In [1]:
```
import numpy as np
import pandas as pd
```

In [2]:
```
titanic_test = pd.read_csv("C:/Users/RAM1/Desktop/Titanic/titanic/test.csv")
```

In [3]:
```
titanic_train = pd.read_csv("C:/Users/RAM1/Desktop/Titanic/titanic/train.csv")
```

In [4]:
```
titanic_train.head()
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nal
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Nal
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Nal

In [5]:
```
titanic_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId    891 non-null int64
Survived       891 non-null int64
Pclass         891 non-null int64
Name           891 non-null object
Sex            891 non-null object
Age            714 non-null float64
SibSp          891 non-null int64
Parch          891 non-null int64
Ticket         891 non-null object
Fare           891 non-null float64
Cabin          204 non-null object
Embarked       889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```

In [6]:
```
titanic_train.isna().sum()
```

Out[6]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [7]:
```
titanic_train.Age.isna().value_counts()
```

Out[7]:
```
False    714
True     177
Name: Age, dtype: int64
```

In [8]:
```
titanic_train.describe()
```

Out[8]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [9]:
```
#Replace the missing value in "Age" column by median
#Count of passanger by cabin number
#Delete Cabin
#Categorical (Object values) Value_count()
```

In [10]:
```
#titanic_train["Age"]= titanic_train["Age"].str.replace
```

In [11]:
```
#titanic_train.Age=titanic_train.Age.fillna()
```

In [12]:
```
titanic_train.Age= titanic_train.Age.fillna(titanic_train.Age.median())
```

In [13]:
```
titanic_train.Age.isna().value_counts()
```

Out[13]:
```
False    891
Name: Age, dtype: int64
```

In [14]:
```
titanic_train.Age.describe()
```

Out[14]:
```
count    891.000000
mean      29.361582
std       13.019697
min        0.420000
25%       22.000000
50%       28.000000
75%       35.000000
max       80.000000
Name: Age, dtype: float64
```

In [15]:
```
titanic_train.head()
```

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	Nal
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C8
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	Nal
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C12
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	Nal

In [16]:
```
titanic_train=titanic_train.drop(["Cabin","PassengerId","Name","Fare","Ticket"],a
```

In [17]:
```
titanic_train.head()
```

Out[17]:

	Survived	Pclass	Sex	Age	SibSp	Parch	Embarked
0	0	3	male	22.0	1	0	S
1	1	1	female	38.0	1	0	C
2	1	3	female	26.0	0	0	S
3	1	1	female	35.0	1	0	S
4	0	3	male	35.0	0	0	S

In [18]:
```
titanic_train=titanic_train.drop(["Embarked"],axis=1)
```

In [19]:
```
#titanic_train[titanic_train["Embarked"].isna()]
```

In [20]:
```
dummy=pd.get_dummies(titanic_train["Sex"])
```

In [21]:
```
titanic_train=pd.concat([titanic_train,dummy],axis=1)
```

In [22]:
```
titanic_train=titanic_train.drop(["Sex"],axis=1)
```

In [23]:
```
titanic_train.head()
```

Out[23]:

	Survived	Pclass	Age	SibSp	Parch	female	male
0	0	3	22.0	1	0	0	1
1	1	1	38.0	1	0	1	0
2	1	3	26.0	0	0	1	0
3	1	1	35.0	1	0	1	0
4	0	3	35.0	0	0	0	1

In [24]:
```
titanic_train.Survived.value_counts()
```

Out[24]:
```
0    549
1    342
Name: Survived, dtype: int64
```

In [25]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [ ]:
```
```

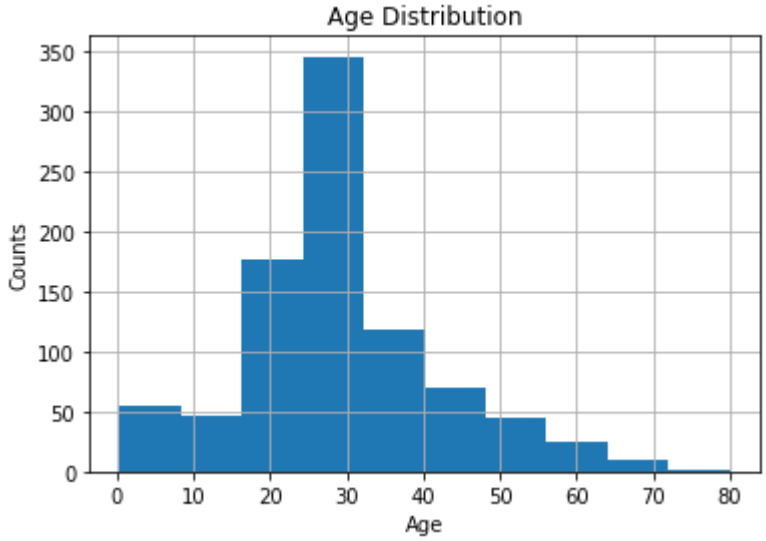
In [26]:
```
titanic_train.Survived.isna().value_counts()
```

Out[26]:
```
False    891
Name: Survived, dtype: int64
```

In [27]:
```
titanic_train['Survived'] = titanic_train['Survived'].fillna((titanic_train['Surv
```

In [28]:
```
#sns.countplot(x="Survived",data=titanic_train,palette="hls")
```

In [29]:
```
titanic_train.Age.hist()
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Counts")
plt.show()
```



In [30]:
```
from sklearn.linear_model import LogisticRegression
```

In [ ]:

In [ ]:

In [31]:
```
titanic_train_x=titanic_train.drop("Survived",axis=1)
titanic_train_y=titanic_train["Survived"]
```

In [32]:
```
titanic_train_x.head()
```

Out[32]:

	Pclass	Age	SibSp	Parch	female	male
0	3	22.0	1	0	0	1
1	1	38.0	1	0	1	0
2	3	26.0	0	0	1	0
3	1	35.0	1	0	1	0
4	3	35.0	0	0	0	1

In [33]:
```
titanic_train_y.head()
```

Out[33]:
```
0    0
1    1
2    1
3    1
4    0
Name: Survived, dtype: int64
```

In [34]:
```
#titanic_model=LogisticRegression.fit(titanic_train_x,titanic_train_y)
```

In [35]:
```
titanic_test.isna().sum()
```

Out[35]:
```
PassengerId      0
Pclass           0
Name             0
Sex              0
Age             86
SibSp            0
Parch            0
Ticket           0
Fare             1
Cabin          327
Embarked         0
Survived         0
dtype: int64
```

In [36]:
```
titanic_test.Age.isna().value_counts()
```

Out[36]:
```
False    332
True      86
Name: Age, dtype: int64
```

In [37]:
```
titanic_test.Age= titanic_test.Age.fillna(titanic_test.Age.median())
```

In [38]:
```
titanic_test.Age.isna().value_counts()
```

Out[38]:
```
False    418
Name: Age, dtype: int64
```

In [39]:
```
titanic_test.head()
```

Out[39]:

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

In [40]:
```
titanic_test=titanic_test.drop(["PassengerId","Name","Ticket","Fare","Cabin"],axi
```

In [41]:
```
titanic_test.head()
```

Out[41]:

	Pclass	Sex	Age	SibSp	Parch	Embarked	Survived
0	3	male	34.5	0	0	Q	0
1	3	female	47.0	1	0	S	1
2	2	male	62.0	0	0	Q	0
3	3	male	27.0	0	0	S	0
4	3	female	22.0	1	1	S	1

In [42]:
```
dummy2=pd.get_dummies(titanic_test["Sex"])
titanic_test=pd.concat([titanic_test,dummy2],axis=1)
```

In [43]:
```
titanic_test=titanic_test.drop("Sex",axis=1)
```

In [44]:
```
titanic_test=titanic_test.drop("Embarked",axis=1)
```

In [45]:
```
titanic_test.head()
```

Out[45]:

	Pclass	Age	SibSp	Parch	Survived	female	male
0	3	34.5	0	0	0	0	1
1	3	47.0	1	0	1	1	0
2	2	62.0	0	0	0	0	1
3	3	27.0	0	0	0	0	1
4	3	22.0	1	1	1	1	0

In [46]:
```
titanic_test_x=titanic_test.drop("Survived",axis=1)
titanic_test_y=titanic_test["Survived"]
```

In [47]:
```
lgr=LogisticRegression()
titanic_model=lgr.fit(titanic_train_x,titanic_train_y)
```

In [48]:
```
titanic_predict=titanic_model.predict(titanic_test_x)
```

In [49]:
```
titanic_test_x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 6 columns):
Pclass    418 non-null int64
Age       418 non-null float64
SibSp     418 non-null int64
Parch     418 non-null int64
female    418 non-null uint8
male      418 non-null uint8
dtypes: float64(1), int64(3), uint8(2)
memory usage: 14.0 KB
```

In [50]:
```
titanic_predict
```

Out[50]:
```
array([0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0,
       1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1,
       1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1,
       1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1,
       0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1,
       1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1,
       0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0,
       1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
       0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1,
       0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0,
       0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,
       0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0,
       0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0,
       1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1,
       0, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0],
      dtype=int64)
```

In [51]:
```
#titanic_test.Age.hist()
#plt.title("Age Distribution")
#plt.xlabel("Age")
#plt.ylabel("Counts")
#plt.show()
```

In [52]:
```
titanic_test_x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 6 columns):
Pclass    418 non-null int64
Age       418 non-null float64
SibSp     418 non-null int64
Parch     418 non-null int64
female    418 non-null uint8
male      418 non-null uint8
dtypes: float64(1), int64(3), uint8(2)
memory usage: 14.0 KB
```

In [53]:
```
#titanic_test_model=pd.DataFrame(["titanic_test_x"],["titanic_predict"])
```

In [54]:
```
titanic_test_model=pd.DataFrame(titanic_test_y)
```

In [55]:
```
titanic_test_model["Predicted_surv"]=titanic_predict
```

In [56]:
```
titanic_test_model
```

Out[56]:

	Survived	Predicted_surv
0	0	0
1	1	0
2	0	0
3	0	0
4	1	1
5	0	0
6	1	1
7	0	0
8	1	1
9	0	0
10	0	0
11	0	0
12	1	1
13	0	0
14	1	1
15	1	1
16	0	0
17	0	0
18	1	1
19	1	0
20	0	0
21	0	0
22	1	1
23	0	1
24	1	1
25	0	0
26	1	1
27	0	0
28	0	0
29	0	0
...	...	...
388	0	0
389	0	0
390	0	1

	Survived	Predicted_surv
391	1	1
392	0	0
393	0	0
394	0	0
395	1	1
396	0	0
397	1	1
398	0	0
399	0	0
400	1	1
401	0	0
402	1	1
403	0	1
404	0	0
405	0	0
406	0	0
407	0	0
408	1	1
409	1	1
410	1	1
411	1	1
412	1	1
413	0	0
414	1	1
415	0	0
416	0	0
417	0	0

418 rows × 2 columns

In [57]:
```
#To find accuracy
titanic_model.score(titanic_test_x,titanic_test_y)
```

Out[57]: 0.9473684210526315

In [58]:
```
from sklearn import metrics
```

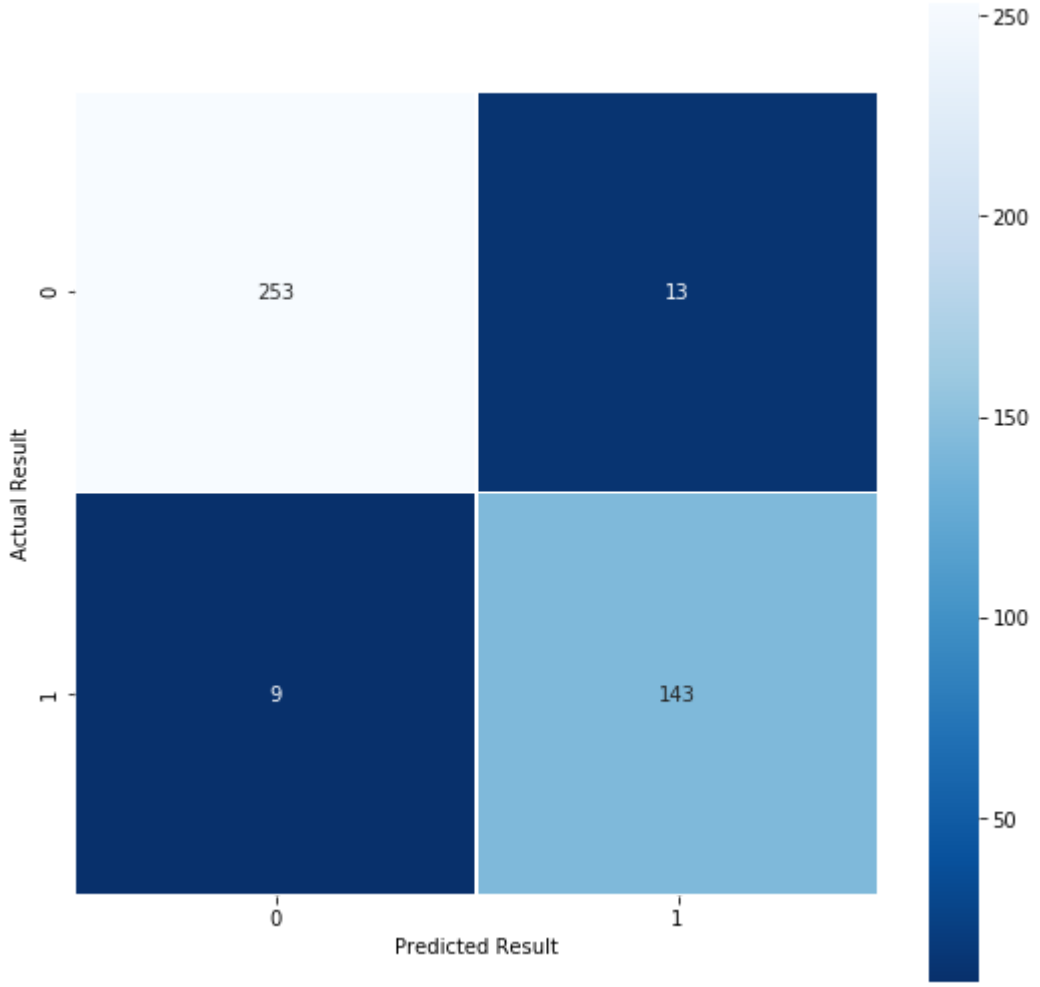
In [59]:
```
conf_metrics = metrics.confusion_matrix(titanic_test_y,titanic_predict)
```

In [60]:
```
#titanic_model["titanic_test_x","Predict_surv"].value_counts()
```

In [61]:
```
print(conf_metrics)
```

```
[[253  13]
 [  9 143]]
```

In [62]:
```
plt.figure(figsize=(9,9))
sns.heatmap(conf_metrics,annot=True,fmt='.0f',linewidth=.5,square =True,cmap='Blu
plt.ylabel('Actual Result')
plt.xlabel('Predicted Result')
plt.show()
```



In [ ]:

In [ ]: