Output :

	Passanger	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	Passanger	Non-null	int 64
1	Survived	891 non-null	int 64
2	pclass	891 non-null	int 64
3	Name	891 non-null	object

Memory Usage : 83.7 KB

Passanger : 0
Survived : 0
pclass : 0
Name : 0
Sex : 0
Age : 177
Sibsh : 0

EX. NO : 2

Data Analytics Lifecycle

[Data Discovery and preparation]

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 = 'world')

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

X = df['Pclass', 'age', 'Fare', 'sisp', 'Patch'

Y = df ['Survived']

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

Training Data Shape : (712, 8)

Testing Data Shape : (179, 8)

```
X_train, x_test ; Y_train, Y_test = train_test_
                                    split (X, Y, test_size =
                                    0.3)
print (" Training data : x_ train, shape = 4)
```

RESULT:

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