Exp no: load the titan dataset & convert it into a data fo

Aim:-

TO perform basic preprocessing & EDA & dataset using pas, dash & stat. tool

Procedure:

1). load the dataset into dataframe
2). Print and chat dolum-
3) handle missing valu
4). remove duplicat
5). nule catgoria doloum
6). standary numeric data
7). cite visualiyle

code:

```
import pandas as pd
import seaborn as ns
import matplotlib-pyplot- as plt
for realon preprocessing layut dalel encal, stat sch
df = sns. load_dataset("titanic")
Print(" first 3 rows :\n', df.head())
Print("\n missing valu :\n', df.isnl-sum())
df[age]- fitna (df [age], - median (), inplace = True)
df [adua], fitna ("num, inplut = True)
model = linen Regry (). fit (n_train , y_ train )
y_ pret= model predict(n_ test)
```

output:

Survived	Pclass	Age	Sex	Sib	Parch	Fare
0	3	M	22	1	7.22	3
1	1	F	29	0	7.86	C
1	3	F	26	0	7.32	S
0	1	M	35	1	7.54	S

who	adult_male	deck	embark_town
man	True	nan	Queens
women	False	C	Cherbourg
man	True	nan	Cork
women	False	2	early

```
Print ("MSE", mean_squared_error( y_test, y_pred)
sns.scatterplot (x= y_test , y_pred)
scale, standard scaler)
df[["age", "fare"]] = scale.fit_transform (df [["age", "fare"]])
sns.pairplot (df [["pclass", "fare", "age", "survived"]])
plt.show()
sns.heatmap (df [["pclass", "age", "sibsp", "parch", "fare"]].corr(),
        annot= True, cmap= "coolwarm")
plt.show()
```

Result)

Thus the data preprocessing EDA in Titanic dataset has been implemented successfully.