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

print("Setup OK ✅")
```

```
Setup OK ✅
```

```
from google.colab import files
files.upload()  # choose train.csv
```

Show hidden output

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

# Load dataset
df = pd.read_csv("train.csv")

# Display first 5 rows
df.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	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

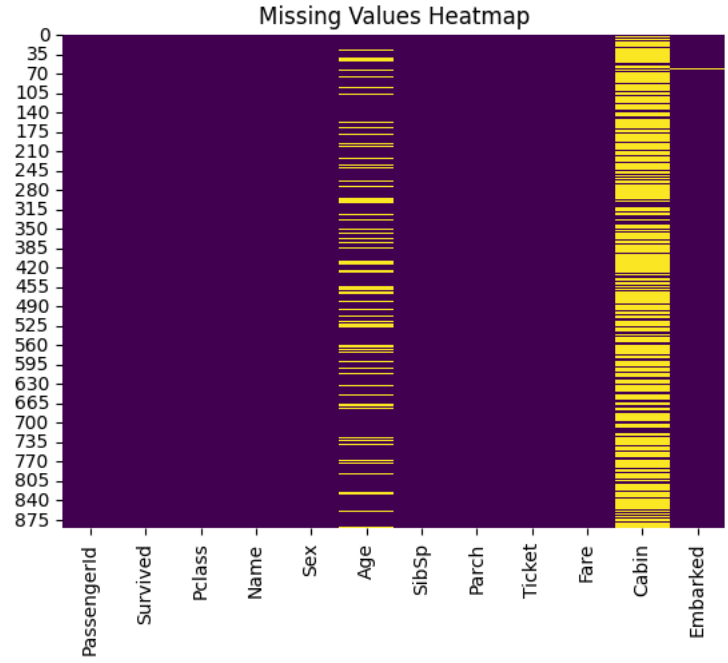
Next steps: Generate code with df | New interactive sheet

```
# Data structure
df.info()

# Descriptive statistics (numerical)
df.describe()

# Frequency of categorical values
print(df['Sex'].value_counts())
print(df['Embarked'].value_counts())
print(df['Pclass'].value_counts())
```

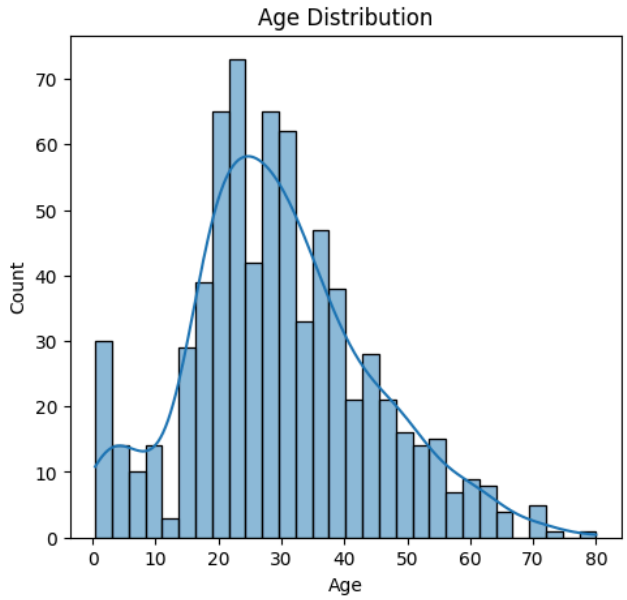
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Survived     891 non-null    int64
 2   Pclass       891 non-null    int64
 3   Name         891 non-null    object
 4   Sex          891 non-null    object
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64
 7   Parch        891 non-null    int64
 8   Ticket       891 non-null    object
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object
 11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
Sex
male      577
female    314
Name: count, dtype: int64
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64
Pclass
3    491
1    216
2    184
```

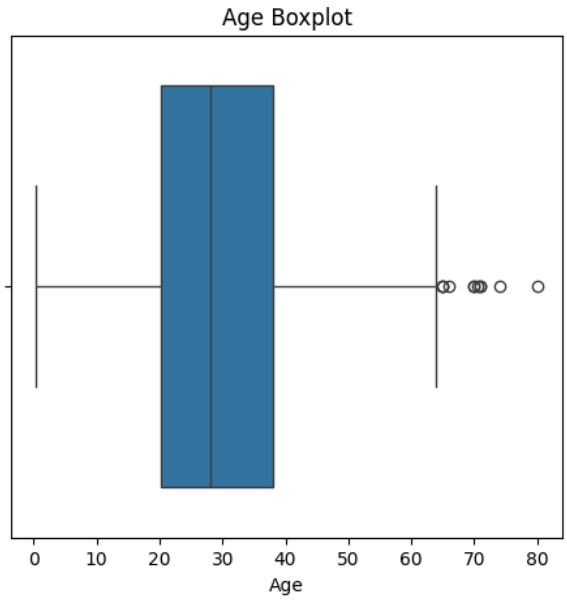
```
Name: count, dtype: int64
```

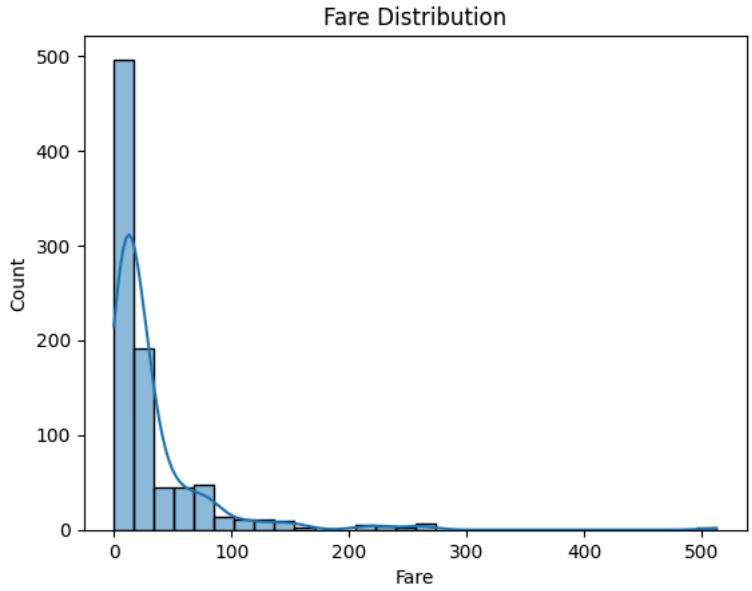
```
df.isnull().sum()
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
plt.title("Missing Values Heatmap")
plt.show()
```



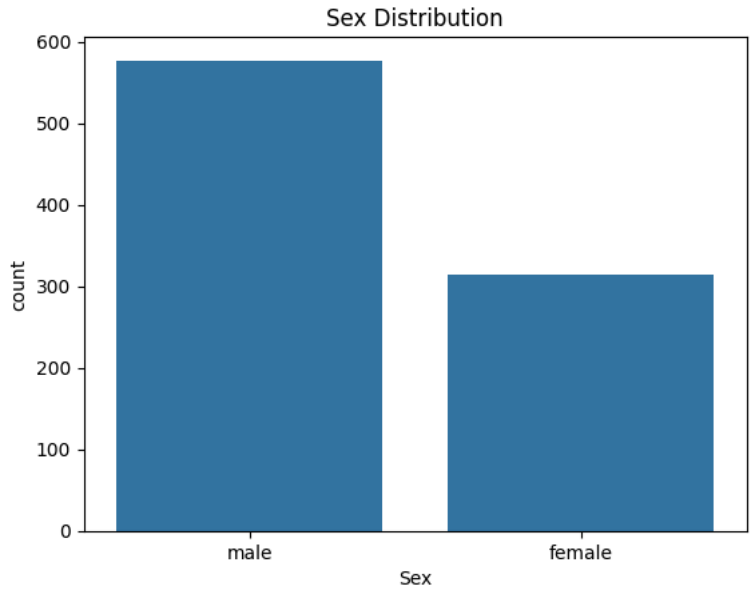

```
fig, axes = plt.subplots(1, 2, figsize=(12,5))

sns.histplot(df['Age'].dropna(), kde=True, bins=30, ax=axes[0])
axes[0].set_title("Age Distribution")

sns.boxplot(x=df['Age'], ax=axes[1])
axes[1].set_title("Age Boxplot")

plt.show()

sns.histplot(df['Fare'], kde=True, bins=30)
plt.title("Fare Distribution")
plt.show()
```

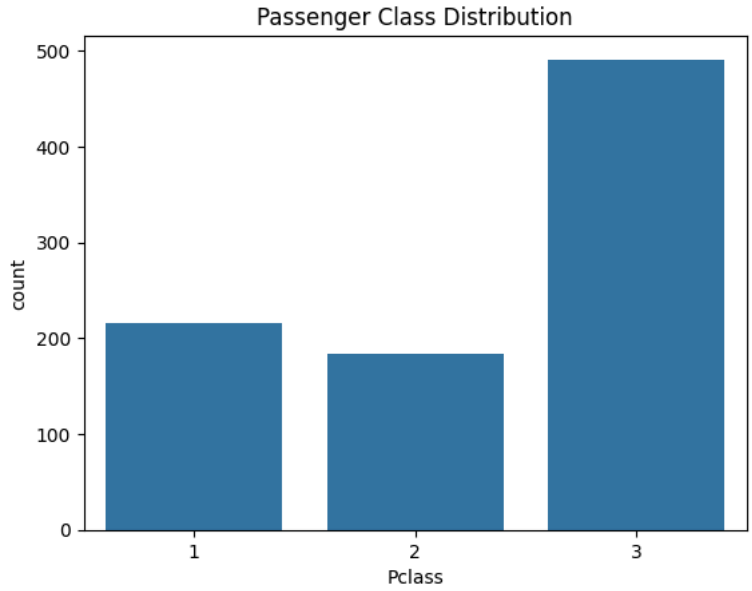


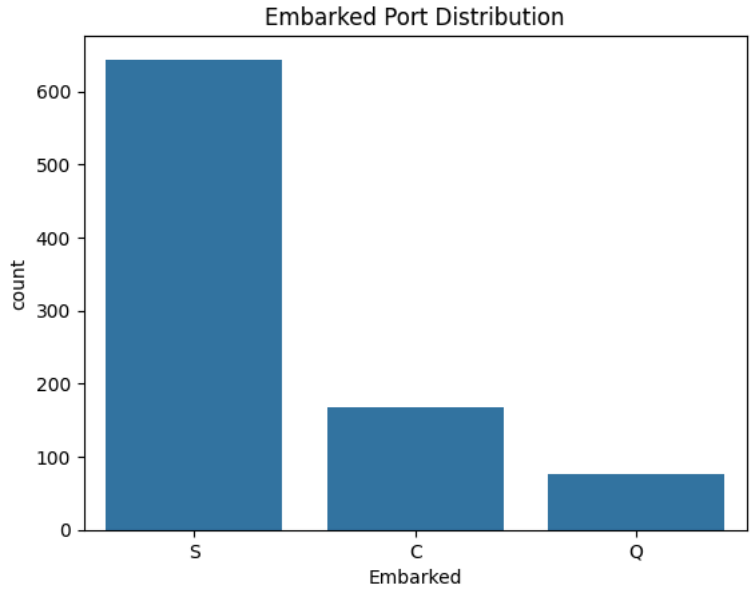





```
sns.countplot(x='Sex', data=df)
plt.title("Sex Distribution")
plt.show()

sns.countplot(x='Pclass', data=df)
plt.title("Passenger Class Distribution")
plt.show()

sns.countplot(x='Embarked', data=df)
plt.title("Embarked Port Distribution")
plt.show()
```

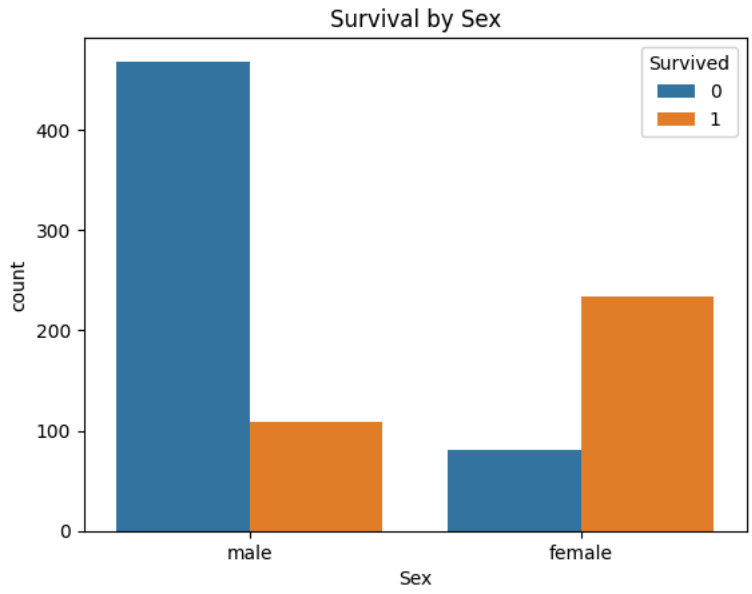


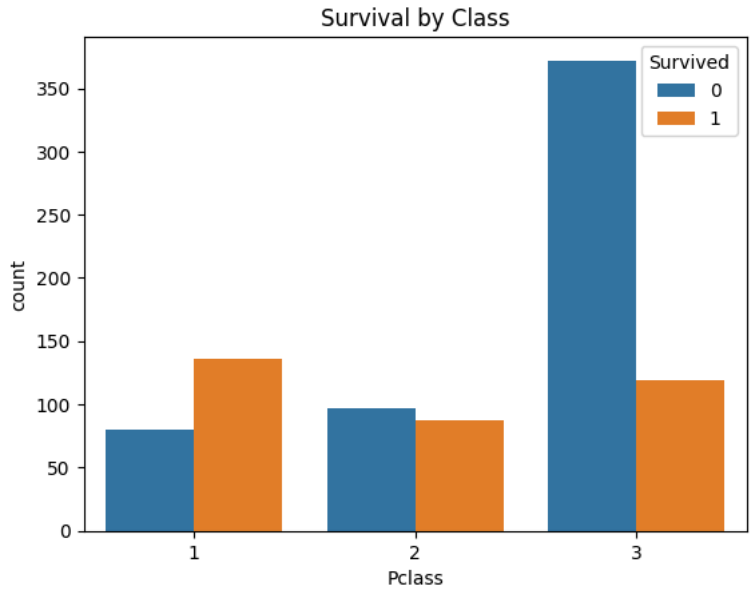


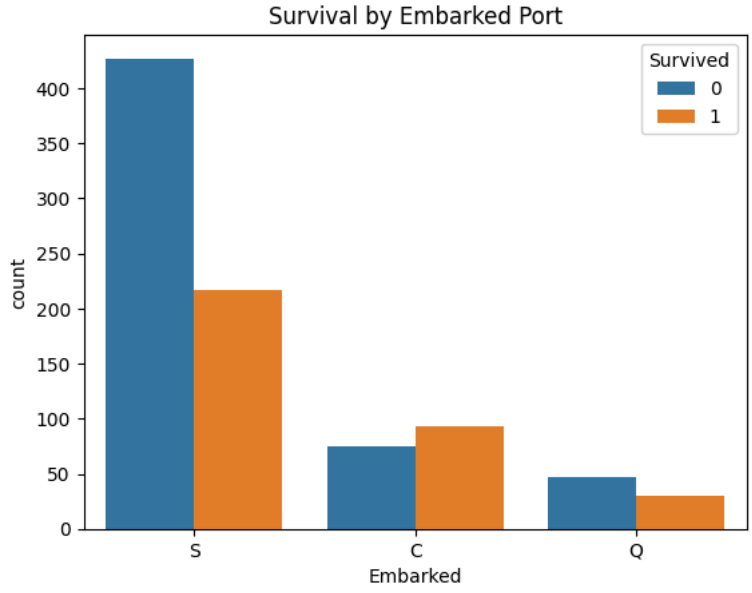



```
sns.countplot(x='Sex', hue='Survived', data=df)
plt.title("Survival by Sex")
plt.show()

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

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

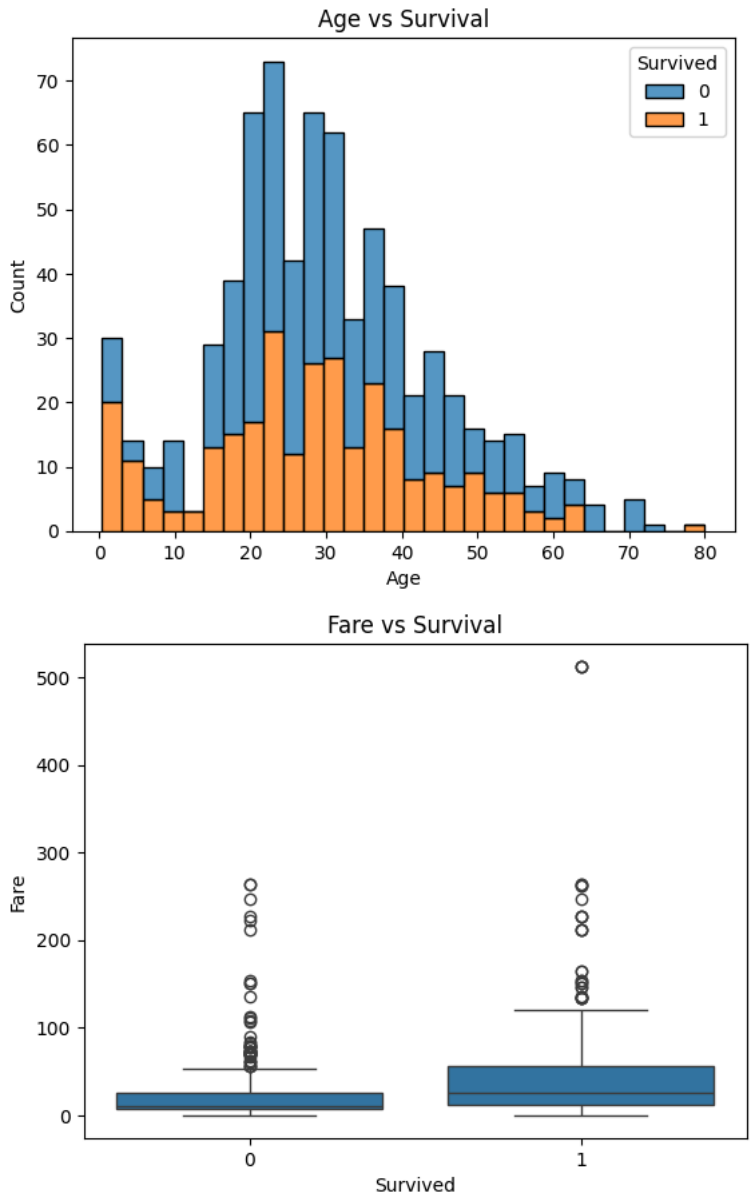




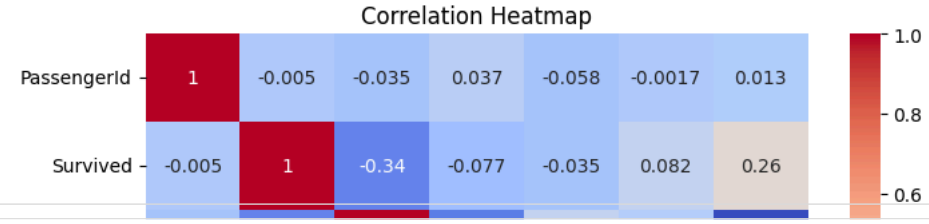



```
sns.histplot(data=df, x="Age", hue="Survived", multiple="stack", bins=30)
plt.title("Age vs Survival")
plt.show()

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

```
plt.title("Fare vs Survival")
plt.show()
```




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




```
sns.pairplot(df[['Survived','Pclass','Age','Fare','SibSp','Parch']], hue='Survived')
plt.show()
```