

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

# Display settings
%matplotlib inline
sns.set_style('whitegrid')
```

```
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

```
df = pd.read_csv('train.csv')
df.head()
```

1 to 5 of 5 entries Filter ?

| index | 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 | 7.25 | NaN | S |
| 1 | 2 | 1 | 1 | Cumings, Mrs. John Bradley (Florence Briggs Thayer) | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.925 | NaN | S |
| 3 | 4 | 1 | 1 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 35.0 | 1 | 0 | 113803 | 53.1 | C123 | S |
| 4 | 5 | 0 | 3 | Allen, Mr. William Henry | male | 35.0 | 0 | 0 | 373450 | 8.05 | NaN | S |

Show 25 per page

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
import pandas as pd

# Replace with your actual filename if it's different
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 | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 3 | 1 | 3 | Heikinen, Miss. Laina | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| | | | | Futrelle, Mrs. Jacques Heath | | | | | | | | |

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
df.shape
df.info()
df.describe()
df.isnull().sum()
df['Survived'].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

```

count

Survived

| | |
|---|-----|
| 0 | 549 |
| 1 | 342 |

dtype: int64

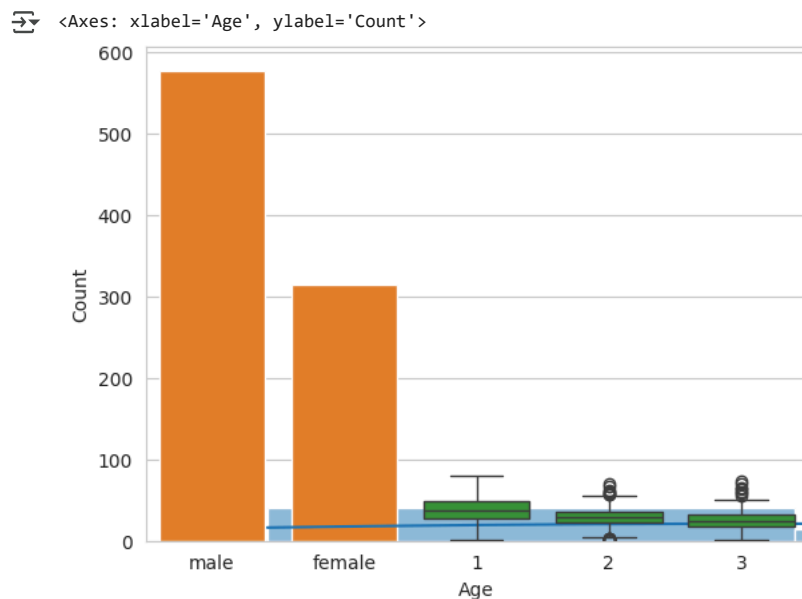
```

# Distribution of age
sns.histplot(df['Age'].dropna(), kde=True)

# Countplot of gender
sns.countplot(x='Sex', data=df)

# Boxplot to see age distribution by class
sns.boxplot(x='Pclass', y='Age', data=df)

```



✓ Observation: Males had a significantly lower survival rate compared to females.


```

# Survival by gender
sns.countplot(x='Survived', hue='Sex', data=df)

# Plot correlation heatmap with only numerical columns
plt.figure(figsize=(10,6))
sns.heatmap(df.select_dtypes(include='number').corr(), annot=True, cmap='coolwarm')

# Pairplot (filtered for important features)
sns.pairplot(df[['Survived', 'Age', 'Fare', 'Pclass']].dropna(), hue='Survived')

```

 <seaborn.axisgrid.PairGrid at 0x78f572afd910>