

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

```
In [2]: dataset=pd.read_csv(r'C:\Users\siyad\AppData\Local\Temp\Temp1_Dataset-20200813T141334Z-001.zip\Dataset\train.csv' )
```

```
In [3]: dataset.columns
```

Out[3]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'], dtype='object')

```
In [4]: dataset['Age'].mean()
```

Out[4]: 32.69985376827896

```
In [5]: dataset.isnull().sum()
```

Out[5]: PassengerId 0
Survived 0
Pclass 0
Name 0
Sex 0
Age 0
SibSp 0
Parch 0
Ticket 0
Fare 0
Cabin 687
Embarked 0
dtype: int64

```
In [6]: from sklearn import preprocessing
```

```
In [7]: from sklearn.model_selection import train_test_split
```

```
In [8]: from sklearn.naive_bayes import GaussianNB
```

```
In [9]: from sklearn.metrics import accuracy_score
```

```
In [10]: from sklearn.metrics import confusion_matrix
```

```
In [11]: le=preprocessing.LabelEncoder()
```

```
In [12]: le.fit(dataset['Sex'])
```

Out[12]: LabelEncoder()

```
In [13]: print(le.classes_)
```

['female' 'male']

```
In [14]: dataset['Sex']=le.transform(dataset['Sex'])
```

```
In [20]: le.fit(dataset['Embarked'])
```

Out[20]: LabelEncoder()

```
In [21]: inverse_transform_dict = {}
for col, d in transform_dict.items():
    inverse_transform_dict[col] = {v:k for k, v in d.items()}

inverse_transform_dict
{'Embarked': {1: 'C', 2: 'Q', 3: 'S'}}
```

NameError Traceback (most recent call last)
<ipython-input-21-4554bcbe92cc> in <module>
1 inverse_transform_dict = {}
----> 2 for col, d in transform_dict.items():
3 inverse_transform_dict[col] = {v:k for k, v in d.items()}
4
5 inverse_transform_dict

NameError: name 'transform_dict' is not defined

```
In [19]: dataset.replace(transform_dict,inplace=True)
```

NameError Traceback (most recent call last)
<ipython-input-19-ac1185f1bac8> in <module>
----> 1 dataset.replace(transform_dict,inplace=True)

NameError: name 'transform_dict' is not defined

```
In [86]: dataset.head()
```

Out[86]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	0	0	2	108	1	28.0	1	0	522	18.0	NaN	2
1	1	1	0	190	0	51.0	1	0	595	207.0	80.0	0
2	2	1	2	353	0	34.0	0	0	668	41.0	NaN	2
3	3	1	0	272	0	47.0	1	0	48	189.0	54.0	2
4	4	0	2	15	1	47.0	0	0	471	43.0	NaN	2

```
In [87]: print(le.classes_)

['C' 'Q' 'S']

In [88]: y=dataset['Survived']

In [89]: X=dataset.drop(['Survived','PassengerId','Cabin','Ticket','Name'],axis=1)

In [90]: y.count()

Out[90]: 889

In [91]: X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=0)

In [92]: X_train.head()

Out[92]:
```

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
350	0	1	59.0	0	0	167.0	2
124	2	1	16.0	1	0	75.0	0
577	2	0	59.0	1	0	95.0	0
422	2	0	36.0	1	1	93.0	2
118	2	0	6.0	4	2	158.0	2

```
In [93]: X_train.isnull().sum()

Out[93]: Pclass      0
Sex            0
Age            0
SibSp          0
Parch          0
Fare           0
Embarked       0
dtype: int64

In [17]: from sklearn.naive_bayes import *

In [95]: clf=BernoulliNB()

In [96]: y_pred=clf.fit(X_train,y_train).predict(X_test)

In [97]: accuracy_score(y_test,y_pred,normalize=True)

Out[97]: 0.7677902621722846
```

In [98]:

confusion_matrix(y_test,y_pred)

Out[98]: array([[125, 32],
 [30, 80]], dtype=int64)

In []: