

```
import pandas as pd

train=pd.read_csv("train.csv")
test=pd.read_csv("test.csv")

train.head(10)
```

	PassengerId	Survived	Pclass	Name	Sex	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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	0	3	Johnson, Mrs. Oscar W (Elisabeth Lucille M. Briggs)	female	27.0	0	0	315166	14.4542	NaN	S

```
id_sub=test["PassengerId"]

train.drop(["Ticket", "Cabin", "Name", "PassengerId"],axis=1,inplace=True)
test.drop(["Ticket", "Cabin", "Name", "PassengerId"],axis=1,inplace=True)

import numpy as np
from sklearn.impute import SimpleImputer
imp=SimpleImputer(missing_values=np.nan,strategy='median')

imp.fit(train[["SibSp", "Parch", "Fare", "Age"]])
train[["SibSp", "Parch", "Fare", "Age"]]=imp.transform(train[["SibSp", "Parch", "Fare", "Age"]])
imp1=SimpleImputer(missing_values=np.nan,strategy='median')
imp1.fit(test[["SibSp", "Parch", "Fare", "Age"]])
test[["SibSp", "Parch", "Fare", "Age"]]=imp1.transform(test[["SibSp", "Parch", "Fare", "Age"]])

train.isnull().sum()

Survived    0
Pclass      0
Sex         0
Age         0
SibSp       0
Parch       0
Fare        0
Embarked    2
dtype: int64

test.isnull().sum()

Pclass      0
Sex         0
Age         0
SibSp       0
Parch       0
Fare        0
Embarked    0
dtype: int64

test
```

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked	
0	3	male	34.5	0.0	0.0	7.8292	Q	
1	3	female	47.0	1.0	0.0	7.0000	S	
2	2	male	62.0	0.0	0.0	9.6875	Q	
3	3	male	27.0	0.0	0.0	8.6625	S	
4	3	female	22.0	1.0	1.0	12.2875	S	
...	...	...	...	...	...	...	...	
413	3	male	27.0	0.0	0.0	8.0500	S	
414	1	female	39.0	0.0	0.0	108.9000	C	

```
train['Embarked'].fillna("S",inplace=True)

from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
cols=["Sex","Embarked"]
for col in cols:
    train[col]=le.fit_transform(train[col])
    test[col]=le.fit_transform(test[col])
    print(le.classes_)

['female' 'male']
['C' 'Q' 'S']

from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
y=train["Survived"]
x=train.drop("Survived",axis=1)
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,random_state=42)

regressor=LogisticRegression(random_state=0,max_iter=1000).fit(x_train,y_train)

pred=regressor.predict(x_test)
from sklearn.metrics import accuracy_score
accuracy_score(y_test,pred)

0.8100558659217877

sub=regressor.predict(test)

df=pd.DataFrame({"PassengerId":id_sub,
                 "Survived":sub})

df
```

	PassengerId	Survived	
0	892	0	
1	893	0	
2	894	0	
3	895	0	
4	896	1	
...	...	...	
413	1305	0	
414	1306	1	
415	1307	0	
416	1308	0	
417	1309	0	

418 rows × 2 columns

```
df.to_csv("submission.csv",index=False)
```

[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 11:46 AM

