

Output :

	Passenger	Survived	pclass
0	1	0	3
1	2	1	1
2	3	1	3
3	4	1	1
4	5	0	3

Data Columns (total 12 columns)

	Column	Non-null count	dtype
0	Passenger	Non-null	uint 64
1	Survived	891 non-null	uint 64
2	pclass	891 non-null	uint 64
3	Name	891 non-null	object

Memory Usage : 83.7 KB

Passenger : 0

Survived : 0

pclass : 0

Name : 0

Sex : 0

Age : 177

SibSp : 0

AIM:

To perform data discovery and exploratory analysis on real world dataset.

CODE:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.model_selection import
    train_test_split

df = pd.read_csv("titanic.csv")
print(df.head())
print(df.info())

sns.heatmap(df.isnull(), char=False,
            cmap='magma')

plt.title("Missing data visualisation")
plt.show()

X = df[['pclass', 'age', 'fare', 'sibsp', 'parch']]
Y = df['Survived']

X = [pd.get_dummies, drop_first=True]
```

Training Data Shape : (712, 3)

Testing Data shape : (179, 8)

by an oblique line.

9th Co. 1st Infantry Reg. Min.

but the original copy

transfer in 1 day. Below, provide more
help. Not sure.

(\cos sinusit) $\cos 20^\circ$ $\sin 70^\circ$

(1) less (2) twice

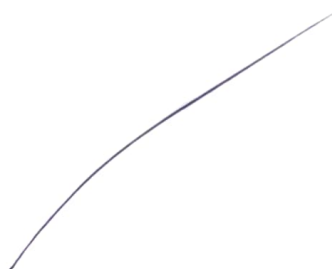
(c) after the time

the "Wine" club. "Glad!" I said. Yes.

$X_{\text{-train}}, X_{\text{-test}}, Y_{\text{-train}}, Y_{\text{-test}} = \text{train_test_}$

$\text{split}(X, Y, \text{test_size} = 0.3)$

$\text{print}(\text{" Training data : } X_{\text{-train}}, \text{shape} = 4)$



Result :

Training data : X_train, shape = 4

RESULT:

Thus, the program to perform data discovery and exploration has been completed successfully.