Titanic

In [1]:
```
import pandas as pd
import numpy as np
tit=pd.read_csv("titanic_data.csv")
tit.head()
```

Out[1]:

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [2]:
```
tit.isnull().sum()
```

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

In [3]:
```
len(tit)
```

Out[3]:
```
891
```

In [4]:
```
tit.drop(['Cabin'],axis=1,inplace=True)
```

In [5]:
```
tit.Embarked.unique()
```

Out[5]:
```
array(['S', 'C', 'Q', nan], dtype=object)
```

In [6]:
```
tit.Age.mean()
```

Out[6]:
```
29.69911764705882
```

In [9]:
```
tit.Age.fillna(tit.Age.mean(),inplace=True)
```

In [10]:
```
tit.isnull().sum()
```

Out[10]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Gender         0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64
```

In [11]:
```
tit.drop(['PassengerId','Name','Ticket','Fare'],axis=1,inplace=True)
```

In [16]:
```
tit['Gen']=tit.Gender.map({'male':0,'female':1})
```

In [18]:
```
tit.head()
```

Out[18]:

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

In [62]:
```
#-Dummies alternate way of of mapping
emb=pd.get_dummies(tit.Embarked)
emb.drop('C',axis=1,inplace=True)
emb.head()
```

Out[62]:

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

In [63]:
```
tit1=pd.concat([tit,emb],axis=1)
tit1.drop(['Gender','Embarked'],axis=1,inplace=True)
tit1.head()
```

Out[63]:

	Survived	Pclass	Age	SibSp	Parch	Gen	Q	S
0	0	3	22.0	1	0	0	0	1
1	1	1	38.0	1	0	1	0	0
2	1	3	26.0	0	0	1	0	1
3	1	1	35.0	1	0	1	0	1
4	0	3	35.0	0	0	0	0	1

In [64]:
```
x=tit1.drop('Survived',axis=1)
y=tit1['Survived']
```

In [65]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

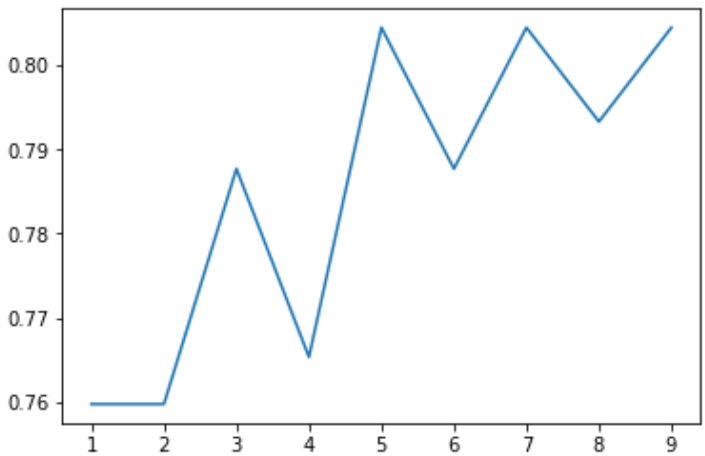
In [73]:
```
from sklearn.neighbors import KNeighborsClassifier
kmodel=KNeighborsClassifier(n_neighbors=8)
kmodel.fit(x_train,y_train)


y_pred=kmodel.predict(x_test)
y_pred
```

Out[73]:
```
array([0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1,
       0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1,
       0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0,
       0, 1, 0], dtype=int64)
```

In [74]:
```
kmodel.score(x_test,y_test)
```

Out[74]:
```
0.7932960893854749
```

In [75]:
```
import numpy as np
import matplotlib.pyplot as plt
score=[]
for i in range(1,10):
    km=KNeighborsClassifier(n_neighbors=i)
    km.fit(x_train,y_train)
    s=km.score(x_test,y_test)
    score.append(s)
score
plt.plot(np.arange(1,10),score)
plt.show()
```



In [81]:
```
from sklearn.model_selection import cross_val_score
cv=cross_val_score(kmodel,x_test,y_test,cv=5)
np.mean(cv)
```

Out[81]:
```
0.6707936507936508
```

In [85]:
```
kmodel.predict([[2,33,1,0,1,0,0]])
```

Out[85]:
```
array([1], dtype=int64)
```