

ASSIGNMENT-3

NAME:M.Sarika

REGNO:21BCE9631

EMAIL:sarika.21bce9631@vitapstudent.ac.in

ASSIGNMENT-3 NAME:Boya Chaitanya REG NO:21BCE9968

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

```
da=pd.read_csv("Titanic-Dataset.csv")
da
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th.)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heldkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows x 12 columns

```
da.head()
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th.)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heldöinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
da.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523000	0.381594	32.204200
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454000
75%	668.500000	1.000000	3.000000	36.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.000000	0	0	112053	30.0000	842	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	29.699118	1	2	W/C 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.000000	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	170376	7.7500	NaN	Q

891 rows x 12 columns

```
da=da.drop("Cabin",axis=1)
```

```
da.isnull().sum()
```

```

PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age             0
SibSp           0
Parch           0
Ticket          0
Fare            0
Embarked        0
dtype: int64

```

```

In [10]: da.Sex.unique()
Out[10]:
array(['male', 'female'], dtype=object)

In [11]: da.Sex.value_counts()
Out[11]:
male      577
female    312
Name: Sex, dtype: int64

```

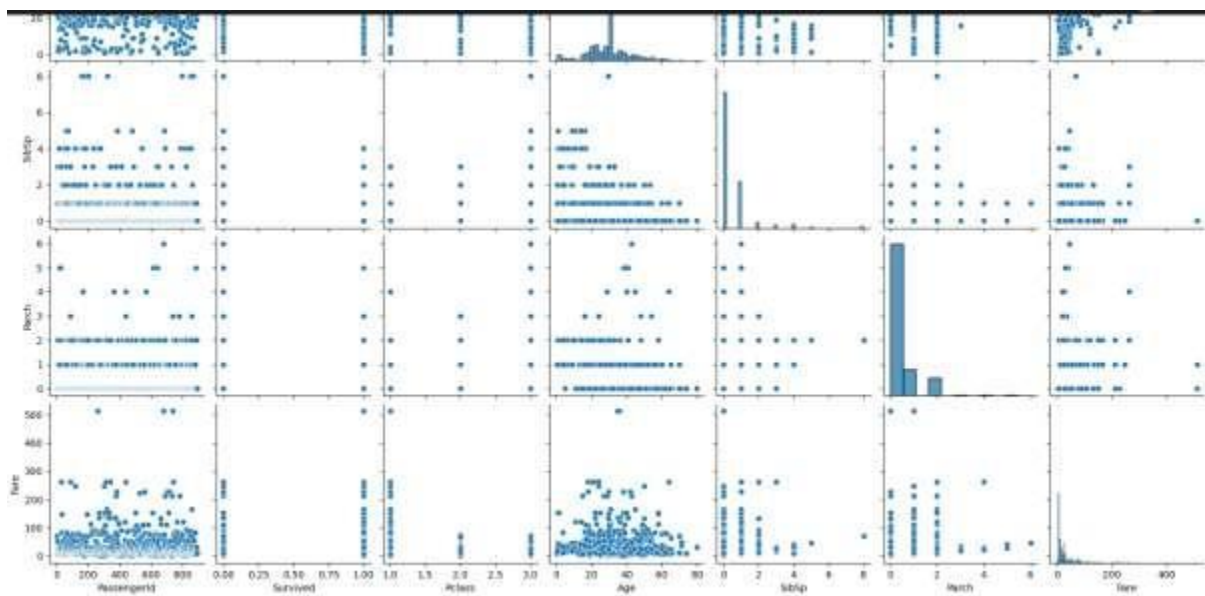
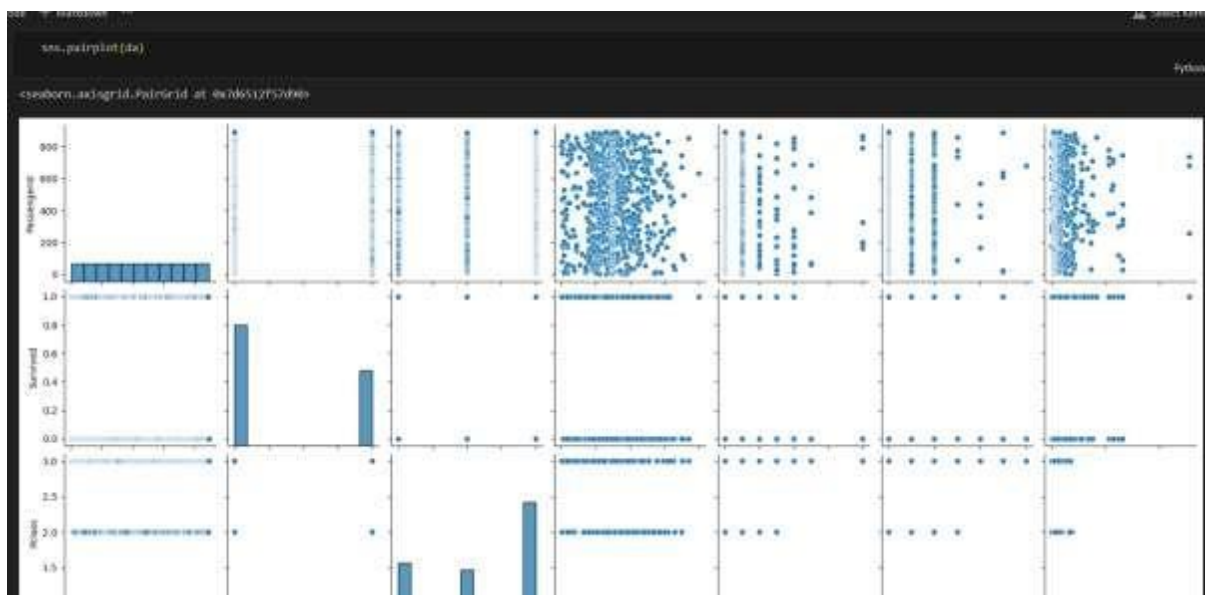
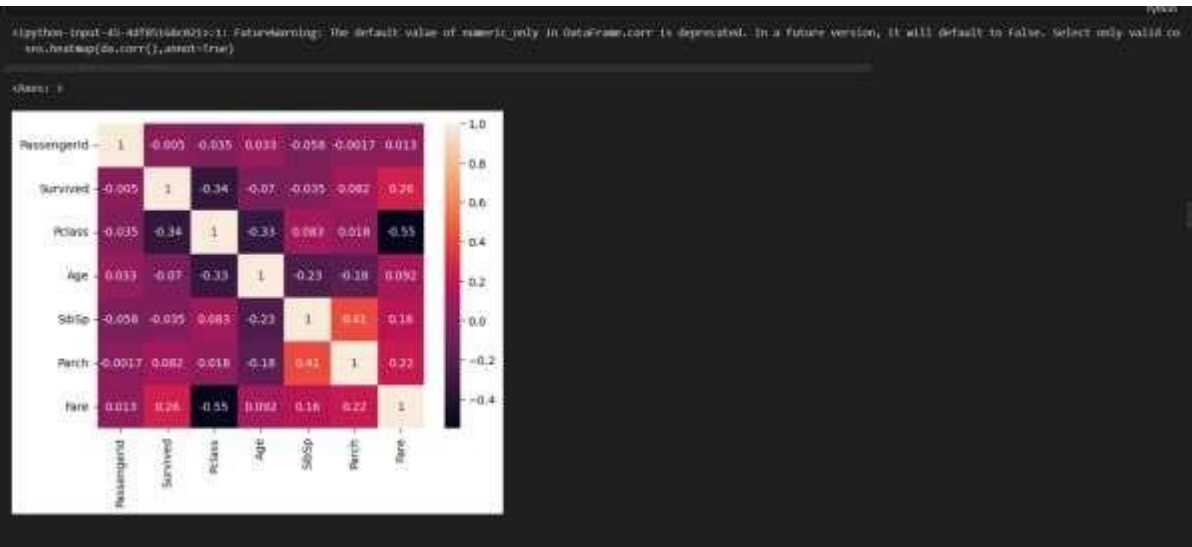
DATA VISUALIZATION

```

In [12]: plt.scatter(da["Fare"],da["Age"])
Out[12]:
<matplotlib.collections.PathCollection at 0x7805138f4c0>

```





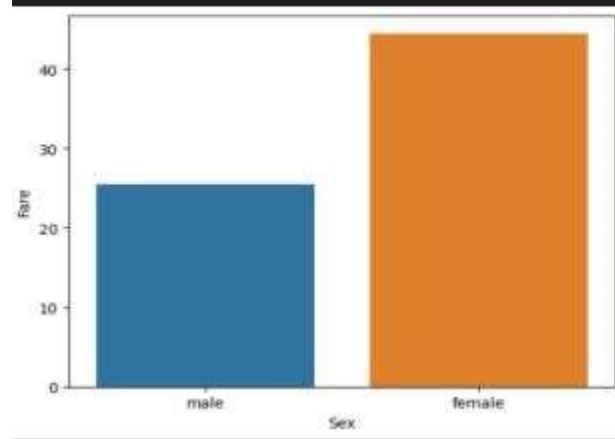
```

sns.barplot(x=da["Sex"],y=da["Fare"],ci=0)

ipython-input-47-fd5f5e31e7ed>1: FutureWarning:
the 'ci' parameter is deprecated. Use 'errorbar=(\'ci\', 0)' for the same effect.

sns.barplot(x=da["Sex"],y=da["Fare"],ci=0)
Axes: xlabel='Sex', ylabel='Fare'>

```



```

da.head()

```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	ParCh	Ticket	Fare	Embarked
0	1	0	3	Brown, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th.)	female	38.0	1	0	PC 17599	53.0000	C
2	3	1	3	Hedderley, Mrs. Laura	female	26.0	0	0	STON/O2. 3101282	7.0000	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.0000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0000	S

```

sns.boxplot(da["Pclass"])

```

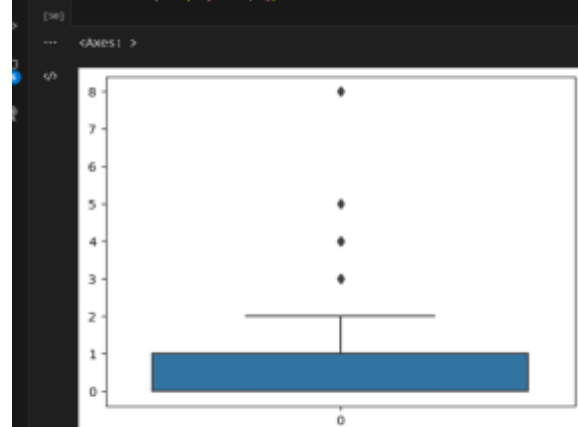


Code Markdown

```

sns.boxplot(da["SibSp"])

```



```

da.head()

```

[10]

```
z0.head()
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	
0	1	0	3	Randall, Mr. Owen Harry	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Heddiowas, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	17450	8.0500	S

```
x=df.drop(columns=["Name"],axis=1)  
x.head()
```

PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Embarked	Sex_female	Sex_male	...	Ticket_STON/O2. 3101290	Ticket_SW/PP 751	Ticket_W/C 14258	Ticket_W/C 14263	Ticket_W/C 6607	Ticket_W/C 6608	Ticket_W/C 6609
0	1	0	3	22.0	1	0	7.2500	S	0	1	-	0	0	0	0	0	0
1	2	1	1	38.0	1	0	71.2833	C	1	0	-	0	0	0	0	0	0
2	3	1	3	26.0	0	0	7.9250	S	1	0	-	0	0	0	0	0	0
3	4	1	1	35.0	1	0	53.1000	S	1	0	-	0	0	0	0	0	0
4	5	0	3	35.0	0	0	8.0500	S	0	1	-	0	0	0	0	0	0

rows = 891 columns

```
x.shape
```

Python						
Ticket_W/C. 14263	Ticket_W/C. 6607	Ticket_W/C. 6608	Ticket_W/C. 6609	Ticket_W.E.P. 5734	Ticket_W/C 14208	Ticket_WE/P 5735
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

```
type(x)
```

gandas.core.frame.DataFrame

```
y=df["fare"]  
y.head()
```

0	7.2500
1	71.2833
2	7.9250
3	53.1000
4	8.0500

Name: fare, dtype: float64

```
x.head()
```

PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Embarked	Sex_female	Sex_male	...	Ticket_STON/O2. 3101290	Ticket_SW/PP 751	Ticket_W/C 14258	Ticket_W/C 14263	Ticket_W/C 6607	Ticket_W/C 6608	Ticket_W/C 6609
0	1	0	3	22.0	1	0	7.2500	S	0	1	-	0	0	0	0	0	0
1	2	1	1	38.0	1	0	71.2833	C	1	0	-	0	0	0	0	0	0
2	3	1	3	26.0	0	0	7.9250	S	1	0	-	0	0	0	0	0	0
3	4	1	1	35.0	1	0	53.1000	S	1	0	-	0	0	0	0	0	0
4	5	0	3	35.0	0	0	8.0500	S	0	1	-	0	0	0	0	0	0

rows = 891 columns


```
da.drop("Name",axis=1)

Python

PassengerId  Survived  Pclass    Age  SibSp  Parch    Fare  Embarked  Sex_female  Sex_male  ...  Ticket  STON/OZ.  Ticket  SW/PP  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.
0           1         0       3  22.0000    1     0   7.2500      S         0         1  ...      0         0         0         0         0         0
1           2         1       1  38.0000    1     0  71.2833      C         1         0  ...      0         0         0         0         0         0
2           3         1       3  26.0000    0     0   7.9250      S         1         0  ...      0         0         0         0         0         0
3           4         1       1  35.0000    1     0  53.0000      S         1         0  ...      0         0         0         0         0         0
4           5         0       3  35.0000    0     0  80.5000      S         0         1  ...      0         0         0         0         0         0
...         ...         ...     ...     ...     ...     ...     ...     ...     ...     ...  ...      ...         ...         ...         ...         ...         ...
886          887         0       2  27.0000    0     0  13.0000      S         0         1  ...      0         0         0         0         0         0
887          888         1       1  19.0000    0     0  30.0000      S         1         0  ...      0         0         0         0         0         0
888          889         0       3  29.699118  1     2  23.4500      S         1         0  ...      0         0         0         0         1         0
889          890         1       1  26.0000    0     0  30.0000      C         0         1  ...      0         0         0         0         0         0
890          891         0       3  32.0000    0     0   7.5000      Q         0         1  ...      0         0         0         0         0         0
891 rows x 691 columns

x_scaled=pd.DataFrame(mo_fit.transform(x),columns=x.columns)
x_scaled.head()

Python

PassengerId  Survived  Pclass    Age  SibSp  Parch    Fare  Embarked  Sex_female  Sex_male  ...  Ticket  STON/OZ.  Ticket  SW/PP  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.
0      0.000000         0.0       1.0  0.271174  0.125     0.0  0.018151      1.0         0.0         1.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
1      0.001134         1.0       0.0  0.472229  0.125     0.0  0.139136      0.0         1.0         0.0  ...      0.0         0.0         0.0         0.0         0.0         0.0

x_scaled=pd.DataFrame(mo_fit.transform(x),columns=x.columns)
x_scaled.head()

Python

PassengerId  Survived  Pclass    Age  SibSp  Parch    Fare  Embarked  Sex_female  Sex_male  ...  Ticket  STON/OZ.  Ticket  SW/PP  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.  Ticket  W/C.
0      0.000000         0.0       1.0  0.271174  0.125     0.0  0.018151      1.0         0.0         1.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
1      0.001134         1.0       0.0  0.472229  0.125     0.0  0.139136      0.0         1.0         0.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
2      0.002227         1.0       1.0  0.321408  0.000     0.0  0.015460      1.0         1.0         0.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
3      0.003379         1.0       0.0  0.434637  0.125     0.0  0.103666      1.0         1.0         0.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
4      0.004494         0.0       1.0  0.434637  0.000     0.0  0.015379      1.0         0.0         1.0  ...      0.0         0.0         0.0         0.0         0.0         0.0
5 rows x 691 columns

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_scaled,y,test_size=0.2,random_state=4)

Python

print(x_train.shape,x_test.shape,y_train.shape,y_test.shape)

Python

(712, 691) (179, 691) (712,) (179,)
```

et_W/C.	Ticket_W/C.	Ticket_W/C.	Ticket_W/C.	Ticket_W.E.P.	Ticket_W/C	Ticket_WE/P
14263	6607	6608	6609	5734	14208	5735
0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0