

## Problem statement:

```
# To predict and analyze which gender has a high chance of survival at the time of disaster
```

In [1]:

```
#import Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

In [2]:

```
df=pd.read_csv(r"C:\Users\DELL\Downloads\heart_disease_data.csv")
df
```

Out[2]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	ta
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	

303 rows × 14 columns



In [3]:

```
df.head()
```

Out[3]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	

In [4]:

```
df.tail()
```

Out[4]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	

In [5]:

```
df.describe
```

Out[5]:

```
<bound method NDFrame.describe of
restecg  thalach  exang  oldpeak
0      63      1      3      145    233      1      0      150      0      2.3
\
1      37      1      2      130    250      0      1      187      0      3.5
2      41      0      1      130    204      0      0      172      0      1.4
3      56      1      1      120    236      0      1      178      0      0.8
4      57      0      0      120    354      0      1      163      1      0.6
..      ...      ...      ..      ...      ...      ...      ...      ...      ...      ...
298     57      0      0      140    241      0      1      123      1      0.2
299     45      1      3      110    264      0      1      132      0      1.2
300     68      1      0      144    193      1      1      141      0      3.4
301     57      1      0      130    131      0      1      115      1      1.2
302     57      0      1      130    236      0      0      174      0      0.0

      slope  ca  thal  target
0          0  0      1        1
1          0  0      2        1
2          2  0      2        1
3          2  0      2        1
4          2  0      2        1
..      ...  ..      ...      ...
298        1  0      3        0
299        1  0      3        0
300        1  2      3        0
301        1  1      3        0
302        1  1      2        0

[303 rows x 14 columns]>
```

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   age         303 non-null   int64  
 1   sex         303 non-null   int64  
 2   cp          303 non-null   int64  
 3   trestbps    303 non-null   int64  
 4   chol        303 non-null   int64  
 5   fbs         303 non-null   int64  
 6   restecg     303 non-null   int64  
 7   thalach     303 non-null   int64  
 8   exang       303 non-null   int64  
 9   oldpeak     303 non-null   float64 
10   slope       303 non-null   int64  
11   ca          303 non-null   int64  
12   thal        303 non-null   int64  
13   target      303 non-null   int64  
dtypes: float64(1), int64(13)
memory usage: 33.3 KB
```

In [7]:

```
df.shape
```

Out[7]:

```
(303, 14)
```

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
target      0
dtype: int64
```

In [10]:



```
df['target'].value_counts()
```

Out[10]:

```
target
1      165
0      138
Name: count, dtype: int64
```

In [11]:



```
df['target'].value_counts()
```

Out[11]:

```
target
1      165
0      138
Name: count, dtype: int64
```

In [ ]:

