

# Exploratory Data Analysis

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.api import sm
from statsmodels.graphics.api import abline_plot
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn import linear_model, preprocessing
from sklearn.model_selection import train_test_split
import warnings
from statsmodels.stats.multicomp import pairwise_tukeyhsd
from statsmodels.stats.multicomp import MultiComparison
cwd = os.getcwd()
print(cwd)
import statsmodels.api as sm

import scipy.stats as stats
import statsmodels.formula.api as smf
from statsmodels.formula.api import ols
from sklearn import preprocessing
from matplotlib import pyplot
from sklearn.metrics import precision_recall_curve
from sklearn.metrics import f1_score
from sklearn.metrics import auc
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix, roc_curve, roc_auc_score
from sklearn.metrics import accuracy_score, log_loss
from matplotlib import pyplot
from sklearn.ensemble import RandomForestClassifier
import itertools
from sklearn import svm, datasets
from sklearn.metrics import confusion_matrix
from sklearn.linear_model import LinearRegression
from sklearn import metrics
from tabulate import tabulate

C:\Users\Tom\Documents\GitHub
```

```
In [2]: loans = pd.read_csv("cleanloans.csv")
```

## Data Description

A total of 5000 observations, 24 total variables that are both demographics and information regarding Credit current and History of the customers. Out of the 24 variables we have 10 numerical and 14 categorical. Except of the Credit Amount, Current Address number of years, Credit amount originally lend, Number of Credit Accounts, Number of dependents, and Age, all the other variables are ordinal that describe ranges of the variable. The reason for the use of the ordinal measurement level is due to legal requirements to protect the privacy of the customers data.

By reviewing the descriptive statistics of the data, we can observe that the loans given to customers have an average value of \$3200, with an average duration of 21 months. The customers are younger with an average age of 35 years old. 1 dependent and an average number of credit amount 1.45, while they reside in their current residency the last 3 years. Of all the customers, we have an average 30 percent defaulted loans.

Overall, 75 percent of the customers is younger than 42 years old and have received a loan of less than \$4000

```
In [3]: loans.describe()
```

	Customer_ID	Duration_in_Months	Credit_Amount	Inst_Rt_Income	Current_Address_Yrs	Age	Num_CC	Dependents	Default
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	102500.500000	20.903000	3271.258000	2.973000	2.845000	35.5397	1.407000	1.155000	0.29906
std	1443.520003	12.063989	2821.607329	1.118267	1.103276	11.379917	0.577423	0.361941	0.447392
min	100001.000000	4.000000	250.000000	1.000000	1.000000	19.000000	1.000000	1.000000	0
25%	101250.750000	12.000000	1365.500000	2.000000	2.000000	27.000000	1.000000	1.000000	0
50%	102500.500000	18.000000	2319.500000	3.000000	3.000000	33.000000	1.000000	1.000000	0
75%	103750.250000	24.000000	3972.250000	4.000000	4.000000	42.000000	2.000000	1.000000	0
max	105000.000000	72.000000	18424.000000	4.000000	4.000000	75.000000	4.000000	2.000000	1

```
In [4]: loans = loans.rename(columns={'Years_At_Present_Employment': 'Employment', 'Status_Checking_Acc': 'Checking',
```

```
In [5]: loans1 = loans.iloc[1:5000,1:22]
table = loans1.describe()
```

```
In [6]: print(tabulate(table,headers="keys",tablefmt="fancy_grid"))
```

	Months	Credit_Amount	Inst_Rt_Income	Address	Age	Num_CC	Dependents
count	4999	4999	4999	4999	4999	4999	4999
mean	20.29906	3271.68	2.97279	2.84477	35.5397	1.40688	1.15503
std	12.0534	2821.73	1.11828	1.10327	11.3633	0.57742	0.36197
min	4	250	1	1	19	1	1
25%	12	1366	2	2	27	1	1
50%	18	2320	3	3	33	1	1
75%	24	3972.5	4	4	42	2	1
max	72	18424	4	4	75	4	2

```
In [7]: loans.dtypes
```

Customer_ID	int64
Checking	object
Months	int64
Credit_History	object
Purpose_Credit_Taken	object
Credit_Amount	int64
Savings_Acc	object
Employment	object
Inst_Rt_Income	int64
Marital_Status_Gender	object
Other_Debtors_Guarantors	object
Address	int64
Property	object
Age	int64
Other_Inst_Plans	object
Housing	object
Num_CC	int64
Job	object
Dependents	int64
Phone	object
Mar	object
Default	int64
Count	int64
Default_On_Payment2	object
dtype:	object

## Description of Categorical Data

For the data set we will be working on, most of our customers are receiving a loan to obtain Electronics, they are Single Males and they have a good Credit history, with no information about their Checking accounts. There majority own a Car, they are employed in their recent company for less than 4 years, they own their home. In their majority hold low Skills with no specialization Job, and they are American Citizens.

```
In [8]: loans1.describe(include='object')
fig=plt.figure(figsize=(10,10))
fig=plt.figure(figsize=(10,10))
```

	Checking	Credit_History	Purpose_Credit_Taken	Savings_Acc	Employment	Marital_Status_Gender	Other_Debtors_Guarantors	Property
count	4999	4999	4999	4999	4999	4999	4999	4999
unique	4	5	10	5	5	4	3	4
top	NoAcc	Current	Radio_TV	Low	Less4	MaleSingle	No	Car
freq	1970	2650	1399	3015	1695	2739	4534	1660

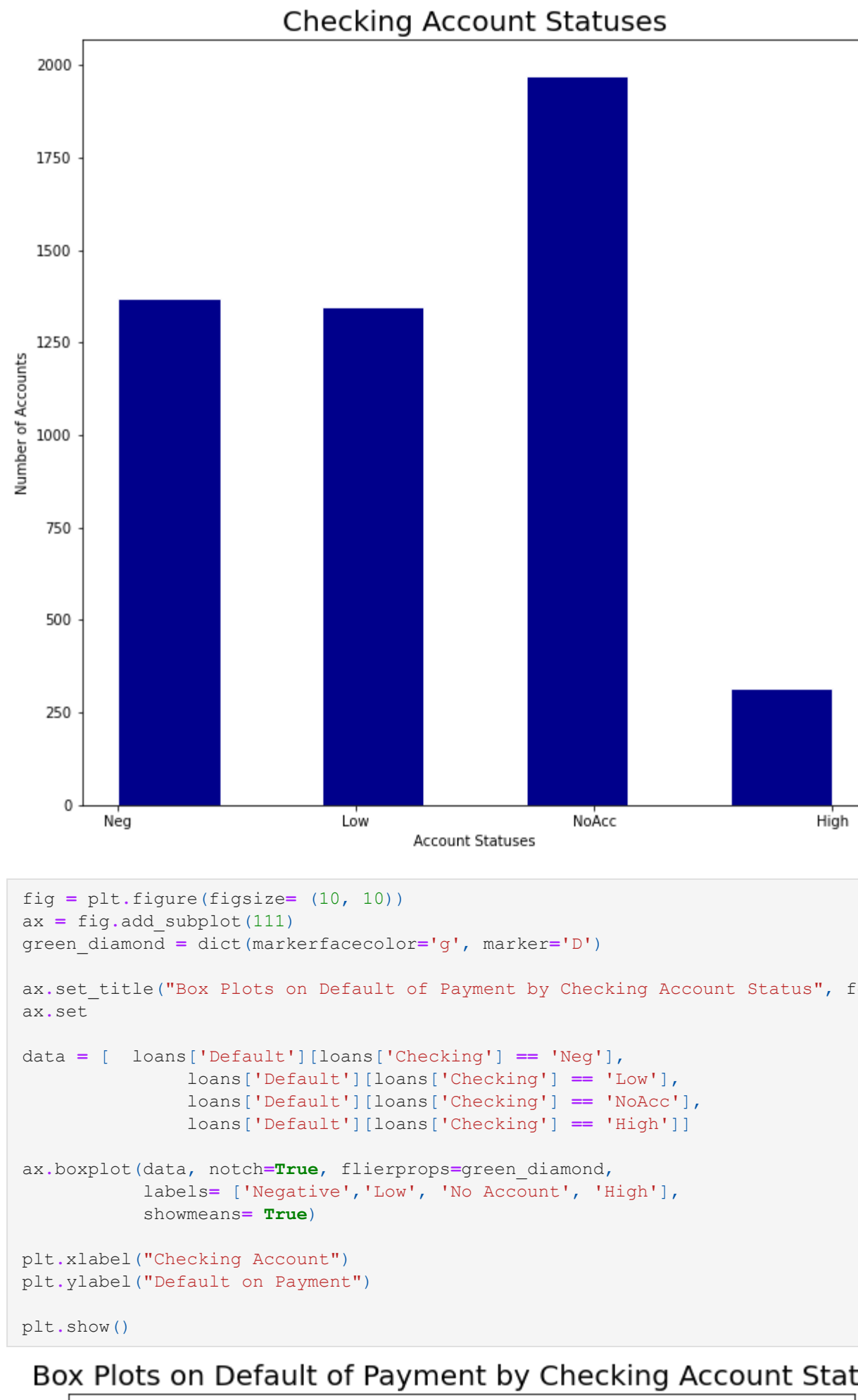
## Visual Representations

Our goal is to understand better our dataset before we work with our model. Disecting the dataset and understanding the relations of the variables.

## Credit History Frequency Graph

We can see that most of the Customers ( more than 50% ) are Current with their payments, while 1/3 of their customers are deliquent for more than 3 payments.

```
In [9]: fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111)
ax.set_title("Credit History", fontsize=20)
ax.set_xlabel("Credit History", color = "darkblue", bins=7, edgecolor="white")
ax.set_ylabel("Number of Accounts")
plt.show()
```



## Checking Accounts Information

Unfortunately, we no have information about the Checking accounts of almost 1/3 of our clients. For the rest the largest majority has either a negative balance or low balance, only 1/8th of our customers have a high balance.

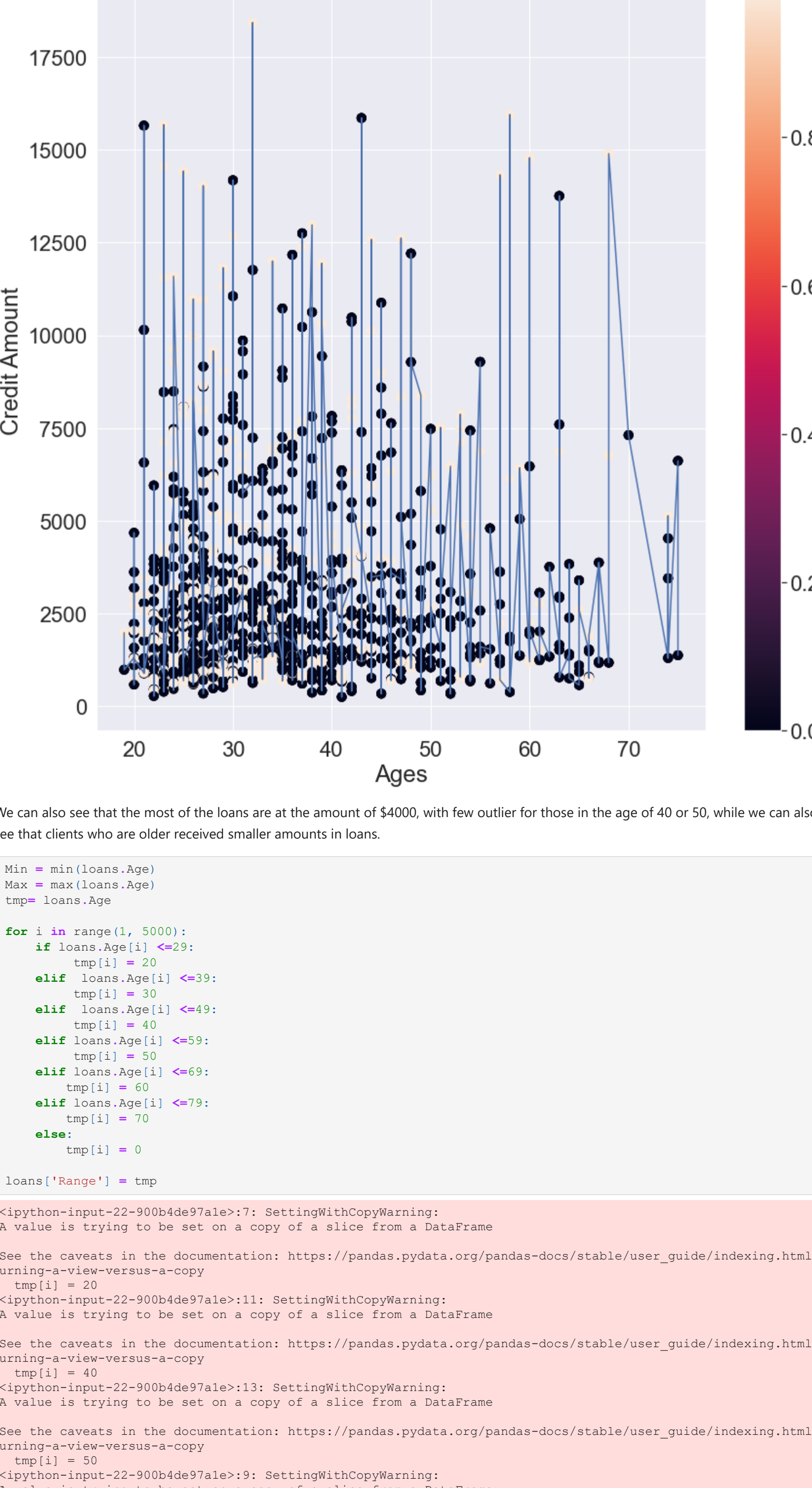
```
In [10]: fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111)
ax.set_title("Checking Account Statuses", fontsize=20)
ax.set_xlabel("Checking Account", color = "darkblue", bins=7, edgecolor="white")
ax.set_ylabel("Number of Accounts")
plt.show()
```



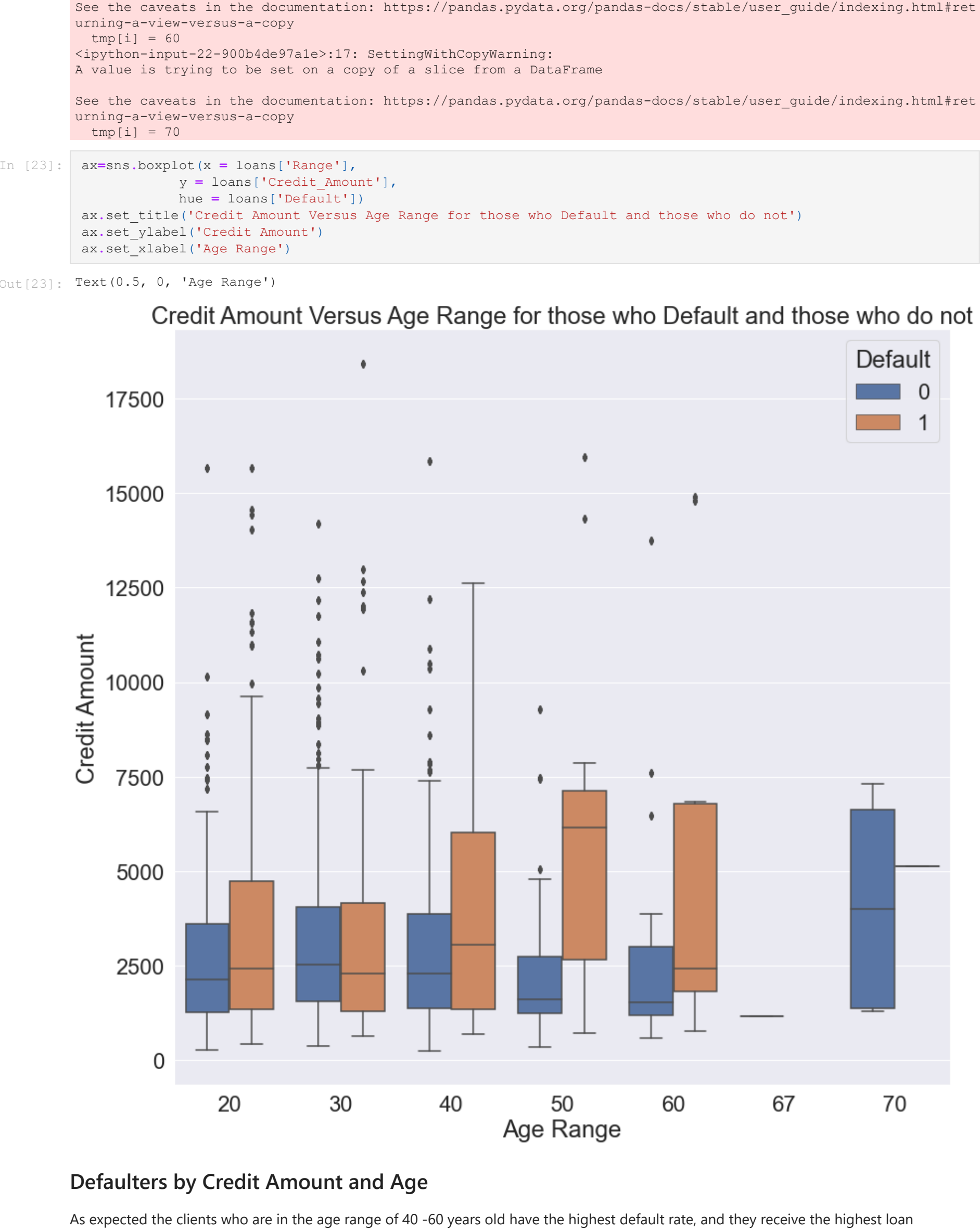
```
In [11]: fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111)
green_diamond = dict(markerfacecolor='g', marker='D')
ax.set_title("Box Plots on Default of Payment by Checking Account Status", fontsize=20)
ax.set_xlabel("Checking Account")
ax.set_ylabel("Default on Payment")
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set_xlabel('Checking Account')
ax.set_xticks(['Negative', 'Low', 'No Account', 'High'])
ax.set_xticklabels(['Negative', 'Low', 'No Account', 'High'])
ax.set_yticks([0.0, 0.2, 0.4, 0.6, 0.8, 1.0])
ax.set_ylim(0, 1.0)
ax.set_xlim(0, 4)
ax.set_ylabel('Default on Payment')
ax.set
```



Credit Amount by age of debtors



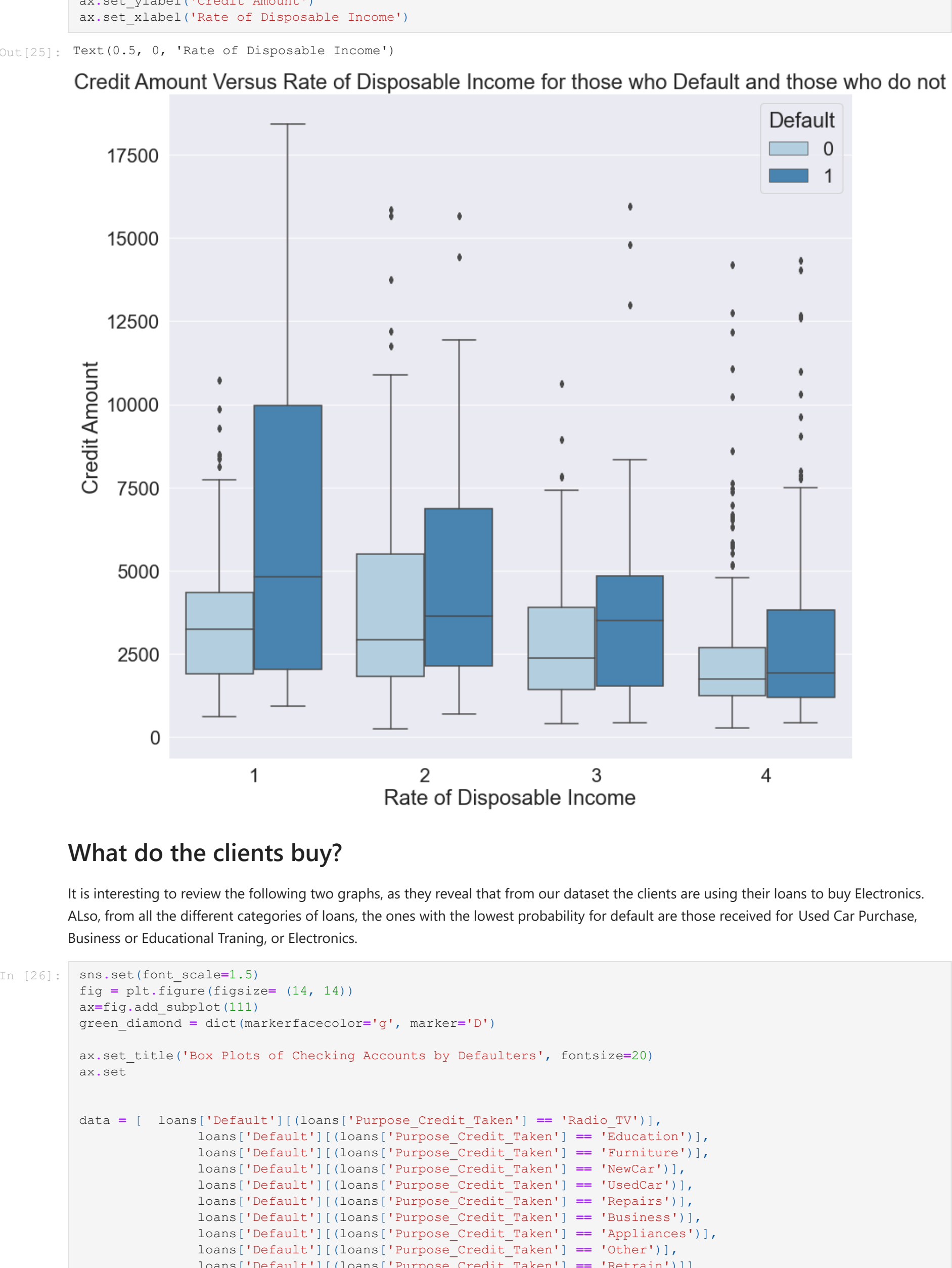
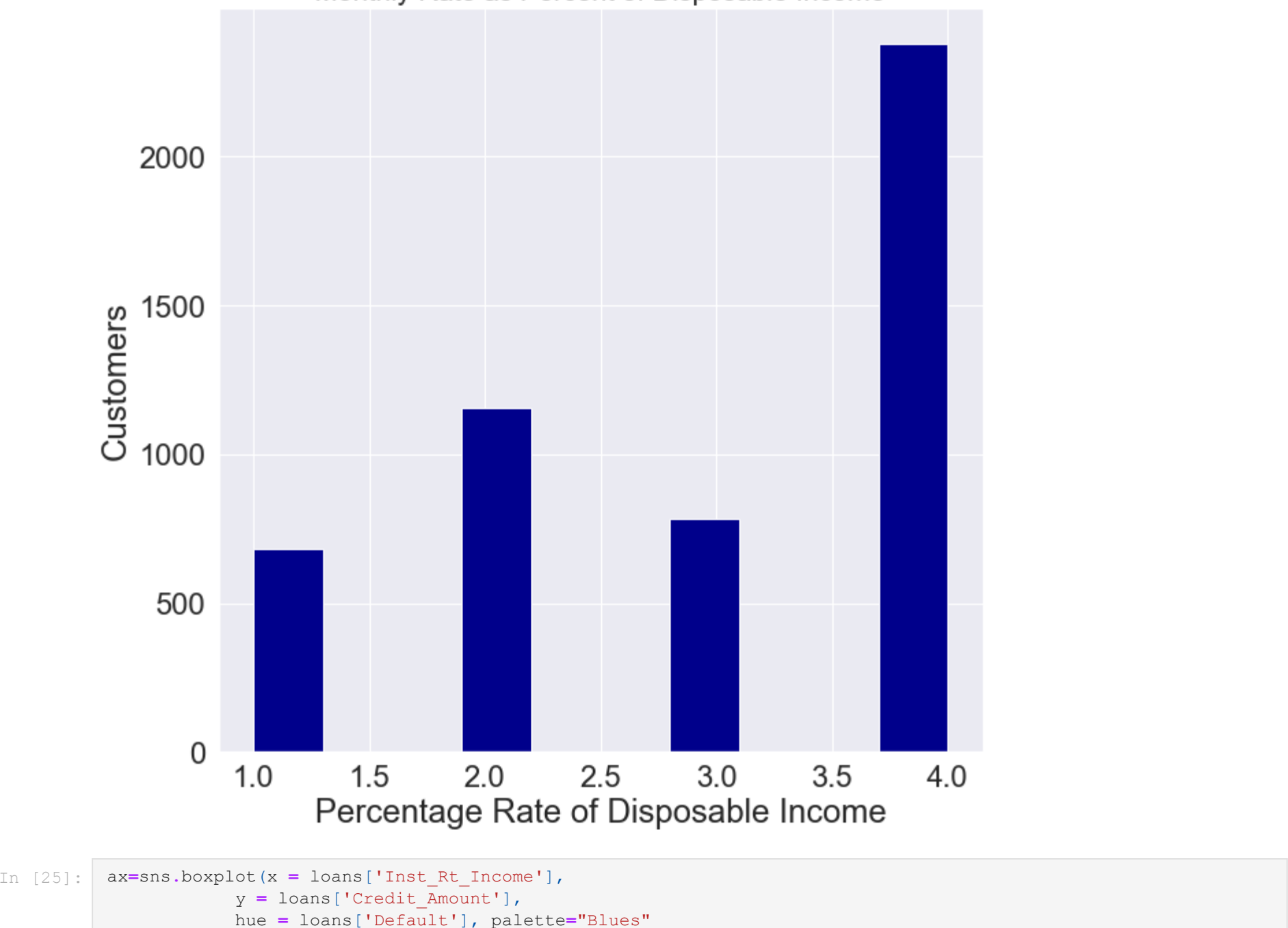
We can also see that the most of the loans are at the amount of \$4000, with few outlier for those in the age of 40 or 50, while we can also see that clients who are older received smaller amounts in loans.



Defaulters by Credit Amount and Age

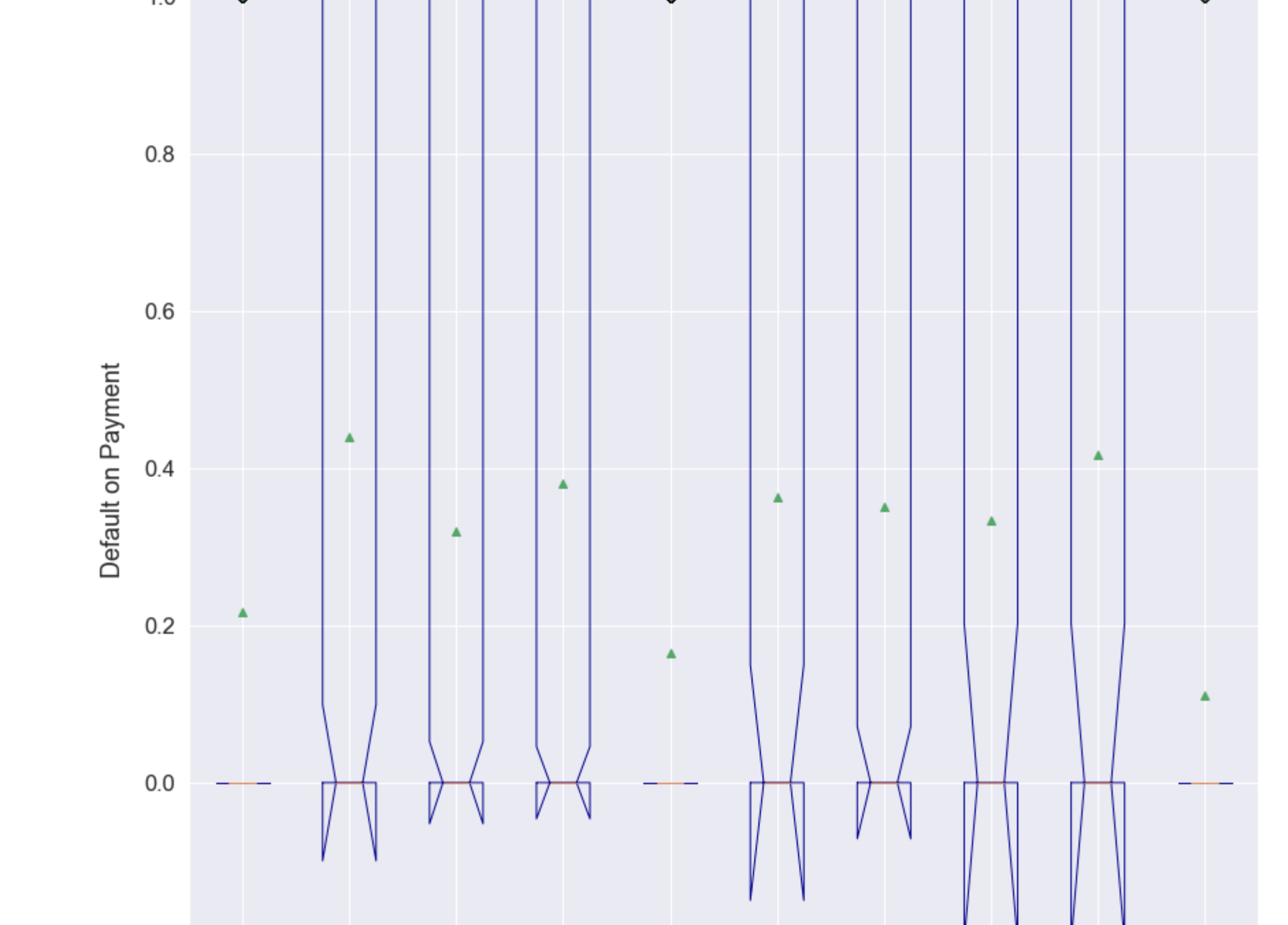
As expected the clients who are in the age range of 40-60 years old have the highest default rate, and they receive the highest loan amounts. While clients who are older than 60 DO NOT default. What is interesting to identify is the outliers in the range of 20-30 years old, in this age range most of the clients pay their loans but there are few outliers who receive very high loan amounts and it would be interesting to see if they are prone to default.

While in the following graph we can see that the probability to default is high for the borrowers who already have accumulate higher level of debt.



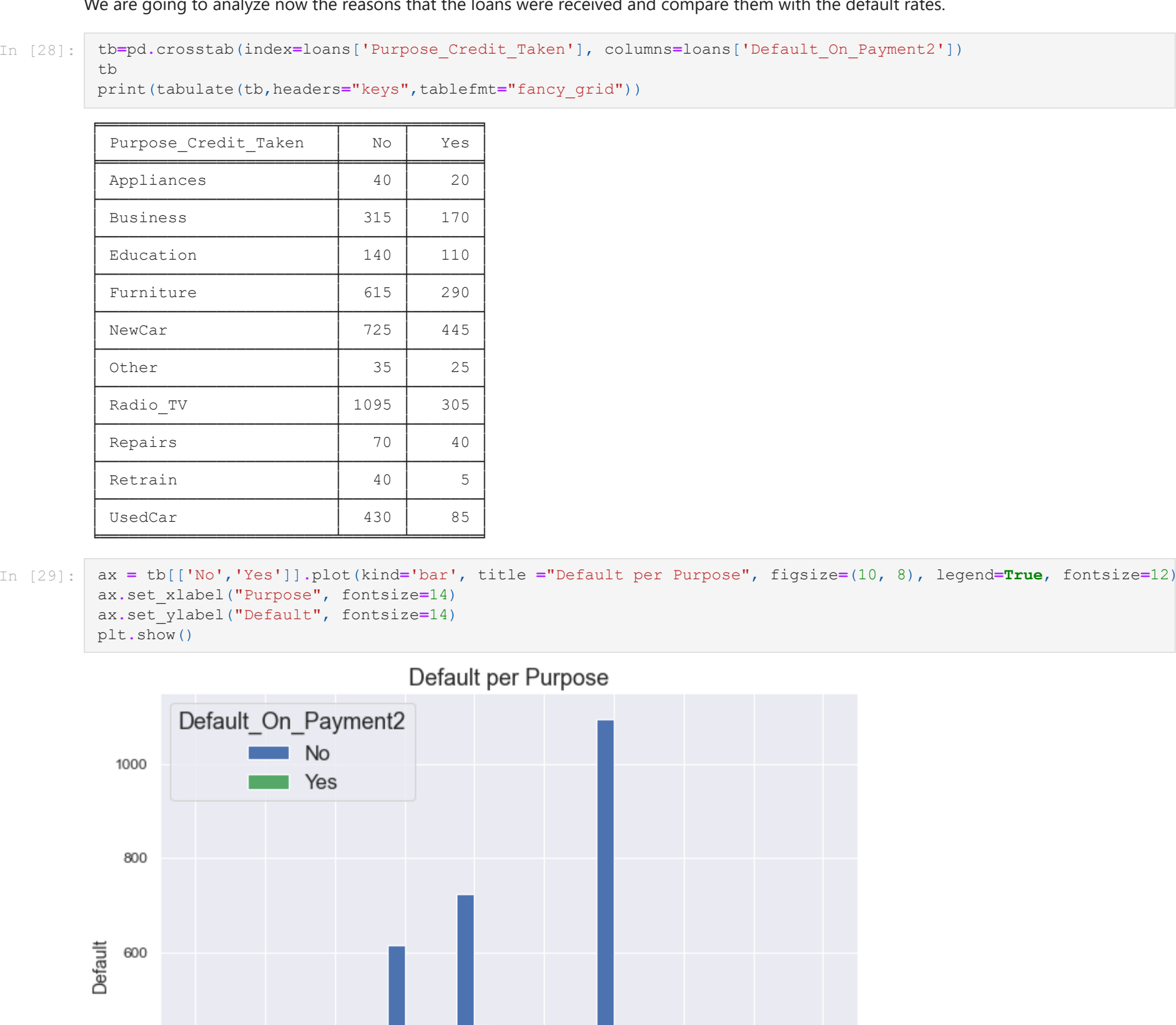
What do the clients buy?

It is interesting to review the following two graphs, as they reveal that from our dataset the clients are using their loans to buy Electronics. Also, from all the different categories of loans, the ones with the lowest probability for default are those received for Used Car Purchase, Business or Educational Training, or Electronics.

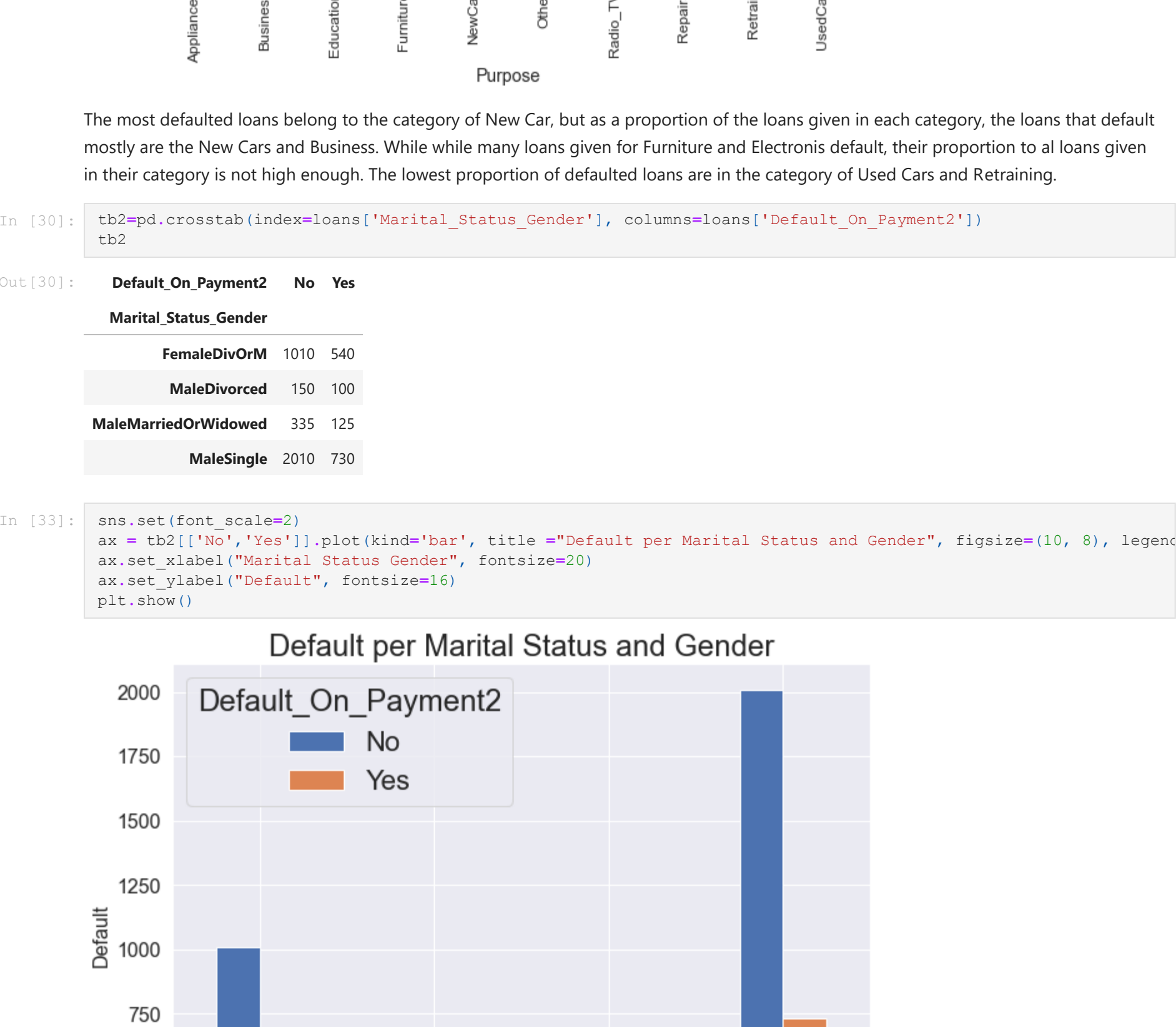


Visualizing the distribution of Defaulted loans.

We are going to analyze now the reasons that the loans were received and compare them with the default rates.



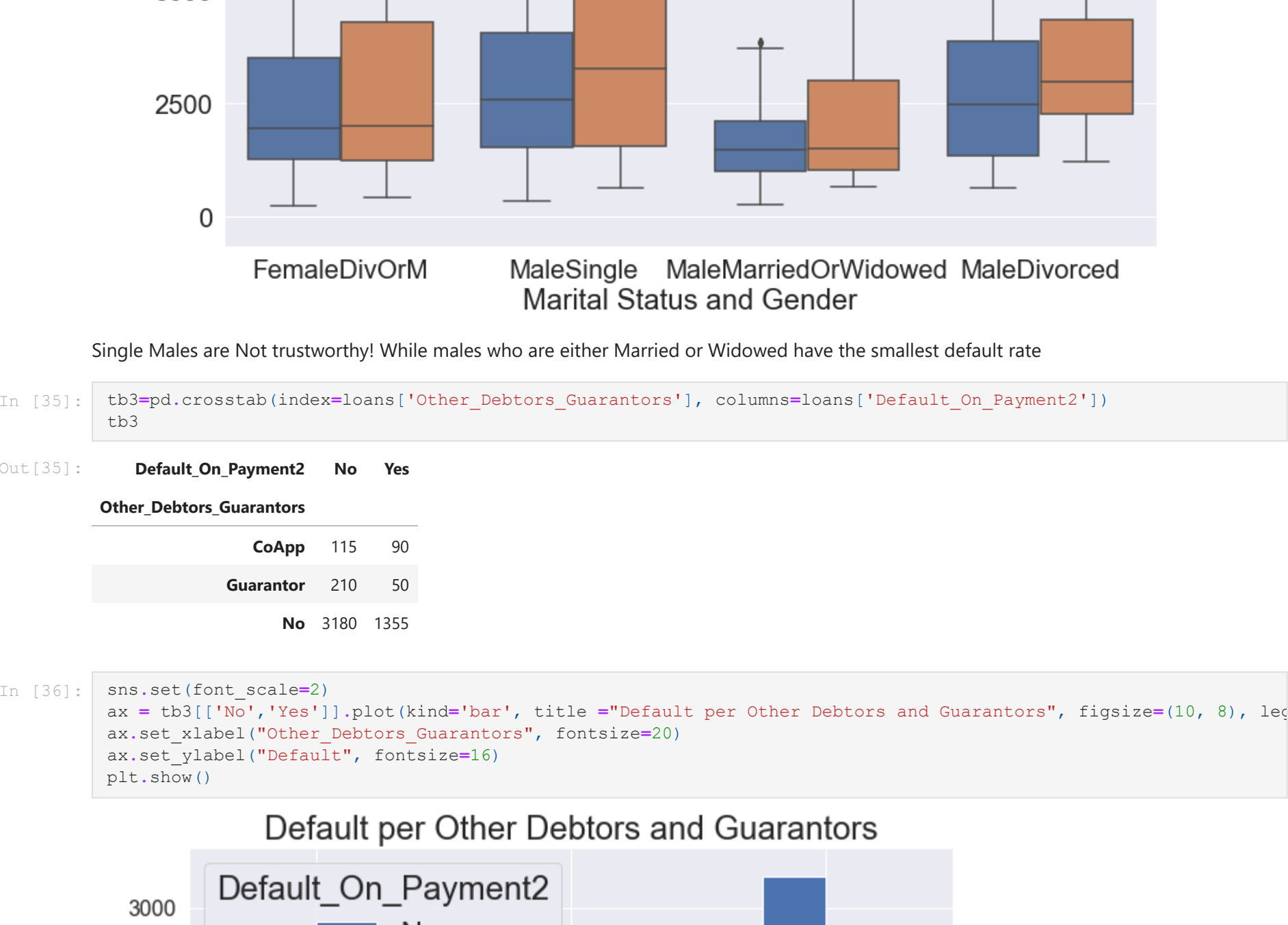
The most defaulted loans belong to the category of New Car, but as a proportion of the loans given in each category, the loans that default mostly are the New Cars and Business. While while many loans given for Furniture and Electronics default, their proportion to all loans given in their category is not high enough. The lowest proportion of defaulted loans are in the category of Used Cars and Retrain.



Marital Status and Gender Versus Credit Amount for those who Default and those who do not



Single Males are not trustworthy! While males who are either Married or Widowed have the smallest default rate



Clearly those who use Guarantors default in the lowest rates



The clients who have no Skills and they are residents/citizens defaulter proportionally more than those with higher level jobs or with low

