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

In [ ]:
```
dataset = pd.read_csv("dirty_irisdata.csv")
```

In [ ]:
```
dataset.head(10)
```

Out[ ]:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	6.4	3.2	4.5	1.5	versicolor
1	6.3	3.3	6.0	2.5	virginica
2	6.2	NaN	5.4	2.3	virginica
3	5.0	3.4	1.6	0.4	setosa
4	5.7	2.6	3.5	1.0	versicolor
5	5.3	NaN	NaN	0.2	setosa
6	6.4	2.7	5.3	NaN	virginica
7	5.9	3.0	5.1	1.8	virginica
8	5.8	2.7	4.1	1.0	versicolor
9	4.8	3.1	1.6	0.2	setosa

In [ ]:
```
new_n = dataset.dropna().shape[0]
n = dataset.shape[0]
print(f"Number of complete records:{new_n}")
print("Percentage of complete records:{:.2f}%".format(float(new_n*100/n)))
```

```
Number of complete records:96
Percentage of complete records:64.00%
```

In [ ]:
```
dataset.dropna(inplace=True)
dataset.reset_index(inplace=True)
dataset.replace(['?',np.inf], 'NA',inplace=True)
```

In [ ]:
```
dataset.head(5)
```

Out[ ]:

	index	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
0	0	6.4	3.2	4.5	1.5	versicolor
1	1	6.3	3.3	6.0	2.5	virginica
2	3	5.0	3.4	1.6	0.4	setosa
3	4	5.7	2.6	3.5	1.0	versicolor
4	7	5.9	3.0	5.1	1.8	virginica

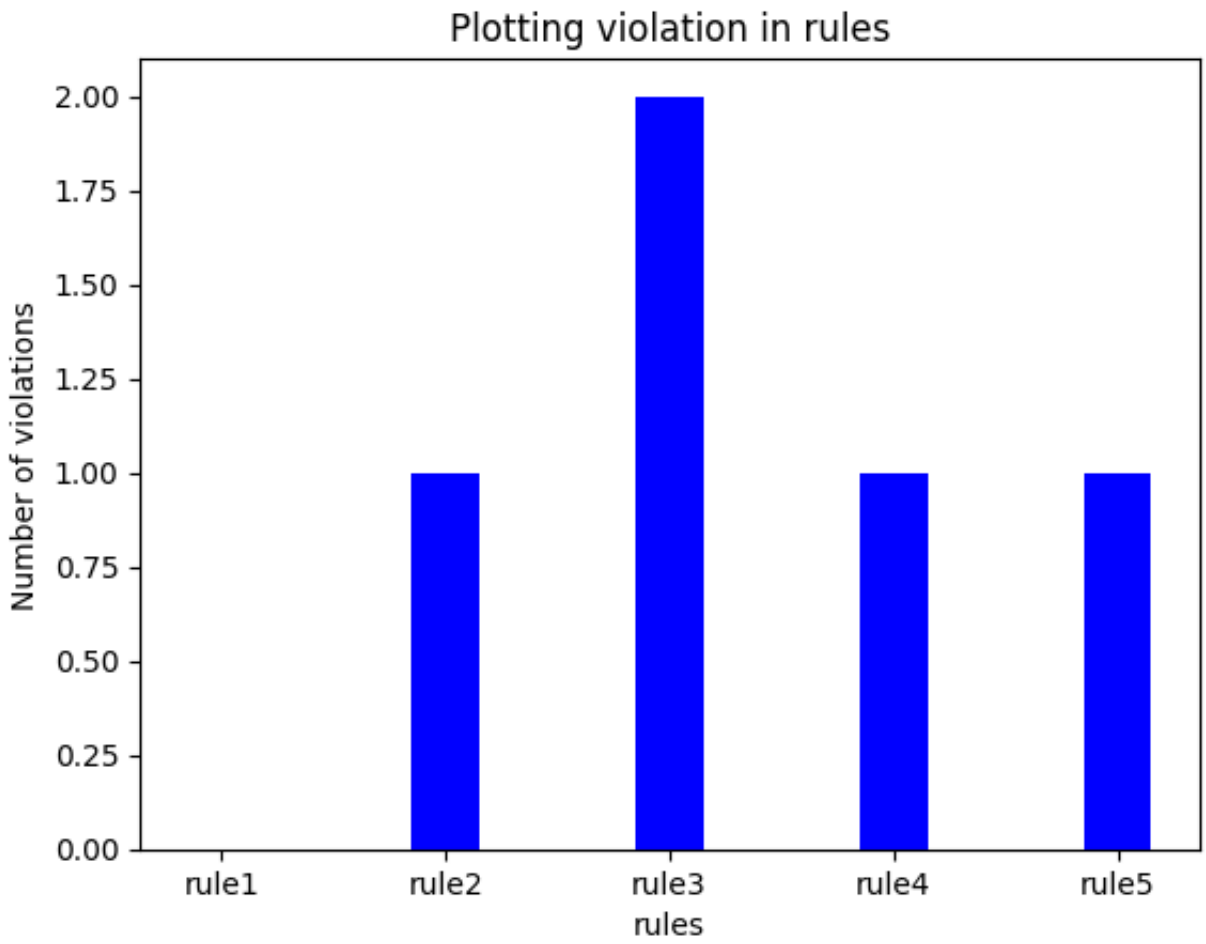
In [ ]:
```
rules = []
```

In [ ]:
```
rules.append(ruleset.check_species)
rules.append(ruleset.check_positive)
rules.append(ruleset.check_petal_length)
rules.append(ruleset.check_sepal_length)
rules.append(ruleset.check_sepal_length2)
```

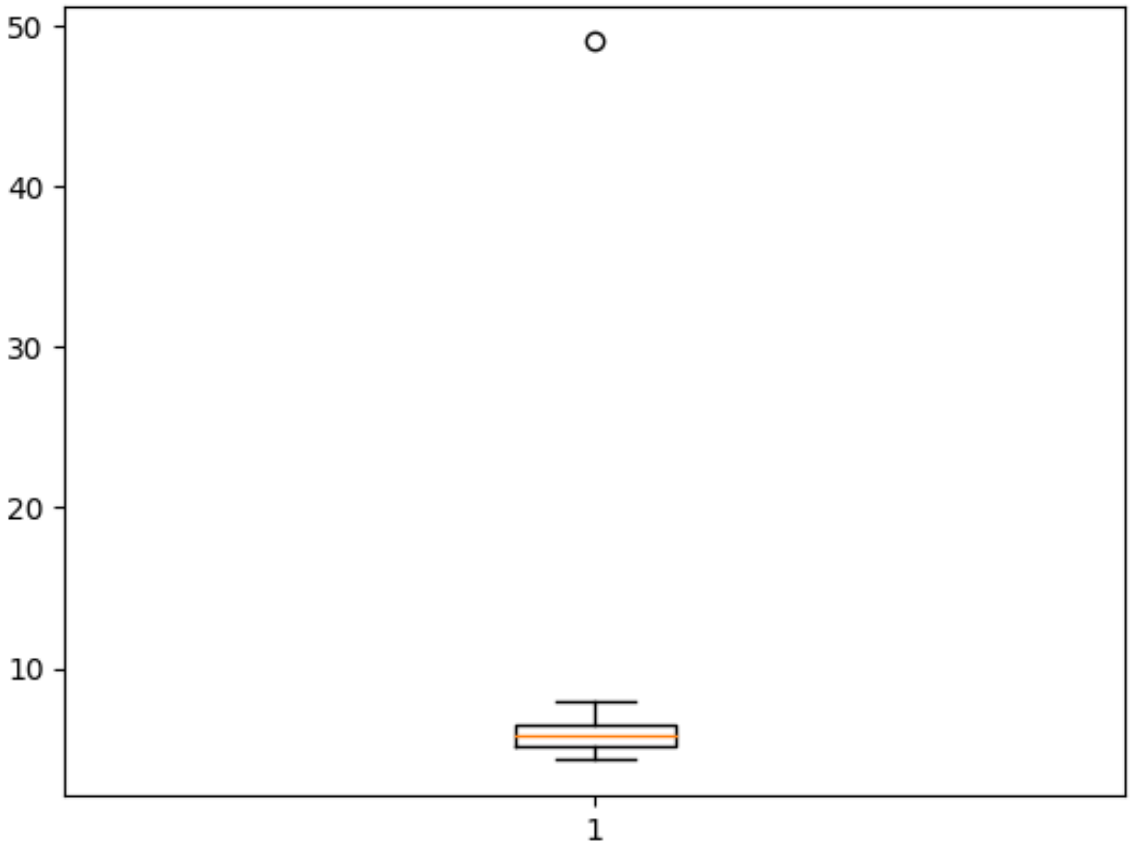
In [ ]:
```
violations = []
rule = list(map(lambda x: "rule"+str(x), range(1,len(rules)+1)))
for i in range(len(rules)):
    violation,desc = rules[i](dataset)
    print(f"{rule[i]}: {desc}\nviolations:{violation}")
    violations.append(violation)
```

```
rule1: Checking if all species consist of setosa, versicolor and virgini
ca
violations:0
rule2: Checking if there all the length values are greater than 0
violations:1
rule3: Checking if petal length is at least twice of petal width
violations:2
rule4: Checking if all sepal lengths are below 30 cm
violations:1
rule5: Checking if sepal length is more than petal length
violations:1
```

In [ ]:
```
fig = plt.figure()
fig.patch.set_facecolor('white')
ax = fig.add_subplot()
ax.bar(rule,violations,0.3,color = 'blue')
plt.ylabel("Number of violations")
plt.xlabel("rules")
plt.title("Plotting violation in rules")
plt.show()
```



In [ ]:
```
fig = plt.figure()
fig.patch.set_facecolor('white')
ax = fig.add_subplot()
ax.boxplot(dataset["Sepal.Length"][dataset["Sepal.Length"]!='NA'])
plt.show()
```



In [ ]: