

```
In [3]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv("heart_failure_clinical_records_dataset.csv")
df.head(5)
```

```
Out[3]:
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets
0	75.0	0	582	0	20	1	265000
1	55.0	0	7861	0	38	0	263300
2	65.0	0	146	0	20	0	162000
3	50.0	1	111	0	20	0	210000
4	65.0	1	160	1	20	0	327000

```
In [5]: df.describe()
```

```
Out[5]:
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets
count	299.000000	299.000000	299.000000	299.000000	299.000000	299.000000	299.000000
mean	60.833893	0.431438	581.839465	0.418060	38.083612	0.418060	263300.000000
std	11.894809	0.496107	970.287881	0.494067	11.834841	0.494067	263300.000000
min	40.000000	0.000000	23.000000	0.000000	14.000000	0.000000	140000.000000
25%	51.000000	0.000000	116.500000	0.000000	30.000000	0.000000	210000.000000
50%	60.000000	0.000000	250.000000	0.000000	38.000000	0.000000	263300.000000
75%	70.000000	1.000000	582.000000	1.000000	45.000000	1.000000	327000.000000
max	95.000000	1.000000	7861.000000	1.000000	80.000000	1.000000	327000.000000

```
In [9]: df.isnull().sum()
```

```
Out[9]: age                0
anaemia                  0
creatinine_phosphokinase  0
diabetes                 0
ejection_fraction        0
high_blood_pressure       0
platelets                 0
serum_creatinine          0
serum_sodium              0
sex                       0
smoking                   0
time                     0
DEATH_EVENT              0
dtype: int64
```

```
In [12]: df.drop("time",axis=1)
```

```
Out[12]:
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	p
0	75.0	0	582	0	20	1	26:
1	55.0	0	7861	0	38	0	26:
2	65.0	0	146	0	20	0	16:
3	50.0	1	111	0	20	0	21:
4	65.0	1	160	1	20	0	32:
...
294	62.0	0	61	1	38	1	15:
295	55.0	0	1820	0	38	0	27:
296	45.0	0	2060	1	60	0	74:
297	45.0	0	2413	0	38	0	14:
298	50.0	0	196	0	45	0	39:

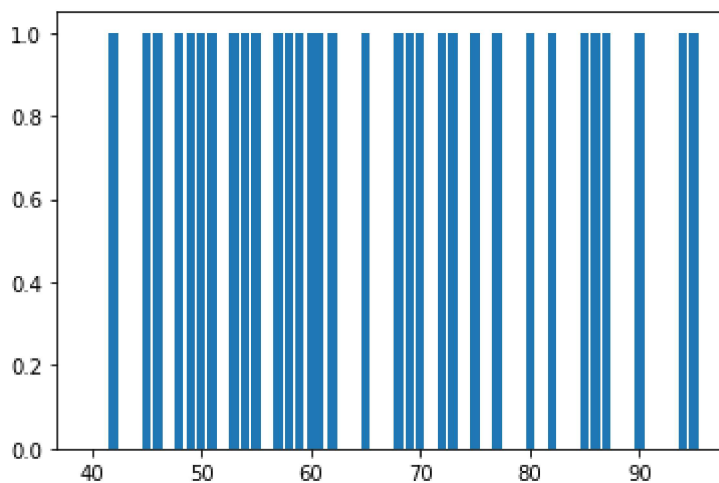
299 rows × 12 columns



```
In [30]: df.rename(columns={"high_blood_pressure":"bp"},inplace= True)
```

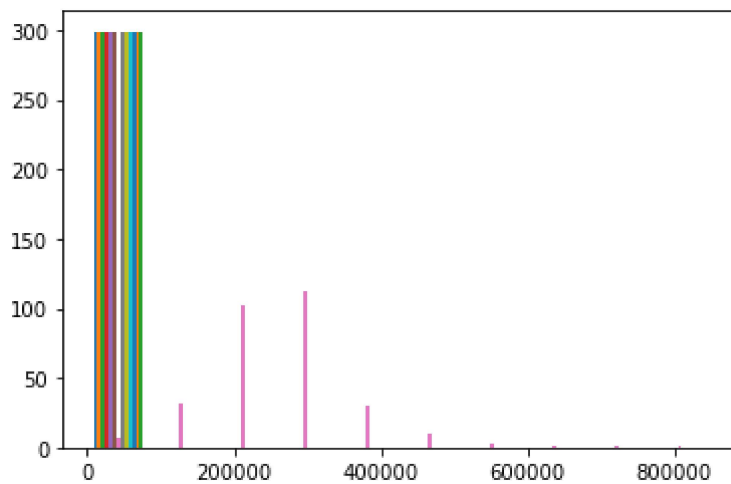
```
In [31]: plt.bar(df.age,df.DEATH_EVENT)
```

```
Out[31]: <BarContainer object of 299 artists>
```



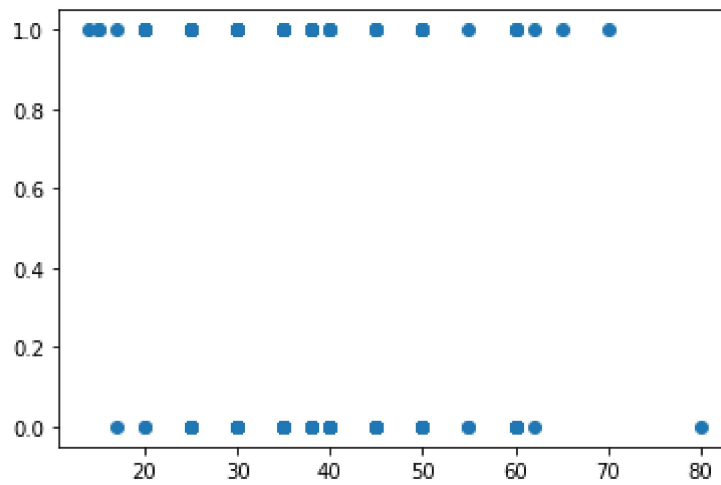
```
In [32]: plt.hist(df)
```

```
Out[32]: (array([[299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [  7., 32., 102., 113., 30., 10.,  2.,  1.,  1.,  1.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.],
 [299.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.]],
 array([ 0.,  85000., 170000., 255000., 340000., 425000., 510000.,
        595000., 680000., 765000., 850000.]),
 <a list of 13 BarContainer objects>)
```



```
In [36]: plt.scatter(df.ejection_fraction,df.DEATH_EVENT)
```

```
Out[36]: <matplotlib.collections.PathCollection at 0x165fcb5a670>
```



```
In [49]: x=df.drop("DEATH_EVENT",axis=1)
x.head()
```

```
Out[49]:
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	bp	platelets	serum_crea
0	75.0	0	582	0	20	1	265000.00	
1	55.0	0	7861	0	38	0	263358.03	
2	65.0	0	146	0	20	0	162000.00	
3	50.0	1	111	0	20	0	210000.00	
4	65.0	1	160	1	20	0	327000.00	

```
In [51]: y=df.DEATH_EVENT
y.head()
```

```
Out[51]: 0    1
         1    1
         2    1
         3    1
         4    1
         Name: DEATH_EVENT, dtype: int64
```

```
In [53]: 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_s
```

```
In [55]: X_test.head()
```

```
Out[55]:
```

	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	bp	platelets	serum_cre
173	50.0	1	115	0	20	0	189000.0	
287	45.0	0	582	1	55	0	543000.0	
51	53.0	1	91	0	20	1	418000.0	
146	52.0	0	132	0	30	0	218000.0	
214	65.0	1	135	0	35	1	290000.0	

```
In [56]: y_test.head()
```

```
Out[56]: 173    0
         287    0
         51    1
         146    0
         214    0
         Name: DEATH_EVENT, dtype: int64
```

```
In [57]: from sklearn import linear_model
LRG = linear_model.LogisticRegression(
    random_state = 0)
```

```
In [61]: LRG.fit(x,y)
```

```
Out[61]: LogisticRegression(random_state=0)
```

```
In [68]: LRG.predict(X_test)
```

```
Out[68]: array([0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1,
                0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0,
                1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0], dtype=int64)
```

In [70]: LRG.score(x,y)

Out[70]: 0.822742474916388

In []:

In []: