

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

[Choose files](#) No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving train.csv to train.csv

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

sns.set(style="whitegrid") # Nice style for plots
```

```
# Replace with the name of your uploaded file
df = pd.read_csv("/content/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	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	

```
# Info about data
df.info()

# Summary statistics for numeric columns
df.describe()

# Value counts for categorical columns
print("Survival:\n", df['Survived'].value_counts(), "\n")
print("Passenger Class:\n", df['Pclass'].value_counts(), "\n")
print("Gender:\n", df['Sex'].value_counts(), "\n")
print("Embarked:\n", df['Embarked'].value_counts(), "\n")
```

```
<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
```

```
Survival:
```

```
Survived
```

```
0    549
```

```
1    342
```

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

```
Passenger Class:
```

```
Pclass
```

```
3    491
```

```
1    216
```

```
2    184
```

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

```
Gender:
```

```
Sex
```

```
male    577
```

```
female  314
```

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

```
Embarked:
```

```
Embarked
```

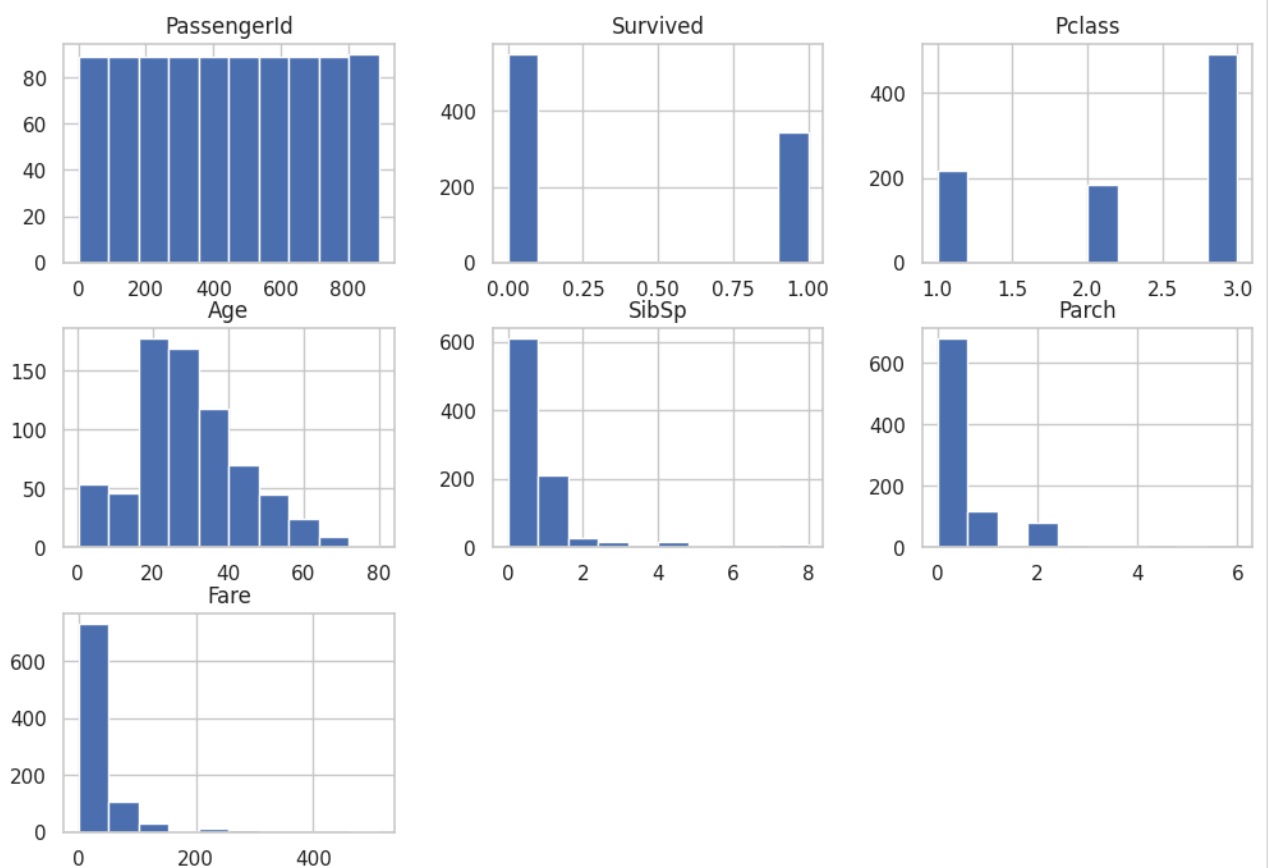
```
S    644
```

```
C    168
```

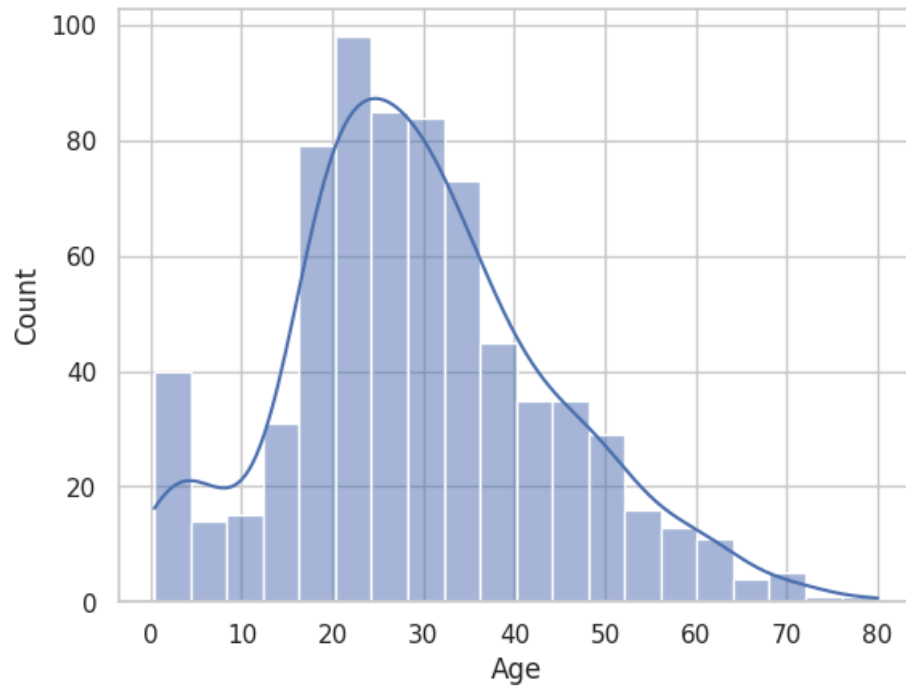
```
Q     77
```

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

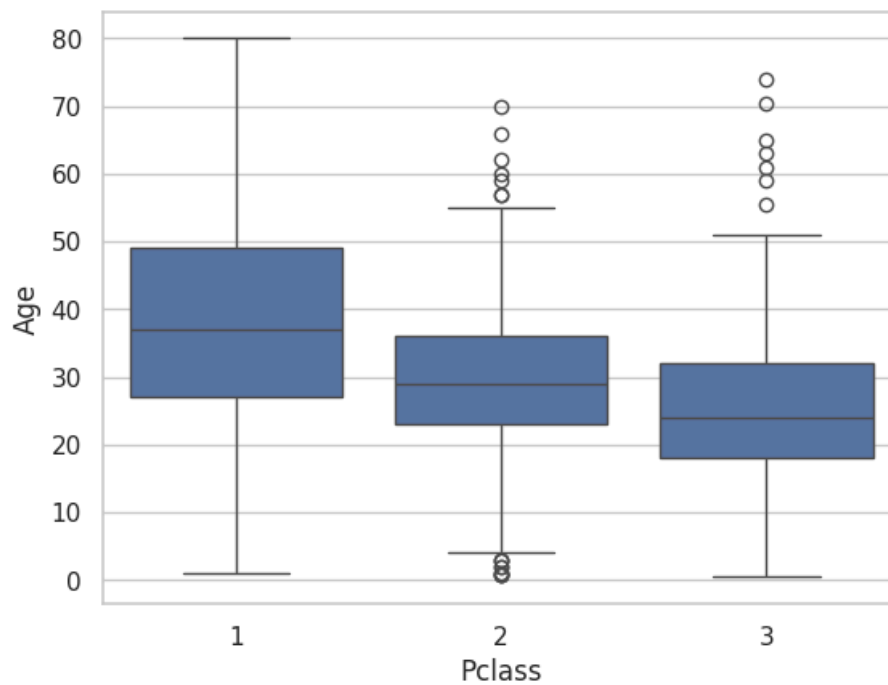
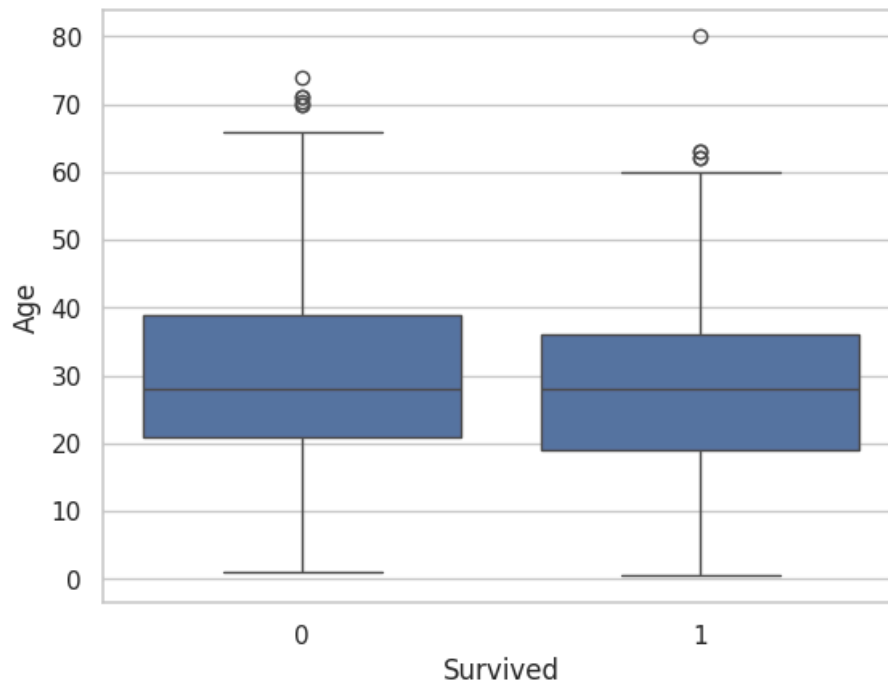
```
df.hist(figsize=(12,8))
plt.show()
```



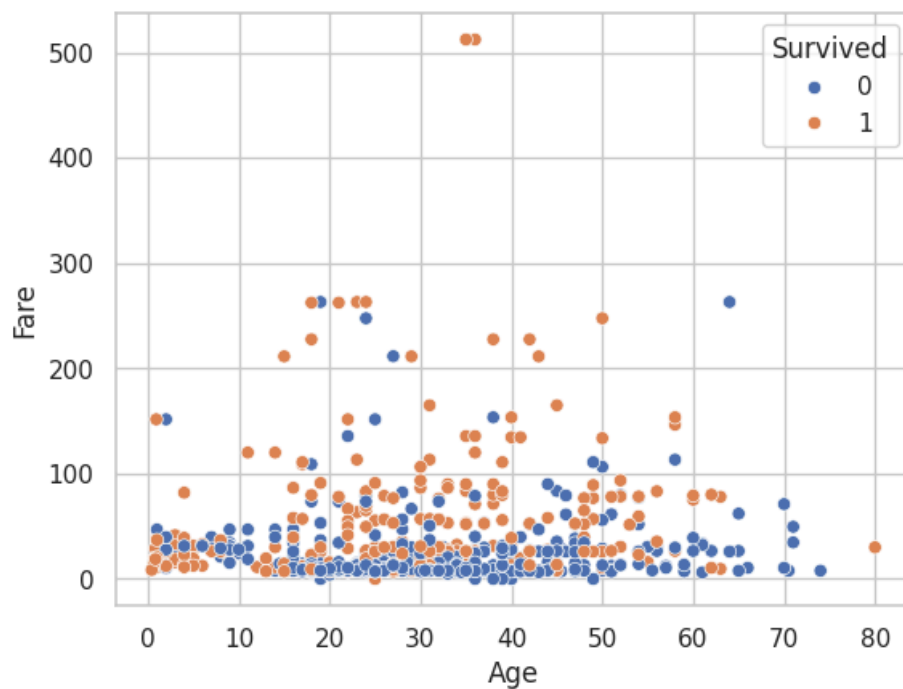
```
sns.histplot(df['Age'].dropna(), kde=True)  
plt.show()
```



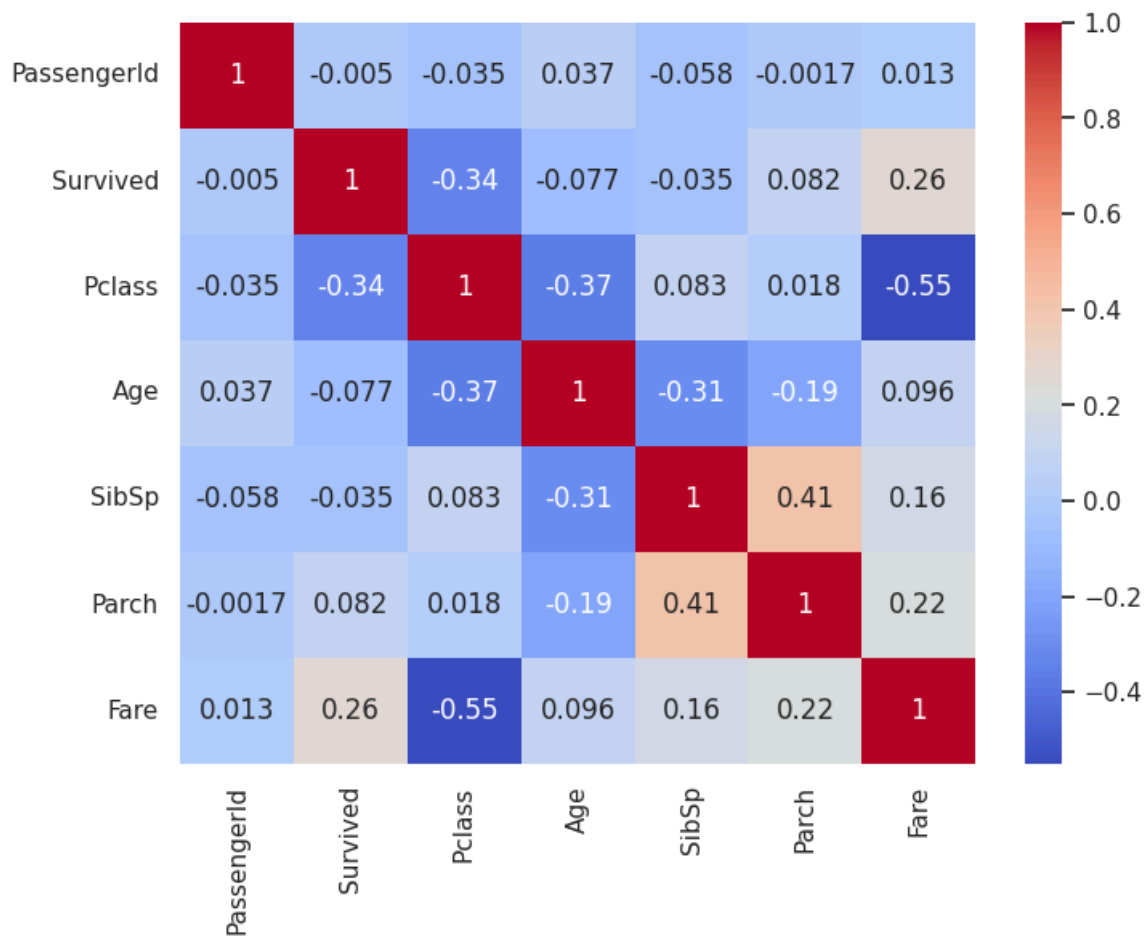
```
# Age distribution by Survival  
sns.boxplot(x="Survived", y="Age", data=df)  
plt.show()  
  
# Age distribution by Pclass  
sns.boxplot(x="Pclass", y="Age", data=df)  
plt.show()
```



```
# Fare vs Age
sns.scatterplot(x="Age", y="Fare", hue="Survived", data=df)
plt.show()
```



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



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

