LendingLoanPredication

August 3, 2023

[12]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

[13]:
```
data = pd.read_csv("loan_data.csv")
```

[14]:
```
data.head()
```

[14]:

	credit.policy	purpose	int.rate	installment	log.annual.inc	dti	fico	days.with.cr.line	revol.bal	revol.util	inq.last.6mths	delinq.2yrs	pub.rec	not.fully.paid
0	1	debt_consolidation	0.1189	829.10	11.350407	19.48	737	5639.958333	28854	52.1	0	0	0	0
1	1	credit_card	0.1071	228.22	11.082143	14.29	707	2760.000000	33623	76.7	0	0	0	0
2	1	debt_consolidation	0.1357	366.86	10.373491	11.63	682	4710.000000	3511	25.6	1	0	0	0
3	1	debt_consolidation	0.1008	162.34	11.350407	8.10	712	2699.958333	33667	73.2	1	0	0	0
4	1	credit_card	0.1426	102.92	11.299732	14.97	667	4066.000000	4740	39.5	0	1	0	0

[15]:
```
data.tail()
```

[15]:

	credit.policy	purpose	int.rate	installment
9573	0	all_other	0.1461	344.76
9574	0	all_other	0.1253	257.70
9575	0	debt_consolidation	0.1071	97.81
9576	0	home_improvement	0.1600	351.58
9577	0	debt_consolidation	0.1392	853.43

```
      log.annual.inc    dti  fico  days.with.cr.line  revol.bal  revol.util  \
9573       12.180755  10.39   672       10474.000000     215372        82.1
9574       11.141862   0.21   722        4380.000000        184         1.1
9575       10.596635  13.09   687        3450.041667      10036        82.9
9576       10.819778  19.18   692        1800.000000          0         3.2
9577       11.264464  16.28   732        4740.000000      37879        57.0

      inq.last.6mths  delinq.2yrs  pub.rec  not.fully.paid
9573               2            0        0               1
9574               5            0        0               1
9575               8            0        0               1
9576               5            0        0               1
9577               6            0        0               1
```

[16]:
```
data.dtypes
```

[16]:
```
credit.policy          int64
purpose               object
int.rate             float64
installment          float64
log.annual.inc       float64
dti                  float64
fico                   int64
days.with.cr.line    float64
revol.bal              int64
revol.util           float64
inq.last.6mths         int64
delinq.2yrs            int64
pub.rec                int64
not.fully.paid         int64
dtype: object
```

[17]:
```
data.value_counts().sum()
```

[17]:
```
9578
```

[7]:
```
data.isnull().sum()
```

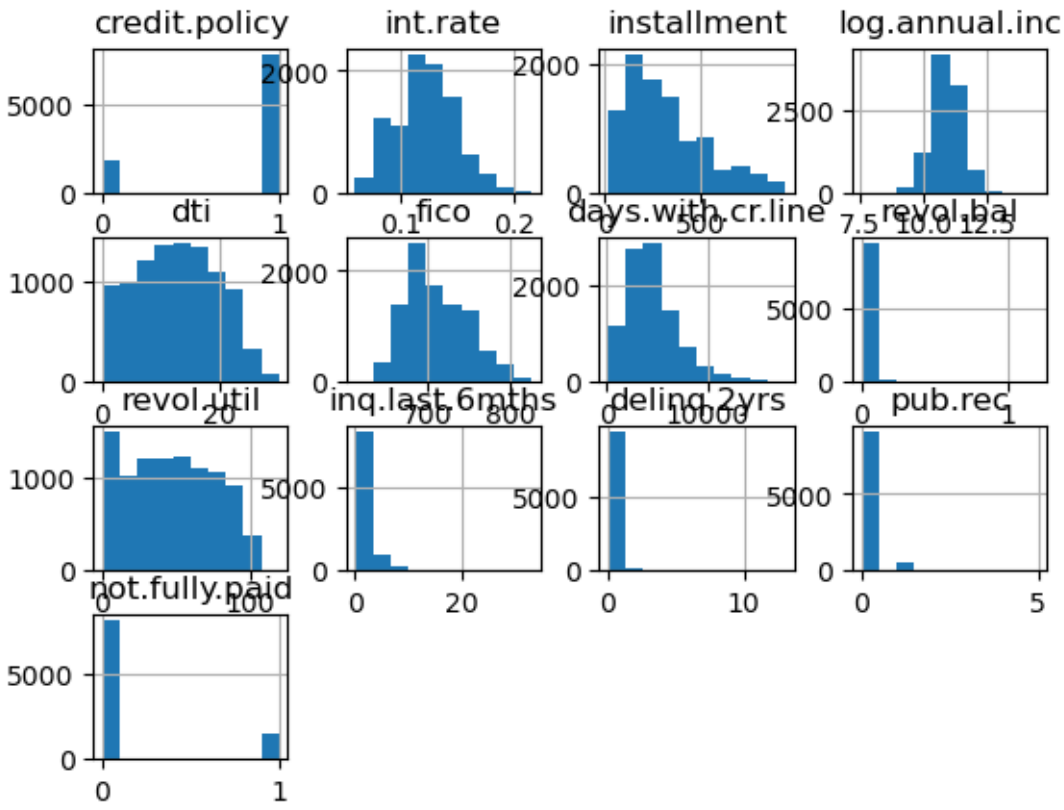
[7]:
```
credit.policy        0
purpose              0
int.rate             0
installment          0
log.annual.inc       0
dti                  0
fico                 0
days.with.cr.line    0
```

```
revol.bal             0
revol.util            0
inq.last.6mths        0
delinq.2yrs           0
pub.rec               0
not.fully.paid        0
dtype: int64
```

[9]:
```
Nums = data.select_dtypes(include=["int64","float64"]).columns.tolist()
```

[10]:
```
data[Nums].hist()
```

[10]:
```
array([[<AxesSubplot:title={'center':'credit.policy'}>,
        <AxesSubplot:title={'center':'int.rate'}>,
        <AxesSubplot:title={'center':'installment'}>,
        <AxesSubplot:title={'center':'log.annual.inc'}>],
       [<AxesSubplot:title={'center':'dti'}>,
        <AxesSubplot:title={'center':'fico'}>,
        <AxesSubplot:title={'center':'days.with.cr.line'}>,
        <AxesSubplot:title={'center':'revol.bal'}>],
       [<AxesSubplot:title={'center':'revol.util'}>,
        <AxesSubplot:title={'center':'inq.last.6mths'}>,
        <AxesSubplot:title={'center':'delinq.2yrs'}>,
        <AxesSubplot:title={'center':'pub.rec'}>],
       [<AxesSubplot:title={'center':'not.fully.paid'}>, <AxesSubplot:>,
        <AxesSubplot:>, <AxesSubplot:>]], dtype=object)
```



[19]:
```
data[data.duplicated()]
```

[19]:
```
Empty DataFrame
Columns: [credit.policy, purpose, int.rate, installment, log.annual.inc, dti,
fico, days.with.cr.line, revol.bal, revol.util, inq.last.6mths, delinq.2yrs,
pub.rec, not.fully.paid]
Index: []
```

[20]:
```
data.describe()
```

[20]:

	credit.policy	int.rate	installment	log.annual.inc	dti	\
count	9578.000000	9578.000000	9578.000000	9578.000000	9578.000000	
mean	0.804970	0.122640	319.089413	10.932117	12.606679	
std	0.396245	0.026847	207.071301	0.614813	6.883970	
min	0.000000	0.060000	15.670000	7.547502	0.000000	
25%	1.000000	0.103900	163.770000	10.558414	7.212500	
50%	1.000000	0.122100	268.950000	10.928884	12.665000	
75%	1.000000	0.140700	432.762500	11.291293	17.950000	
max	1.000000	0.216400	940.140000	14.528354	29.960000	

	fico	days.with.cr.line	revol.bal	revol.util	\

```
count  9578.000000         9578.000000  9.578000e+03  9578.000000
mean    710.846314         4560.767197  1.691396e+04    46.799236
std      37.970537         2496.930377  3.375619e+04    29.014417
min     612.000000          178.958333  0.000000e+00     0.000000
25%     682.000000         2820.000000  3.187000e+03    22.600000
50%     707.000000         4139.958333  8.596000e+03    46.300000
75%     737.000000         5730.000000  1.824950e+04    70.900000
max     827.000000        17639.958330  1.207359e+06   119.000000

       inq.last.6mths  delinq.2yrs      pub.rec  not.fully.paid
count     9578.000000  9578.000000  9578.000000     9578.000000
mean         1.577469     0.163708     0.062122        0.160054
std          2.200245     0.546215     0.262126        0.366676
min          0.000000     0.000000     0.000000        0.000000
25%          0.000000     0.000000     0.000000        0.000000
50%          1.000000     0.000000     0.000000        0.000000
75%          2.000000     0.000000     0.000000        0.000000
max         33.000000    13.000000     5.000000        1.000000
```

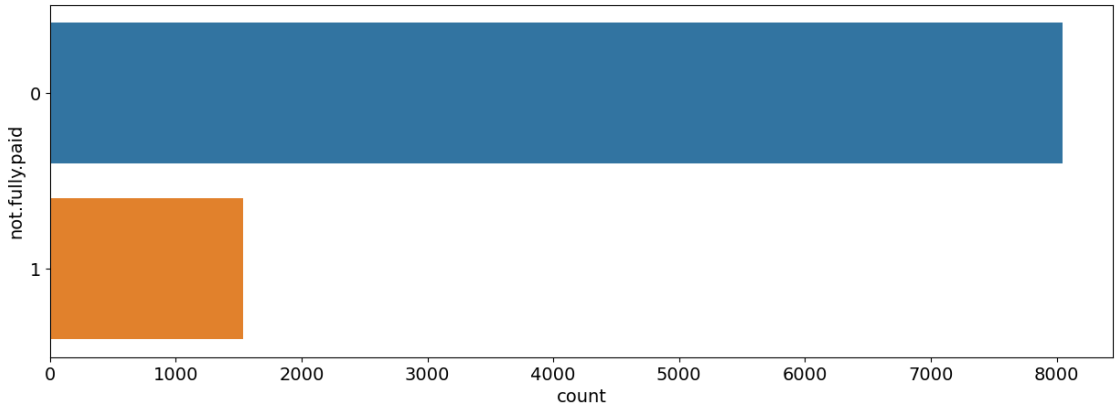
[22]:
```
# Check duplicate

print(data.duplicated().value_counts())
print(len(data))
```

```
False    9578
dtype: int64
9578
```

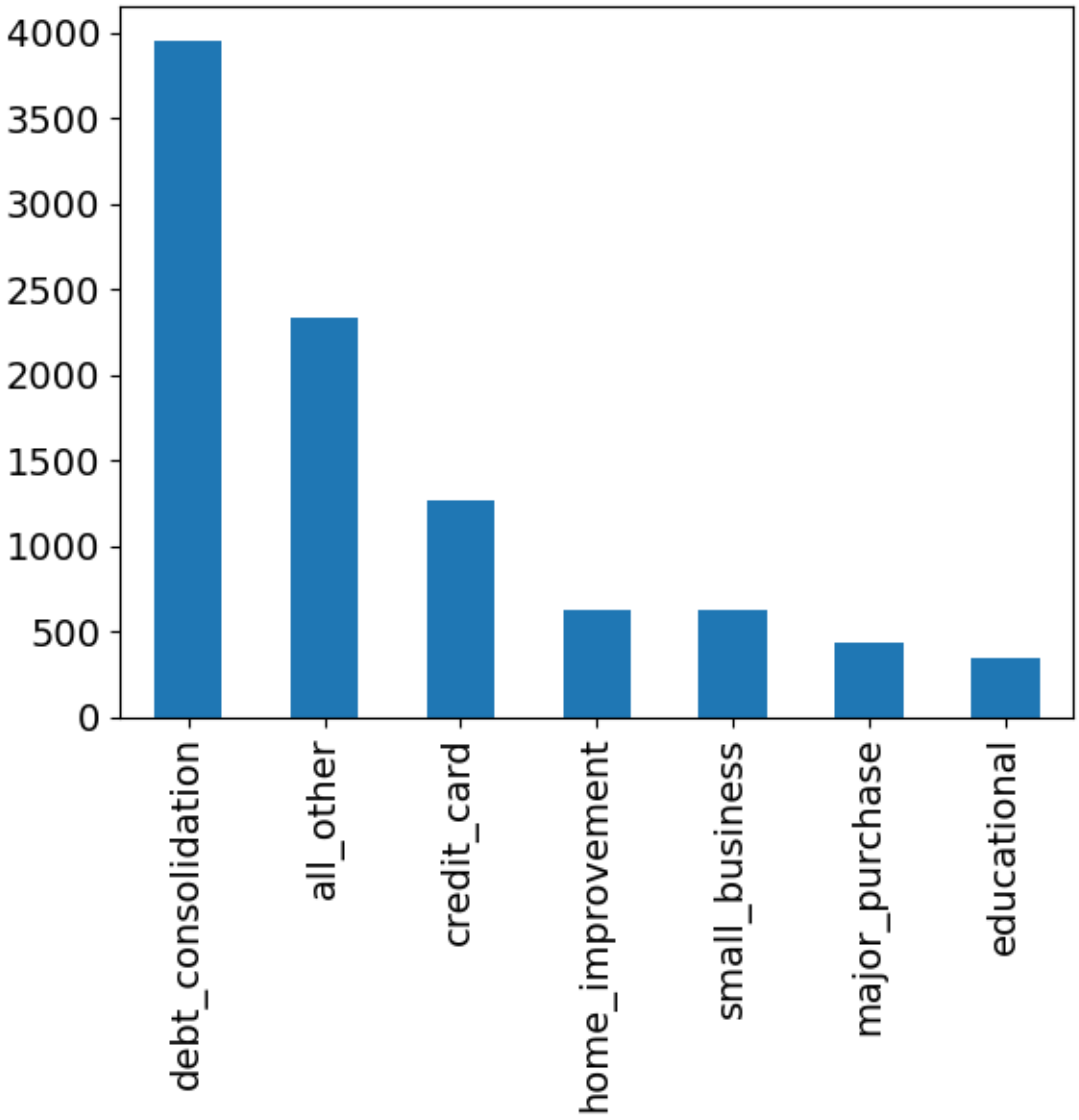
[24]:
```
# Target Variable

plt.figure(figsize=(15,5))
plt.rc("font",size=14)
sns.countplot(y="not.fully.paid",data=data)
plt.show()
```



[26]:
```
data["purpose"].value_counts().plot.bar()
```

[26]: <AxesSubplot:>



[28]:
```
data[Nums].corr()
```

[28]:

	credit.policy	int.rate	installment	log.annual.inc	\
credit.policy	1.000000	-0.294089	0.058770	0.034906	
int.rate	-0.294089	1.000000	0.276140	0.056383	
installment	0.058770	0.276140	1.000000	0.448102	
log.annual.inc	0.034906	0.056383	0.448102	1.000000	

```
dti                   -0.090901  0.220006     0.050202       -0.054065
fico                   0.348319 -0.714821     0.086039        0.114576
days.with.cr.line      0.099026 -0.124022     0.183297        0.336896
revol.bal             -0.187518  0.092527     0.233625        0.372140
revol.util            -0.104095  0.464837     0.081356        0.054881
inq.last.6mths        -0.535511  0.202780    -0.010419        0.029171
delinq.2yrs           -0.076318  0.156079    -0.004368        0.029203
pub.rec               -0.054243  0.098162    -0.032760        0.016506
not.fully.paid        -0.158119  0.159552     0.049955       -0.033439

                          dti      fico  days.with.cr.line  revol.bal  \
credit.policy     -0.090901  0.348319           0.099026  -0.187518
int.rate           0.220006 -0.714821          -0.124022   0.092527
installment        0.050202  0.086039           0.183297   0.233625
log.annual.inc    -0.054065  0.114576           0.336896   0.372140
dti                1.000000 -0.241191           0.060101   0.188748
fico              -0.241191  1.000000           0.263880  -0.015553
days.with.cr.line  0.060101  0.263880           1.000000   0.229344
revol.bal          0.188748 -0.015553           0.229344   1.000000
revol.util         0.337109 -0.541289          -0.024239   0.203779
inq.last.6mths     0.029189 -0.185293          -0.041736   0.022394
delinq.2yrs       -0.021792 -0.216340           0.081374  -0.033243
pub.rec            0.006209 -0.147592           0.071826  -0.031010
not.fully.paid     0.037362 -0.149666          -0.029237   0.053699

                   revol.util  inq.last.6mths  delinq.2yrs   pub.rec  \
credit.policy       -0.104095       -0.535511    -0.076318 -0.054243
int.rate             0.464837        0.202780     0.156079  0.098162
installment          0.081356       -0.010419    -0.004368 -0.032760
log.annual.inc       0.054881        0.029171     0.029203  0.016506
dti                  0.337109        0.029189    -0.021792  0.006209
fico                -0.541289       -0.185293    -0.216340 -0.147592
days.with.cr.line   -0.024239       -0.041736     0.081374  0.071826
revol.bal            0.203779        0.022394    -0.033243 -0.031010
revol.util           1.000000       -0.013880    -0.042740  0.066717
inq.last.6mths      -0.013880        1.000000     0.021245  0.072673
delinq.2yrs         -0.042740        0.021245     1.000000  0.009184
pub.rec              0.066717        0.072673     0.009184  1.000000
not.fully.paid       0.082088        0.149452     0.008881  0.048634

                   not.fully.paid
credit.policy          -0.158119
int.rate                0.159552
installment             0.049955
log.annual.inc         -0.033439
dti                     0.037362
fico                   -0.149666
```

```
days.with.cr.line       -0.029237
revol.bal                0.053699
revol.util               0.082088
inq.last.6mths           0.149452
delinq.2yrs              0.008881
pub.rec                  0.048634
not.fully.paid           1.000000
```

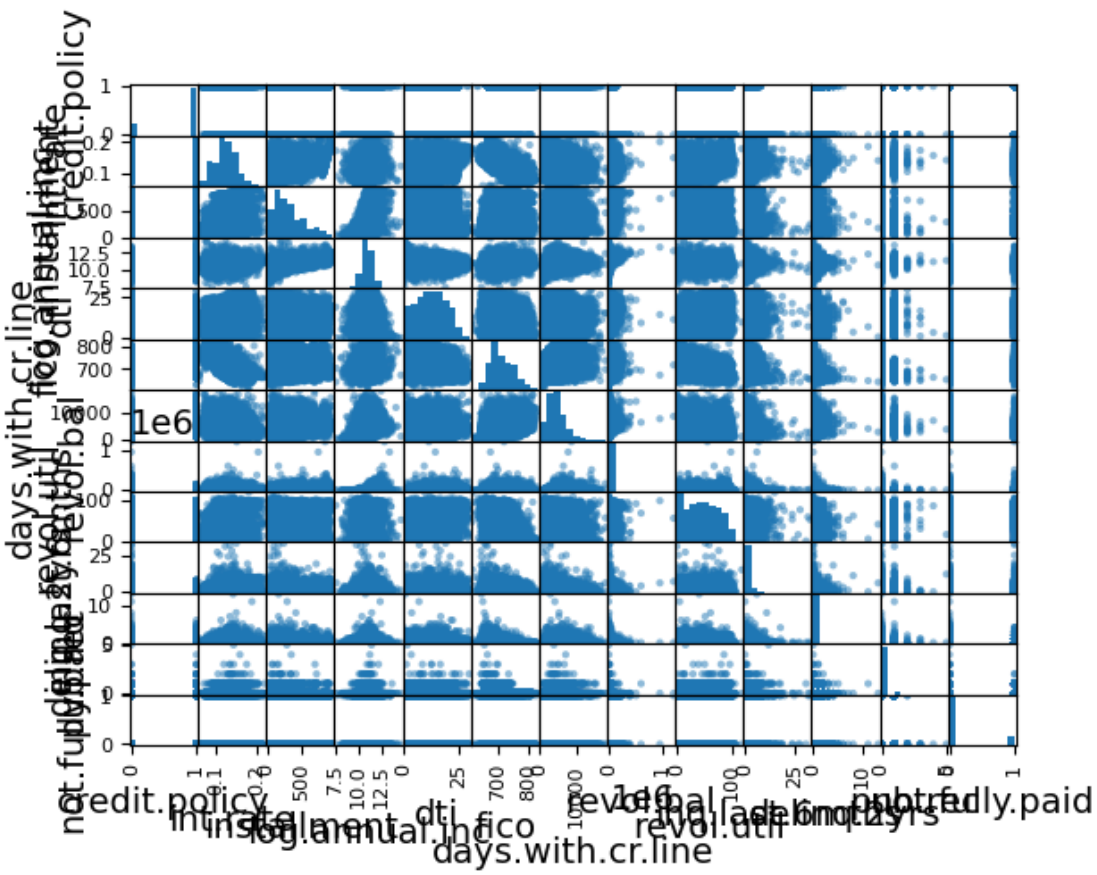
[33]:
```
from pandas.plotting import scatter_matrix
scatter_matrix(data)
```

[33]:
```
array([[<AxesSubplot:xlabel='credit.policy', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='int.rate', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='installment', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='log.annual.inc', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='dti', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='fico', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='days.with.cr.line', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='revol.bal', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='revol.util', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='inq.last.6mths', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='delinq.2yrs', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='pub.rec', ylabel='credit.policy'>,
        <AxesSubplot:xlabel='not.fully.paid', ylabel='credit.policy'>],
       [<AxesSubplot:xlabel='credit.policy', ylabel='int.rate'>,
        <AxesSubplot:xlabel='int.rate', ylabel='int.rate'>,
        <AxesSubplot:xlabel='installment', ylabel='int.rate'>,
        <AxesSubplot:xlabel='log.annual.inc', ylabel='int.rate'>,
        <AxesSubplot:xlabel='dti', ylabel='int.rate'>,
        <AxesSubplot:xlabel='fico', ylabel='int.rate'>,
        <AxesSubplot:xlabel='days.with.cr.line', ylabel='int.rate'>,
        <AxesSubplot:xlabel='revol.bal', ylabel='int.rate'>,
        <AxesSubplot:xlabel='revol.util', ylabel='int.rate'>,
        <AxesSubplot:xlabel='inq.last.6mths', ylabel='int.rate'>,
        <AxesSubplot:xlabel='delinq.2yrs', ylabel='int.rate'>,
        <AxesSubplot:xlabel='pub.rec', ylabel='int.rate'>,
        <AxesSubplot:xlabel='not.fully.paid', ylabel='int.rate'>],
       [<AxesSubplot:xlabel='credit.policy', ylabel='installment'>,
        <AxesSubplot:xlabel='int.rate', ylabel='installment'>,
        <AxesSubplot:xlabel='installment', ylabel='installment'>,
        <AxesSubplot:xlabel='log.annual.inc', ylabel='installment'>,
        <AxesSubplot:xlabel='dti', ylabel='installment'>,
        <AxesSubplot:xlabel='fico', ylabel='installment'>,
        <AxesSubplot:xlabel='days.with.cr.line', ylabel='installment'>,
        <AxesSubplot:xlabel='revol.bal', ylabel='installment'>,
        <AxesSubplot:xlabel='revol.util', ylabel='installment'>,
        <AxesSubplot:xlabel='inq.last.6mths', ylabel='installment'>,
```

```
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='installment'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='installment'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='installment'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='int.rate', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='installment', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='dti', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='fico', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='revol.util', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='log.annual.inc'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='log.annual.inc'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='dti'>,
 <AxesSubplot:xlabel='int.rate', ylabel='dti'>,
 <AxesSubplot:xlabel='installment', ylabel='dti'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='dti'>,
 <AxesSubplot:xlabel='dti', ylabel='dti'>,
 <AxesSubplot:xlabel='fico', ylabel='dti'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='dti'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='dti'>,
 <AxesSubplot:xlabel='revol.util', ylabel='dti'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='dti'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='dti'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='dti'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='dti'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='fico'>,
 <AxesSubplot:xlabel='int.rate', ylabel='fico'>,
 <AxesSubplot:xlabel='installment', ylabel='fico'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='fico'>,
 <AxesSubplot:xlabel='dti', ylabel='fico'>,
 <AxesSubplot:xlabel='fico', ylabel='fico'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='fico'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='fico'>,
 <AxesSubplot:xlabel='revol.util', ylabel='fico'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='fico'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='fico'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='fico'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='fico'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='int.rate', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='installment', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='dti', ylabel='days.with.cr.line'>,
```

```
 <AxesSubplot:xlabel='fico', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='revol.util', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='days.with.cr.line'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='days.with.cr.line'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='int.rate', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='installment', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='dti', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='fico', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='revol.util', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='revol.bal'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='revol.bal'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='revol.util'>,
 <AxesSubplot:xlabel='int.rate', ylabel='revol.util'>,
 <AxesSubplot:xlabel='installment', ylabel='revol.util'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='revol.util'>,
 <AxesSubplot:xlabel='dti', ylabel='revol.util'>,
 <AxesSubplot:xlabel='fico', ylabel='revol.util'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='revol.util'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='revol.util'>,
 <AxesSubplot:xlabel='revol.util', ylabel='revol.util'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='revol.util'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='revol.util'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='revol.util'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='revol.util'>],
[<AxesSubplot:xlabel='credit.policy', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='int.rate', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='installment', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='log.annual.inc', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='dti', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='fico', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='days.with.cr.line', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='revol.bal', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='revol.util', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='inq.last.6mths', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='delinq.2yrs', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='pub.rec', ylabel='inq.last.6mths'>,
 <AxesSubplot:xlabel='not.fully.paid', ylabel='inq.last.6mths'>],
```

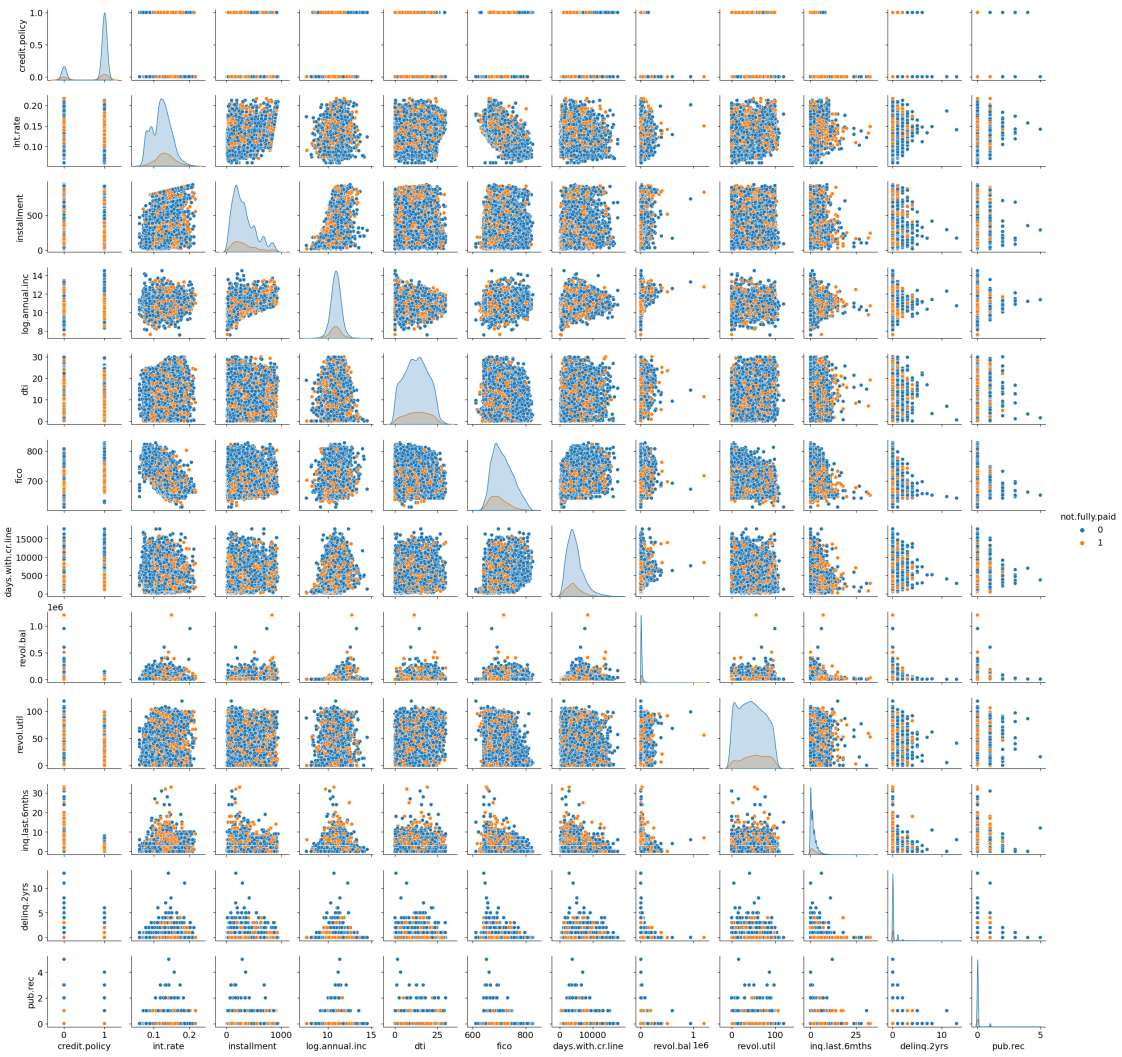
```
 [<AxesSubplot:xlabel='credit.policy', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='int.rate', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='installment', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='log.annual.inc', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='dti', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='fico', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='days.with.cr.line', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='revol.bal', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='revol.util', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='inq.last.6mths', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='delinq.2yrs', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='pub.rec', ylabel='delinq.2yrs'>,
  <AxesSubplot:xlabel='not.fully.paid', ylabel='delinq.2yrs'>],
 [<AxesSubplot:xlabel='credit.policy', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='int.rate', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='installment', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='log.annual.inc', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='dti', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='fico', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='days.with.cr.line', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='revol.bal', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='revol.util', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='inq.last.6mths', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='delinq.2yrs', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='pub.rec', ylabel='pub.rec'>,
  <AxesSubplot:xlabel='not.fully.paid', ylabel='pub.rec'>],
 [<AxesSubplot:xlabel='credit.policy', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='int.rate', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='installment', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='log.annual.inc', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='dti', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='fico', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='days.with.cr.line', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='revol.bal', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='revol.util', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='inq.last.6mths', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='delinq.2yrs', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='pub.rec', ylabel='not.fully.paid'>,
  <AxesSubplot:xlabel='not.fully.paid', ylabel='not.fully.paid'>]],
dtype=object)
```



[34]:
```
# exploratory data analysis
```

[35]:
```
sns.pairplot(data,hue="not.fully.paid",height=2)
```

[35]: <seaborn.axisgrid.PairGrid at 0x17d43343e50>

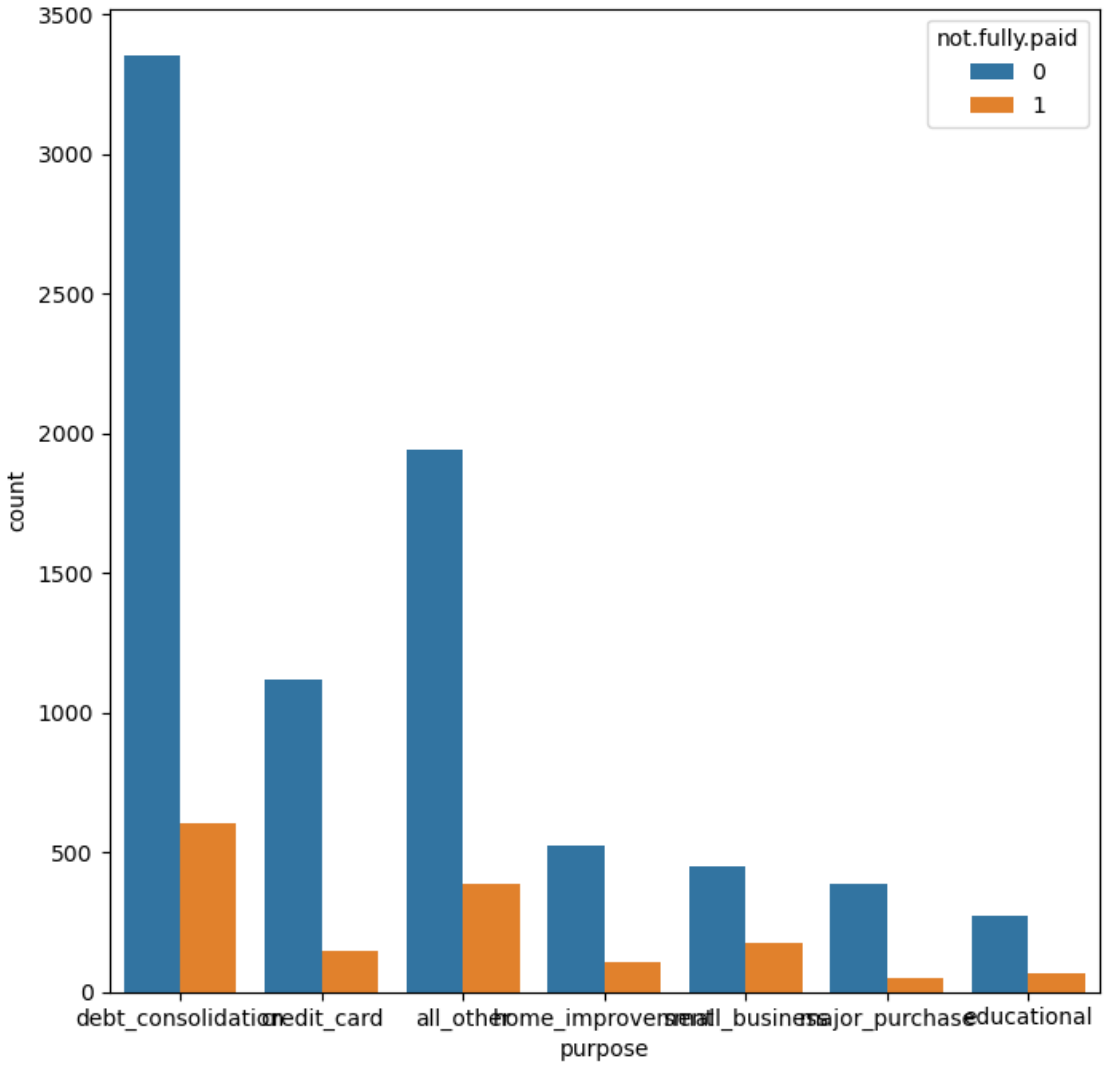


[10]:
```
len(data.loc[data["fico"] < 300])
```

[10]: 0

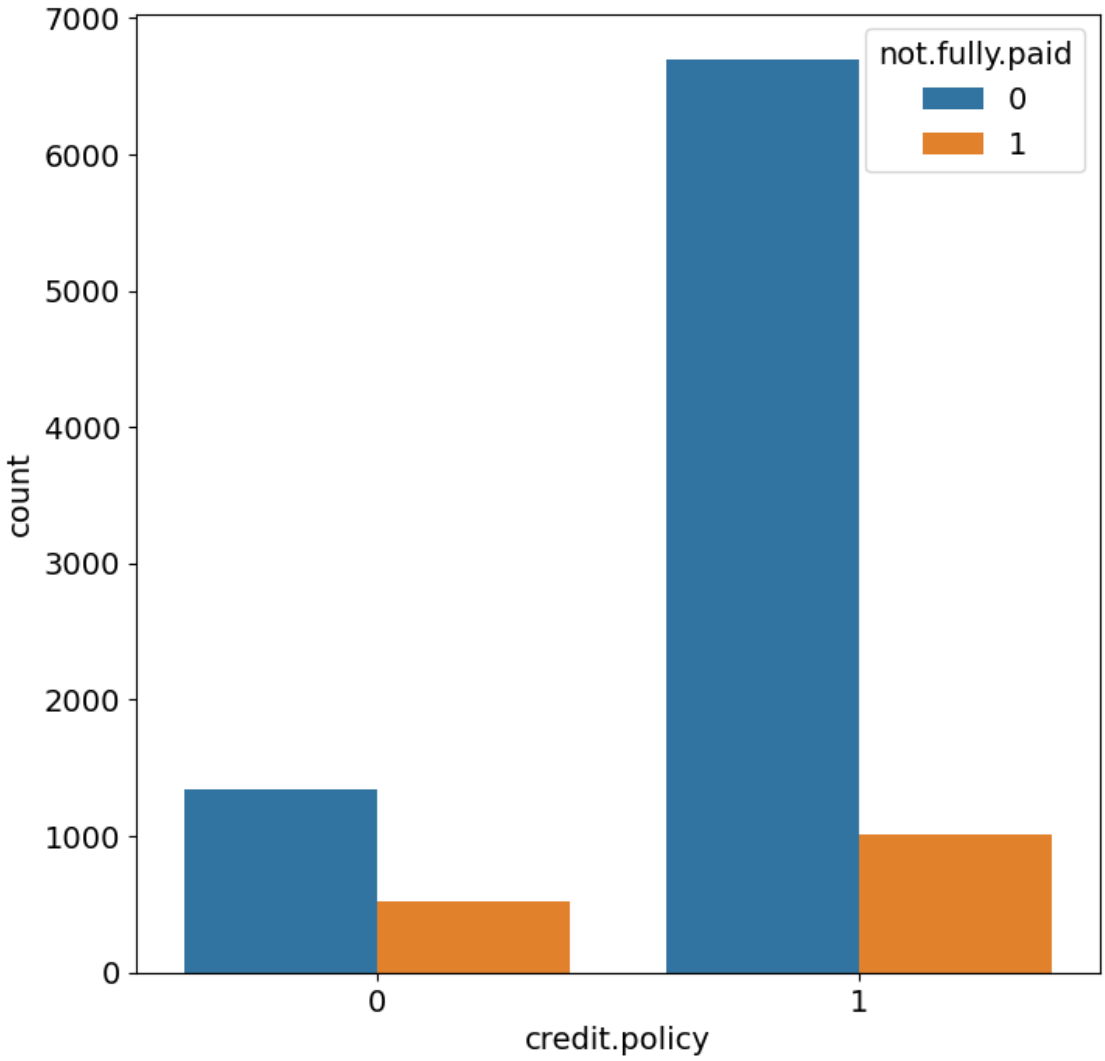
[11]:
```
plt.figure(figsize=(8,8))
sns.countplot(x="purpose",hue="not.fully.paid",data=data)
```

[11]: <AxesSubplot:xlabel='purpose', ylabel='count'>



[46]:
```
plt.figure(figsize=(8,8))
sns.countplot(x="credit.policy",hue="not.fully.paid",data=data)
```

[46]: <AxesSubplot:xlabel='credit.policy', ylabel='count'>



[48]:
```
data.columns
```

[48]:
```
Index(['credit.policy', 'purpose', 'int.rate', 'installment', 'log.annual.inc',
       'dti', 'fico', 'days.with.cr.line', 'revol.bal', 'revol.util',
       'inq.last.6mths', 'delinq.2yrs', 'pub.rec', 'not.fully.paid'],
      dtype='object')
```

[20]:
```
df = pd.get_dummies(data)
```

[12]:
```
df
```

[12]:

	credit.policy	int.rate	installment	log.annual.inc	dti	fico	\
0	1	0.1189	829.10	11.350407	19.48	737	
1	1	0.1071	228.22	11.082143	14.29	707	

```
2                1    0.1357       366.86       10.373491  11.63   682
3                1    0.1008       162.34       11.350407   8.10   712
4                1    0.1426       102.92       11.299732  14.97   667
…              …        …            …              …     …   …
9573             0    0.1461       344.76       12.180755  10.39   672
9574             0    0.1253       257.70       11.141862   0.21   722
9575             0    0.1071        97.81       10.596635  13.09   687
9576             0    0.1600       351.58       10.819778  19.18   692
9577             0    0.1392       853.43       11.264464  16.28   732

      days.with.cr.line  revol.bal  revol.util  inq.last.6mths  delinq.2yrs  \
0           5639.958333      28854        52.1               0            0
1           2760.000000      33623        76.7               0            0
2           4710.000000       3511        25.6               1            0
3           2699.958333      33667        73.2               1            0
4           4066.000000       4740        39.5               0            1
…                   …         …           …              …           …
9573       10474.000000     215372        82.1               2            0
9574        4380.000000        184         1.1               5            0
9575        3450.041667      10036        82.9               8            0
9576        1800.000000          0         3.2               5            0
9577        4740.000000      37879        57.0               6            0

      pub.rec  not.fully.paid  purpose_all_other  purpose_credit_card  \
0           0               0                  0                    0
1           0               0                  0                    1
2           0               0                  0                    0
3           0               0                  0                    0
4           0               0                  0                    1
…       …             …                …                  …
9573        0               1                  1                    0
9574        0               1                  1                    0
9575        0               1                  0                    0
9576        0               1                  0                    0
9577        0               1                  0                    0

      purpose_debt_consolidation  purpose_educational  \
0                              1                    0
1                              0                    0
2                              1                    0
3                              1                    0
4                              0                    0
…                            …                  …
9573                           0                    0
9574                           0                    0
9575                           1                    0
9576                           0                    0
```

```
9577                            1                    0

      purpose_home_improvement  purpose_major_purchase  purpose_small_business
0                            0                       0                       0
1                            0                       0                       0
2                            0                       0                       0
3                            0                       0                       0
4                            0                       0                       0
…                          …                       …                       …
9573                         0                       0                       0
9574                         0                       0                       0
9575                         0                       0                       0
9576                         1                       0                       0
9577                         0                       0                       0

[9578 rows x 20 columns]
```

[21]:
```
x = df.drop(columns=["not.fully.paid"],axis=1)
y = df["not.fully.paid"]
```

[22]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.
 ↪2,random_state=42)
```

[10]:
```
from sklearn.linear_model import LogisticRegression
```

[11]:
```
clf = LogisticRegression(random_state=0)
clf.fit(x_train,y_train)
```

[11]:
```
LogisticRegression(random_state=0)
```

[12]:
```
y_pred = clf.predict(x_test)
```

[25]:
```
from sklearn.metrics import accuracy_score
```

[26]:
```
acc = accuracy_score(y_test,y_pred)
```

[27]:
```
print("Logistic Regression model accuracy (in %):", acc*100)
```

```
Logistic Regression model accuracy (in %): 84.13361169102296
```

[28]:
```
from sklearn.metrics import confusion_matrix
cm_lr = confusion_matrix(y_test,y_pred)
```

[23]:
```
print(cm_lr)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
```

```
~\AppData\Local\Temp\ipykernel_7444\3515638039.py in <module>
----> 1 print(cm_lr)

NameError: name 'cm_lr' is not defined
```

1 XGboost Model

[18]:
```
from sklearn.datasets import load_iris
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=0)
cross_val_score(clf,x,y,cv=20)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_7444\1880895377.py in <module>
      3 from sklearn.tree import DecisionTreeClassifier
      4 clf = DecisionTreeClassifier(random_state=0)
----> 5 cross_val_score(clf,x,y,cv=20)

NameError: name 'x' is not defined
```

[9]:
```
tree = DecisionTreeClassifier(random_state=0).fit(x,y)
tree.predict(x_test)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_7444\3931284864.py in <module>
----> 1 tree = DecisionTreeClassifier(random_state=0).fit(x,y)
      2 tree.predict(x_test)

NameError: name 'DecisionTreeClassifier' is not defined
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: