

# 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\smb06\Downloads\heart_disease_data.csv")
df
```

Out[2]:

	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
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

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
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

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	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

In [5]:

```
df.shape
```

Out[5]:

```
(303, 14)
```

In [6]:

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

Out[6]:

```
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 [7]:

```
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 [8]:

```
df.describe()
```

Out[8]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000

In [9]:

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

Out[9]:

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

In [10]:

```
x=df.drop(columns='target',axis=1)
y=df['target']
```

In [11]:

```
print(x)
```

	age	sex	cp	trestbps	chol	fbs	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
0	0	0	1
1	0	0	2
2	2	0	2
3	2	0	2
4	2	0	2
..	...	..	...
298	1	0	3
299	1	0	3
300	1	2	3
301	1	1	3
302	1	1	2

[303 rows x 13 columns]

In [12]:

```
print(y)
```

```
0    1
1    1
2    1
3    1
4    1
..
298  0
299  0
300  0
301  0
302  0
Name: target, Length: 303, dtype: int64
```

In [13]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, stratify=y, random_state=2
```

In [14]:

```
print(x.shape, x_train.shape, x_test.shape)
```

```
(303, 13) (242, 13) (61, 13)
```

## MODEL TRAINING FOR LOGISTIC REGRESSION

In [15]:

```
model = LogisticRegression()
```

In [16]:

```
model.fit(x_train,y_train)
```

C:\Users\smb06\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\linear\_model\\_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=1):  
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max\_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html> (<https://scikit-learn.org/stable/modules/preprocessing.html>)

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression) ([https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression))

```
n_iter_i = _check_optimize_result(
```

Out[16]:

```
LogisticRegression()  
LogisticRegression()
```

### #MODEL PREDICTION

In [17]:

```
# accuracy on training data
```

```
x_train_prediction = model.predict(x_train)
```

```
training_data_accuracy = accuracy_score(x_train_prediction, y_train)
```

In [18]:

```
# accuracy on test data
```

```
x_test_prediction = model.predict(x_test)
```

```
test_data_accuracy = accuracy_score(x_test_prediction, y_test)
```

In [19]:

```
print('Accuracy on Test data : ', test_data_accuracy)
```

```
Accuracy on Test data : 0.819672131147541
```

In [ ]: