ted2-1

June 7, 2017

1 Exercice 2: Part 1

2 Import the necessary libraries and open the data set

import pandas as pd import numpy as np from sklearn.tree import DecisionTreeClassifier from sklearn.ensemble import RandomForestClassifier from sklearn.cross_validation import train_test_split, ShuffleSplit, cross_val_score from sklearn import preprocessing import matplotlib.pyplot as plt import plotly.plotly as py import plotly.graph_objs as go

data_set = pd.read_csv('train.tsv', sep='') good = data_set[data_set.Label== 1] bad = data_set[data_set.Label ==2]

3 Function to Create Qualitive Plots

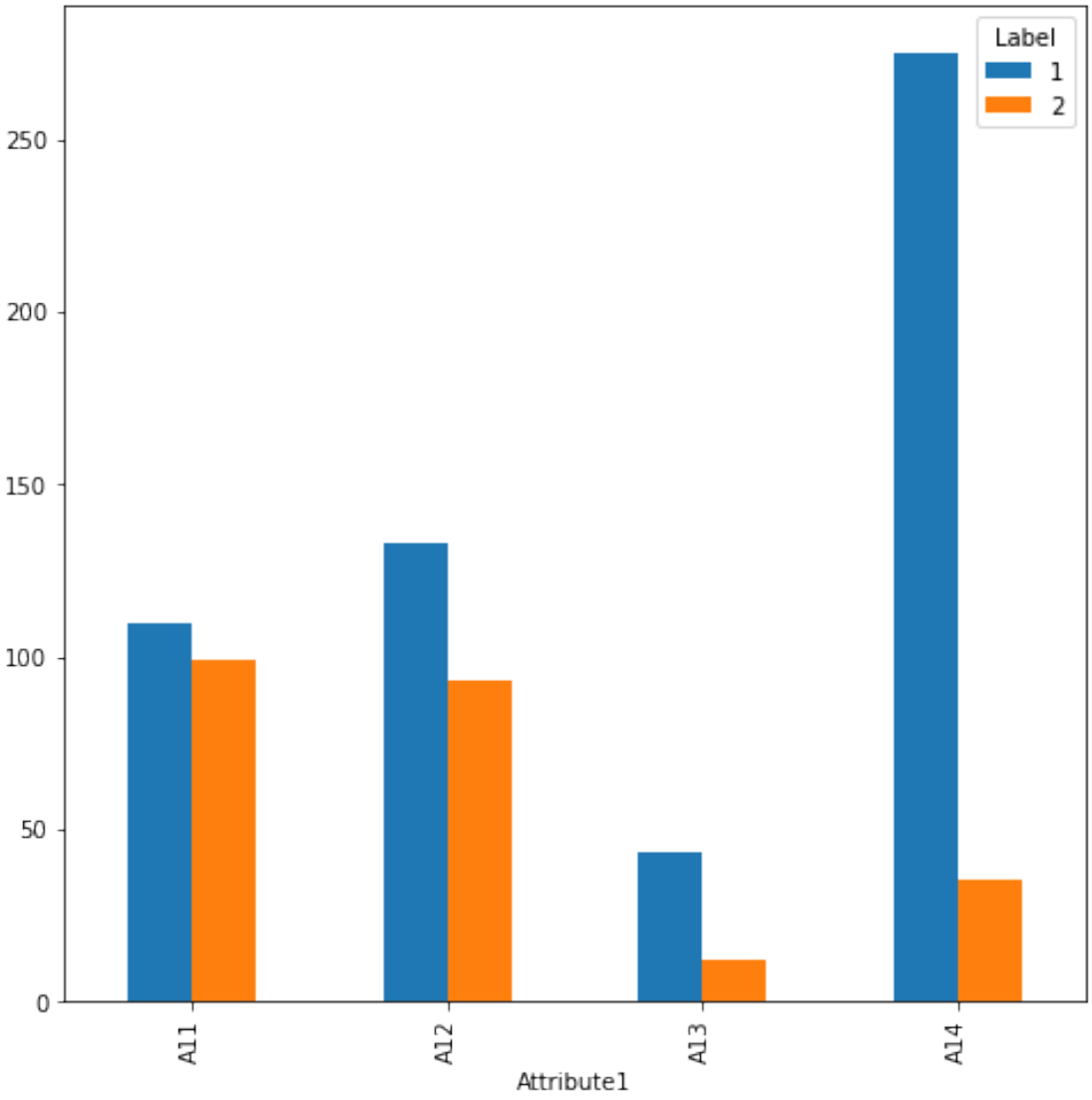
```
In [11]: def QualitivePlot(data_set, No):
             string = "Attribute"+No.__str__()
             attr1 = pd.crosstab(index=data_set[string], columns=data_set['Label'])
             attr1.plot(kind='bar', figsize=(8,8))
             plt.show()
```

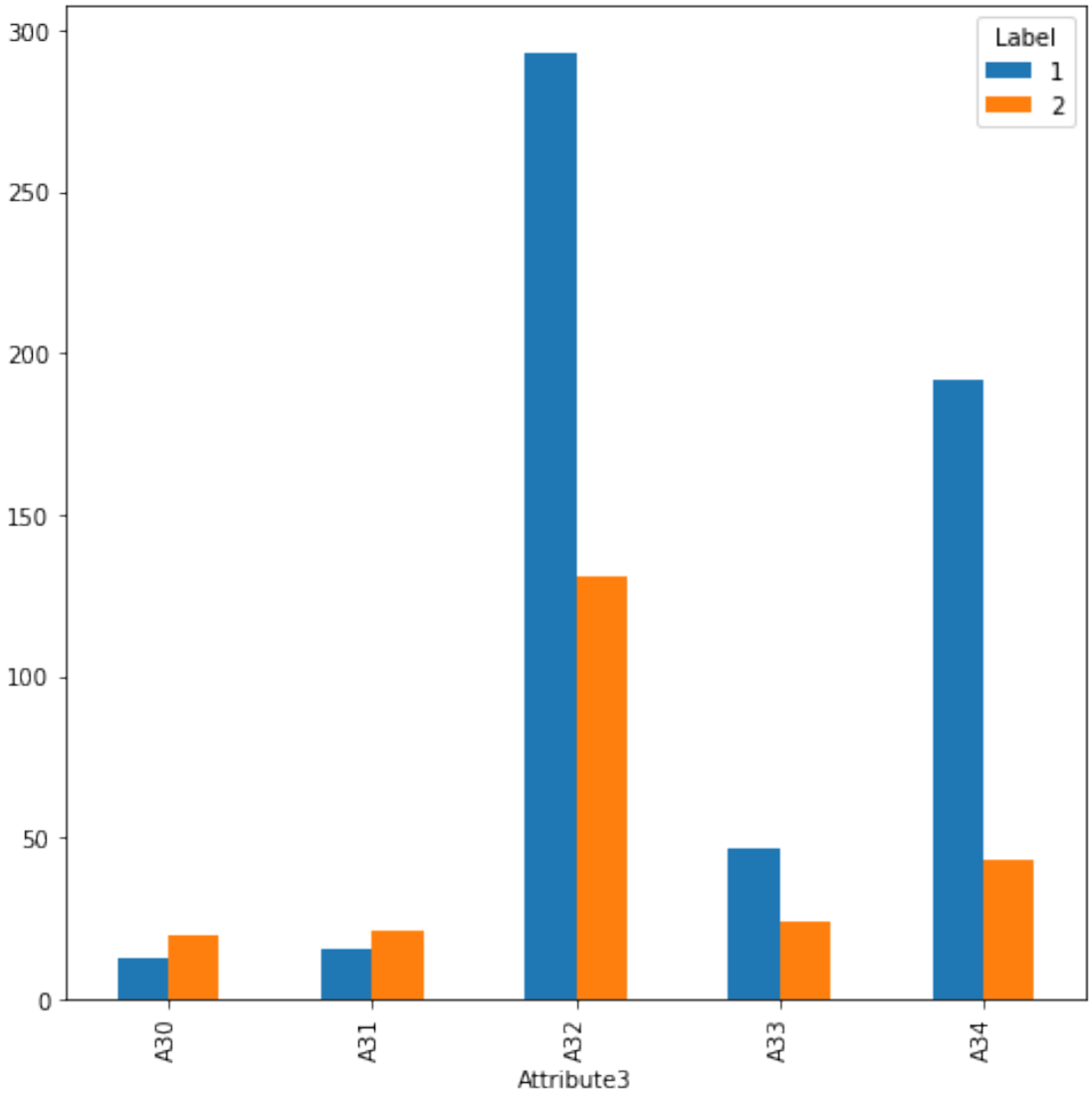
4 Function to Create Numerical Plots

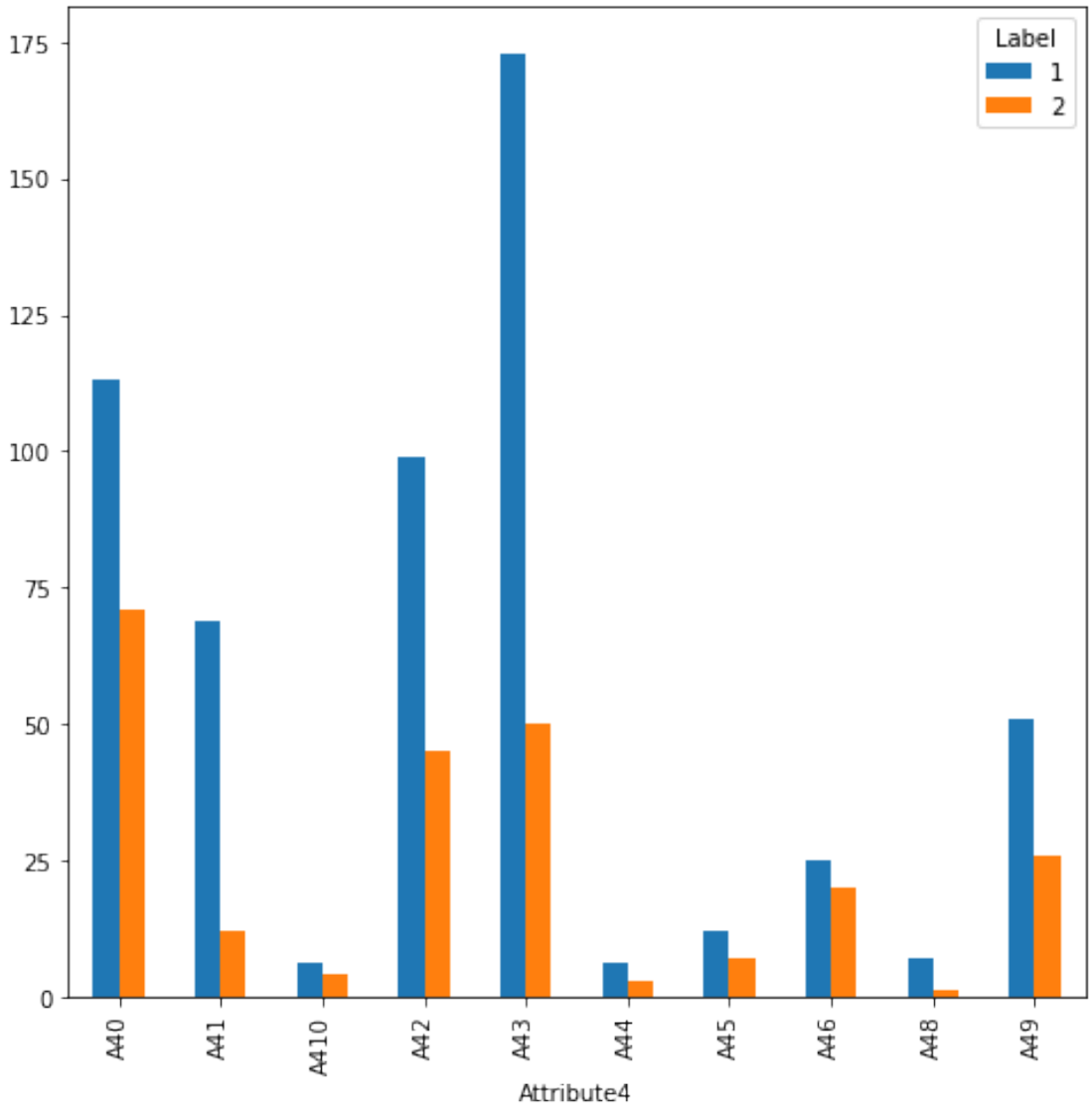
```
In [12]: def NumericalPlot(good, bad, Num, tit, ylab):
             Attribute = "Attribute"+Num.__str__()
             df = pd.DataFrame({'Good': good[Attribute], 'Bad':bad[Attribute] }, columns=['Good'
             df.plot.box()
             plt.title(Attribute + tit)
             plt.ylabel(ylab)
             plt.show()
```

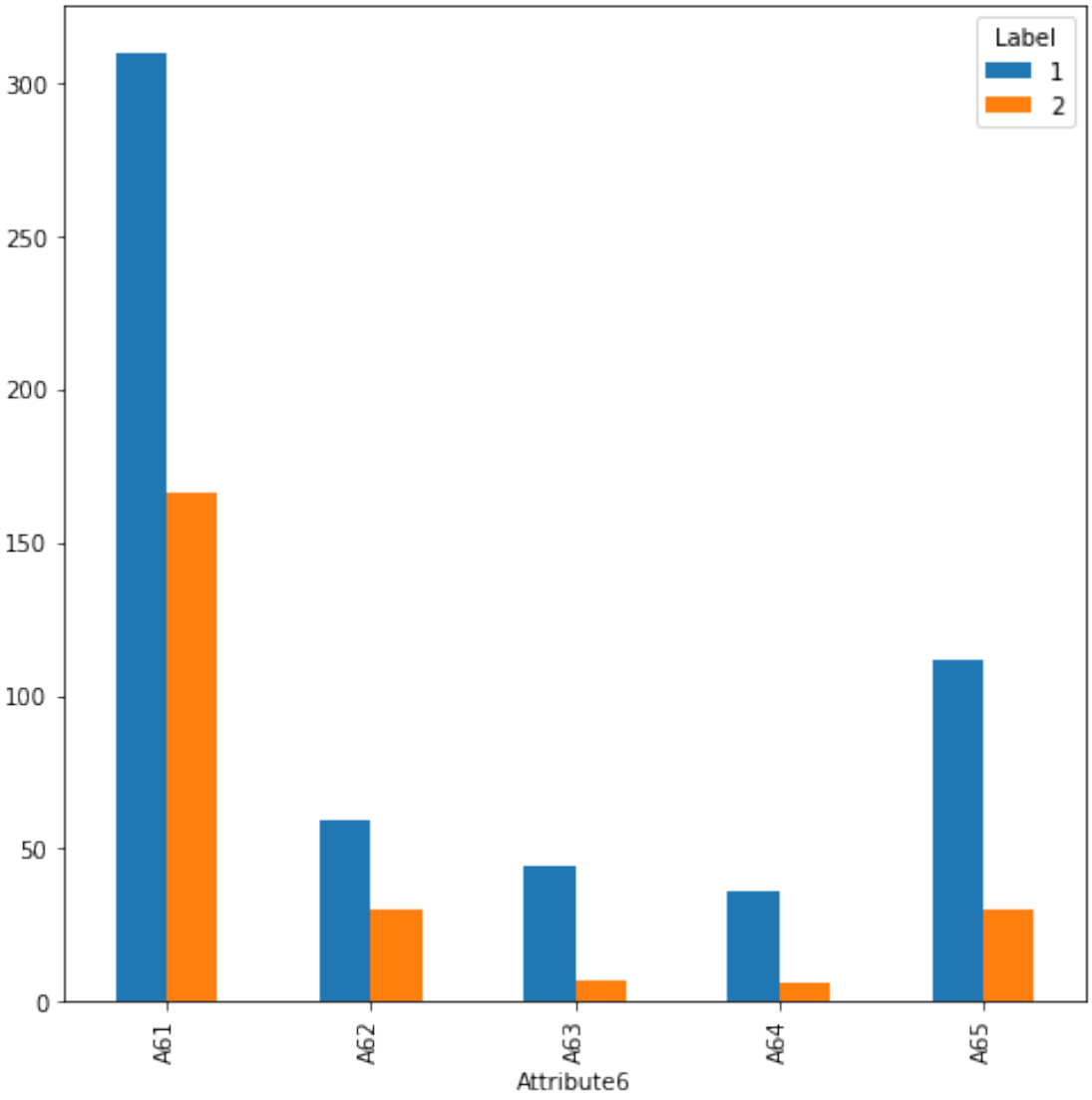
5 Call of the Function to Create all the Qualitive Plots

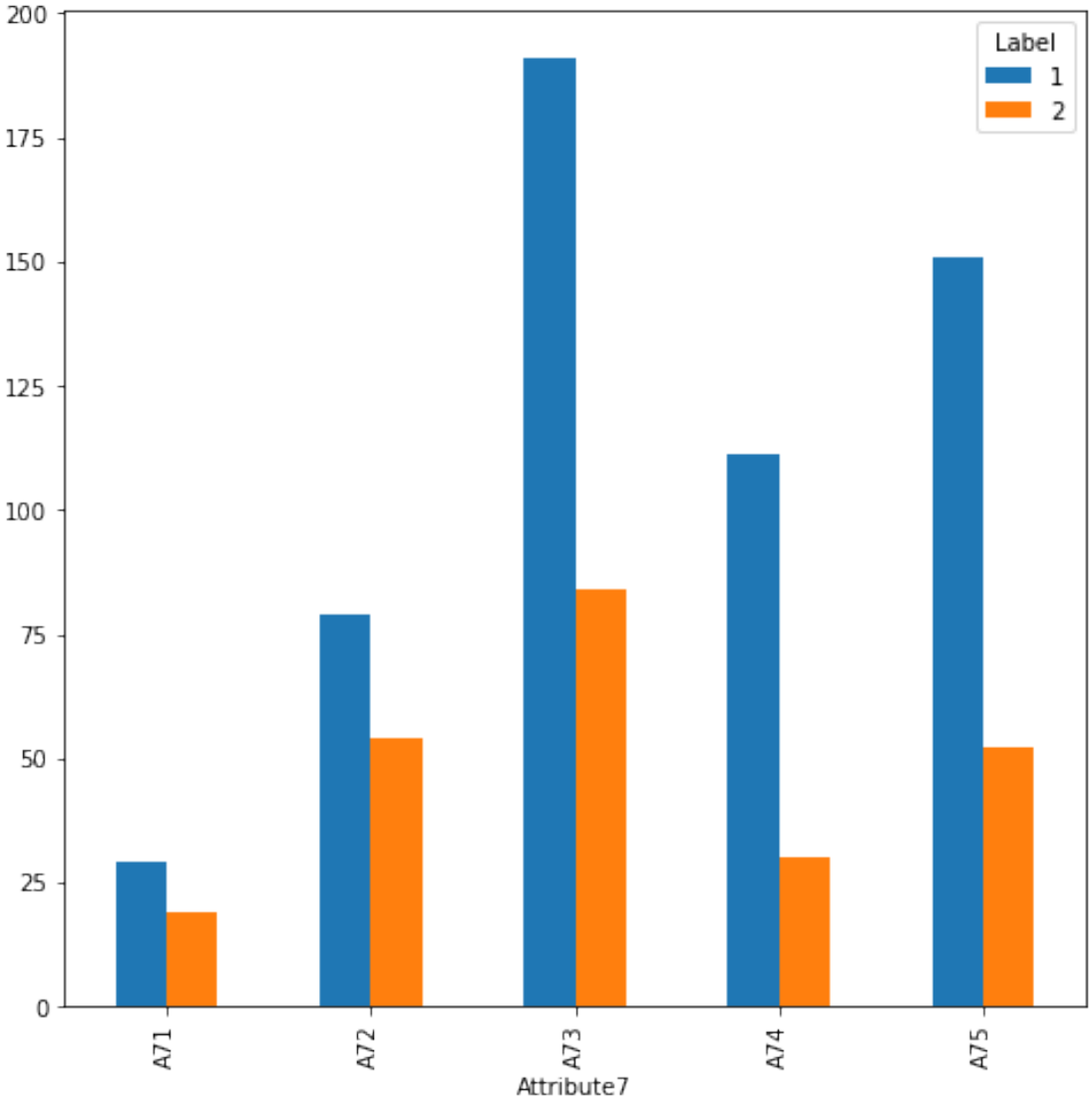
```
In [13]: qualitativeList = [1,3,4,6,7,9,10,12,14,15,17,19,20]
         for x in qualitativeList:
             QualitivePlot(data_set, x)
```

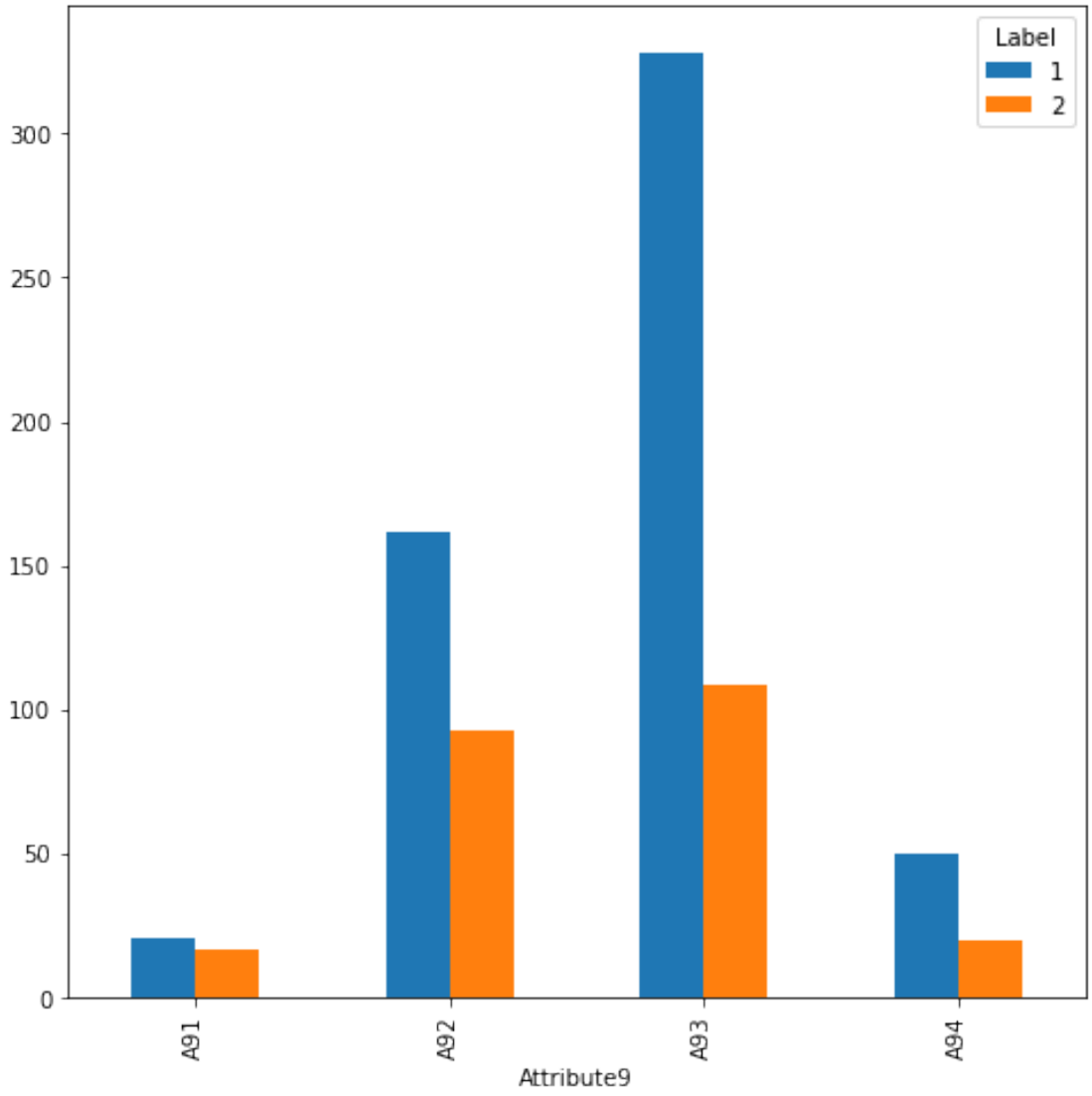

Label
1
2
250
200
150
100
50
0
A11
A12
A13
A14
Attribute1

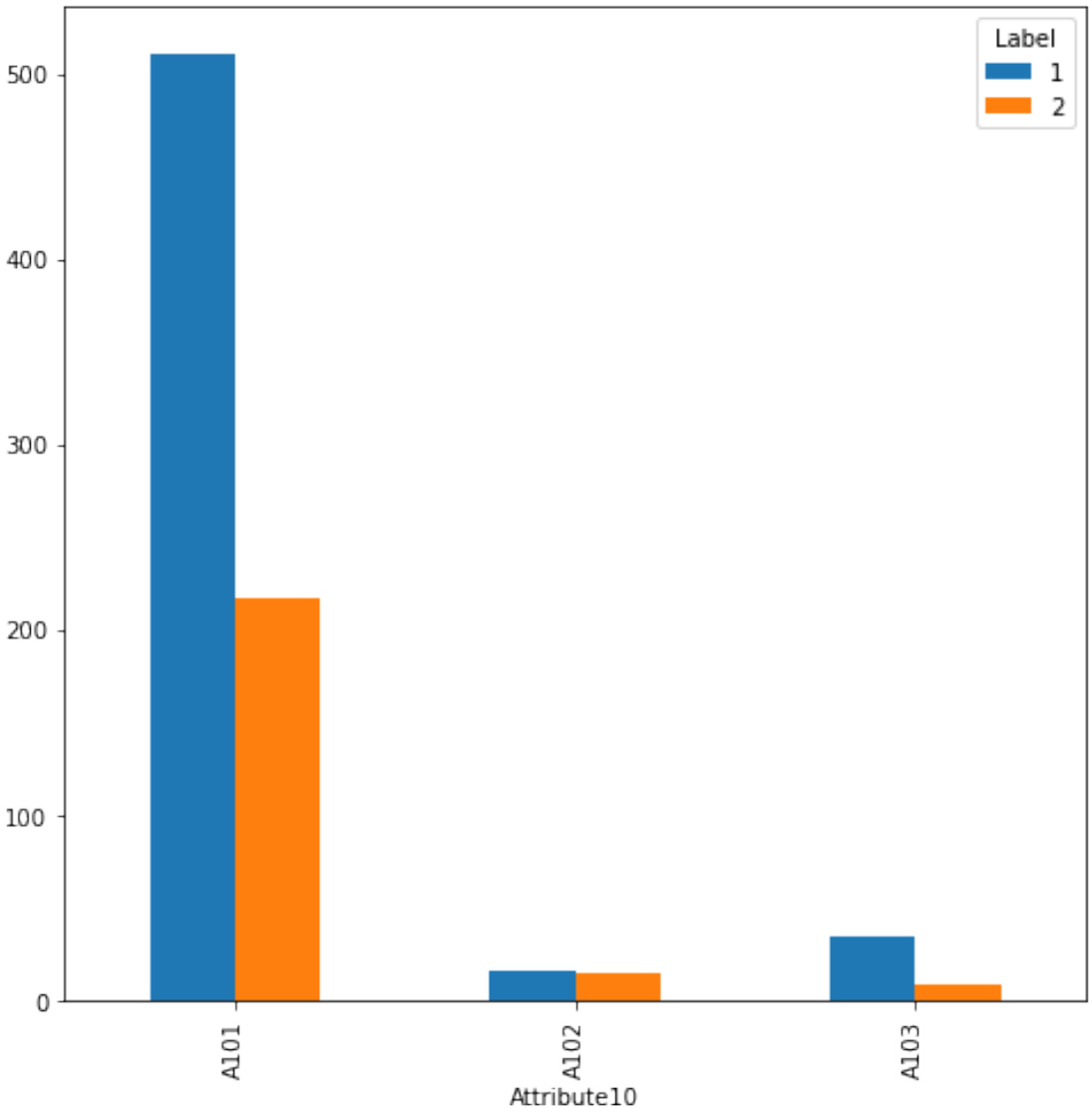

Label
1
2
300
250
200
150
100
50
0
A30
A31
A32
A33
A34
Attribute3

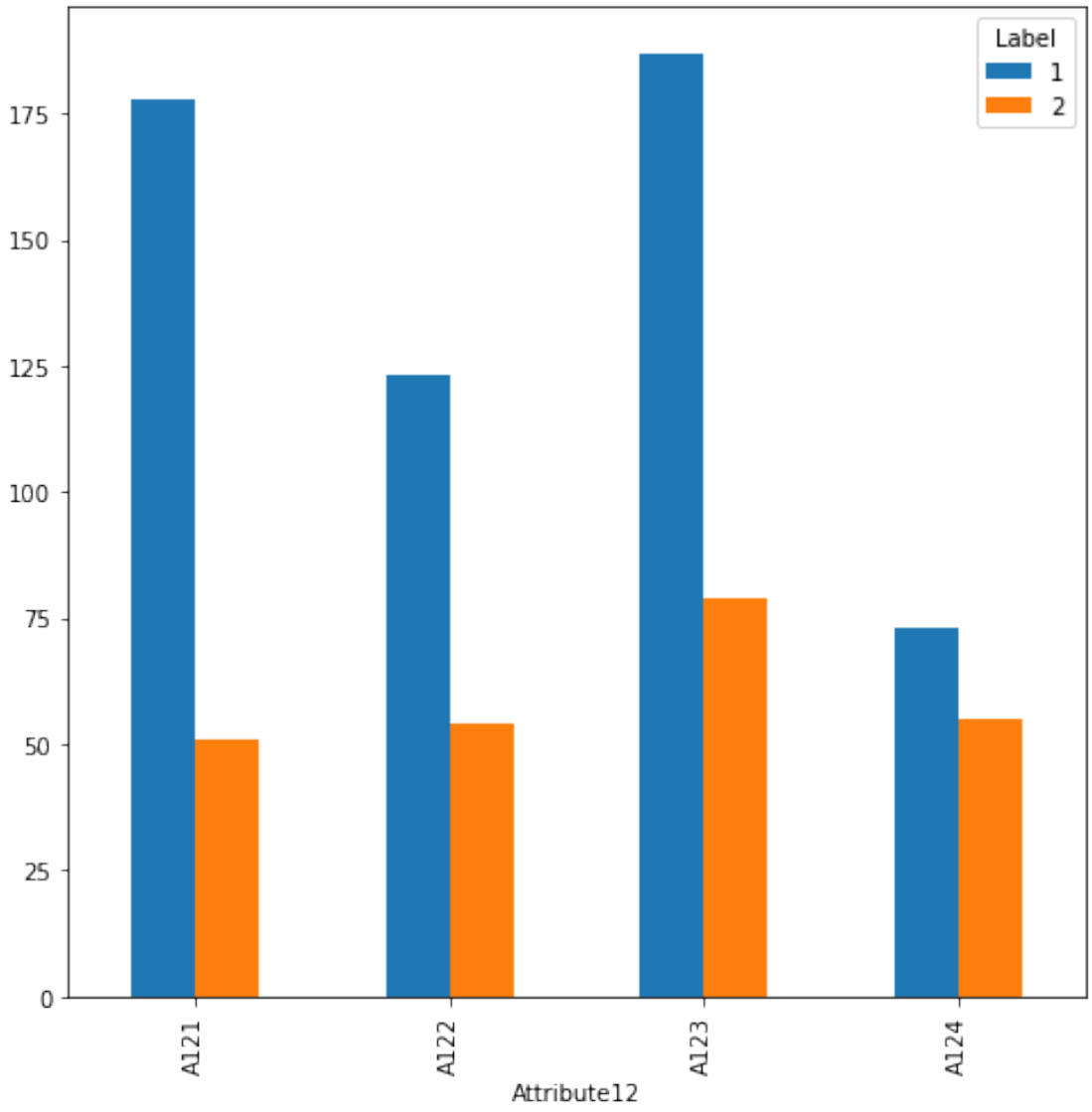

Label
1
2
175
150
125
100
75
50
25
0
A40
A41
A410
A42
A43
A44
A45
A46
A48
A49
Attribute4

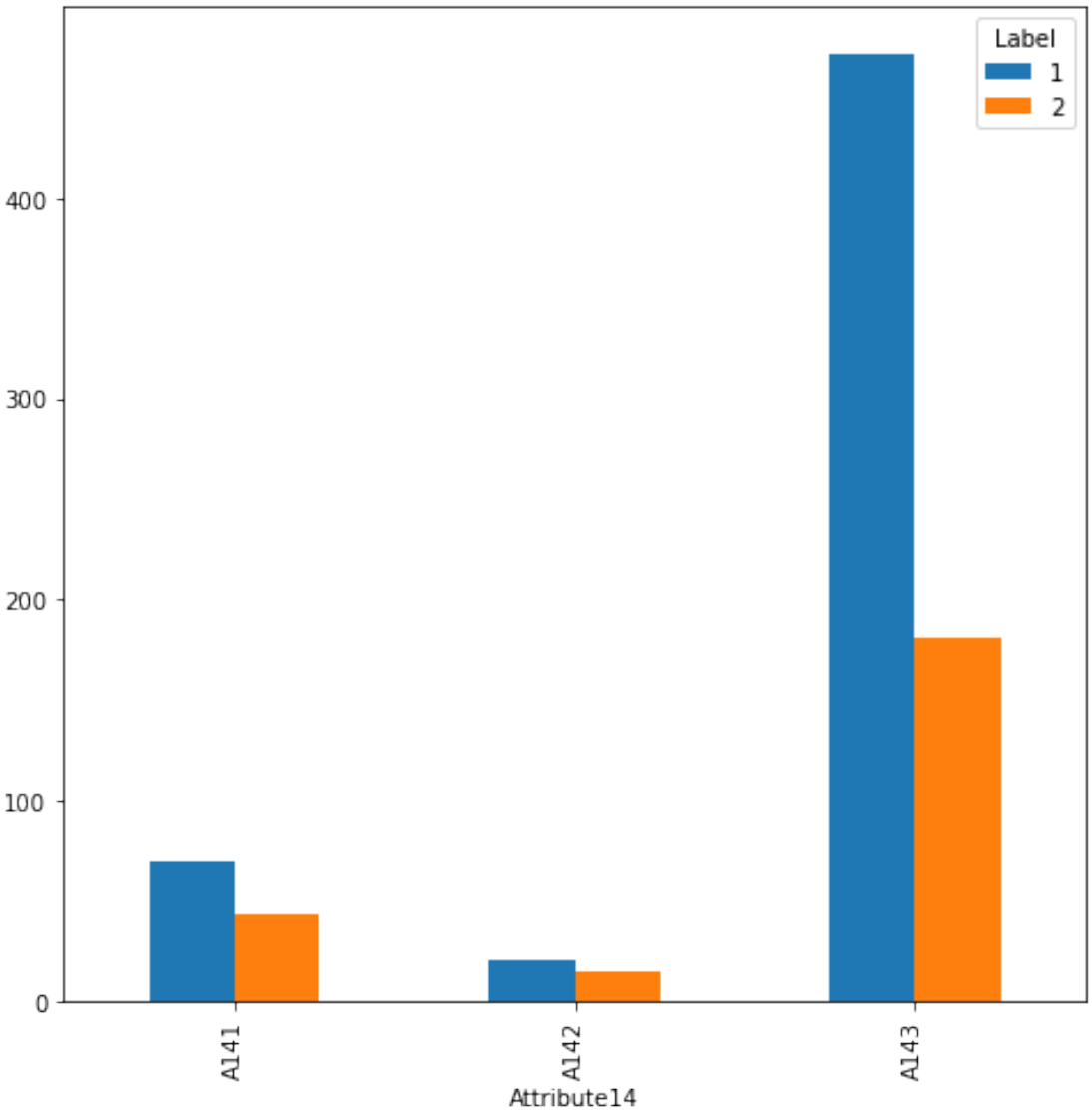

Label
1
2
300
250
200
150
100
50
0
A61
A62
A63
A64
A65
Attribute6

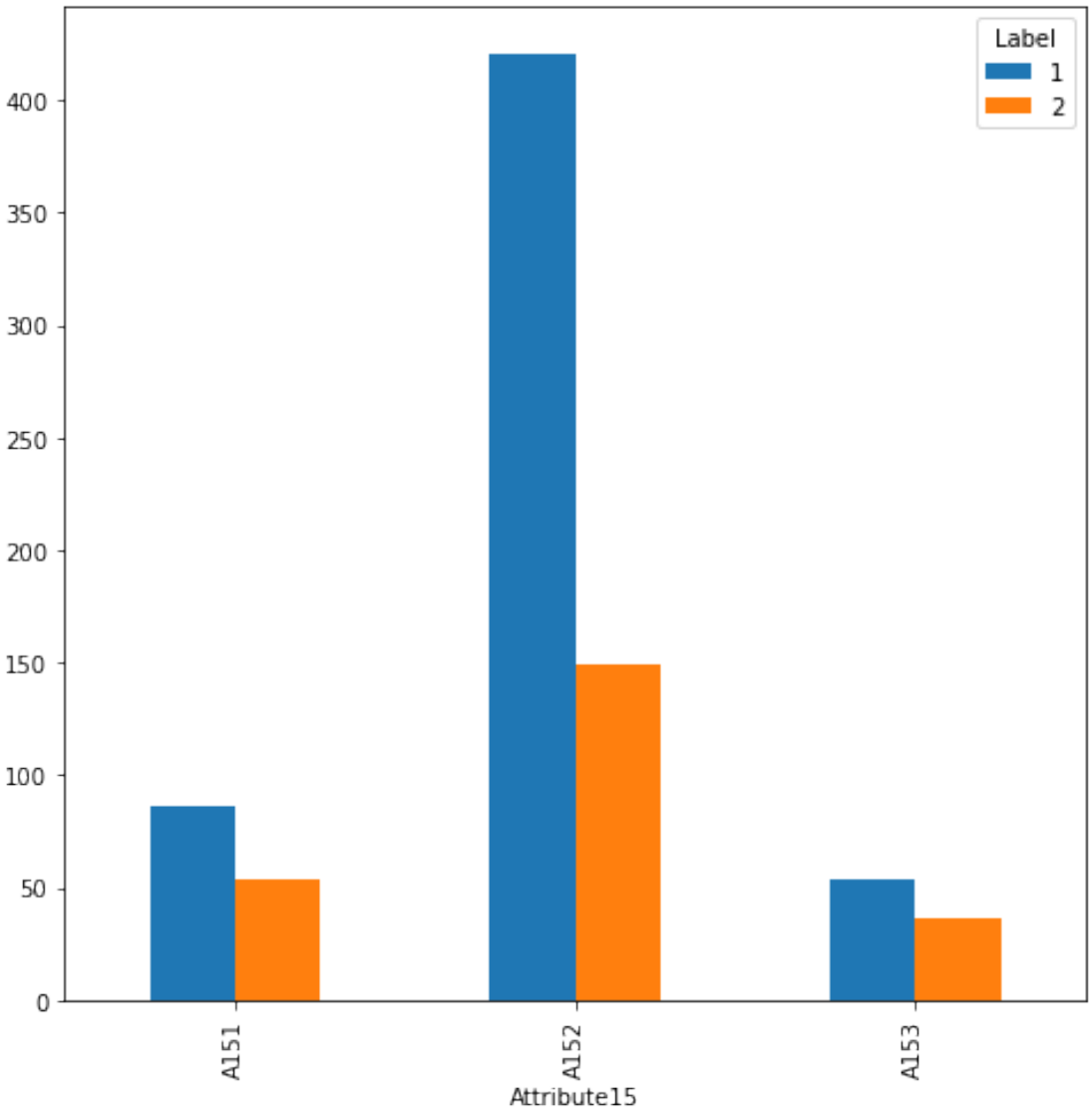

Label
1
2
200
175
150
125
100
75
50
25
0
A71
A72
A73
A74
A75
Attribute7

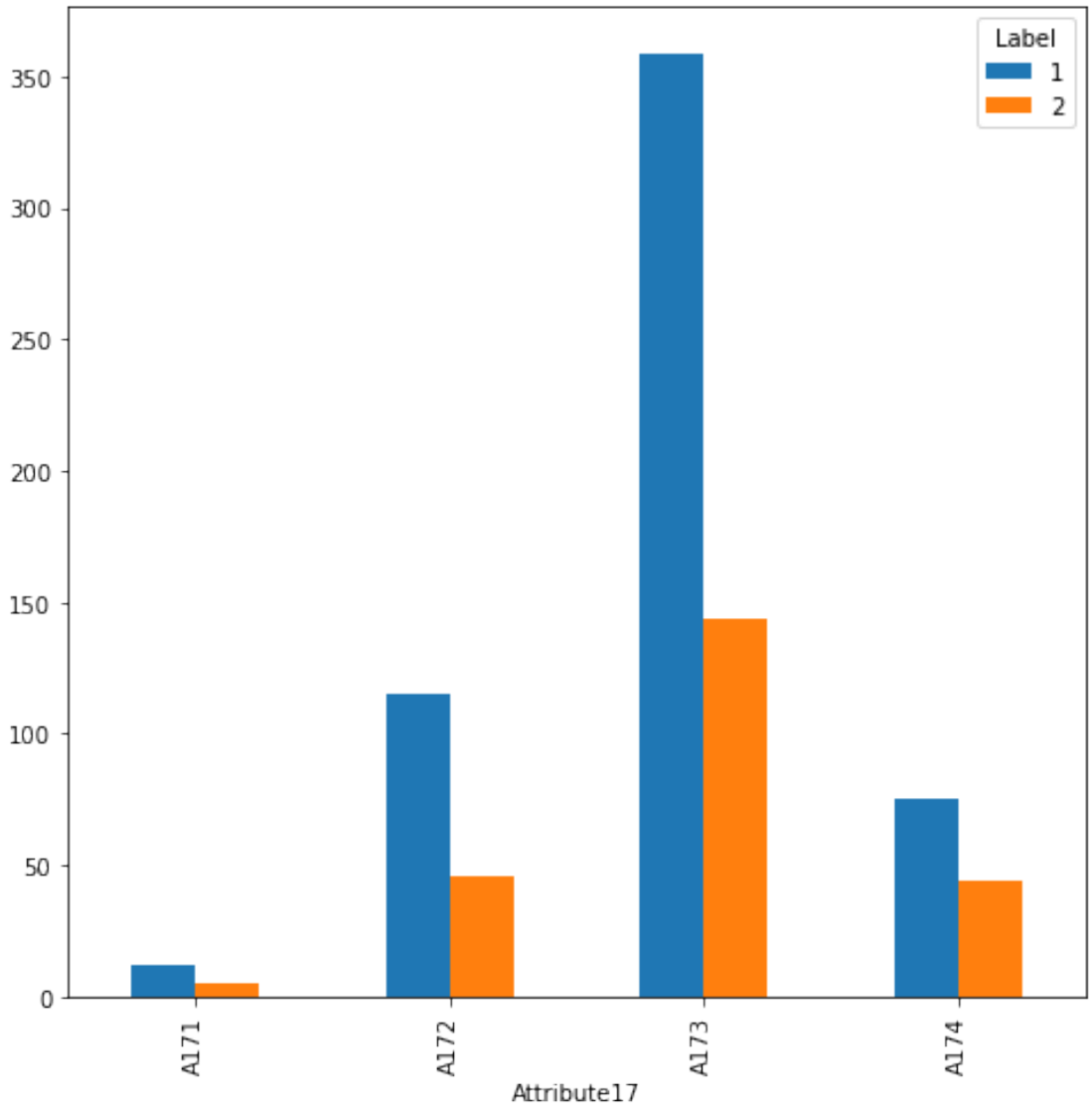

Label
1
2
300
250
200
150
100
50
0
A91
A92
A93
A94
Attribute9

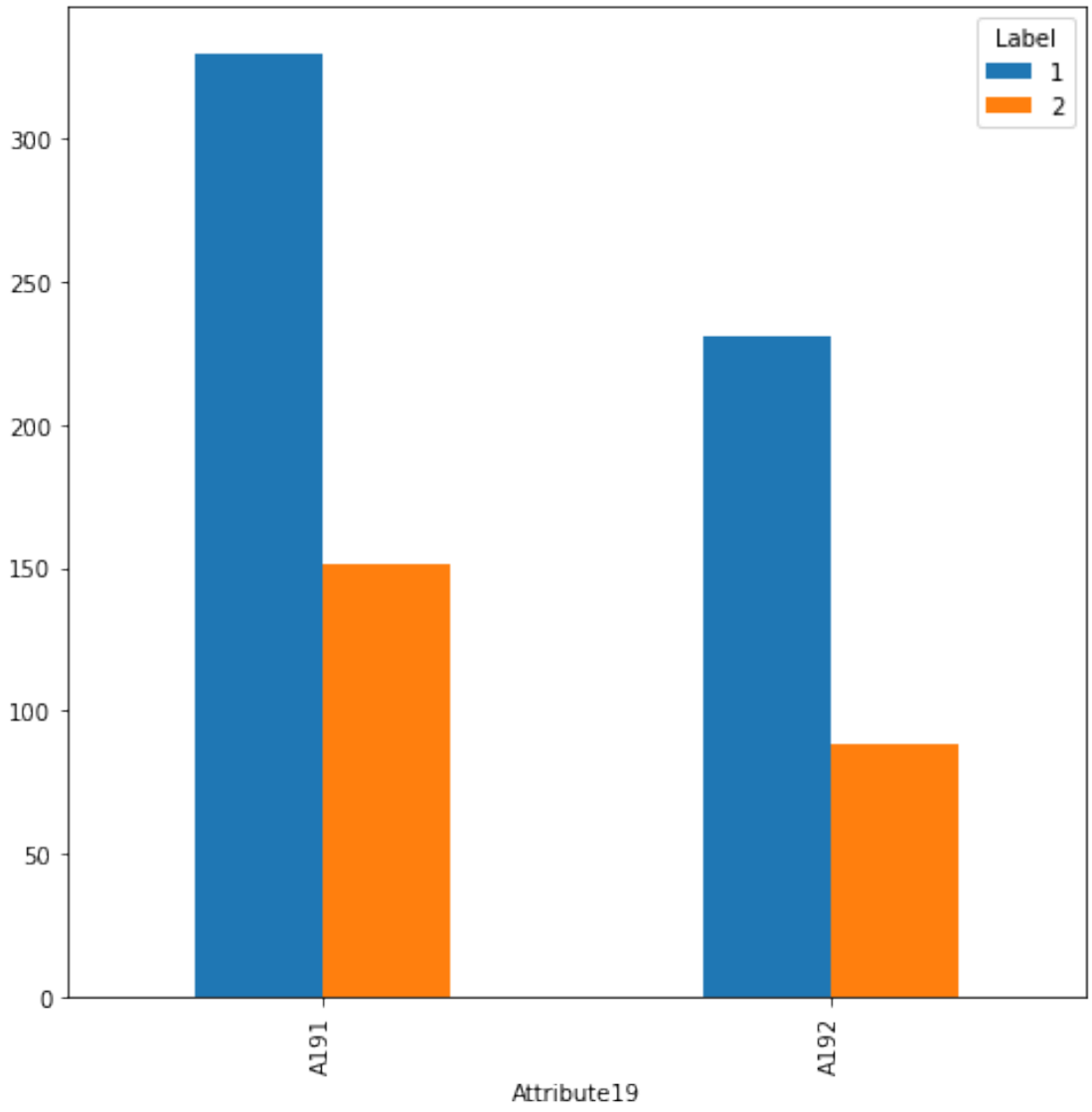

Label
1
2
500
400
300
200
100
0
A101
A102
A103
Attribute10

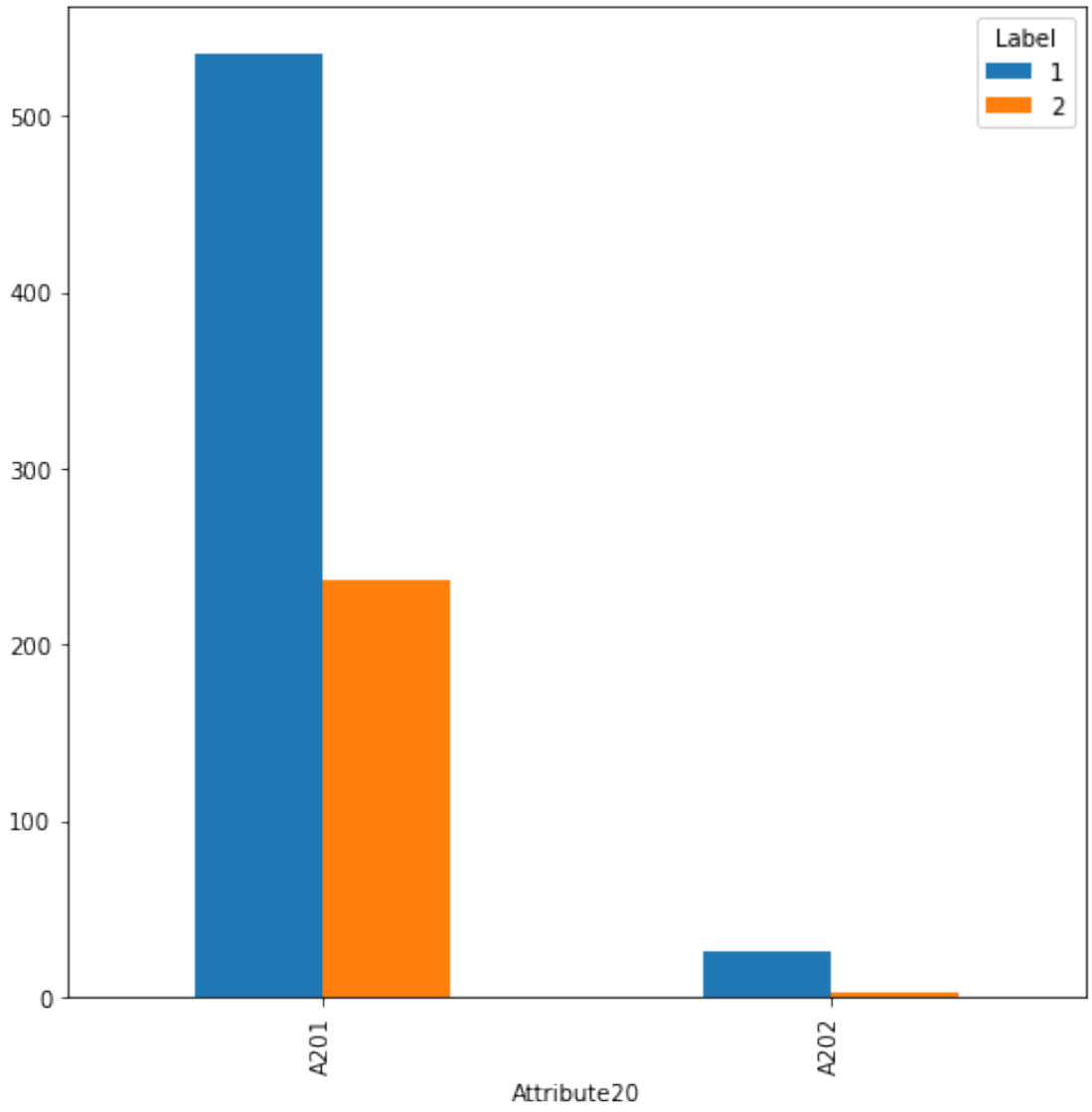

Label
1
2
175
150
125
100
75
50
25
0
A121
A122
A123
A124
Attribute12


Label
1
2
400
300
200
100
0
A141
A142
A143
Attribute14


Label
1
2
400
350
300
250
200
150
100
50
0
A151
A152
A153
Attribute15

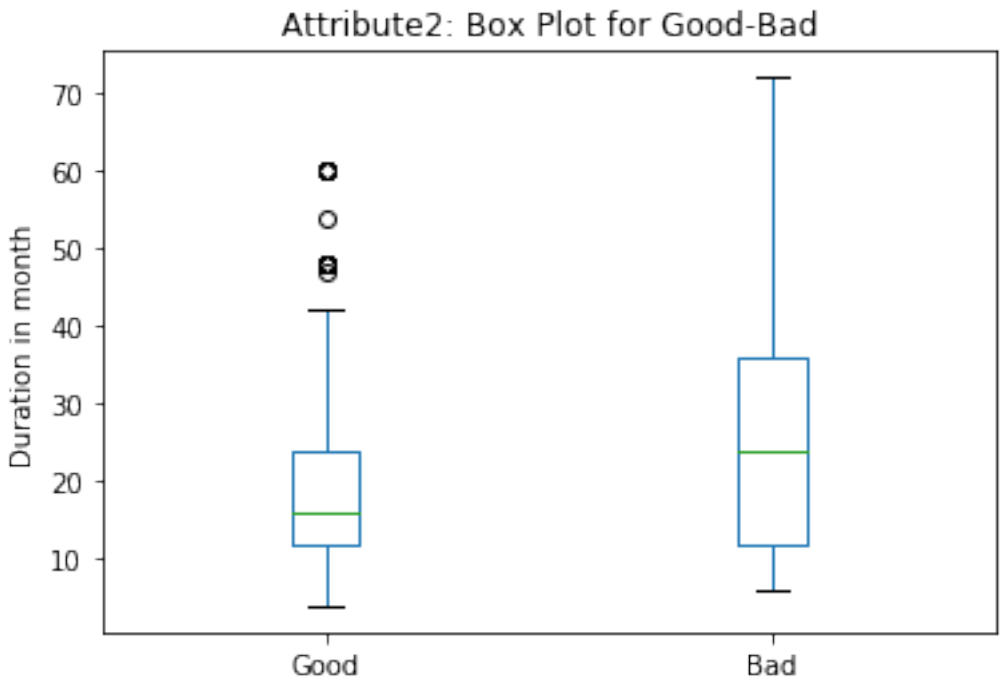

Label
1
2
350
300
250
200
150
100
50
0
A171
A172
A173
A174
Attribute17

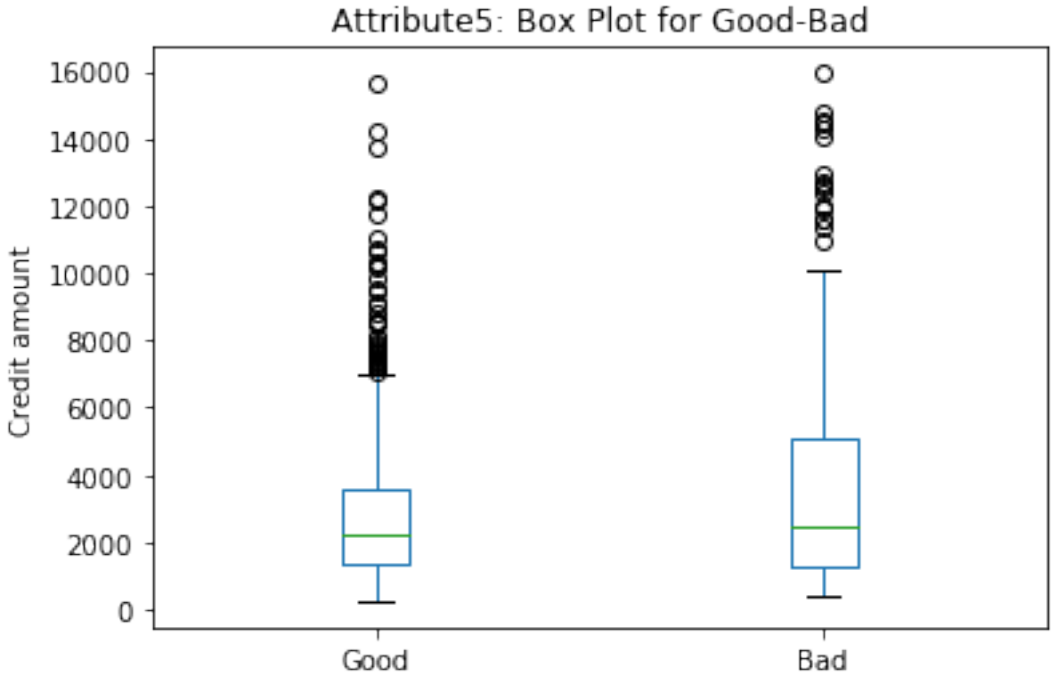

Label
1
2
300
250
200
150
100
50
0
A191
A192
Attribute19

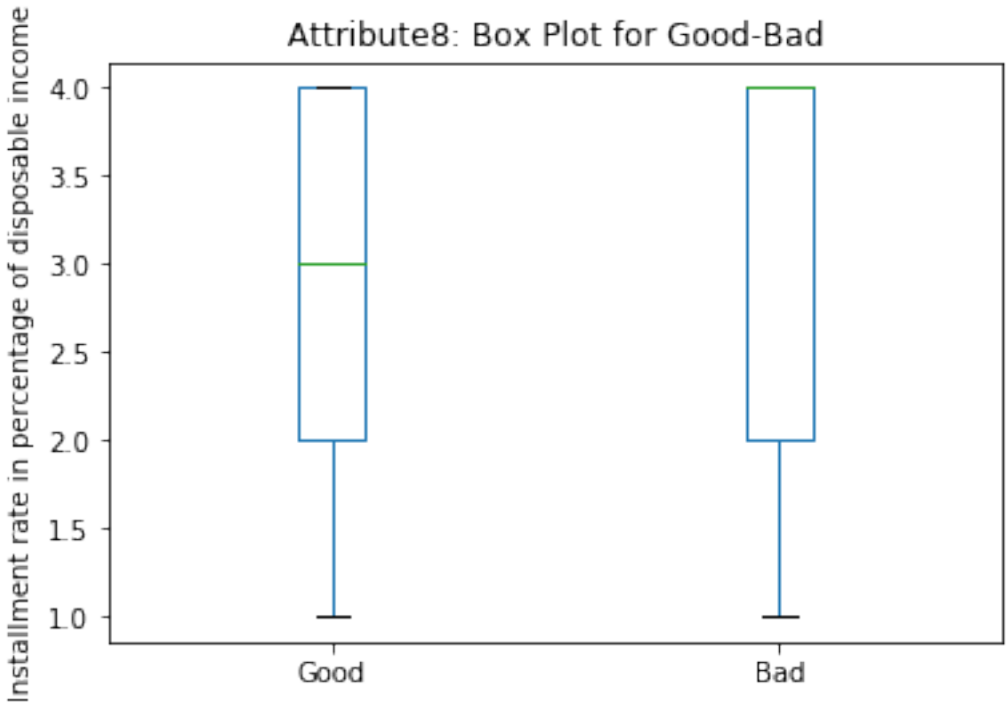


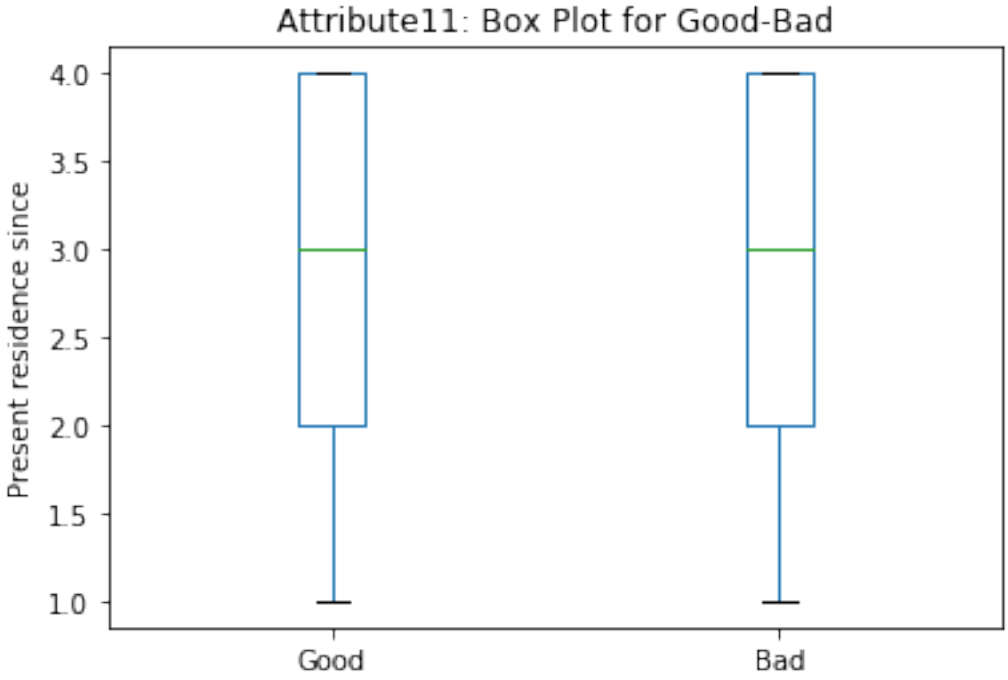
6 Call of the Function to Create all the Numerical Plots

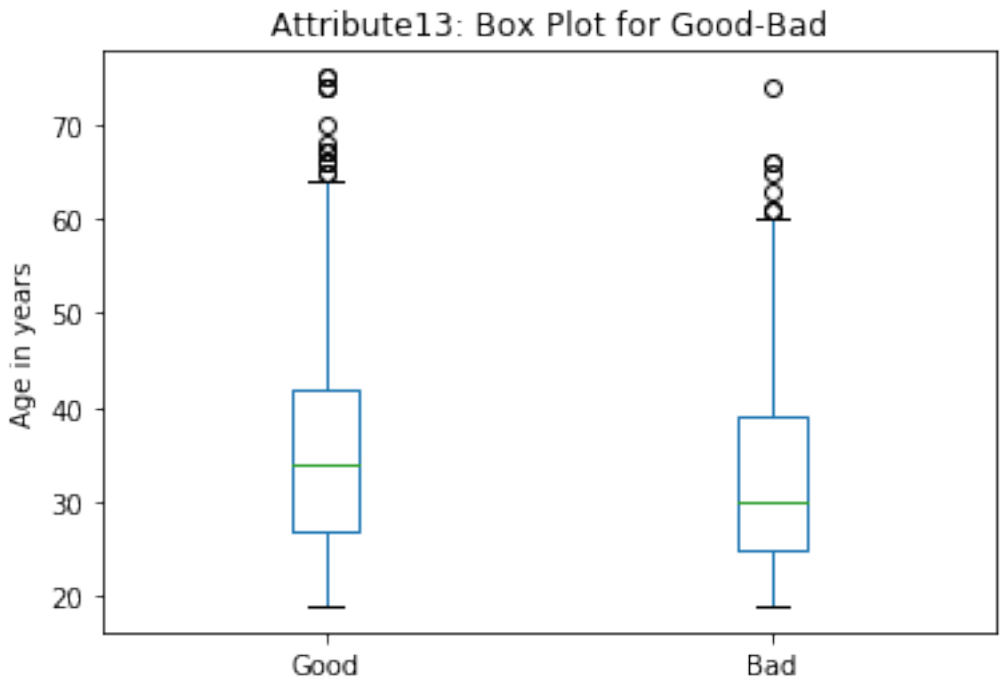
```
In [14]: numericalList = [2,5,8,11,13,16,18]
         title = ": Box Plot for Good-Bad"
         ylab = ['Duration in month', 'Credit amount','Installment rate in percentage of disposa
         'Age in years','Number of existing credits at this bank','Number of people being liable

         for x in range(0,len(numericalList)):
             NumericalPlot(good, bad, numericalList[x], title, ylab[x])
```

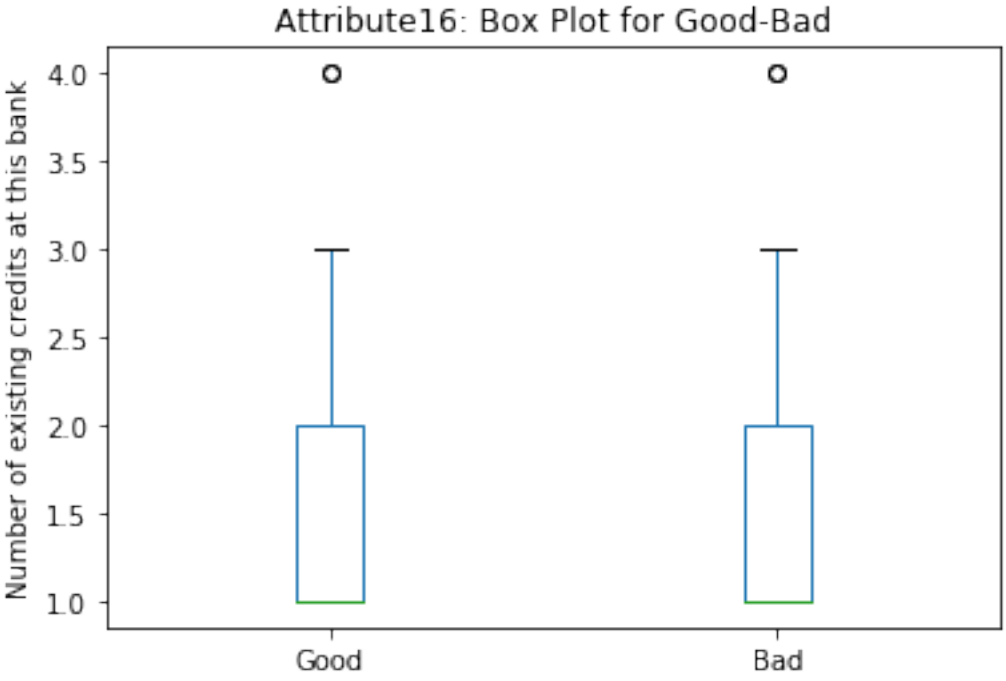

Attribute2: Box Plot for Good-Bad
Duration in month
70
60
50
40
30
20
10
Good
Bad

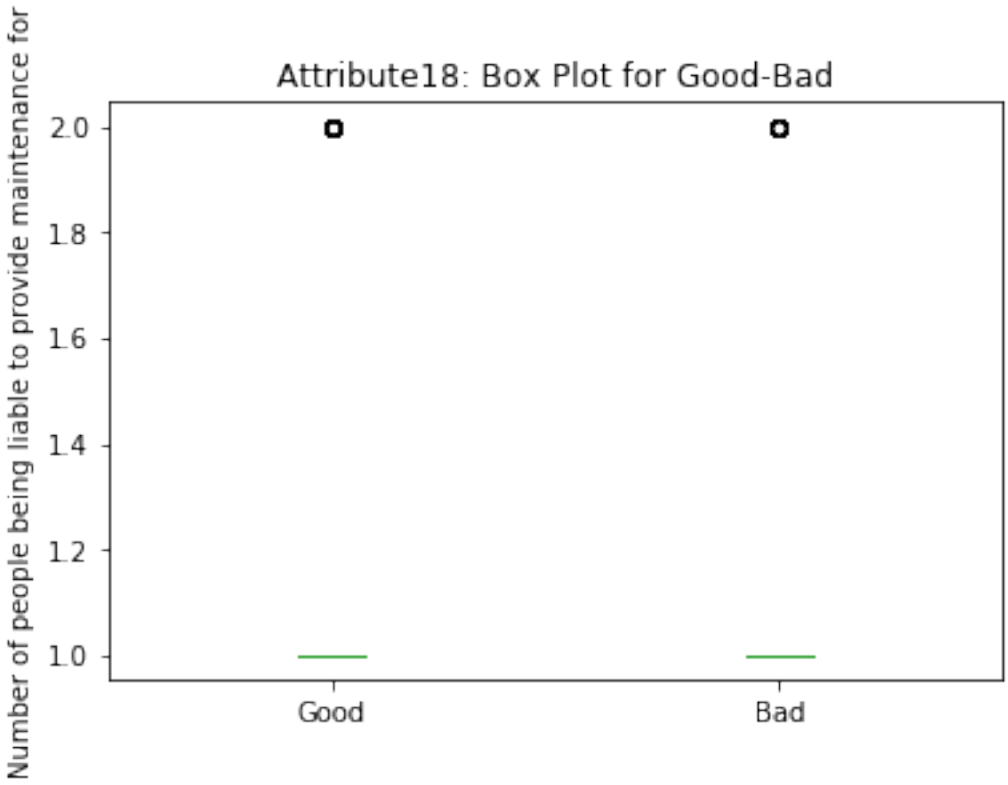

Attribute5: Box Plot for Good-Bad
Credit amount
16000
14000
12000
10000
8000
6000
4000
2000
0
Good
Bad


Attribute8: Box Plot for Good-Bad
Installment rate in percentage of disposable income
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Good
Bad


Attribute11: Box Plot for Good-Bad
Present residence since
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Good
Bad


Attribute13: Box Plot for Good-Bad
Age in years
70
60
50
40
30
20
Good
Bad


Attribute16: Box Plot for Good-Bad
Number of existing credits at this bank
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Good
Bad


Attribute18: Box Plot for Good-Bad
Number of people being liable to provide maintenance for
2.0
1.8
1.6
1.4
1.2
1.0
Good
Bad