In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix,accuracy_score,classification_report
from sklearn.preprocessing import scale
train_data='C:/Users/ramya/Desktop/New folder/train_new.csv'
full_data=pd.read_csv(train_data)
full_data.describe()
test_data='C:/Users/ramya/Desktop/New folder/test_new.csv'
full_test=pd.read_csv(test_data)
full_test.describe()
```

Out[2]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare
count	418.000000	418.000000	332.000000	418.000000	418.000000	417.000000
mean	1100.500000	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.841838	14.181209	0.896760	0.981429	55.907576
min	892.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	1.000000	21.000000	0.000000	0.000000	7.895800
50%	1100.500000	3.000000	27.000000	0.000000	0.000000	14.454200
75%	1204.750000	3.000000	39.000000	1.000000	0.000000	31.500000
max	1309.000000	3.000000	76.000000	8.000000	9.000000	512.329200

In [3]:

```
a=full_data[['Age', 'Survived']].groupby(['Age'], as_index=False).mean().sort_values(by='Survived'
, ascending=False)
a.head(10)
```

Out[3]:

	Age	Survived
0	0.42	1.0
9	5.00	1.0
79	63.00	1.0
68	53.00	1.0
1	0.67	1.0
17	13.00	1.0
16	12.00	1.0
87	80.00	1.0
2	0.75	1.0
4	0.92	1.0

In [4]:

```
full_data[['Sex', 'Survived']].groupby(['Sex'], as_index=False).mean().sort_values(by='Survived',
ascending=True)
```

Out[4]:

	Sex	Survived
1	male	0.188908
0	female	0.742038

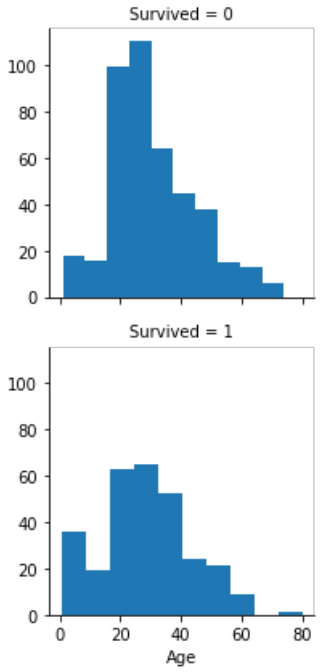
	Sex	Survived
0	female	0.742038

In [5]:

```
g=sns.FacetGrid(full_data,'Survived')
g.map(plt.hist,'Age',bins=10)
```

Out[5]:

```
<seaborn.axisgrid.FacetGrid at 0x1c89c1c61d0>
```



In [6]:

```
full_data[['Parch', 'Survived']].groupby(['Parch'],
as_index=False).mean().sort_values(by='Survived', ascending=True)
```

Out[6]:

	Parch	Survived
4	4	0.000000
6	6	0.000000
5	5	0.200000
0	0	0.343658
2	2	0.500000
1	1	0.550847
3	3	0.600000

In [7]:

```
full_data[['Pclass', 'Survived']].groupby(['Pclass'],
as_index=False).mean().sort_values(by='Survived', ascending=True)
```

Out[7]:

	Pclass	Survived
2	3	0.242363
1	2	0.472826
0	1	0.629630

In [8]:

```
full_data[['Cabin', 'Survived']].groupby(['Cabin'],
as_index=False).mean().sort_values(by='Survived', ascending=True)
```

Out[8]:

	Cabin	Survived
0	A10	0.0
45	B86	0.0
46	B94	0.0
52	C110	0.0
53	C111	0.0
54	C118	0.0
56	C124	0.0
59	C128	0.0
64	C30	0.0
67	C46	0.0
69	C49	0.0
79	C82	0.0
82	C86	0.0
83	C87	0.0
85	C91	0.0
44	B82 B84	0.0
88	C95	0.0
100	D30	0.0
106	D46	0.0
108	D48	0.0
110	D50	0.0
112	D6	0.0
122	E31	0.0
126	E38	0.0
129	E46	0.0
132	E58	0.0
133	E63	0.0
136	E77	0.0
139	F G63	0.0
140	F G73	0.0
...	...	...
68	C47	1.0
22	B3	1.0
66	C45	1.0
65	C32	1.0
57	C125	1.0
58	C126	1.0
48	C101	1.0
81	C85	1.0
19	B20	1.0
47	B96 B98	1.0
101	D33	1.0
30	B42	1.0
99	D28	1.0

	Cabin	Survived
15	B101	1.0
97	D21	1.0
96	D20	1.0
95	D19	1.0
94	D17	1.0
103	D36	1.0
93	D15	1.0
91	D10 D12	1.0
33	B50	1.0
89	C99	1.0
29	B41	1.0
87	C93	1.0
86	C92	1.0
17	B18	1.0
84	C90	1.0
92	D11	1.0
73	C62 C64	1.0

147 rows × 2 columns

In [9]:

```
full_data[['SibSp', 'Survived']].groupby(['SibSp'],
as_index=False).mean().sort_values(by='Survived', ascending=True)
```

Out[9]:

	SibSp	Survived
5	5	0.000000
6	8	0.000000
4	4	0.166667
3	3	0.250000
0	0	0.345395
2	2	0.464286
1	1	0.535885

In [10]:

```
full_data[['Embarked', 'Survived']].groupby(['Embarked'], as_index=False).mean().sort_values(by='S
urvived', ascending=True)
```

Out[10]:

	Embarked	Survived
2	S	0.336957
1	Q	0.389610
0	C	0.553571

In [11]:

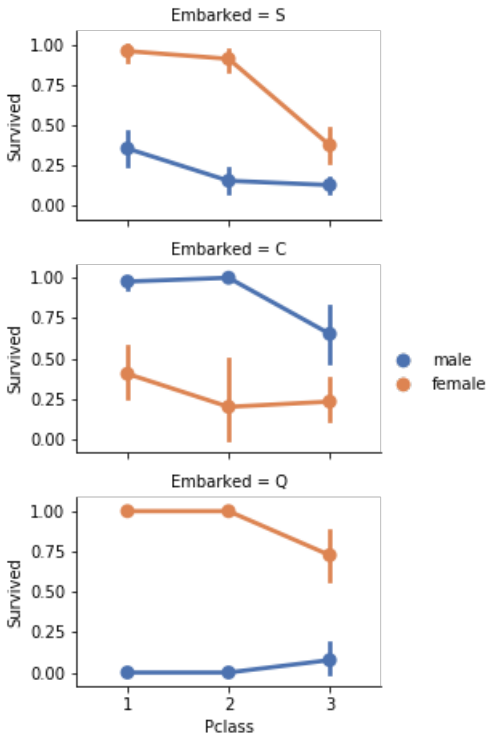
```
grid = sns.FacetGrid(full_data, row='Embarked', size=2.2, aspect=1.6)
grid.map(sns.pointplot, 'Pclass', 'Survived', 'Sex', palette='deep')
grid.add_legend()
```

```
C:\Users\ramya\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The `size`
paramter has been renamed to `height`; please update your code.
```

```
  warnings.warn(msg, UserWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\seaborn\axisgrid.py:715: UserWarning: Using the
pointplot function without specifying `order` is likely to produce an incorrect plot.
  warnings.warn(warning)
C:\Users\ramya\Anaconda3\lib\site-packages\seaborn\axisgrid.py:720: UserWarning: Using the
pointplot function without specifying `hue_order` is likely to produce an incorrect plot.
  warnings.warn(warning)
C:\Users\ramya\Anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Using a non-t
uple sequence for multidimensional indexing is deprecated; use `arr[tuple(seq)]` instead of `arr[s
eq]`. In the future this will be interpreted as an array index, `arr[np.array(seq)]`, which will r
esult either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

Out[11]:

```
<seaborn.axisgrid.FacetGrid at 0x1c8a18092b0>
```



In [12]:

```
full_data[['Age', 'Survived']].groupby(['Age'], as_index=False).mean().sort_values(by='Survived',
ascending=True)
```

Out[12]:

	Age	Survived
77	61.00	0.000000
54	40.50	0.000000
30	23.50	0.000000
26	20.50	0.000000
37	28.50	0.000000
86	74.00	0.000000
40	30.50	0.000000
60	45.50	0.000000
19	14.50	0.000000
61	46.00	0.000000
32	24.50	0.000000
14	10.00	0.000000
80	64.00	0.000000

	Age	Survived
49	36.50	0.000000
85	71.00	0.000000
46	34.50	0.000000
81	65.00	0.000000
71	55.50	0.000000
73	57.00	0.000000
84	70.50	0.000000
75	59.00	0.000000
82	66.00	0.000000
83	70.00	0.000000
62	47.00	0.111111
50	37.00	0.166667
57	43.00	0.200000
25	20.00	0.200000
27	21.00	0.208333
13	9.00	0.250000
15	11.00	0.250000
...	...	...
78	62.00	0.500000
67	52.00	0.500000
12	8.00	0.500000
65	50.00	0.500000
18	14.00	0.500000
31	24.00	0.500000
76	60.00	0.500000
48	36.00	0.500000
43	32.50	0.500000
74	58.00	0.600000
35	27.00	0.611111
47	35.00	0.611111
63	48.00	0.666667
64	49.00	0.666667
10	6.00	0.666667
8	4.00	0.700000
5	1.00	0.714286
20	15.00	0.800000
7	3.00	0.833333
79	63.00	1.000000
0	0.42	1.000000
17	13.00	1.000000
16	12.00	1.000000
9	5.00	1.000000
4	0.92	1.000000
3	0.83	1.000000
2	0.75	1.000000
1	0.67	1.000000
68	53.00	1.000000
87	80.00	1.000000

88 rows × 2 columns

In [13]:

```
full_data.drop("Cabin", axis=1, inplace=True)
```

In [14]:

```
full_data.isnull().sum()
```

Out[14]:

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         2
dtype: int64
```

In [15]:

```
full_data.Age.fillna(full_data.Age.mean(),axis = 0,inplace = True)
```

In [16]:

```
full_data.isnull().sum()
```

Out[16]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64
```

In [17]:

```
full_data.Embarked.fillna(method = 'ffill',axis=0,inplace = True)
```

In [18]:

```
full_data.isnull().sum()
```

Out[18]:

```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       0
dtype: int64
```

In [19]:

In [19]:

```
sex=pd.get_dummies(full_data["Sex"],drop_first=True)
sex.head(5)
```

Out[19]:

	male
0	1
1	0
2	0
3	0
4	1

In [20]:

```
full_data.Embarked.fillna(method = 'ffill',axis=0,inplace = True)
```

In [21]:

```
embark=pd.get_dummies(full_data["Embarked"],drop_first=True)
embark.head(5)
```

Out[21]:

	Q	S
0	0	1
1	0	0
2	0	1
3	0	1
4	0	1

In [22]:

```
full_data=pd.concat([full_data,sex,embark],axis=1)
full_data.head(5)
```

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	male	Q	S
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S	1	0	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C	0	0	0
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S	0	0	1
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S	0	0	1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S	1	0	1

In [23]:

```
full_data.drop(['Embarked','Sex','Ticket','Name'],axis=1,inplace=True)
full_data.head()
```

Out[23]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1	0	7.2500	1	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
2	3	1	3	26.0	0	0	7.9250	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	1
4	5	0	3	35.0	0	0	8.0500	1	0	1

In [24]:

```
from sklearn.feature_selection import RFE
from sklearn.feature_selection import RFECV
cols=['Pclass','Fare','male','Parch']
X = full_data[cols]
y = full_data['Survived']
model1=LogisticRegression()
rfe = RFE(model1, 8)
rfe = rfe.fit(X, y)
rfecv = RFECV(estimator=LogisticRegression(), step=1, cv=10, scoring='accuracy')
rfecv.fit(X, y)
```

```
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

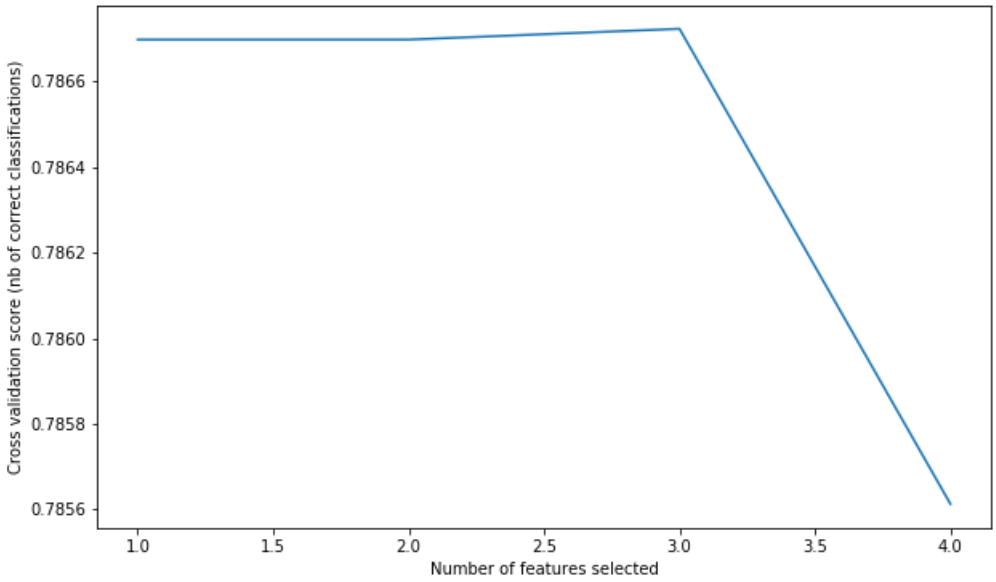
```
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

Out[24]:

```
RFECV(cv=10,
   estimator=LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='warn',
          n_jobs=None, penalty='l2', random_state=None, solver='warn',
          tol=0.0001, verbose=0, warm_start=False),
   min_features_to_select=1, n_jobs=None, scoring='accuracy', step=1,
   verbose=0)
```

In [25]:

```
plt.figure(figsize=(10,6))
plt.xlabel("Number of features selected")
plt.ylabel("Cross validation score (nb of correct classifications)")
plt.plot(range(1, len(rfecv.grid_scores_) + 1), rfecv.grid_scores_)
plt.show()
```



In [26]:

```
X = full_data.drop("Survived",axis =1)
y = full_data["Survived"]
```

In [32]:

```
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest = train_test_split(X,y,test_size = 0.2,random_state = 10)
```

In [33]:

```
model1.fit(xtrain,ytrain)
```

```
C:\Users\ramya\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWarning: De
fault solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)
```

Out[33]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='warn',
          n_jobs=None, penalty='l2', random_state=None, solver='warn',
          tol=0.0001, verbose=0, warm_start=False)
```

In [ ]:

In [34]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
          penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
          verbose=0, warm_start=False)
```

Out[34]:

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
          intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=1,
          penalty='l2', random_state=None, solver='liblinear', tol=0.0001,
          verbose=0, warm_start=False)
```

In [35]:

```
ypred = model1.predict(xtest)
accuracy_score(ytest,ypred)
```

Out[35]:

```
0.8212290502793296
```

In [36]:

```
full_test.isnull().sum()
```

Out[36]:

```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

In [37]:

```
full_test.Age.fillna(full_data.Age.mean(),axis = 0,inplace = True)
```

In [38]:

```
full_test.isnull().sum()
```

Out[38]:

```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
dtype: int64
```

In [39]:

```
full_test.Fare.fillna(full_data.Fare.mean(),axis = 0,inplace = True)
```

In [40]:

In [40]:

```
full_test.isnull().sum()
```

Out[40]:

```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          327
Embarked         0
dtype: int64
```

In [41]:

```
full_test.drop("Cabin", axis=1, inplace=True)
```

In [42]:

```
full_test.head(5)
```

Out[42]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	S

In [43]:

```
sex=pd.get_dummies(full_test["Sex"],drop_first=True)
sex.head(5)
```

Out[43]:

	male
0	1
1	0
2	1
3	1
4	0

In [44]:

```
embark=pd.get_dummies(full_test["Embarked"],drop_first=True)
embark.head(5)
```

Out[44]:

	Q	S
0	1	0
1	0	1
2	1	0

	0	1
3	Q	S
4	0	1

In [45]:

```
full_test=pd.concat([full_test,sex,embark],axis=1)
full_test.head(5)
```

Out[45]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	male	Q	S
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	Q	1	1	0
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	S	0	0	1
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	Q	1	1	0
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	S	1	0	1
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	S	0	0	1

In [46]:

```
full_test.drop(["Ticket","Name","Sex","Embarked"], axis=1, inplace=True)
full_test.head()
```

Out[46]:

	PassengerId	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	892	3	34.5	0	0	7.8292	1	1	0
1	893	3	47.0	1	0	7.0000	0	0	1
2	894	2	62.0	0	0	9.6875	1	1	0
3	895	3	27.0	0	0	8.6625	1	0	1
4	896	3	22.0	1	1	12.2875	0	0	1

In [47]:

```
full_test.isnull().sum()
```

Out[47]:

```
PassengerId    0
Pclass         0
Age            0
SibSp          0
Parch          0
Fare           0
male           0
Q              0
S              0
dtype: int64
```

In [48]:

```
y_pred = model1.predict(xtest)
```

In [49]:

```
y_pred
```

Out[49]:

```
array([0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0,
       0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1,
```

```
       0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0,
       1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0,
       0, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0,
       0, 0, 1], dtype=int64)
```

In [54]:

```
output=full_test.iloc[:, 0]
```

In [55]:

```
ypred = pd.DataFrame(ypred)
```

In [56]:

```
output= pd.concat([output,ypred],axis=1)
y_pred_prob.columns = ['PassengerId','Survived']
y_pred_prob.head()
```

Out[56]:

	PassengerId	Survived
0	892	0.0
1	893	0.0
2	894	0.0
3	895	1.0
4	896	1.0

In [60]:

```
output.to_csv('output.csv',index=False)
```