

```
from google.colab import files
uploaded = files.upload()
```

Choose Files train.csv

- train.csv(text/csv) - 61194 bytes, last modified: 6/9/2025 - 100% done

Saving train.csv to train.csv

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

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

```
df = pd.read_csv("train.csv")
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	Cummings, 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 (Lilv Mav Peel)	female	35.0	1	0	113803	53.1000	C123	S

Next steps:

Generate code with df

View recommended plots

New interactive sheet

```
print("=== Dataset Info ===")
print(df.info())

print("\n=== Statistical Summary ===")
print(df.describe())

print("\n=== Missing Values ===")
print(df.isnull().sum())
```

=== Dataset Info ===

<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

None

=== Statistical Summary ===

	PassengerId	Survived	Pclass	Age	SibSp
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400

```

50%    0.000000    14.454200
75%    0.000000    31.000000
max     6.000000   512.329200

```

```
=== Missing Values ===
```

```

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

```

```

print(df['Sex'].value_counts())
print(df['Embarked'].value_counts())
print(df['Pclass'].value_counts())

```

```

↩ 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

```

```

plt.figure(figsize=(8,5))
sns.histplot(df['Age'].dropna(), kde=True, bins=30)
plt.title("Age Distribution")

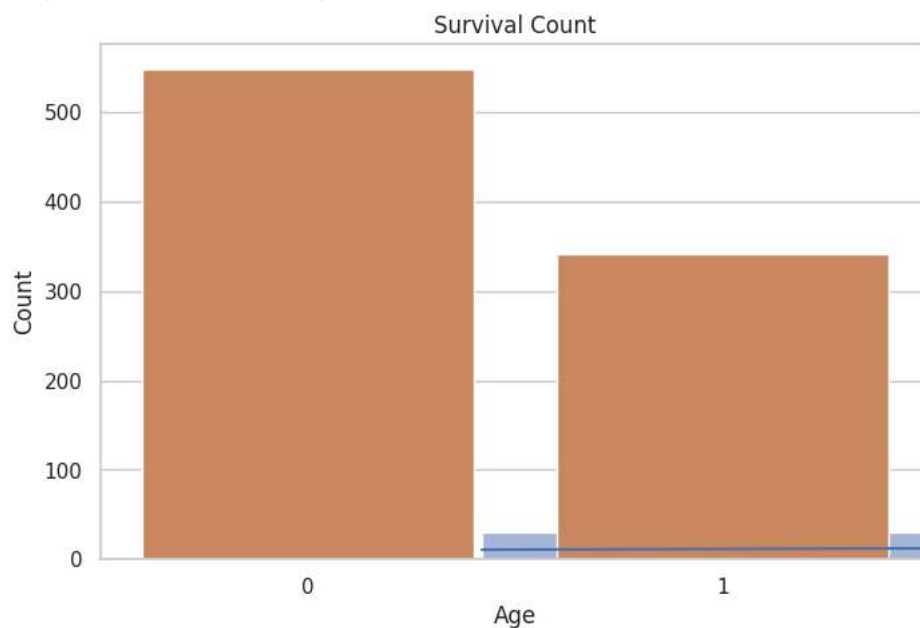
```

```

sns.countplot(x='Survived', data=df)
plt.title("Survival Count")

```

```
↩ Text(0.5, 1.0, 'Survival Count')
```



```

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

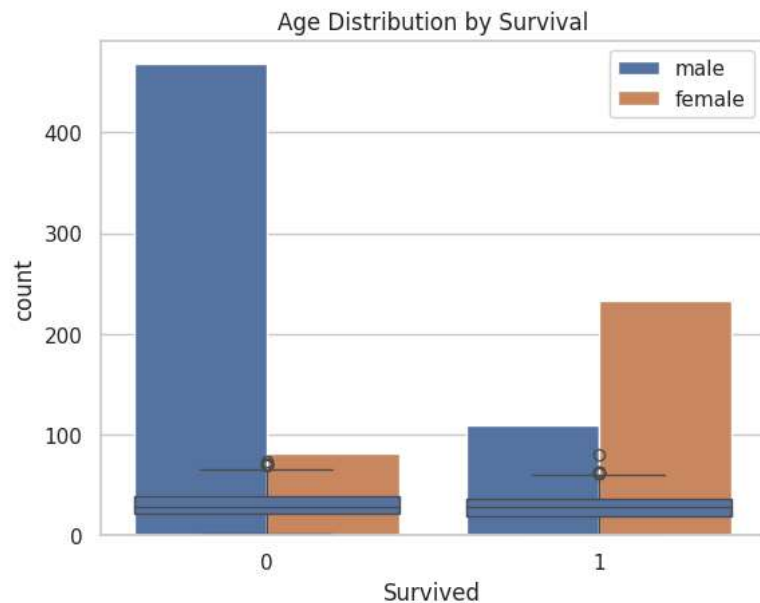
```

```

sns.boxplot(x='Survived', y='Age', data=df)
plt.title("Age Distribution by Survival")

```

Text(0.5, 1.0, 'Age Distribution by Survival')

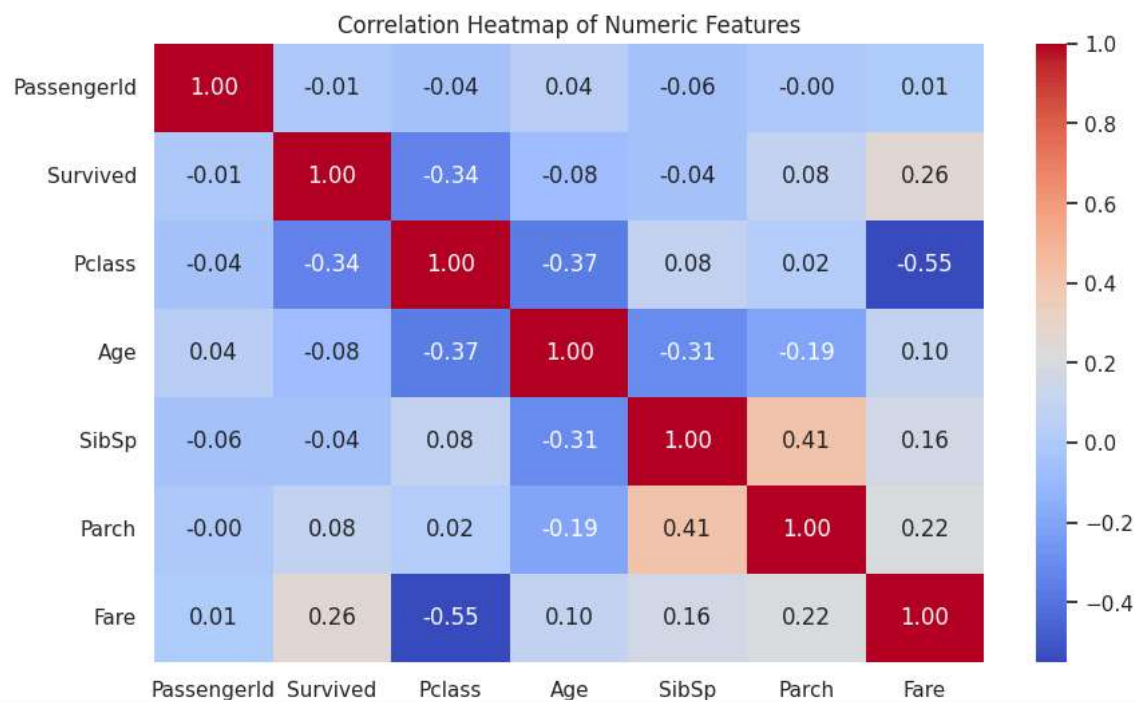


```
plt.figure(figsize=(10, 6))

numeric_df = df.select_dtypes(include='number')

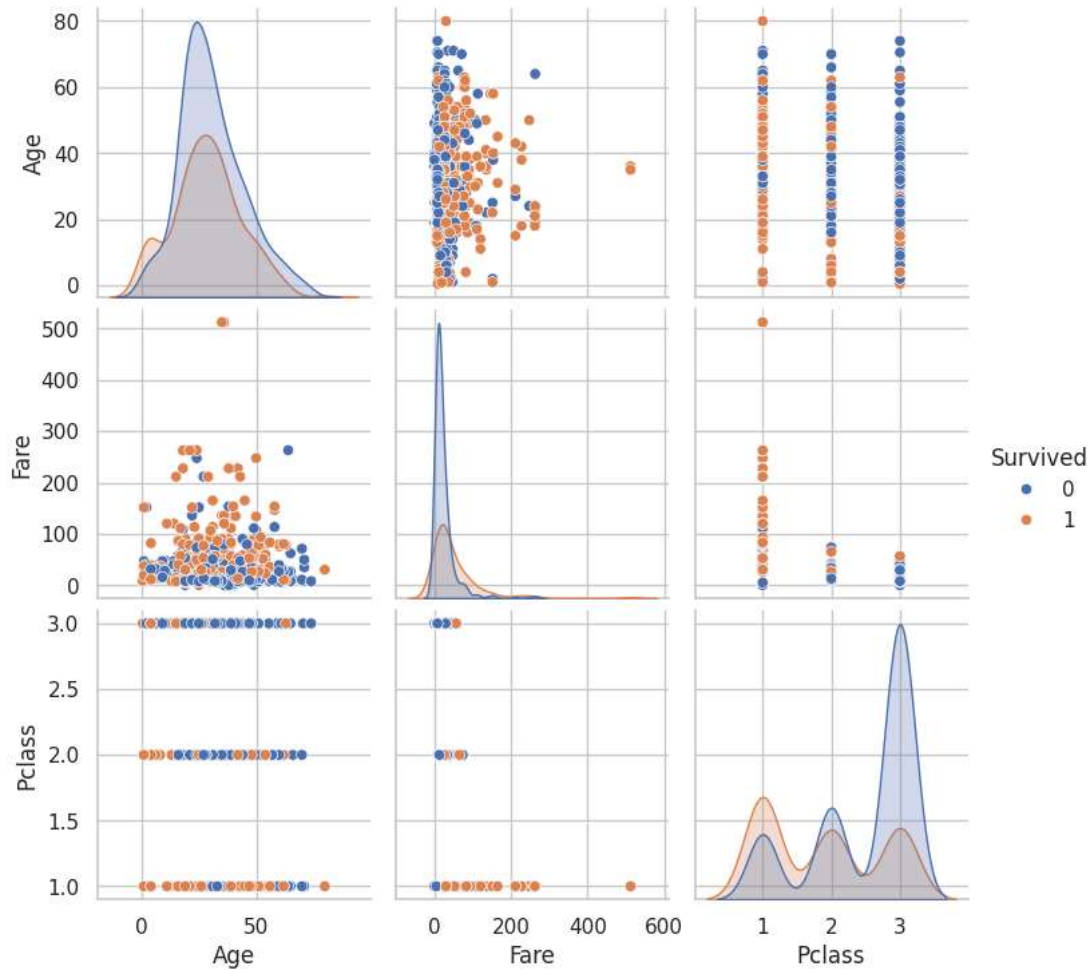
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm', fmt=".2f")

plt.title("Correlation Heatmap of Numeric Features")
plt.show()
```



```
sns.pairplot(df[['Age', 'Fare', 'Pclass', 'Survived']].dropna(), hue='Survived')
```

```
<seaborn.axisgrid.PairGrid at 0x7afd4acf0990>
```



```
print("""
```

```
FINAL OBSERVATIONS:
```

```
- Females had a much higher survival rate than males.
- Passengers in 1st class survived more than those in lower classes.
- Younger passengers tended to survive more.
- Fare and Pclass are correlated – higher fares were mostly paid by 1st class passengers.
""")
```



```
FINAL OBSERVATIONS:
```

```
- Females had a much higher survival rate than males.
- Passengers in 1st class survived more than those in lower classes.
- Younger passengers tended to survive more.
- Fare and Pclass are correlated – higher fares were mostly paid by 1st class passengers.
```