1                             93 non-null    object
10  Unnamed: 10                          0 non-null     float64
11  Amount_in_1000$_2                   93 non-null    object
12  Number_2                             93 non-null    object
13  CC_Amount_BY_TA1                    93 non-null    object
dtypes: float64(1), object(13)
memory usage: 11.2+ KB
```

In [10]:

df.pop('Unnamed: 10')

Out[10]:

```
1    NaN
2    NaN
3    NaN
4    NaN
5    NaN
..
92   NaN
93   NaN
94   NaN
95   NaN
96   NaN
Name: Unnamed: 10, Length: 96, dtype: float64
```

In [11]:

```
df.shape
```

Out[11]:

```
(96, 13)
```

In [12]:

```
dict(df['Rank'].value_counts())
```

Out[12]:

```
{'NR': 5,  
'19': 3,  
'13': 2,  
'40': 2,  
'67': 2,  
'34': 2,  
'8': 2,  
'11': 2,  
'88': 1,  
'10': 1,  
'59': 1,  
'48': 1,  
'80': 1,  
'54': 1,  
'63': 1,  
'4': 1,  
'71': 1,  
'45': 1,  
'7': 1,  
'49': 1,  
'58': 1,  
'1': 1,  
'27': 1,  
'29': 1,  
'79': 1,  
'53': 1,  
'51': 1,  
'46': 1,  
'55': 1,  
'42': 1,  
'6': 1,  
'43': 1,  
'66': 1,  
'31': 1,  
'56': 1,  
'65': 1,  
'22': 1,  
'38': 1,  
'69': 1,  
'78': 1,  
'33': 1,  
'52': 1,  
'60': 1,  
'76': 1,  
'18': 1,  
'30': 1,  
'77': 1,  
'50': 1,  
'26': 1,  
'47': 1,  
'15': 1,  
'73': 1,  
'70': 1,  
'25': 1,  
'16': 1,
```

```
'5': 1,  
'28': 1,  
'32': 1,  
'24': 1,  
'86': 1,  
'75': 1,  
'17': 1,  
'85': 1,  
'37': 1,  
'72': 1,  
'61': 1,  
'64': 1,  
'44': 1,  
'82': 1,  
'36': 1,  
'2': 1,  
'3': 1,  
'83': 1,  
'57': 1,  
'23': 1,  
'39': 1,  
'84': 1,  
'62': 1,  
'74': 1,  
'81': 1,  
'87': 1}
```

In [13]:

```
df.isnull().sum()
```

Out[13]:

```
Name of Lending Institution    0  
HQ State                      3  
Rank                          3  
TA Ratio1                     3  
TBL Ratio1                    3  
Amount_in_1000$              3  
Number                        3  
Lender_Asset_Size             3  
Amount_in_1000$_1             3  
Number_1                      3  
Amount_in_1000$_2             3  
Number_2                      3  
CC_Amount_BY_TA1              3  
dtype: int64
```

In [14]:

```
df['Rank'].replace({'NR':0,np.nan:0},inplace = True)
```

In [15]:

```
df['Rank'] = pd.to_numeric(df['Rank'])
```

In [16]:

```
df['TA Ratio1'].replace({'-':0,np.nan:0},inplace = True)
```

In [17]:

```
df['TA Ratio1'] = pd.to_numeric(df['TA Ratio1'])
```

In [18]:

```
df['TBL Ratio1'].replace({'-':0,np.nan:0},inplace = True)
```

In [19]:

```
df['TBL Ratio1'] = pd.to_numeric(df['TBL Ratio1'])
```

In [20]:

```
df['Amount_in_1000$'].replace({'-':0,np.nan:0,' - ':0},inplace = True)
```

In [21]:

```
df['Amount_in_1000$'] = pd.to_numeric(df['Amount_in_1000$'])
```

In [22]:

```
df['Number'].replace({'-':0,np.nan:0},inplace = True)
```

In [23]:

```
df['Number'] = pd.to_numeric(df['Number'])
```

In [24]:

```
df['Amount_in_1000$_1'].replace({'-':0,np.nan:0,' - ':0},inplace = True)
```

In [25]:

```
df['Amount_in_1000$_1'] = pd.to_numeric(df['Amount_in_1000$_1'])
```

In [26]:

```
df['Number_1'].replace({'-':0,np.nan:0,' - ':0},inplace = True)
```

In [27]:

```
df['Number_1'] = pd.to_numeric(df['Number_1'])
```

In [28]:

```
df['Amount_in_1000$_2'].replace({'-':0,np.nan:0,' - ':0},inplace = True)
```

In [29]:

```
df['Amount_in_1000$_2'] = pd.to_numeric(df['Amount_in_1000$_2'])
```

In [30]:

```
df['Number_2'].replace({'-':0,np.nan:0,' - ':0},inplace = True)
```

In [31]:

```
df['Number_2'] = pd.to_numeric(df['Number_2'])
```

In [32]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 96 entries, 1 to 96
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Name of Lending Institution           96 non-null     object
1   HQ State                             93 non-null     object
2   Rank                                  96 non-null     int64
3   TA Ratio1                            96 non-null     float64
4   TBL Ratio1                           96 non-null     float64
5   Amount_in_1000$                      96 non-null     int64
6   Number                               96 non-null     int64
7   Lender_Asset_Size                    93 non-null     object
8   Amount_in_1000$_1                   96 non-null     int64
9   Number_1                             96 non-null     int64
10  Amount_in_1000$_2                   96 non-null     int64
11  Number_2                             96 non-null     int64
12  CC_Amount_BY_TA1                    93 non-null     object
dtypes: float64(2), int64(7), object(4)
memory usage: 10.5+ KB
```

In [33]:

```
df.head()
```

Out[33]:

| | Name of Lending Institution | HQ State | Rank | TA Ratio1 | TBL Ratio1 | Amount_in_1000\$ | Number | Lender_Asset_Size | Amc |
|---|-----------------------------|----------|------|-----------|------------|------------------|---------|-------------------|-----|
| 1 | American Express Co. | NY | 1 | 0.238 | 1.000 | 16350038 | 3603226 | >\$50B | |
| 2 | First Citizens Banchares | NC | 2 | 0.187 | 0.491 | 3914507 | 117702 | 10B–50B | |
| 3 | Wintrust Financial Corp. | IL | 3 | 0.169 | 0.384 | 2842793 | 136801 | 10B–50B | |
| 4 | Zions Bancorp | UT | 4 | 0.118 | 0.261 | 6231770 | 52612 | >\$50B | |
| 5 | Synovus Financial Corp. | GA | 5 | 0.162 | 0.381 | 4227168 | 22675 | 10B–50B | |



In [34]:

```
df.describe()
```

Out[34]:

| | Rank | TA Ratio1 | TBL Ratio1 | Amount_in_1000\$ | Number | Amount_in_1000\$_1 | |
|-------|-----------|-----------|------------|------------------|--------------|--------------------|---|
| count | 96.000000 | 96.000000 | 96.000000 | 9.600000e+01 | 9.600000e+01 | 9.600000e+01 | 9 |
| mean | 40.697917 | 0.045094 | 0.230313 | 2.952567e+06 | 2.006384e+05 | 9.712989e+05 | 1 |
| std | 27.446837 | 0.045020 | 0.230731 | 5.812881e+06 | 6.639000e+05 | 2.792406e+06 | 6 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 0 |
| 25% | 16.750000 | 0.011000 | 0.087250 | 2.437332e+05 | 1.676250e+03 | 1.676275e+04 | 4 |
| 50% | 40.000000 | 0.036000 | 0.185500 | 1.120101e+06 | 1.232800e+04 | 1.134570e+05 | 6 |
| 75% | 64.250000 | 0.065250 | 0.261250 | 2.840470e+06 | 4.831825e+04 | 5.269600e+05 | 3 |
| max | 88.000000 | 0.238000 | 1.000000 | 3.457039e+07 | 3.603226e+06 | 1.631809e+07 | 3 |

In [35]:

```
Rm = df['Rank'].mean()
TAm = df['TA Ratio1'].mean()
TBLm = df['TBL Ratio1'].mean()
Am = df['Amount_in_1000$'].mean()
Nm = df['Number'].mean()
A_1m = df['Amount_in_1000$_1'].mean()
N_1m = df['Number_1'].mean()
A_2m = df['Amount_in_1000$_2'].mean()
N_2m = df['Number_2'].mean()
```

In [36]:

```
#df['Rank'] = df['Rank'].replace(0,Rm)
```

In [37]:

```
df['TA Ratio1'] = df['TA Ratio1'].replace(0,TAm)
df['TBL Ratio1'] = df['TBL Ratio1'].replace(0,TBLm)
df['Amount_in_1000$'] = df['Amount_in_1000$'].replace(0,Am)
df['Number'] = df['Number'].replace(0,Nm)
df['Amount_in_1000$_1'] = df['Amount_in_1000$_1'].replace(0,A_1m)
df['Number_1'] = df['Number_1'].replace(0,N_1m)
df['Amount_in_1000$_2'] = df['Amount_in_1000$_2'].replace(0,A_2m)
df['Number_2'] = df['Number_2'].replace(0,N_2m)
```

In [38]:

```
df.head()
```

Out[38]:

| | Name of Lending Institution | HQ State | Rank | TA Ratio1 | TBL Ratio1 | Amount_in_1000\$ | Number | Lender_Asset_Size | Ar |
|---|-----------------------------|----------|------|-----------|------------|------------------|-----------|-------------------|----|
| 1 | American Express Co. | NY | 1 | 0.238 | 1.000 | 16350038.0 | 3603226.0 | >\$50B | |
| 2 | First Citizens Bancharcs | NC | 2 | 0.187 | 0.491 | 3914507.0 | 117702.0 | 10B–50B | |
| 3 | Wintrust Financial Corp. | IL | 3 | 0.169 | 0.384 | 2842793.0 | 136801.0 | 10B–50B | |
| 4 | Zions Bancorp | UT | 4 | 0.118 | 0.261 | 6231770.0 | 52612.0 | >\$50B | |
| 5 | Synovus Financial Corp. | GA | 5 | 0.162 | 0.381 | 4227168.0 | 22675.0 | 10B–50B | |

In [39]:

```
dict(df['Lender_Asset_Size'].value_counts())
# df['Lender_Asset_Size'].replace({'NR':0,np.nan:0},inplace = True)
```

Out[39]:

```
{' $10B-$50B': 43,
' >$50B': 31,
' >$10B': 15,
'>$10B': 3,
'>$50B': 1}
```

In [40]:

```
'''
0 == >$10B,
1 == $10B-$50B,
2 == >$50B == 2
'''

# df['Lender_Asset_Size'].loc[df['Lender_Asset_Size'] == '>$10B'] = '0' or
# df['Lender_Asset_Size'] == '>$10B'

df.loc[(df.Lender_Asset_Size == '>$10B'), 'Lender_Asset_Size'] = '0'
df.loc[(df.Lender_Asset_Size == '>$10B'), 'Lender_Asset_Size'] = '0'

df.loc[(df.Lender_Asset_Size == '$10B-$50B'), 'Lender_Asset_Size'] = '1'

df.loc[(df.Lender_Asset_Size == '>$50B'), 'Lender_Asset_Size'] = '2'
df.loc[(df.Lender_Asset_Size == '>$50B'), 'Lender_Asset_Size'] = '2'
```

In [41]:

```
dict(df['Lender_Asset_Size'].value_counts())
```

Out[41]:

```
{'1': 43, '2': 32, '0': 18}
```

In [42]:

```
dict(df['CC_Amount_BY_TA1'].value_counts())
```

Out[42]:

```
{'': 47,
'0': 30,
'0.01': 7,
'0.03': 2,
'0.25': 1,
'0.06': 1,
'0.02': 1,
'0.07': 1,
'0.1': 1,
'0.05': 1,
'0.28': 1}
```

In [43]:

```
df['CC_Amount_BY_TA1'].replace({'-':0,np.nan:0,' ':0,'.':0},inplace = True)
```

In [44]:

```
df['CC_Amount_BY_TA1'] = pd.to_numeric(df['CC_Amount_BY_TA1'])
```

In [45]:

```
CC_m = df['CC_Amount_BY_TA1'].mean()
```

In [46]:

```
df['CC_Amount_BY_TA1'] = df['CC_Amount_BY_TA1'].replace(0,CC_m)
```

In [47]:

```
df.to_csv('lending_cleaned.csv', index=False)
```

In []:

In []: