

task 2.ipynb

File Edit View Insert Runtime Tools Help

Commands + Code + Text

45s from google.colab import files  
uploaded = files.upload()

Choose Files cleaned\_titanic.csv  
• cleaned\_titanic.csv(text/csv) - 89081 bytes, last modified: 5/26/2025 - 100% done  
Saving cleaned\_titanic.csv to cleaned\_titanic.csv

[3] df = pd.read\_csv('cleaned\_titanic.csv')  
df.head()

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked_Q	Embarked_S
0	1	0	3	Braund, Mr. Owen Harris	0	-0.565736	1	0	A/5 21171	-0.502445	False	True
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Th...	1	0.663861	1	0	PC 17599	0.786845	False	False
2	3	1	3	Heikkinen, Miss. Laina	1	-0.258337	0	0	STON/O2. 3101282	-0.488854	False	True
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	0.433312	1	0	113803	0.420730	False	True
4	5	0	3	Allen, Mr. William Henry	0	0.433312	0	0	373450	-0.486337	False	True

Activate Windows  
Go to PC settings to activate Windows

Variables Terminal Python ✓ 12:12 PM

Commands + Code + Text

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

# Optional: interactive plots
import plotly.express as px

# Load cleaned dataset
df = pd.read_csv('cleaned_titanic.csv')
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked_Q	Embarked_S
0	1	0	3	Braund, Mr. Owen Harris	0	-0.565736	1	0	A/5 21171	-0.502445	False	True
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... ...	1	0.663861	1	0	PC 17599	0.786845	False	False
2	3	1	3	Heikkinen, Miss. Laina	1	-0.258337	0	0	STON/O2 3101282	-0.488854	False	True
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	0.433312	1	0	113803	0.420730	False	True
4	5	0	3	Allen, Mr. William Henry	0	0.433312	0	0	373450	-0.486337	False	True

Activate Windows

Go to PC settings to activate Windows

✓ 12:12 PM

Py

Next steps: Generate code with df

View recommended plots

New interactive sheet



Commands + Code + Text

✓ RAM Disk

```
[6] # Summary of numerical features
df.describe()
```

```
# Summary of all columns (including object types)
df.describe(include='all')
```

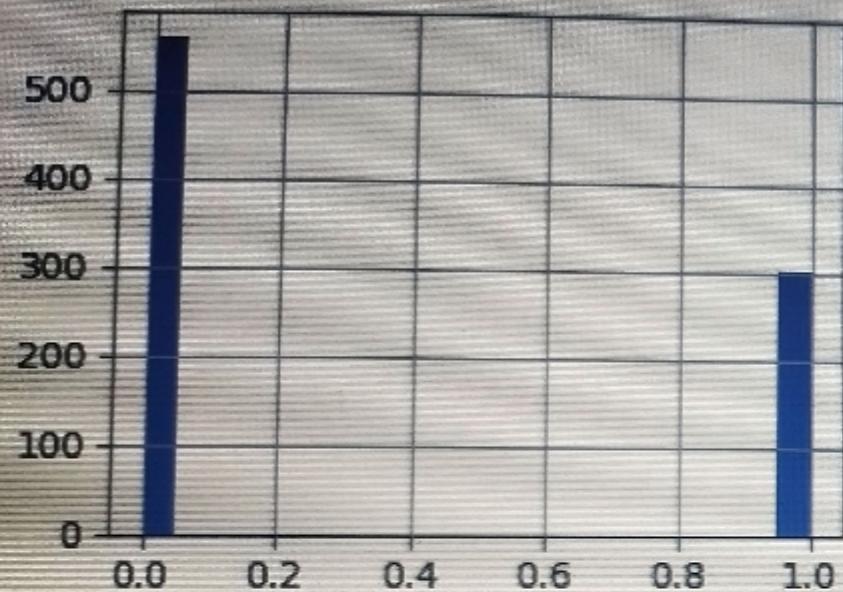
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked_Q	Embarked_S
count	864.000000	864.000000	864.000000	864	864.000000	864.000000	864.000000	864.000000	864	864.000000	864	864
unique	Nan	Nan	Nan	864	Nan	Nan	Nan	Nan	667	Nan	2	2
top	Nan	Nan	Nan	Abelson, Mrs. Samuel (Hannah Wizosky)	Nan	Nan	Nan	Nan	1601	Nan	False	True
freq	Nan	Nan	Nan	1	Nan	Nan	Nan	Nan	7	Nan	788	634
mean	444.748843	0.378472	2.343750	Nan	0.349537	-0.029418	0.520833	0.368056	Nan	-0.114839	Nan	Nan
std	257.517259	0.485287	0.819028	Nan	0.477100	0.960528	1.104937	0.794651	Nan	0.591964	Nan	Nan
min	1.000000	0.000000	1.000000	Nan	0.000000	-2.224156	0.000000	0.000000	Nan	-0.648422	Nan	Nan
25%	221.750000	0.000000	2.000000	Nan	0.000000	-0.565736	0.000000	0.000000	Nan	-0.489442	Nan	Nan
50%	444.500000	0.000000	3.000000	Nan	0.000000	-0.104637	0.000000	0.000000	Nan	-0.369347	Nan	Nan
75%	664.250000	1.000000	3.000000	Nan	1.000000	0.433312	1.000000	0.000000	Nan	-0.048911	Nan	Nan
max	891.000000	1.000000	3.000000	Nan	1.000000	2.815657	8.000000	6.000000	Nan	2.671118	Activate Windows Copy settings to activate Win	Nan

✓ 12:12 PM

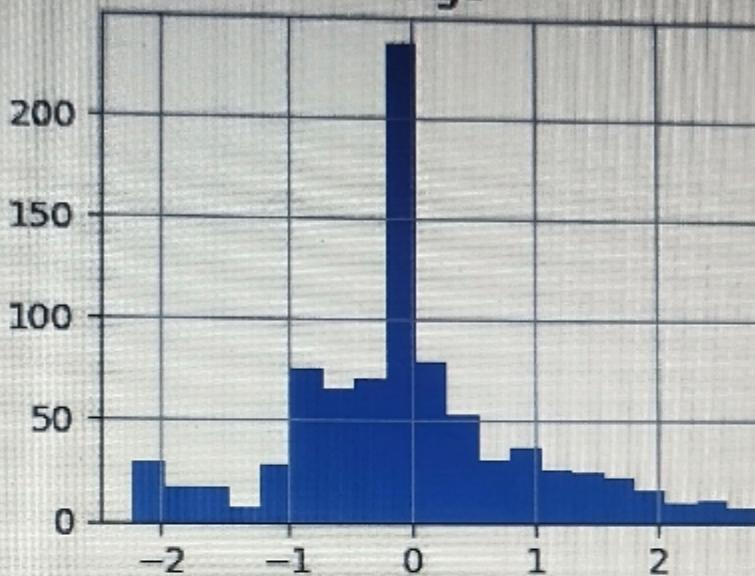
Py

Jobs + Code + Text

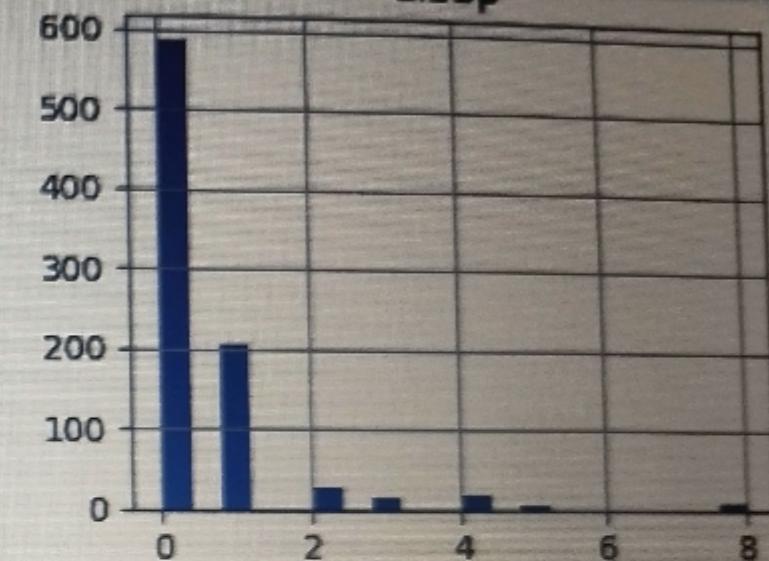
Sex



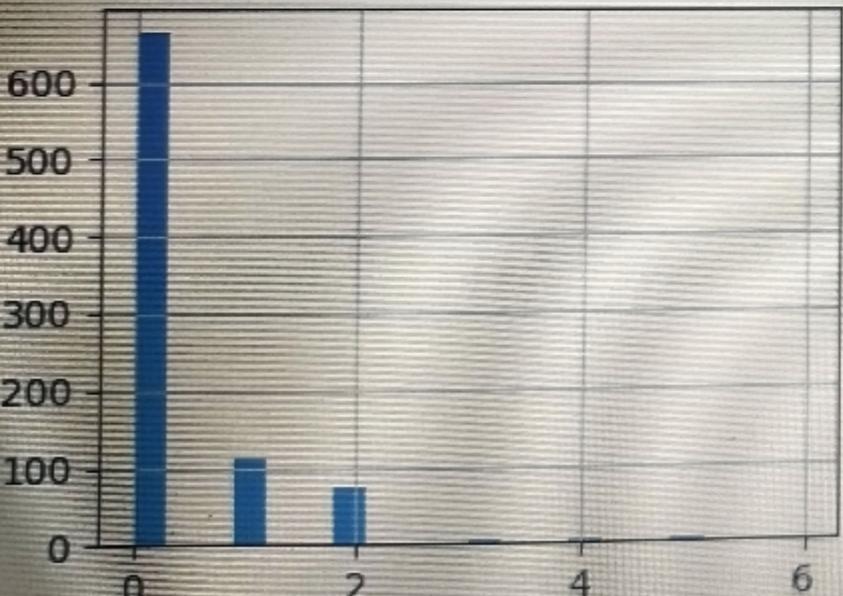
Age



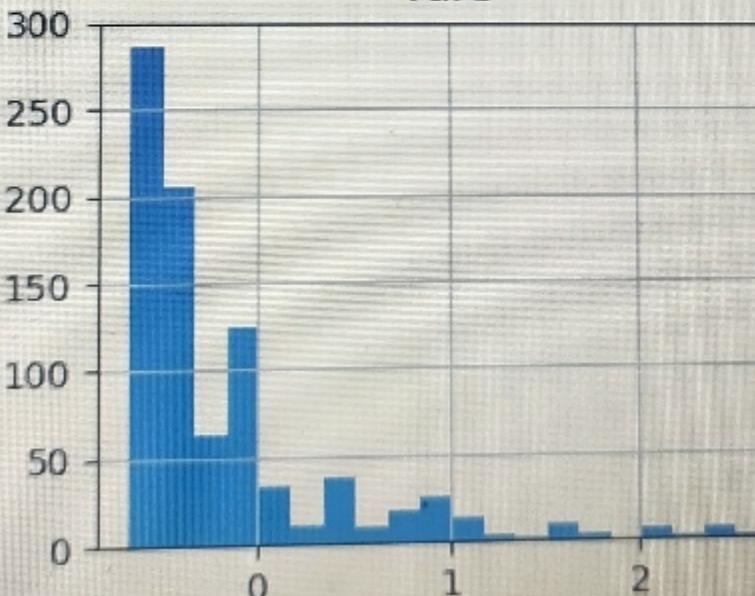
SibSp



Parch

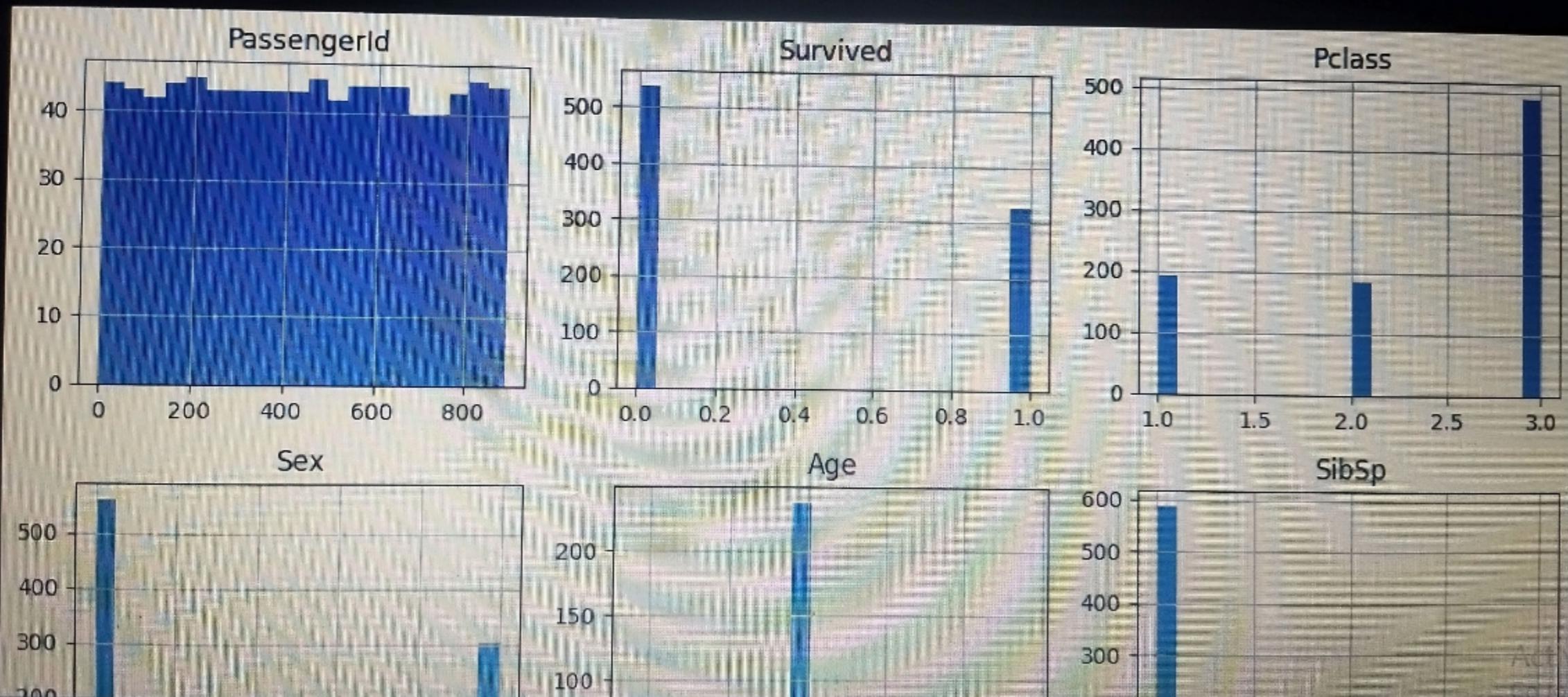


Fare



Commands + Code + Text

2s [7] df.hist(figsize=(10, 8), bins=20)  
plt.tight\_layout()  
plt.show()

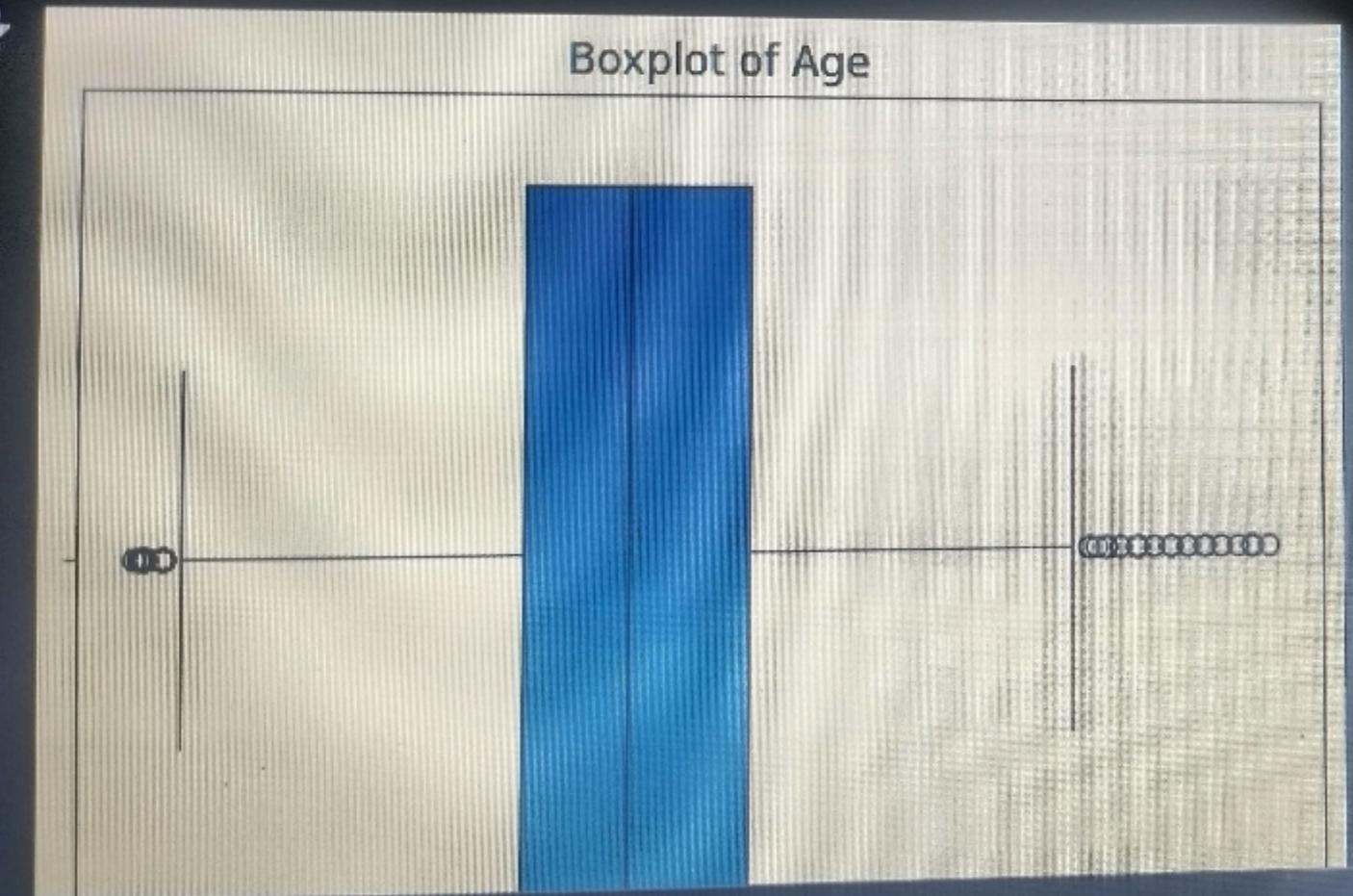
Activate Window  
Change PC settings



Commands

+ Code + Text

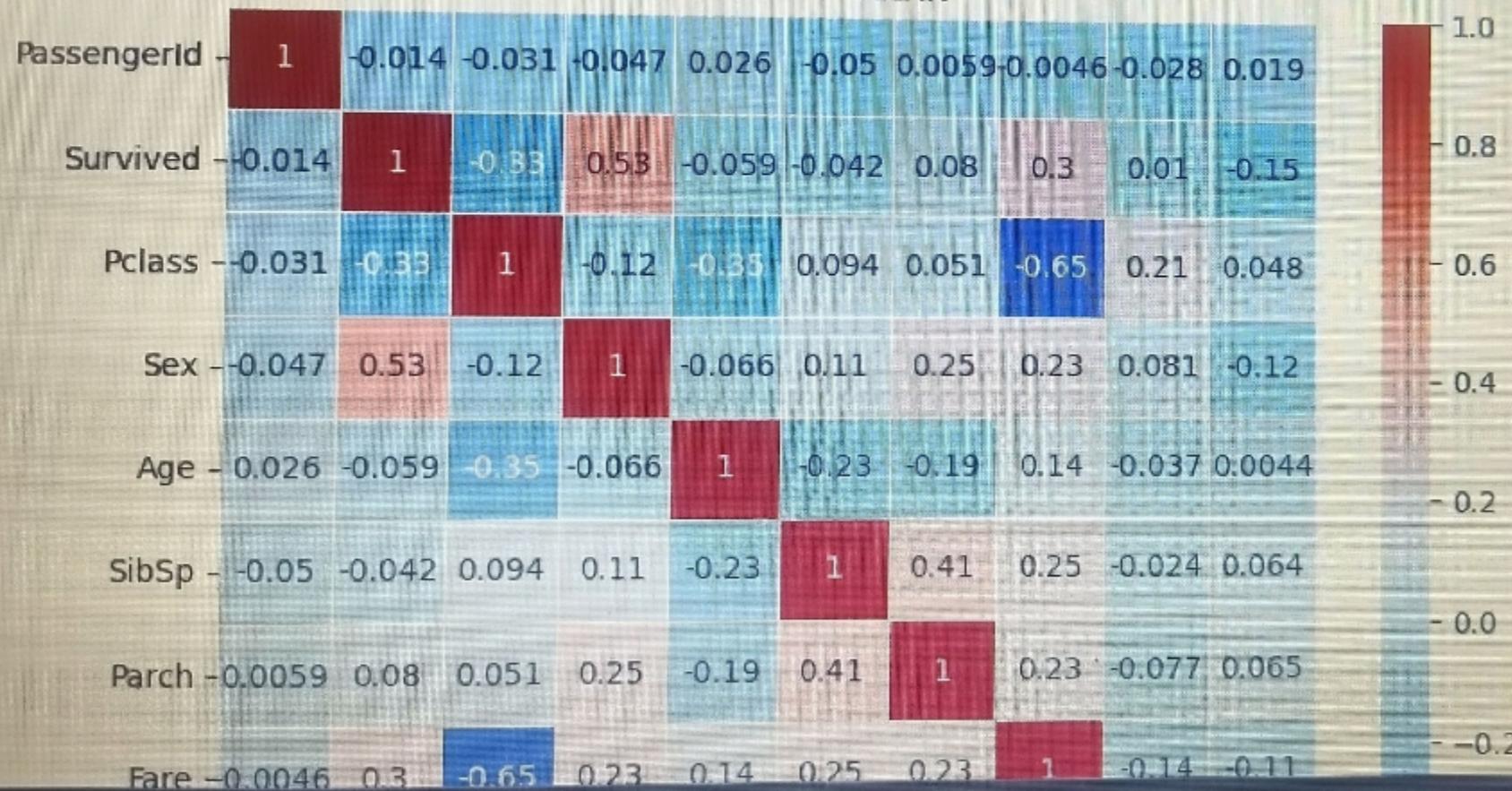
```
[8] sns.boxplot(x=df['Age'])  
plt.title('Boxplot of Age')  
plt.show()  
  
sns.boxplot(x=df['Fare'])  
plt.title('Boxplot of Fare')  
plt.show()
```

Acti  
Go to

Commands + Code + Text

```
corr_matrix = df.corr(numeric_only=True)
plt.figure(figsize=(8,6))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', linewidths=0.5)
plt.title("Correlation Matrix")
plt.show()
```

Correlation Matrix

Activate W  
Go to PC setti

-2

-1

0

1

2

3

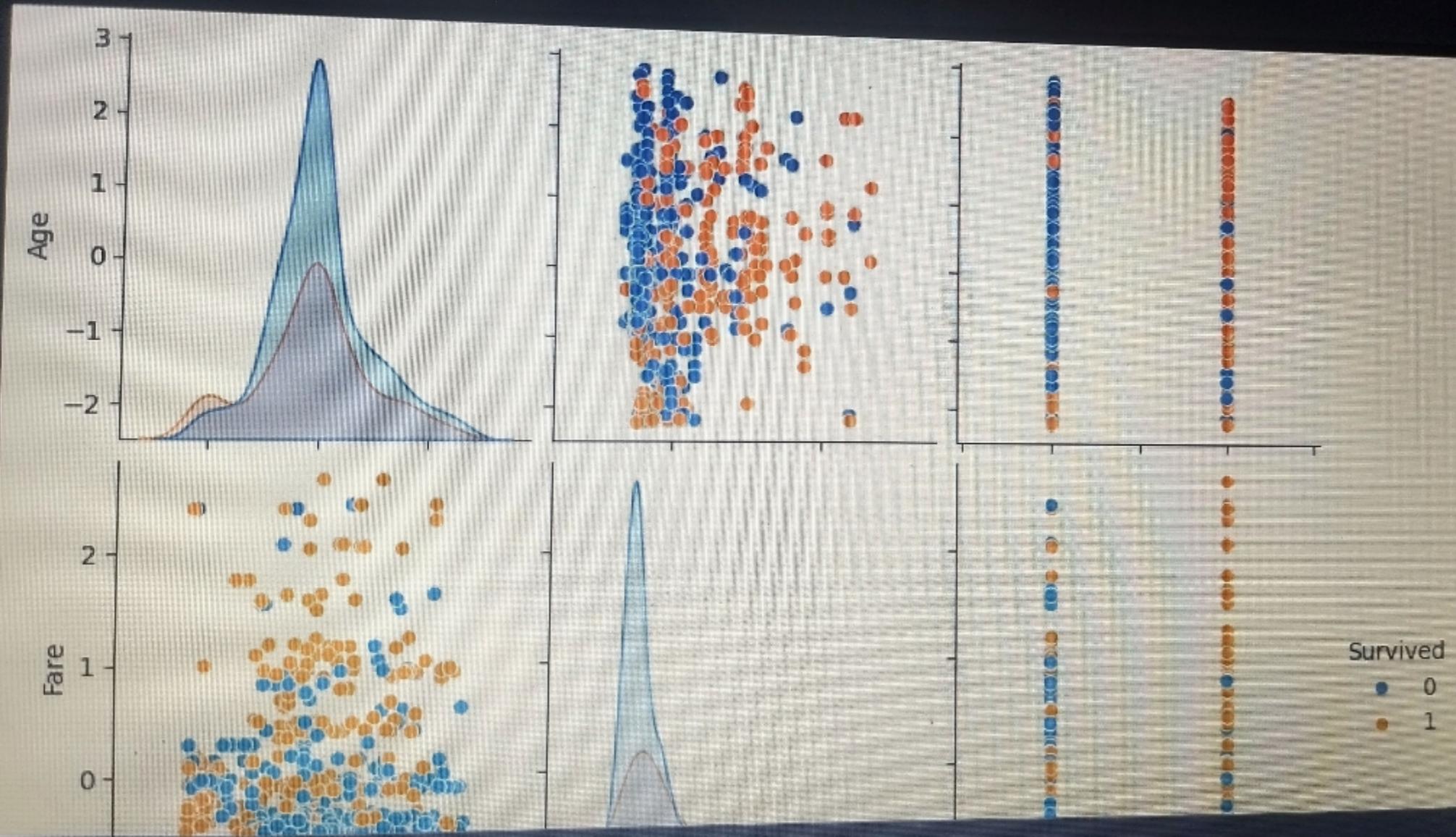
Age

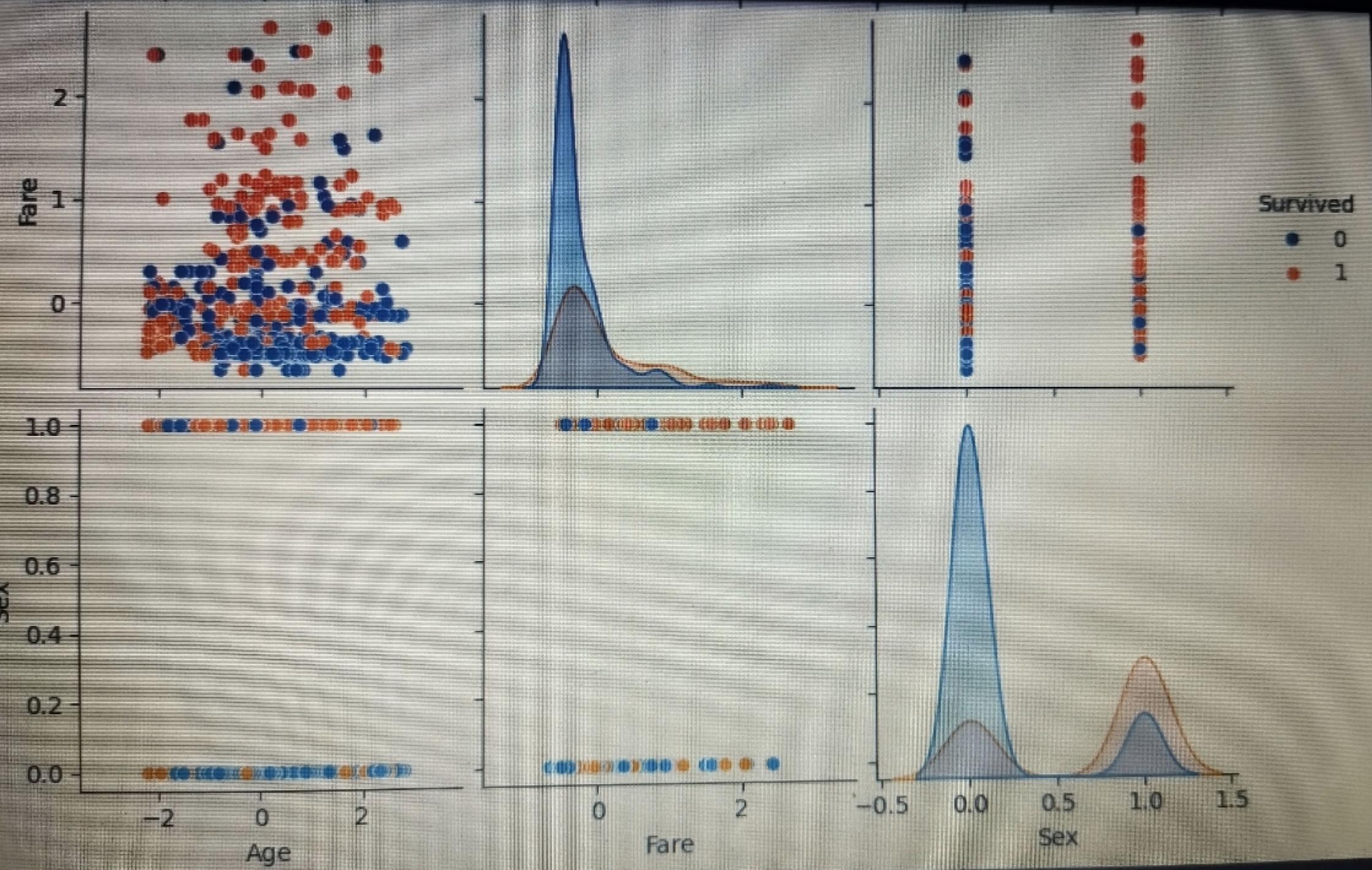
Boxplot of Fare





```
2s 2s
sns.pairplot(df[['Age', 'Fare', 'Sex', 'Survived']], hue='Survived')
plt.show()
```

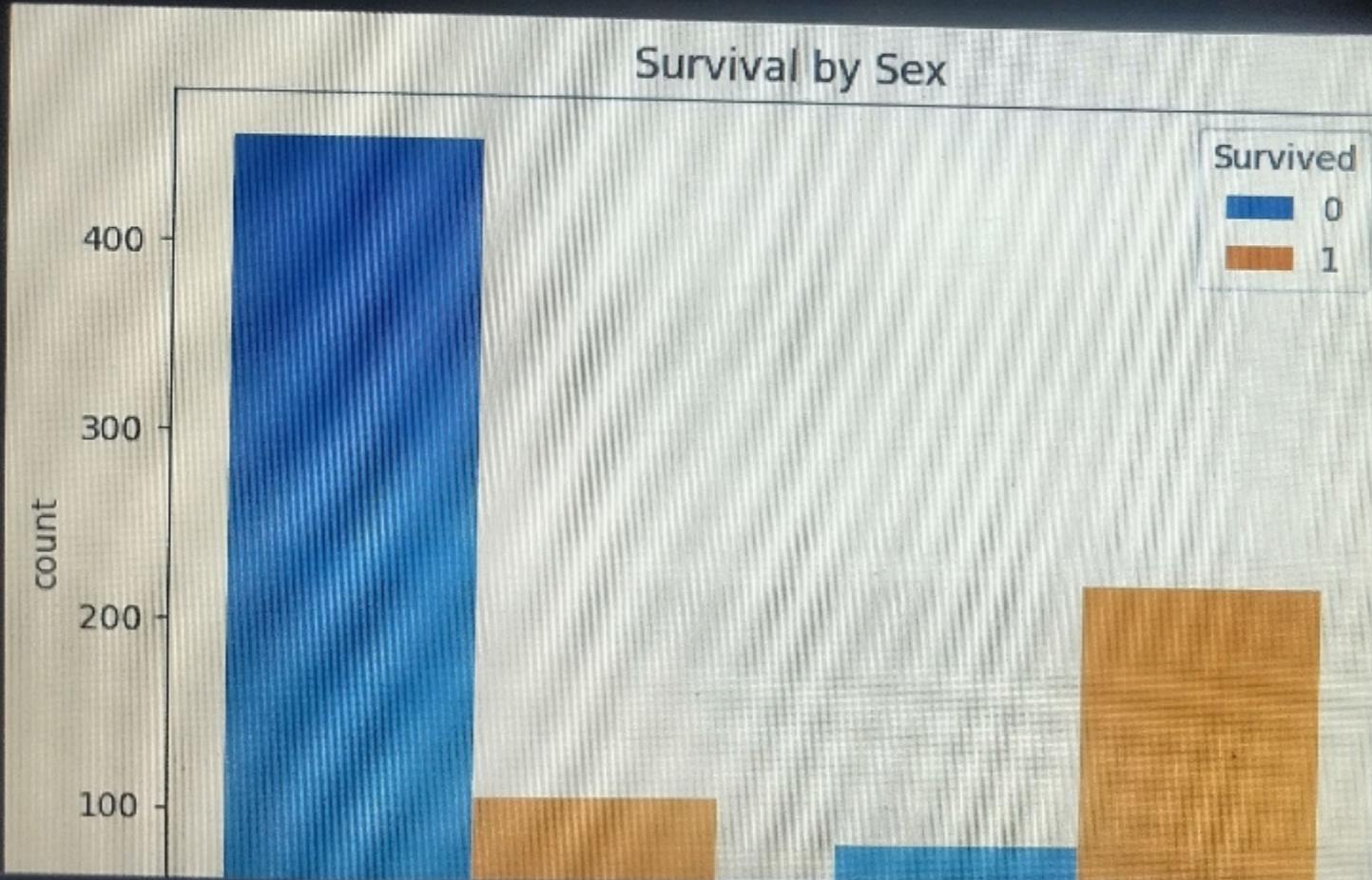


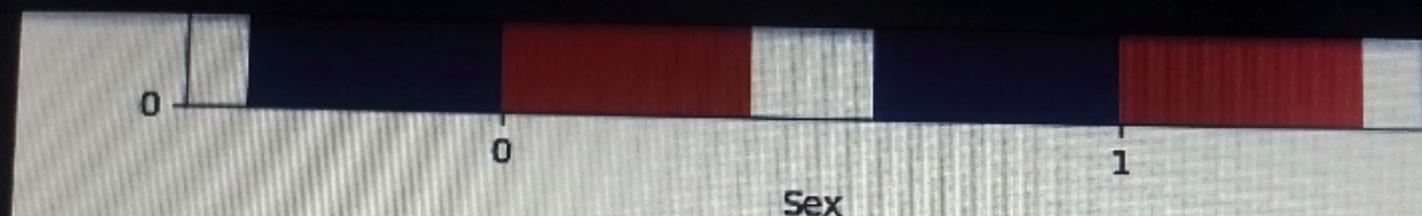


Commands + Code + Text

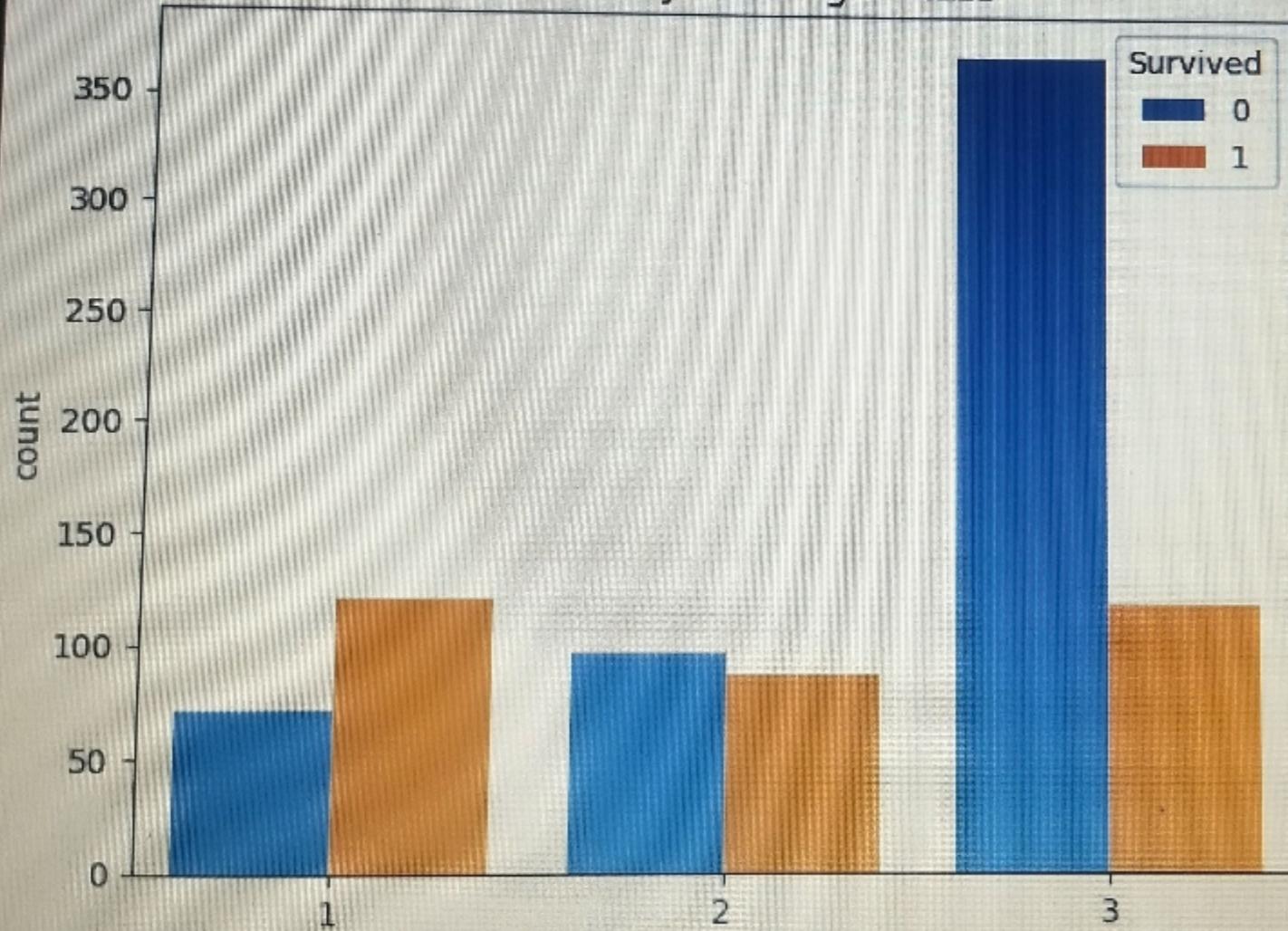
```
[11] sns.countplot(x='Sex', hue='Survived', data=df)
      plt.title("Survival by Sex")
      plt.show()
```

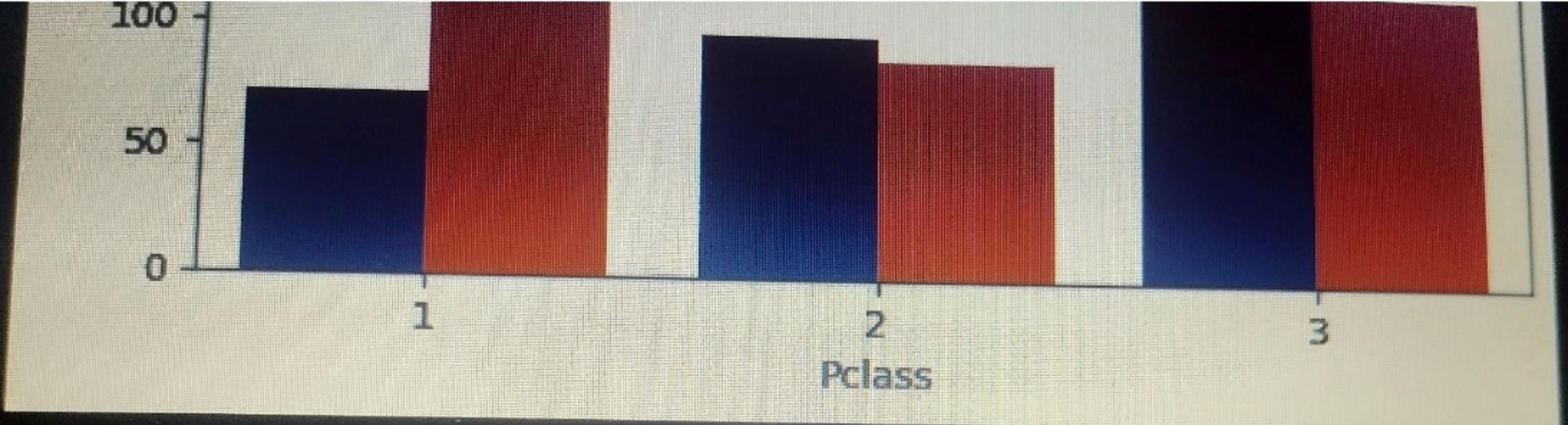
```
sns.countplot(x='Pclass', hue='Survived', data=df)
plt.title("Survival by Passenger Class")
plt.show()
```





Survival by Passenger Class





```
[12] print("Skewness of Fare:", df['Fare'].skew())
     print("Skewness of Age:", df['Age'].skew())
```

→ Skewness of Fare: 2.2571777277456295  
Skewness of Age: 0.32109598624805535