```
from google.colab import files
uploaded = files.upload() # Choose titanic.zip
```

```
\overline{\Sigma}
```

Choose files titanic.zip

• titanic.zip(application/x-zip-compressed) - 34877 bytes, last modified: 12/08/2025 - 100% done Saving titanic.zip to titanic (1).zip

import zipfile

with zipfile.ZipFile("titanic.zip", 'r') as zip_ref: print(zip_ref.namelist()) # Shows files inside

['gender_submission.csv', 'test.csv', 'train.csv']

with zipfile.ZipFile("titanic.zip", 'r') as zip_ref: zip_ref.extractall() # Extracts to current folder

import pandas as pd
df = pd.read_csv("train.csv") # Or use the name you saw in step 2
df.head()

		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Ħ
	0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	th
	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.2833	C85	С	
	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	

Next steps: Generate code with df

View recommended plots

New interactive sheet

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")
plt.rcParams['figure.figsize'] = (10,6)
```

```
df = pd.read_csv("train.csv") # or use the URL method
```

df.head()
df.shape
df.info()

df.describe()

df.isnull().sum()

```
<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
                      891 non-null
          Survived
                                      int64
     1
                       891 non-null
          Pclass.
                                      int64
                      891 non-null
                                      object
     3
         Name
                      891 non-null
     4
         Sex
                                      object
                      714 non-null
     5
         Age
                                      float64
     6
         SibSp
                      891 non-null
                                       int64
         Parch
                       891 non-null
                                       int64
         Ticket
                       891 non-null
     8
                                      object
         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
     PassengerId
                    0
       Survived
        Pclass
         Name
                    0
         Sex
                    0
                  177
         Age
         SibSp
                    0
         Parch
        Ticket
                    0
         Fare
                    0
         Cabin
                  687
       Embarked
                    2
     dtype: int64
df['Age'].fillna(df['Age'].median(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
df.drop(columns=['Cabin'], inplace=True)
    /tmp/ipython-input-224943107.py:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series t
     The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on whi
     For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' o
       df['Age'].fillna(df['Age'].median(), inplace=True)
     /tmp/ipython-input-224943107.py:2: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series t
     The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on whi
     For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' o
       df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
sns.countplot(x='Survived', data=df)
plt.title("Survival Count")
plt.show()
sns.histplot(df['Age'], bins=30, kde=True)
plt.title("Age Distribution")
plt.show()
sns.countplot(x='Pclass', data=df)
```

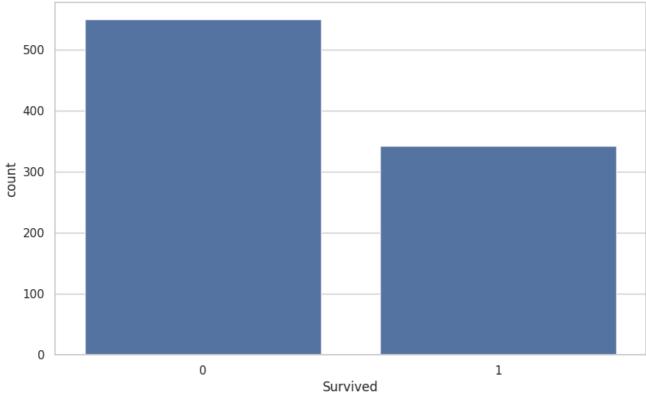
nlt title ("Dacconger Class Distribution")

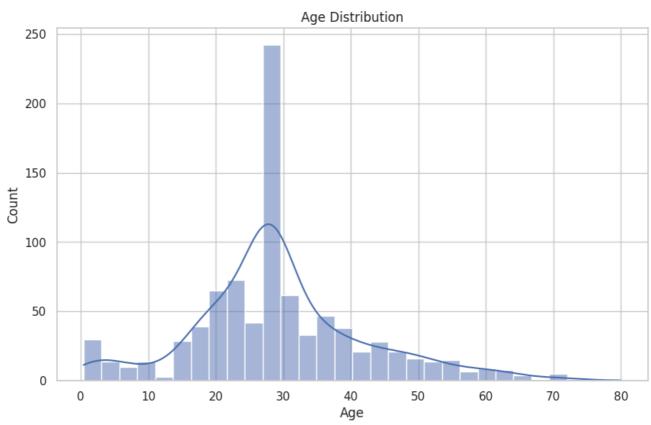
```
plt.show()

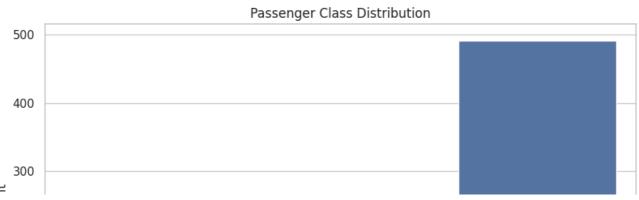
sns.countplot(x='Sex', data=df)
plt.title("Gender Distribution")
plt.show()
```

 $\overline{2}$









Sex

male

200

100

0

female