Loan_Data

October 17, 2018

In [3]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [4]:
```
#This dataset contains loan characteristics of a bank's customers.
#The goal is to clean the dataset, create a model that predicts loan outcome(paid or c
#find the characteristics that are highly correlated with safe loan applicantions.
```

In [5]:
```
#Creating the dataset
data = pd.read_csv('LoansTrainingSet.csv')
```

```
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/IPython/core/interactiveshell.py:2
  interactivity=interactivity, compiler=compiler, result=result)
```

In [6]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 256984 entries, 0 to 256983
Data columns (total 19 columns):
Loan ID                         256984 non-null object
Customer ID                     256984 non-null object
Loan Status                     256984 non-null object
Current Loan Amount             256984 non-null int64
Term                            256984 non-null object
Credit Score                    195308 non-null float64
Years in current job            245508 non-null object
Home Ownership                  256984 non-null object
Annual Income                   195308 non-null float64
Purpose                         256984 non-null object
Monthly Debt                    256984 non-null object
Years of Credit History         256984 non-null float64
Months since last delinquent    116601 non-null float64
Number of Open Accounts         256984 non-null int64
Number of Credit Problems       256984 non-null int64
Current Credit Balance          256984 non-null int64
Maximum Open Credit             256984 non-null object
Bankruptcies                    256455 non-null float64
```

```
Tax Liens                       256961 non-null float64
dtypes: float64(6), int64(4), object(9)
memory usage: 37.3+ MB
```

```
In [7]: data.head()
```

Out[7]:

	Loan ID	Customer ID
0	000025bb-5694-4cff-b17d-192b1a98ba44	5ebc8bb1-5eb9-4404-b11b-a6eebc401a19
1	00002c49-3a29-4bd4-8f67-c8f8fbc1048c	927b388d-2e01-423f-a8dc-f7e42d668f46
2	00002d89-27f3-409b-aa76-90834f359a65	defce609-c631-447d-aad6-1270615e89c4
3	00005222-b4d8-45a4-ad8c-186057e24233	070bcecb-aae7-4485-a26a-e0403e7bb6c5
4	0000757f-a121-41ed-b17b-162e76647c1f	dde79588-12f0-4811-bab0-e2b07f633fcd

	Loan Status	Current Loan Amount	Term	Credit Score
0	Fully Paid	11520	Short Term	741.0
1	Fully Paid	3441	Short Term	734.0
2	Fully Paid	21029	Short Term	747.0
3	Fully Paid	18743	Short Term	747.0
4	Fully Paid	11731	Short Term	746.0

	Years in current job	Home Ownership	Annual Income	Purpose
0	10+ years	Home Mortgage	33694.0	Debt Consolidation
1	4 years	Home Mortgage	42269.0	other
2	10+ years	Home Mortgage	90126.0	Debt Consolidation
3	10+ years	Own Home	38072.0	Debt Consolidation
4	4 years	Rent	50025.0	Debt Consolidation

	Monthly Debt	Years of Credit History	Months since last delinquent
0	$584.03	12.3	41.0
1	$1,106.04	26.3	NaN
2	$1,321.85	28.8	NaN
3	$751.92	26.2	NaN
4	$355.18	11.5	NaN

	Number of Open Accounts	Number of Credit Problems	Current Credit Balance
0	10	0	6760
1	17	0	6262
2	5	0	20967
3	9	0	22529
4	12	0	17391

	Maximum Open Credit	Bankruptcies	Tax Liens
0	16056	0.0	0.0
1	19149	0.0	0.0
2	28335	0.0	0.0
3	43915	0.0	0.0
4	37081	0.0	0.0

```
In [8]: #Get rid of non-integer units in series
        for ch in ['$', ',']:
            data['Monthly Debt'] = [i.replace(ch, '') for i in data['Monthly Debt']]

In [9]: #Make Monthly Debt numeric through list comprehension
        data['Monthly Debt'] = [float(i) for i in data['Monthly Debt']]

In [10]: data['Monthly Debt'].head()

Out[10]: 0     584.03
         1    1106.04
         2    1321.85
         3     751.92
         4     355.18
         Name: Monthly Debt, dtype: float64

In [11]: data['Loan Status'].unique()

Out[11]: array(['Fully Paid', 'Charged Off'], dtype=object)

In [12]: #Change 'Loan Status' unique values to integers, e.g. create dummy variable
         data['Loan Status'] = data['Loan Status'].replace('Fully Paid', 1).replace('Charged O

In [13]: data['Loan Status'].head()

Out[13]: 0    1
         1    1
         2    1
         3    1
         4    1
         Name: Loan Status, dtype: int64

In [14]: #The columns 'Term', 'Home Ownership', and 'Purpose' are all categorical
         #Here we create dummy variables and merge them with the dataset
         data2 = pd.merge(data, (pd.get_dummies(data['Term'], drop_first = True)),left_index =

In [15]: data3 = pd.merge(data2, (pd.get_dummies(data['Home Ownership'], drop_first = True)),

In [16]: data4 = pd.merge(data3, (pd.get_dummies(data['Purpose'], drop_first = True)), left_in

In [17]: del data4['Term']
         del data4['Home Ownership']
         del data4['Purpose']

In [18]: data4.corr()

Out[18]:                              Loan Status  Current Loan Amount  Credit Score  \
         Loan Status                     1.000000             0.269804     -0.440307
         Current Loan Amount             0.269804             1.000000     -0.139743
         Credit Score                   -0.440307            -0.139743      1.000000
```

Annual Income	0.070328	0.024069	-0.033221
Monthly Debt	-0.016609	-0.001928	0.005518
Years of Credit History	0.035548	0.014725	-0.011658
Months since last delinquent	0.023541	0.003488	-0.007994
Number of Open Accounts	-0.020787	-0.003094	0.008124
Number of Credit Problems	-0.010179	-0.000062	0.000777
Current Credit Balance	0.008838	0.003138	-0.003263
Bankruptcies	0.001154	0.003576	-0.003426
Tax Liens	-0.012585	-0.003069	0.004381
Short Term	0.183811	0.048366	-0.063124
Home Mortgage	0.066485	0.022415	-0.027572
Own Home	-0.008300	-0.005950	0.002765
Rent	-0.062959	-0.019448	0.026512
Buy House	0.003706	0.001675	-0.000005
Buy a Car	0.024949	0.007060	-0.010635
Debt Consolidation	-0.001793	-0.000173	0.003534
Educational Expenses	0.000505	0.002249	-0.000422
Home Improvements	0.022265	0.005638	-0.008781
Medical Bills	-0.005133	-0.002055	0.003071
Other	0.014854	0.005424	-0.007807
Take a Trip	-0.000259	-0.001048	-0.002141
other	-0.018092	-0.005161	0.003812

	Annual Income	Monthly Debt	\
Loan Status	0.070328	-0.016609	
Current Loan Amount	0.024069	-0.001928	
Credit Score	-0.033221	0.005518	
Annual Income	1.000000	0.454638	
Monthly Debt	0.454638	1.000000	
Years of Credit History	0.146859	0.188646	
Months since last delinquent	-0.059675	-0.056377	
Number of Open Accounts	0.140463	0.410416	
Number of Credit Problems	-0.013672	-0.052792	
Current Credit Balance	0.292165	0.472581	
Bankruptcies	-0.044837	-0.078440	
Tax Liens	0.038185	0.020557	
Short Term	-0.064219	-0.151294	
Home Mortgage	0.170834	0.205076	
Own Home	-0.035239	-0.034933	
Rent	-0.152658	-0.187406	
Buy House	0.008286	-0.020118	
Buy a Car	-0.014758	-0.052909	
Debt Consolidation	-0.031492	0.098217	
Educational Expenses	-0.008967	-0.024400	
Home Improvements	0.072842	-0.003684	
Medical Bills	-0.003047	-0.012430	
Other	-0.004447	-0.062845	
Take a Trip	-0.011189	-0.016935	

other	-0.020287	-0.054886

	Years of Credit History \
Loan Status	0.035548
Current Loan Amount	0.014725
Credit Score	-0.011658
Annual Income	0.146859
Monthly Debt	0.188646
Years of Credit History	1.000000
Months since last delinquent	-0.039695
Number of Open Accounts	0.128033
Number of Credit Problems	0.061251
Current Credit Balance	0.201001
Bankruptcies	0.062049
Tax Liens	0.020915
Short Term	-0.035773
Home Mortgage	0.176821
Own Home	0.032031
Rent	-0.197944
Buy House	-0.009164
Buy a Car	-0.007641
Debt Consolidation	0.009894
Educational Expenses	0.003037
Home Improvements	0.034114
Medical Bills	0.002957
Other	-0.033322
Take a Trip	-0.011536
other	-0.014512

	Months since last delinquent \
Loan Status	0.023541
Current Loan Amount	0.003488
Credit Score	-0.007994
Annual Income	-0.059675
Monthly Debt	-0.056377
Years of Credit History	-0.039695
Months since last delinquent	1.000000
Number of Open Accounts	-0.035803
Number of Credit Problems	0.088612
Current Credit Balance	-0.024292
Bankruptcies	0.112907
Tax Liens	0.002730
Short Term	0.012107
Home Mortgage	-0.048589
Own Home	0.003055
Rent	0.048513
Buy House	-0.005753
Buy a Car	0.005373

Debt Consolidation	0.015807
Educational Expenses	-0.006835
Home Improvements	-0.006413
Medical Bills	-0.008626
Other	-0.004765
Take a Trip	-0.003855
other	-0.003744

	Number of Open Accounts \
Loan Status	-0.020787
Current Loan Amount	-0.003094
Credit Score	0.008124
Annual Income	0.140463
Monthly Debt	0.410416
Years of Credit History	0.128033
Months since last delinquent	-0.035803
Number of Open Accounts	1.000000
Number of Credit Problems	-0.013731
Current Credit Balance	0.222763
Bankruptcies	-0.022805
Tax Liens	0.005754
Short Term	-0.080506
Home Mortgage	0.135632
Own Home	-0.009351
Rent	-0.131451
Buy House	-0.015354
Buy a Car	-0.039338
Debt Consolidation	0.107607
Educational Expenses	-0.014989
Home Improvements	-0.018276
Medical Bills	-0.017824
Other	-0.059107
Take a Trip	-0.018508
other	-0.059669

	Number of Credit Problems \
Loan Status	-0.010179
Current Loan Amount	-0.000062
Credit Score	0.000777
Annual Income	-0.013672
Monthly Debt	-0.052792
Years of Credit History	0.061251
Months since last delinquent	0.088612
Number of Open Accounts	-0.013731
Number of Credit Problems	1.000000
Current Credit Balance	-0.103814
Bankruptcies	0.755866
Tax Liens	0.584917

Short Term	0.019326
Home Mortgage	-0.001143
Own Home	0.006744
Rent	-0.003146
Buy House	-0.000180
Buy a Car	-0.007838
Debt Consolidation	-0.003844
Educational Expenses	-0.007293
Home Improvements	0.023186
Medical Bills	0.000867
Other	-0.017318
Take a Trip	-0.000315
other	0.002627

	Current Credit Balance	...	Rent	\
Loan Status	0.008838	...	-0.062959	
Current Loan Amount	0.003138	...	-0.019448	
Credit Score	-0.003263	...	0.026512	
Annual Income	0.292165	...	-0.152658	
Monthly Debt	0.472581	...	-0.187406	
Years of Credit History	0.201001	...	-0.197944	
Months since last delinquent	-0.024292	...	0.048513	
Number of Open Accounts	0.222763	...	-0.131451	
Number of Credit Problems	-0.103814	...	-0.003146	
Current Credit Balance	1.000000	...	-0.157313	
Bankruptcies	-0.117995	...	0.000976	
Tax Liens	-0.011118	...	-0.003432	
Short Term	-0.090077	...	0.094092	
Home Mortgage	0.156387	...	-0.831889	
Own Home	-0.000808	...	-0.268603	
Rent	-0.157313	...	1.000000	
Buy House	-0.016330	...	0.014430	
Buy a Car	-0.030486	...	0.009219	
Debt Consolidation	0.077174	...	0.038240	
Educational Expenses	-0.010286	...	0.015823	
Home Improvements	-0.012794	...	-0.171867	
Medical Bills	-0.013754	...	0.007377	
Other	-0.050726	...	0.050376	
Take a Trip	-0.019384	...	0.018998	
other	-0.040505	...	0.042624	

	Buy House	Buy a Car	Debt Consolidation	\
Loan Status	0.003706	0.024949	-0.001793	
Current Loan Amount	0.001675	0.007060	-0.000173	
Credit Score	-0.000005	-0.010635	0.003534	
Annual Income	0.008286	-0.014758	-0.031492	
Monthly Debt	-0.020118	-0.052909	0.098217	
Years of Credit History	-0.009164	-0.007641	0.009894	

Months since last delinquent	-0.005753	0.005373	0.015807
Number of Open Accounts	-0.015354	-0.039338	0.107607
Number of Credit Problems	-0.000180	-0.007838	-0.003844
Current Credit Balance	-0.016330	-0.030486	0.077174
Bankruptcies	-0.000671	-0.007643	0.005102
Tax Liens	0.000495	-0.003965	-0.004241
Short Term	-0.003370	0.011237	-0.035947
Home Mortgage	-0.017014	-0.014146	-0.007765
Own Home	0.005417	0.009709	-0.037329
Rent	0.014430	0.009219	0.038240
Buy House	1.000000	-0.008794	-0.151696
Buy a Car	-0.008794	1.000000	-0.222735
Debt Consolidation	-0.151696	-0.222735	1.000000
Educational Expenses	-0.002496	-0.003665	-0.063214
Home Improvements	-0.019210	-0.028206	-0.486548
Medical Bills	-0.008222	-0.012072	-0.208237
Other	-0.015301	-0.022466	-0.387527
Take a Trip	-0.006068	-0.008909	-0.153678
other	-0.018764	-0.027551	-0.475243

	Educational Expenses	Home Improvements	\
Loan Status	0.000505	0.022265	
Current Loan Amount	0.002249	0.005638	
Credit Score	-0.000422	-0.008781	
Annual Income	-0.008967	0.072842	
Monthly Debt	-0.024400	-0.003684	
Years of Credit History	0.003037	0.034114	
Months since last delinquent	-0.006835	-0.006413	
Number of Open Accounts	-0.014989	-0.018276	
Number of Credit Problems	-0.007293	0.023186	
Current Credit Balance	-0.010286	-0.012794	
Bankruptcies	-0.007093	0.022360	
Tax Liens	-0.003560	0.006115	
Short Term	0.013622	-0.013564	
Home Mortgage	-0.013856	0.147775	
Own Home	-0.002889	0.040847	
Rent	0.015823	-0.171867	
Buy House	-0.002496	-0.019210	
Buy a Car	-0.003665	-0.028206	
Debt Consolidation	-0.063214	-0.486548	
Educational Expenses	1.000000	-0.008005	
Home Improvements	-0.008005	1.000000	
Medical Bills	-0.003426	-0.026370	
Other	-0.006376	-0.049075	
Take a Trip	-0.002528	-0.019461	
other	-0.007819	-0.060183	

	Medical Bills	Other	Take a Trip	other

```
Loan Status                          -0.005133  0.014854    -0.000259 -0.018092
Current Loan Amount                  -0.002055  0.005424    -0.001048 -0.005161
Credit Score                          0.003071 -0.007807    -0.002141  0.003812
Annual Income                        -0.003047 -0.004447    -0.011189 -0.020287
Monthly Debt                         -0.012430 -0.062845    -0.016935 -0.054886
Years of Credit History               0.002957 -0.033322    -0.011536 -0.014512
Months since last delinquent         -0.008626 -0.004765    -0.003855 -0.003744
Number of Open Accounts              -0.017824 -0.059107    -0.018508 -0.059669
Number of Credit Problems             0.000867 -0.017318    -0.000315  0.002627
Current Credit Balance               -0.013754 -0.050726    -0.019384 -0.040505
Bankruptcies                          0.000251 -0.020573     0.000649 -0.003704
Tax Liens                            -0.001961 -0.003823    -0.000748  0.001861
Short Term                            0.017970  0.030435     0.031253  0.029255
Home Mortgage                        -0.009205 -0.051419    -0.075989 -0.048490
Own Home                              0.004181  0.004356     0.000343  0.013014
Rent                                  0.007377  0.050376     0.018998  0.042624
Buy House                            -0.008222 -0.015301    -0.006068 -0.018764
Buy a Car                            -0.012072 -0.022466    -0.008909 -0.027551
Debt Consolidation                   -0.208237 -0.387527    -0.153678 -0.475243
Educational Expenses                 -0.003426 -0.006376    -0.002528 -0.007819
Home Improvements                    -0.026370 -0.049075    -0.019461 -0.060183
Medical Bills                         1.000000 -0.021004    -0.008329 -0.025758
Other                                -0.021004  1.000000    -0.015501 -0.047935
Take a Trip                          -0.008329 -0.015501     1.000000 -0.019009
other                                -0.025758 -0.047935    -0.019009  1.000000

[25 rows x 25 columns]
```

In [19]:
```
#Next we deal with missing values for 'Bankruptcies' and 'Tax Liens'
#Correlation map shows both variables are correlated with 'Number of Credit Problems'
#We fill missing values depending on how many credit problems that individual has had
```

In [20]:
```
#Knowing how many credit problems someone has can help us predict how many bankruptci
data4.groupby('Number of Credit Problems')['Bankruptcies'].mean()
```

Out[20]:
```
Number of Credit Problems
0     0.000000
1     0.822141
2     0.900234
3     1.039191
4     1.058182
5     1.008000
6     1.214286
7     1.187500
8     0.916667
9     0.100000
10    0.666667
11    0.000000
Name: Bankruptcies, dtype: float64
```

In [21]:
```
#Filling in missing 'Bankruptcies' values by the number of their credit problems
for i in data4[data4['Bankruptcies'].isnull()].index:
    data4['Bankruptcies'][i] = data['Bankruptcies'][data4['Number of Credit Problems']
```

```
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:3: Deprecatio
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated
  This is separate from the ipykernel package so we can avoid doing imports until
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWi
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  This is separate from the ipykernel package so we can avoid doing imports until
```

In [22]:
```
#Filling in missing 'Tax Liens' values by the number of their credit problems
for i in data4[data4['Tax Liens'].isnull()].index:
    data4['Tax Liens'][i] = data4['Tax Liens'][data4['Number of Credit Problems'] ==
```

```
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:3: Deprecatio
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated
  This is separate from the ipykernel package so we can avoid doing imports until
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWi
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  This is separate from the ipykernel package so we can avoid doing imports until
```

In [23]:
```
#Each Loan is suppose to be unique, therefore duplicates are dropped
data5 = data4.drop_duplicates(['Loan ID'], keep = 'last')
```

In [24]:
```
#Max credit score an individual can have is 800 based off of this particular credit s
#It appears an extra zero was added to these credit scores
data5['Credit Score'][data5['Credit Score'] > 800].head(10)
```

Out[24]:
```
341     6600.0
349     6760.0
420     7460.0
522     7320.0
```

```
623     7270.0
846     6690.0
926     7230.0
1240    7380.0
1306    7440.0
1317    7390.0
Name: Credit Score, dtype: float64
```

```
In [25]: #We divide these values by ten to get rid of the extra zero
         #This is done through list comprehension
         data5['Credit Score'][data5['Credit Score'] > 800] = [i/10 for i in data5['Credit Sco
```

```
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWi
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  This is separate from the ipykernel package so we can avoid doing imports until
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/pandas/core/generic.py:7620: Settin
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  self._update_inplace(new_data)
/Users/charliecarrera/anaconda3/lib/python3.6/site-packages/IPython/core/interactiveshell.py:29
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  exec(code_obj, self.user_global_ns, self.user_ns)
```

```
In [27]: #There are now no more values above 800
         #Distribution of credit score
         data5['Credit Score'].describe()
```

```
Out[27]: count    167406.000000
         mean        723.028022
         std          26.648780
         min         585.000000
         25%         713.000000
         50%         732.000000
         75%         742.000000
         max         751.000000
         Name: Credit Score, dtype: float64
```

```
In [28]: #Some individuals had extremely high loan amounts yet low income
         #Delete outliers in 'Current Loan Amount'
         data6 = data5.drop(data5[data5['Current Loan Amount'] > 1000000].index)
```

```
In [29]: #Next 'Years in current job' is made purely numeric
         data6['Years in current job'].head()
```

```
Out[29]: 0    10+ years
         1      4 years
         2    10+ years
         3    10+ years
         4      4 years
         Name: Years in current job, dtype: object

In [30]: data6['Years in current job'] = [str(i) for i in data6['Years in current job']]

In [31]: h = [i.split(' ')[0] for i in data6['Years in current job']]

In [32]: h = [i.replace('<', '.5') for i in h]
         h = [i.replace('n/a', '0') for i in h]
         h = [i.replace('10+', '10') for i in h]
         h = [float(i) for i in h]

In [33]: data6['Years in current job'] = h

In [34]: hh = list(data6['Maximum Open Credit'])

In [140]: #Now 'Maximum Open Credit is cleaned by getting rid of strings
          hh = [str(i).replace('#VALUE!', '0') for i in hh]
          hh = [float(i) for i in hh]
          data6['Maximum Open Credit'] = hh

In [141]: data6['Maximum Open Credit'] = [float(i) for i in data6['Maximum Open Credit']]

In [36]: #Everytime 'Annual Income is null, so is 'Credit Score'
         #These null rows are deleted
         data[data['Annual Income'].isnull()].head(3)

Out[36]:                                    Loan ID  \
         7   0000afa6-8902-4f8f-b870-25a8fdad0aeb
         8   00011dfc-31c1-4178-932a-fbeb3f341efb
         12  00029f9f-0cc5-4d4e-aabc-ea4a7fe74e12

                                    Customer ID  Loan Status  Current Loan Amount  \
         7   e49c1a82-a0f7-45e8-9f46-2f75c43f9fbc            0                24613
         8   ef6e098c-6c83-4752-8d00-ff793e476b8c            1                10036
         12  afbc2fa3-3bad-4d48-b691-829aed78bad5            0                17980

                Term  Credit Score Years in current job Home Ownership  \
         7    Long Term          NaN              6 years          Rent
         8   Short Term          NaN              5 years          Rent
         12  Short Term          NaN            < 1 year       Own Home

             Annual Income             Purpose  Monthly Debt  Years of Credit History  \
         7             NaN       Business Loan        542.29                     17.6
         8             NaN  Debt Consolidation        386.36                     17.7
```

```
12            NaN  Debt Consolidation        597.50                      9.9

    Months since last delinquent  Number of Open Accounts  \
7                           73.0                        7
8                            NaN                        7
12                          43.0                        7

    Number of Credit Problems  Current Credit Balance Maximum Open Credit  \
7                           0                    14123               16954
8                           0                    11970               16579
12                          0                     6817               22800

    Bankruptcies  Tax Liens
7            0.0        0.0
8            0.0        0.0
12           0.0        0.0
```

In [192]:
```
data7 = data6
del data7['Months since last delinquent']
data8 = data7.drop(data7[data7['Annual Income'].isnull()].index)
```

In [194]:
```
del data8['Loan ID']
del data8['Customer ID']
```

In [196]:
```
#Predictions are created for 'Loan Status' aka whether or not the loan was paid off
#'Loan Status' is categorical, calls for a classification algorithm
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.ensemble import GradientBoostingClassifier
gbc = GradientBoostingClassifier()
```

In [ ]:
```
gbc.fit(data8.drop('Loan Status', axis = 1), data8['Loan Status'])
```

In [770]:
```
#75% accuracy score means
accuracy_score(data8['Loan Status'], gbc.predict(data8.drop('Loan Status', axis = 1))
```

Out[770]:
```
0.75739054131743777
```

In [779]:
```
gbc.feature_importances_
```

Out[779]:
```
array([ 0.10438967,  0.18943464,  0.04284989,  0.18065884,  0.10127577,
        0.0714165 ,  0.03295017,  0.00468396,  0.07069356,  0.09599793,
        0.01642792,  0.00726124,  0.01304626,  0.01581257,  0.00846362,
        0.        ,  0.02838365,  0.        ,  0.0027543 ,  0.00769752,
        0.00249725,  0.00094808,  0.        ,  0.00115457,  0.        ,
        0.0012021 ])
```