

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: data = pd.read_csv('breast_cancer.csv')
x = data.iloc[:,1:-1].values
y = data.iloc[:, -1].values
```

```
In [3]: x
```

```
Out[3]: array([[ 5,  1,  1, ...,  3,  1,  1],
               [ 5,  4,  4, ...,  3,  2,  1],
               [ 3,  1,  1, ...,  3,  1,  1],
               ...,
               [ 5, 10, 10, ...,  8, 10,  2],
               [ 4,  8,  6, ..., 10,  6,  1],
               [ 4,  8,  8, ..., 10,  4, 11]], dtype=int64)
```

```
In [4]: y
```

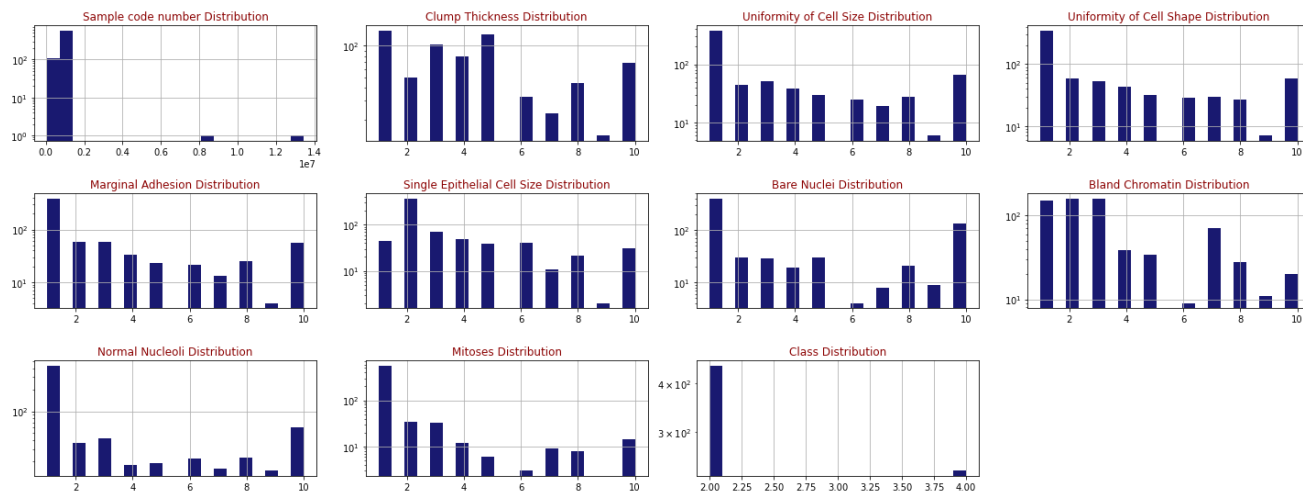
```
Out[4]: array([[2, 2, 2, 2, 2, 4, 2, 2, 2, 2, 2, 2, 4, 2, 4, 4, 2, 2, 4, 2, 4, 4,
2, 2, 4, 2, 2, 2, 2, 2, 2, 4, 2, 2, 2, 4, 2, 4, 4, 4, 4, 4, 2,
4, 2, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 2, 4, 4, 2, 4, 2, 4, 2, 4,
4, 2, 2, 4, 2, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 2, 4, 2, 4,
2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 4, 4, 2, 4, 4, 4, 4, 4, 2, 4, 2, 4,
4, 4, 2, 2, 2, 2, 4, 4, 2, 2, 2, 2, 2, 4, 4, 4, 2, 4, 2, 4, 2, 2,
2, 4, 4, 2, 4, 4, 4, 2, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 2,
2, 4, 4, 2, 2, 2, 4, 4, 2, 4, 4, 4, 2, 2, 4, 2, 2, 4, 4, 4, 2,
4, 4, 2, 4, 4, 4, 2, 4, 2, 4, 4, 4, 4, 2, 2, 2, 2, 2, 2, 4, 4, 2,
4, 4, 2, 4, 4, 4, 2, 2, 2, 2, 4, 4, 4, 4, 4, 4, 2, 4, 4, 4, 2, 4, 2,
4, 4, 2, 2, 2, 2, 4, 2, 4, 4, 4, 4, 4, 4, 4, 4, 2, 4, 4, 4, 2, 4,
2, 4, 4, 2, 4, 2, 4, 4, 2, 4, 4, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 4,
2, 4, 2, 2, 4, 2, 4, 4, 4, 4, 2, 2, 2, 4, 2, 2, 2, 4, 4, 2, 2, 4,
4, 4, 2, 4, 2, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
2, 2, 4, 2, 2, 4, 4, 2, 2, 2, 4, 4, 4, 2, 4, 2, 4, 2, 2, 2, 2,
2, 2, 2, 2, 2, 2, 2, 4, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 2, 4, 2,
2, 4, 2, 4, 4, 4, 4, 2, 2, 4, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 2, 4,
2, 4, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 2, 4,
2, 4, 2, 2, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 2, 2, 2, 2, 2,
2, 2, 2, 2, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 4, 2, 2, 2, 2,
2, 2, 2, 2, 4, 4, 2, 2, 2, 2, 2, 2, 2, 2, 2, 4, 2, 2, 2, 4, 4,
4], dtype=int64)
```

```
In [16]: data['Class'].value_counts()
```

```
Out[16]: 2    444
         4    239
         Name: Class, dtype: int64
```

```
In [13]: def draw_histograms(dataframe, features, rows, cols):
fig=plt.figure(figsize=(20,20))
for i, feature in enumerate(features):
    ax=fig.add_subplot(rows,cols,i+1)
    dataframe[feature].hist(bins=20,ax=ax,facecolor='midnightblue')
    ax.set_title(feature+" Distribution",color='DarkRed')
    ax.set_yscale('log')
fig.tight_layout()
plt.show()
```

```
draw_histograms(data,data.columns,8,4)
```



```
In [5]: 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 = 0)
```

```
In [6]: x_train
```

```
Out[6]: array([[10, 1, 1, ..., 5, 4, 1],
               [ 1, 1, 1, ..., 3, 1, 1],
               [ 5, 1, 1, ..., 3, 1, 1],
               ...,
               [ 1, 1, 1, ..., 1, 1, 1],
               [ 3, 1, 1, ..., 2, 1, 1],
               [10, 9, 7, ..., 7, 7, 1]], dtype=int64)
```

```
In [7]: y_train
```

[illegible]

In [8]: x\_test

Out[8]: array([[ 1, 1, 1, ..., 1, 1, 1],  
[ 3, 1, 1, ..., 2, 1, 1],  
[ 5, 5, 5, ..., 4, 3, 1],  
...,  
[ 4, 1, 1, ..., 1, 1, 1],  
[ 4, 10, 4, ..., 9, 10, 1],  
[ 2, 1, 1, ..., 2, 1, 1]], dtype=int64)

In [9]: y\_test

Out[9]: array([2, 2, 4, 4, 2, 2, 2, 4, 2, 2, 4, 2, 4, 2, 2, 4, 4, 4, 2, 2, 2,  
4, 2, 4, 4, 2, 2, 2, 4, 2, 4, 4, 2, 2, 2, 4, 4, 2, 4, 2, 2, 2, 2,  
2, 2, 2, 4, 2, 2, 4, 2, 4, 2, 2, 2, 4, 2, 2, 4, 2, 2, 2, 2, 2, 2,  
2, 2, 4, 4, 2, 2, 2, 2, 2, 2, 4, 2, 2, 2, 4, 2, 4, 2, 2, 4, 2, 2,  
4, 2, 4, 2, 4, 4, 4, 2, 4, 4, 4, 2, 2, 2, 4, 4, 2, 2, 4, 4, 2, 2,  
4, 2, 2, 4, 2, 2, 2, 4, 2, 2, 2, 4, 2, 2, 4, 4, 2, 4, 2, 4, 2, 2,  
4, 2, 2, 4, 2], dtype=int64)

In [10]: from sklearn.linear\_model import LogisticRegression  
lr = LogisticRegression(random\_state = 0)  
lr.fit(x\_train,y\_train)

Out[10]: LogisticRegression(random\_state=0)

In [11]: y\_pred = lr.predict(x\_test)

In [14]: import seaborn as sns

In [15]: from sklearn.metrics import confusion\_matrix , accuracy\_score  
cm = confusion\_matrix(y\_test,y\_pred)  
conf\_matrix=pd.DataFrame(data=cm,columns=['Predicted:0', 'Predicted:1'],index=['Actual:0', 'Actual:1'])  
plt.figure(figsize = (8,5))  
sns.heatmap(conf\_matrix, annot=True,fmt='d',cmap="YlGnBu");  
ac = accuracy\_score(y\_test,y\_pred)  
print(ac\*100)

95.62043795620438

