```
#Predicting if a person would buy life insurance
#based on his age using logistic regression
```

Predicting if a person would buy life insurance based on his age using logistic regression

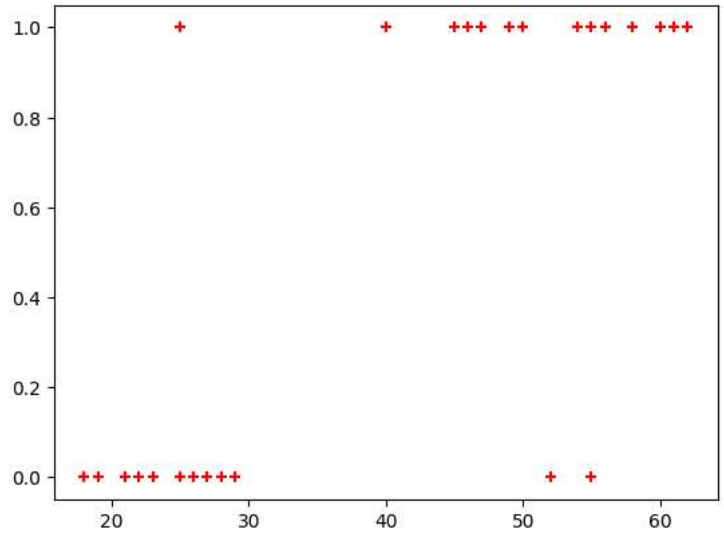
```
import pandas as  pd
from matplotlib import pyplot as plt
```

```
df=pd.read_csv("/content/insurance_data.csv")
df.head()
```

	age	bought_insurance
0	22	0
1	25	0
2	47	1
3	52	0
4	46	1

```
plt.scatter(df.age,df.bought_insurance,marker='+',color='red')
```

```
<matplotlib.collections.PathCollection at 0x789fe466ee00>
```




```
from sklearn.model_selection import train_test_split
```

```
X_train,X_test,y_train,y_test=train_test_split(df[['age']],df.bought_insurance,train_size=0.8)
```

```
X_test
```

	age
0	22
8	62
12	27
25	54
24	50
20	21

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

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

```
LogisticRegression()
```

```
X_test
```

	age
0	22
8	62
12	27
25	54
24	50
20	21

```
y_predicted = model.predict(X_test)
```

```
model.predict_proba(X_test)
```

```
array([[0.87037399, 0.12962601],
       [0.06508584, 0.93491416],
       [0.79135995, 0.20864005],
       [0.14792814, 0.85207186],
       [0.21516935, 0.78483065],
       [0.8827253 , 0.1172747 ]])
```

```
model.score(X_test,y_test)
```

```
1.0
```

```
y_predicted
```

```
array([0, 1, 0, 1, 1, 0])
```

```
X_test
```

	age
0	22
8	62
12	27
25	54
24	50
20	21

```
#model.coef_indicates value of m in y=m*x+b equation
model.coef_
```

```
array([[0.11422543]])
```

```
#model.intercept_indicates value of b in y=m*x+b equation
model.intercept_
```

```
array([-4.41722911])
```

```
#Lets defined sigmoid function now and
#do the math with hand
import math
def sigmoid(x):
  return 1 / (1+math.exp(-x))
```

```
#0.11422543~0.142 and -4.41722911~-4.42
def prediction_function(age):
  z = 0.142*age-4.42
  y = sigmoid(z)
  return y
```

```
#0.634 is more than 0.5 which means
#with 35 age will by insurance
age = 35
prediction_function(age)
```

```
0.6341355910108007
```

```
#0.843 is more than 0.5 which means
#with 43 age will by insurance
age = 43
prediction_function(age)
```

```
0.8436973963686705
```

Exercise

Download employee retention dataset from

here:https://www.kaggle.com/giripujar/hr-analytics.

Now do some exploratory data analysis

to figure out which variables have direct

and clear impact on employee retention

(ie.whether they leave the company or continue to work)

plot bar charts showing impact of employee salaries on retention

plot bar charts showing corelation between department and employee retention

Now build logistic regression model using variables that were narrowed down in step1

Measure the accuracy of the model