```
pip install pandas numpy matplotlib seaborn statsmodels scikit-learn
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)
Requirement already satisfied: numpy in /usr/local/lib/python3.11/dist-packages (2.0.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)
Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.13.2)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.11/dist-packages (0.14.5)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas) (2025.2)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (4.59.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (1.4.8)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (11.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.3)
Requirement already satisfied: scipy!=1.9.2,>=1.8 in /usr/local/lib/python3.11/dist-packages (from statsmodels) (1.16.1)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.11/dist-packages (from statsmodels) (1.0.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.5.1)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

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

```
df = pd.read_csv("/content/drive/MyDrive/ELEVATELABS_TASKS/Titanic-Dataset.csv")
df
```

	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	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, 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

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
sns.set(style="whitegrid")
%matplotlib inline
```

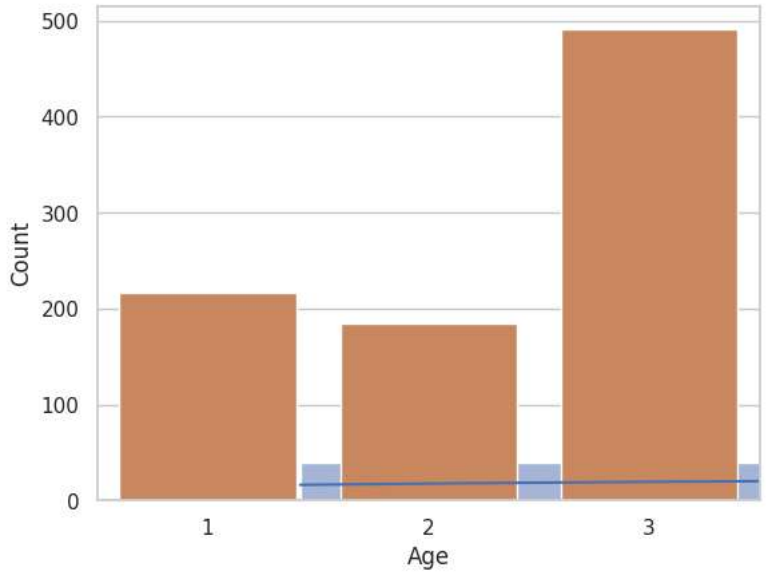
```
df.isnull().sum()
```

	0
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

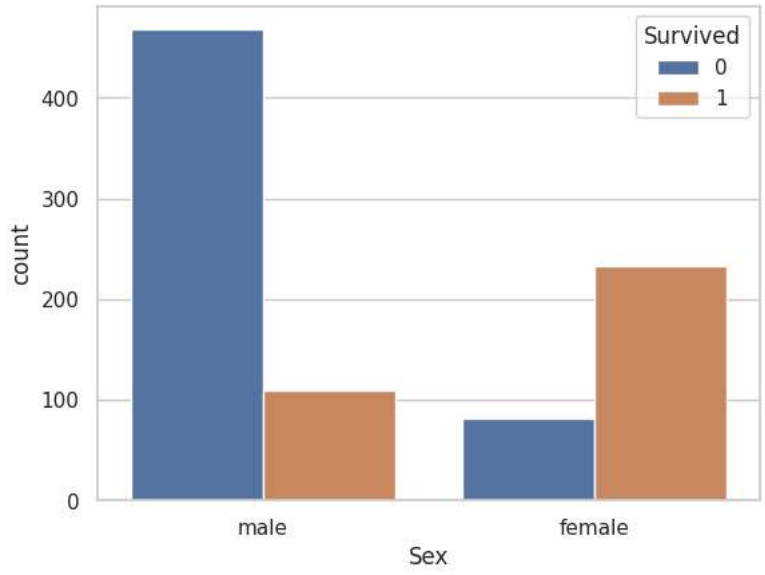
```
sns.histplot(df['Age'], kde=True)
sns.countplot(x='Pclass', data=df)
```

```
<Axes: xlabel='Age', ylabel='Count'>
```



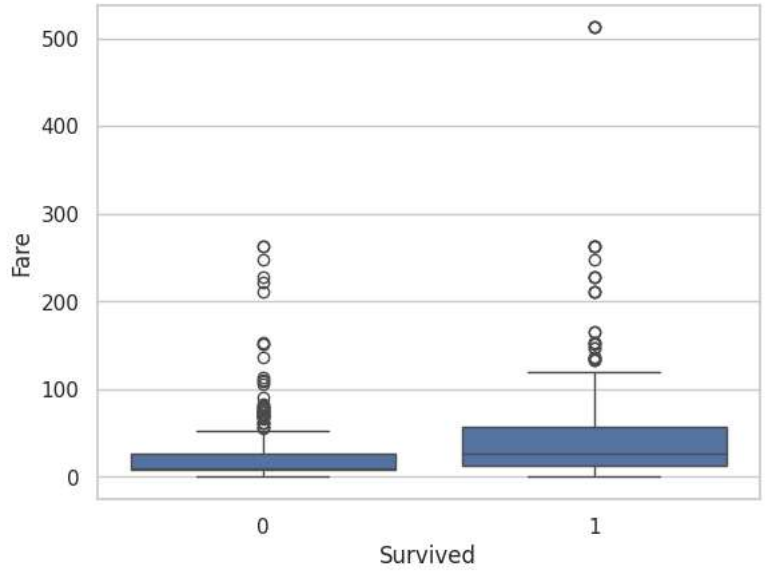

```
sns.countplot(x='Sex', hue='Survived', data=df)
```

```
<Axes: xlabel='Sex', ylabel='count'>
```



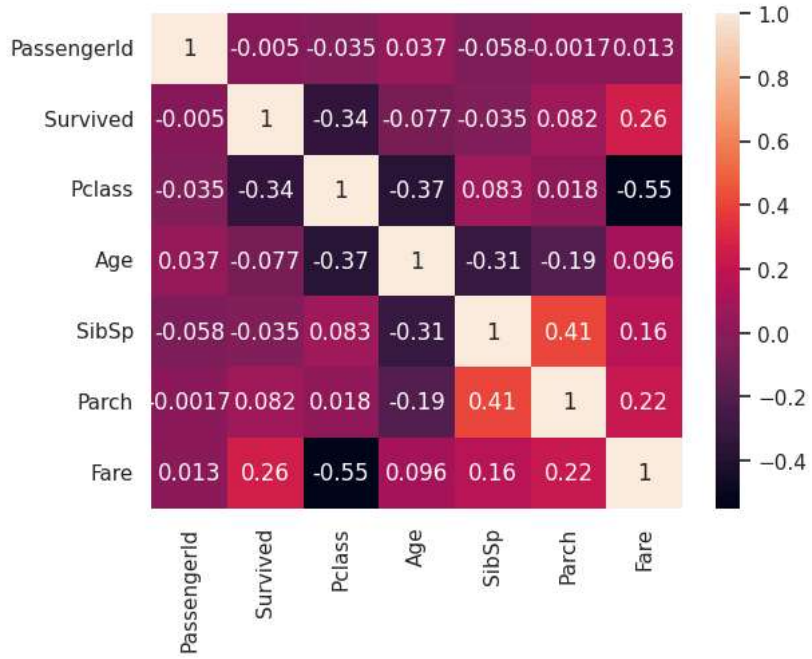

```
sns.boxplot(x='Survived', y='Fare', data=df)
```

```
<Axes: xlabel='Survived', ylabel='Fare'>
```




```
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=True)
```

<Axes: >




```
from statsmodels.stats.outliers_influence import variance_inflation_factor

# Create 'FamilySize' column
df['FamilySize'] = df['SibSp'] + df['Parch']

# Fill missing Age values with the mean
df['Age'].fillna(df['Age'].mean(), inplace=True)

X = df[['Age', 'Fare', 'FamilySize', 'Pclass']]
pd.DataFrame({
    'Feature': X.columns,
    'VIF': [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]
})
```

```
/tmp/ipython-input-4190858814.py:7: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assigr
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are setting valu

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].me

  df['Age'].fillna(df['Age'].mean(), inplace=True)
```

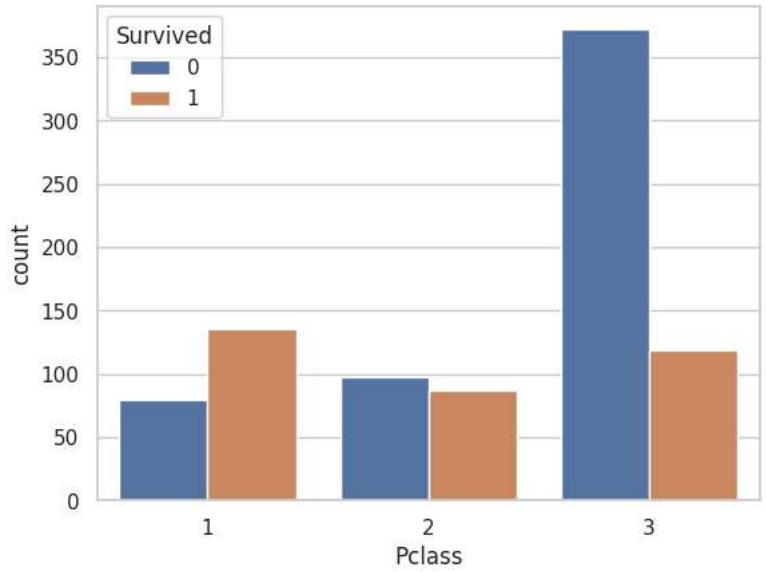
	Feature	VIF
0	Age	3.943261
1	Fare	1.660411
2	FamilySize	1.519252
3	Pclass	3.704240

```
import os

if not os.path.exists('images'):
    os.makedirs('images')

plt.savefig("images/plot_name.png", dpi=150, bbox_inches="tight")
```

<Figure size 640x480 with 0 Axes>

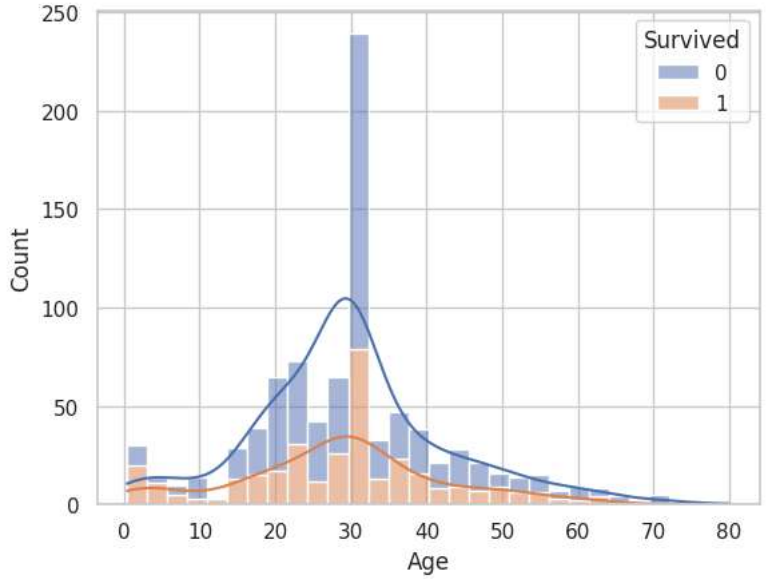
```
sns.countplot(x='Pclass', hue='Survived', data=df)
```

```
<Axes: xlabel='Pclass', ylabel='count'>
```



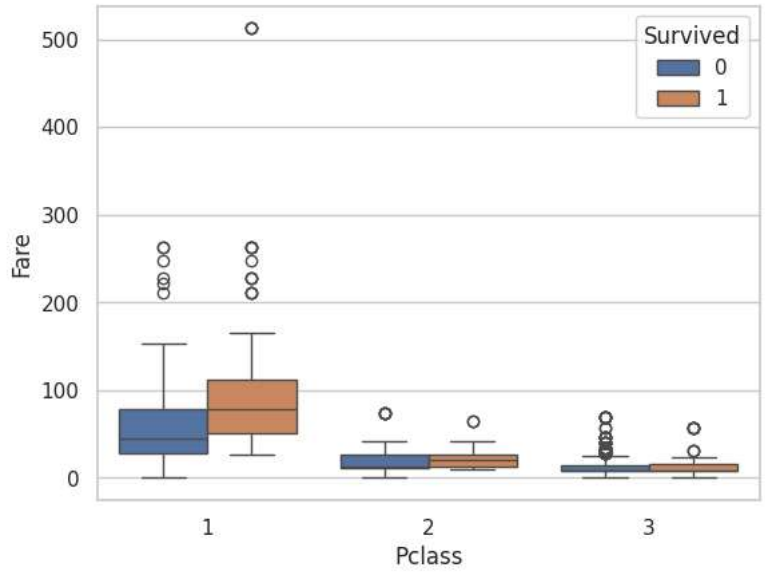

```
sns.histplot(data=df, x='Age', hue='Survived', kde=True, multiple='stack')
```

```
<Axes: xlabel='Age', ylabel='Count'>
```



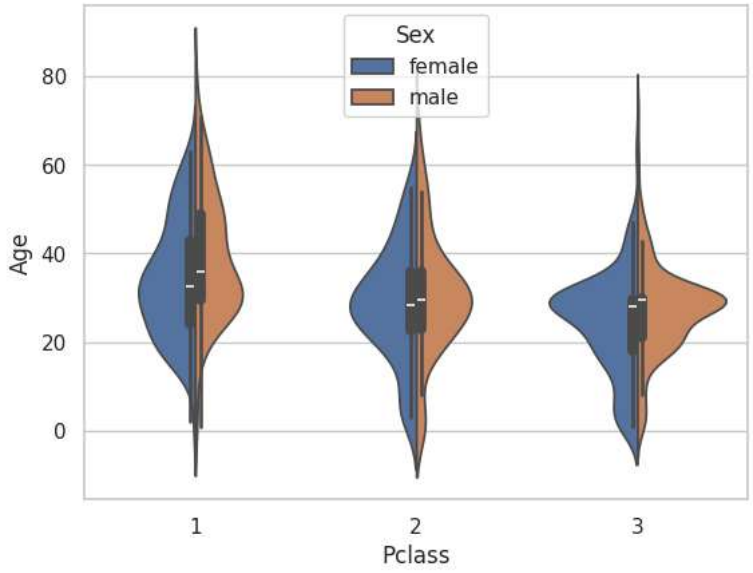

```
sns.boxplot(x='Pclass', y='Fare', hue='Survived', data=df)
```

```
<Axes: xlabel='Pclass', ylabel='Fare'>
```



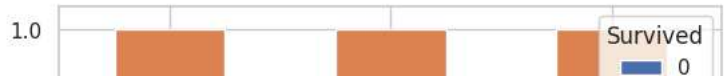

```
sns.violinplot(x='Pclass', y='Age', hue='Sex', split=True, data=df)
```

```
<Axes: xlabel='Pclass', ylabel='Age'>
```



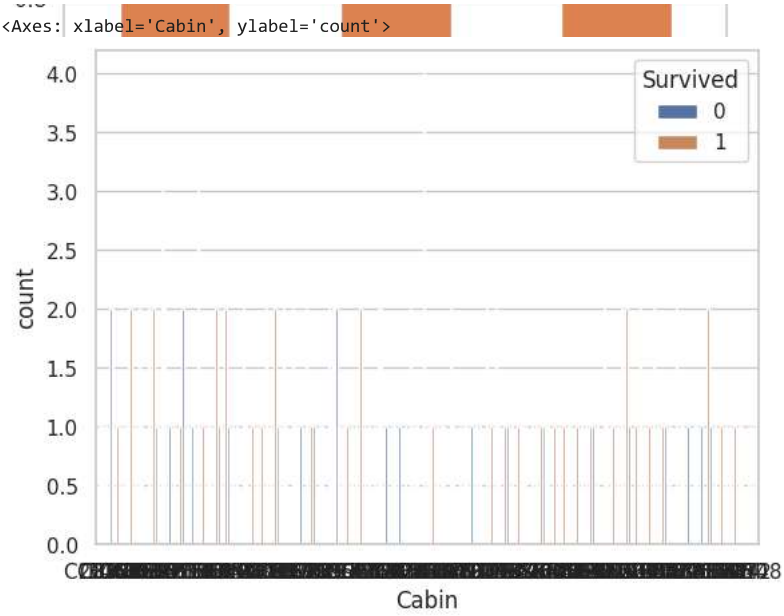

```
(pd.crosstab(df['Embarked'], df['Survived'], normalize='index')
.plot(kind='bar', stacked=True))
```

```
<Axes: xlabel='Embarked'>
```




```
sns.countplot(x='Cabin', hue='Survived', data=df)
```

```
<Axes: xlabel='Cabin', ylabel='count'>
```




```
g = sns.FacetGrid(df, col="Sex", row="Pclass", hue="Survived")
g.map_dataframe(sns.histplot, x="Age")
g.add_legend()
```

```
<seaborn.axisgrid.FacetGrid at 0x786f64df1610>
```