21/8/25

Data Analytics Lifecycle

[Data discovery & preparation]

Aim:

To perform data discovery and exploratory analysis on world data set

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
df = pd.read_csv('titanic.csv')
print(df.head())
print(df.info())
print(df.isnull().sum())
print(df.describe())
sns.heatmap(df.isnull(), cbar = False)
plt.title('missing data')
plt.show()
```

		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)

#		non null	cost
0	Passenger	non-null	int64
1	Survived	non-null	int64
2	Pclass	non-null	int64
3	name	non-null	object

Passenger : 0

survived : 0

P class : 0

name : 0

sex : 0

Age : 177

Train data shape (712, 8)

Test . data shape (179, 8)

```
x = df ['Pclass', 'Age', 'Fare', 'Sharp', 'Pclass']
y = df ['Survived']
X = pd.get_dummies (x, drop_test = True)
x_tr, x_t, y_tr, y_t = train_txt (x, y)
```

Result: Thus the data discovery and exploratory analysis on titanic executed successfully