In [25]:
```
#importing the required libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [26]:
```
#reading the dataset

file=pd.read_csv('titanic3.csv')
```

In [27]:
```
file.shape
```

Out[27]:
```
(1310, 14)
```

In [28]:
```
file.describe()
```

Out[28]:

| | pclass | survived | age | sibsp | parch | fare | body |
|---|---|---|---|---|---|---|---|
| **count** | 1309.000000 | 1309.000000 | 1046.000000 | 1309.000000 | 1309.000000 | 1308.000000 | 121.000000 |
| **mean** | 2.294882 | 0.381971 | 29.881135 | 0.498854 | 0.385027 | 33.295479 | 160.809917 |
| **std** | 0.837836 | 0.486055 | 14.413500 | 1.041658 | 0.865560 | 51.758668 | 97.696922 |
| **min** | 1.000000 | 0.000000 | 0.166700 | 0.000000 | 0.000000 | 0.000000 | 1.000000 |
| **25%** | 2.000000 | 0.000000 | 21.000000 | 0.000000 | 0.000000 | 7.895800 | 72.000000 |
| **50%** | 3.000000 | 0.000000 | 28.000000 | 0.000000 | 0.000000 | 14.454200 | 155.000000 |
| **75%** | 3.000000 | 1.000000 | 39.000000 | 1.000000 | 0.000000 | 31.275000 | 256.000000 |
| **max** | 3.000000 | 1.000000 | 80.000000 | 8.000000 | 9.000000 | 512.329200 | 328.000000 |

In [29]:
```
file.columns
```

Out[29]:
```
Index(['pclass', 'survived', 'name', 'sex', 'age', 'sibsp', 'parch', 'ticket',
       'fare', 'cabin', 'embarked', 'boat', 'body', 'home.dest'],
      dtype='object')
```

In [30]:
```
file.dtypes
```

Out[30]:
```
pclass        float64
survived      float64
name           object
sex            object
age           float64
sibsp         float64
parch         float64
ticket         object
fare          float64
cabin          object
embarked       object
boat           object
body          float64
home.dest      object
dtype: object
```

In [31]:
```
file.head()
```

Out[31]:

| | pclass | survived | name | sex | age | sibsp | parch | ticket | fare | cabin | embarked | boat | body | home.dest |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1.0 | 1.0 | Allen, Miss. Elisabeth Walton | female | 29.0000 | 0.0 | 0.0 | 24160 | 211.3375 | B5 | S | 2 | NaN | St Louis, MO |
| **1** | 1.0 | 1.0 | Allison, Master. Hudson Trevor | male | 0.9167 | 1.0 | 2.0 | 113781 | 151.5500 | C22 C26 | S | 11 | NaN | Montreal, PQ / Chesterville, ON |
| **2** | 1.0 | 0.0 | Allison, Miss. Helen Loraine | female | 2.0000 | 1.0 | 2.0 | 113781 | 151.5500 | C22 C26 | S | NaN | NaN | Montreal, PQ / Chesterville, ON |
| **3** | 1.0 | 0.0 | Allison, Mr. Hudson Joshua Creighton | male | 30.0000 | 1.0 | 2.0 | 113781 | 151.5500 | C22 C26 | S | NaN | 135.0 | Montreal, PQ / Chesterville, ON |
| **4** | 1.0 | 0.0 | Allison, Mrs. Hudson J.C. (Bessie Waldo Daniels) | female | 25.0000 | 1.0 | 2.0 | 113781 | 151.5500 | C22 C26 | S | NaN | NaN | Montreal, PQ / Chesterville, ON |

In [32]:
```
#checking for any missing value

file.isnull().sum()
```

Out[32]:
```
pclass       1
survived     1
name         1
sex          1
age        264
sibsp        1
parch        1
ticket       1
fare         2
cabin     1015
embarked     2
boat       824
body      1189
home.dest  565
dtype: int64
```

In [33]:
```
#okay so there are many missing values
#let's impute these missing values

file['age']
```

Out[33]:
```
0       29.0000
1        0.9167
2        2.0000
3       30.0000
4       25.0000
5       48.0000
6       63.0000
7       39.0000
8       53.0000
9       71.0000
10      47.0000
11      18.0000
12      24.0000
13      26.0000
14      80.0000
15          NaN
16      24.0000
17      50.0000
18      32.0000
19      36.0000
20      37.0000
21      47.0000
22      26.0000
23      42.0000
24      29.0000
25      25.0000
26      25.0000
27      19.0000
28      35.0000
29      28.0000
         ...
1280    22.0000
1281    22.0000
1282        NaN
1283        NaN
1284        NaN
1285    32.5000
1286    38.0000
1287    51.0000
1288    18.0000
1289    21.0000
1290    47.0000
1291        NaN
1292        NaN
1293        NaN
1294    28.5000
1295    21.0000
1296    27.0000
1297        NaN
1298    36.0000
1299    27.0000
1300    15.0000
1301    45.5000
1302        NaN
1303        NaN
1304    14.5000
1305        NaN
1306    26.5000
1307    27.0000
1308    29.0000
1309        NaN
Name: age, Length: 1310, dtype: float64
```

In [34]:
```
#imputing the missing value with the mean of the age variable

file['age'].fillna(np.mean(file['age']),inplace=True)
```

In [35]:
```
file['age'].isnull().sum()
```

Out[35]:
```
0
```

In [36]:
```
#no there are no missing values in the age column
```

In [37]:
```
pd.crosstab(file['sex'],file['survived'])
```

Out[37]:

| survived | 0.0 | 1.0 |
|---|---|---|
| **sex** | | |
| **female** | 127 | 339 |
| **male** | 682 | 161 |

In [38]:
```
#okay so more females survived compared to men
```

In [39]:
```
file.corr()
```

Out[39]:

| | pclass | survived | age | sibsp | parch | fare | body |
|---|---|---|---|---|---|---|---|
| **pclass** | 1.000000 | -0.312469 | -0.366370 | 0.060832 | 0.018322 | -0.558629 | -0.034642 |
| **survived** | -0.312469 | 1.000000 | -0.050199 | -0.027825 | 0.082660 | 0.244265 | NaN |
| **age** | -0.366370 | -0.050199 | 1.000000 | -0.190747 | -0.130872 | 0.171892 | 0.059059 |
| **sibsp** | 0.060832 | -0.027825 | -0.190747 | 1.000000 | 0.373587 | 0.160238 | -0.099961 |
| **parch** | 0.018322 | 0.082660 | -0.130872 | 0.373587 | 1.000000 | 0.221539 | 0.051099 |
| **fare** | -0.558629 | 0.244265 | 0.171892 | 0.160238 | 0.221539 | 1.000000 | -0.043110 |
| **body** | -0.034642 | NaN | 0.059059 | -0.099961 | 0.051099 | -0.043110 | 1.000000 |

In [40]:
```
file.groupby('survived')['fare'].mean().plot.bar()
```

Out[40]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x9c1b438>
```

In [42]:
```
#people having high fare survived more
```

In [43]:
```
#creating numeric values for categorical variables
```

In [44]:
```
file=pd.get_dummies(file)
```

In [45]:
```
file.fillna(0,inplace=True)
```

In [46]:
```
file.isnull().sum()
```

Out[46]:
```
pclass                                                 0
survived                                               0
age                                                    0
sibsp                                                  0
parch                                                  0
fare                                                   0
body                                                   0
name_Abbing, Mr. Anthony                               0
name_Abbott, Master. Eugene Joseph                     0
name_Abbott, Mr. Rossmore Edward                       0
name_Abbott, Mrs. Stanton (Rosa Hunt)                  0
name_Abelseth, Miss. Karen Marie                       0
name_Abelseth, Mr. Olaus Jorgensen                     0
name_Abelson, Mr. Samuel                               0
name_Abelson, Mrs. Samuel (Hannah Wizosky)             0
name_Abrahamson, Mr. Abraham August Johannes           0
name_Abrahim, Mrs. Joseph (Sophie Halaut Easu)         0
name_Adams, Mr. John Bertram Wilkinson Martin          0
name_Adams, Mr. John                                   0
name_Ahlin, Mrs. Johan (Johanna Persdotter Larsson)    0
name_Aks, Master. Philip Frank                         0
name_Aks, Mrs. Sam (Leah Rosen)                        0
name_Akihama, Mr. Hassef Cassem                        0
name_Aldworth, Mr. Charles Augustus                    0
name_Alexander, Mr. William                            0
name_Alhomaki, Mr. Ilmari Rudolf                       0
name_Ali, Mr. Ahmed                                    0
name_Ali, Mr. William                                  0
name_Allen, Miss. Elizabeth Walton                     0
name_Allen, Mr. William Henry                          0
                                                      ..
home.dest_Trenton, NJ                                  0
home.dest_Tuxedo Park, NY                              0
home.dest_Union Hill, NJ                               0
home.dest_Upper Burma, India Pittsburgh, PA            0
home.dest_Vadsbro, Sweden Minneiwa, MI                 0
home.dest_Vancouver, BC                                0
home.dest_Victoria, BC                                 0
home.dest_Walthamstow, England                         0
home.dest_Warwick, England                             0
home.dest_Washington, DC                               0
home.dest_Washington, Chicago, IL                      0
home.dest_West Bromwich, England Pontiac, MI           0
home.dest_West Hampstead, London / Bangawa, NB         0
home.dest_West Haven, CT                               0
home.dest_West Roxbury, MA                             0
home.dest_West Side Ave. Jersey, London                0
home.dest_Westcliff-on-Sea, Essex                      0
home.dest_Weston-Super-Mare / Moose Jaw, SK            0
home.dest_Weston-Super-Mare, Somerset                  0
home.dest_Wiltshire, England Niagara Falls, NY         0
home.dest_Wimbledon Park, London / Bayling Island, Hants 0
home.dest_Windsor, England New York, NY                0
home.dest_Winnipeg, MB                                 0
home.dest_Winnipeg, MN                                 0
home.dest_Woodford County, KY                          0
home.dest_Worcester, England                           0
home.dest_Worcester, MA                                0
home.dest_Yorril, England / Cottage Grove, OR          0
home.dest_Youngstown, OH                               0
home.dest_Zurich, Switzerland                          0
Length: 2830, dtype: int64
```

In [47]:
```
#creating train and test model

file.shape
```

Out[47]:
```
(1310, 2830)
```

In [48]:
```
#we'll use first 1000 rows for training our model and rest for the testing purpose
```

In [49]:
```
train=file[0:999]
test=file[1000:]
```

In [53]:
```
train.shape,test.shape
```

Out[53]:
```
(999, 2830), (310, 2830))
```

In [54]:
```
#separating the dependent and the independent variables

x_train=train.drop('survived',axis=1)
y_train=train['survived']
x_test=test.drop('survived',axis=1)
true_p=test['survived']
```

In [55]:
```
#importing the required libraries for building our model

from sklearn.linear_model import LogisticRegression
```

In [52]:
```
logreg=LogisticRegression()
```

In [55]:
```
#fitting our model

logreg.fit(x_train,y_train)
```

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

Out[55]:
```
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 [56]:
```
#now that our model is fitted it's time to make predictions

pred=logreg.predict(x_test)
```

In [57]:
```
pred
```

Out[57]:
```
array([0., 1., 1., 0., 1., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
       1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0.,
       0., 0., 1., 0., 0., 1., 1., 0., 1., 1., 1., 0., 0., 0., 0., 1.,
       0., 0., 0., 0., 0., 1., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
       0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1.,
       0., 1., 1., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 1.])
```

In [58]:
```
#checking the score of our model

logreg.score(x_train,y_train)
```

Out[58]:
```
0.9919191991992
```