

Ex 12

use titanic .csv and convert
it into dataframe.

AIM:

To use the titanic dataset and do the
data science process in it.

procedure:

Load the dataset.
display & explore column datatypes
Apply forward to fill missing values
Analyse the passenger features.
Split the dataset and observe
the output.

program:

```
import pandas as pd
```

```
data = pd.read_csv('Titanic')
```

```
data.shape
```

```
data.info()
```

```
data.describe()
```

```
from sklearn.impute import SimpleImputer
```

```
imp = SimpleImputer()
```

male - passenger = data[data['sex'] == 'male']
& data['fare'] > 100]

Male - passenger.

Emb - passenger = data[data['embarked']

Sib - passenger = data[data['sibsp'] > 2]

passenger not survived = data[(data
['sibsp'] == 0) & data['parch'] == 0]

0 (dest - passenger)

zero - fare - passenger = data[data['fare']
== 0]

zero - fare - passenger.

from sklearn.model_selection: import

train - test - split

x = data.drop('survived', axis=1)

y = data['survived']

x_train, y_train, x_test, y_test =

train - test - split(x, y, test_size=0.3)

print(x_train.shape, y_train.shape)

print(x_test.shape, y_test.shape)

১/১০

passenger

49

Subst used

0 1 2 3 4 5

Order columns:

Columb

Non-nally count

ofo

0 par-1d

89, non-weak

at 60,

Survived

89, Non-Null

int60

2008

891 non-new

100

3 Name

sci non-well

Object

100% 4

Object

1

0

70016

11.11.20

891 non-m

161

100

ni-van-168

16

10

89, non-va

10/10/10

24

Result

running values are filled, categorical features are encoded, & date split into training & testing sets for modeling.