

```
In [2]: import pandas as pd
```

```
In [3]: url = "https://raw.githubusercontent.com/datasciencedojo/datasets/master/tit
```

```
In [4]: df = pd.read_csv(url)
```

```
In [5]: df.head()
```

```
Out[5]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	l
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0

```
In [8]: print(df.shape)
```

```
(891, 12)
```

```
In [9]: df.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
```

```
In [10]: df.isnull().sum()
```

```
Out[10]: 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
```

```
In [12]: df['Age'] = df['Age'].fillna(df['Age'].median())df['Embarked'].fillna(df['Em
```

```
In [13]: df['Embarked'].fillna(df['Embarked'].mode()[0])
```

```
Out[13]: 0      S
1      C
2      S
3      S
4      S
..
886    S
887    S
888    S
889    C
890    Q
Name: Embarked, Length: 891, dtype: object
```

```
In [14]: df.drop('Cabin', axis=1)
```

Out[14]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
<b>0</b>	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	
<b>1</b>	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	7
<b>2</b>	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
<b>3</b>	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
<b>4</b>	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
...	...	...	...	...	...	...	...	...	...	
<b>886</b>	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
<b>887</b>	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
<b>888</b>	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	W./C. 6607	2
<b>889</b>	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
<b>890</b>	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 11 columns

In [15]: `df.isnull().sum()`

```
Out[15]: PassengerId      0
         Survived        0
         Pclass         0
         Name           0
         Sex            0
         Age            0
         SibSp          0
         Parch          0
         Ticket         0
         Fare           0
         Cabin         687
         Embarked       2
         dtype: int64
```

```
In [16]: df.drop('Cabin', axis=1, inplace=True)
```

```
In [17]: df.isnull().sum()
```

```
Out[17]: PassengerId      0
         Survived        0
         Pclass         0
         Name           0
         Sex            0
         Age            0
         SibSp          0
         Parch          0
         Ticket         0
         Fare           0
         Embarked       2
         dtype: int64
```

```
In [18]: df.describe()
```

```
Out[18]:
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
<b>count</b>	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.0
<b>mean</b>	446.000000	0.383838	2.308642	29.361582	0.523008	0.381594	32.2
<b>std</b>	257.353842	0.486592	0.836071	13.019697	1.102743	0.806057	49.6
<b>min</b>	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.0
<b>25%</b>	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.9
<b>50%</b>	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.4
<b>75%</b>	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.0
<b>max</b>	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.3

```
In [19]: df.describe(include=['O'])
```

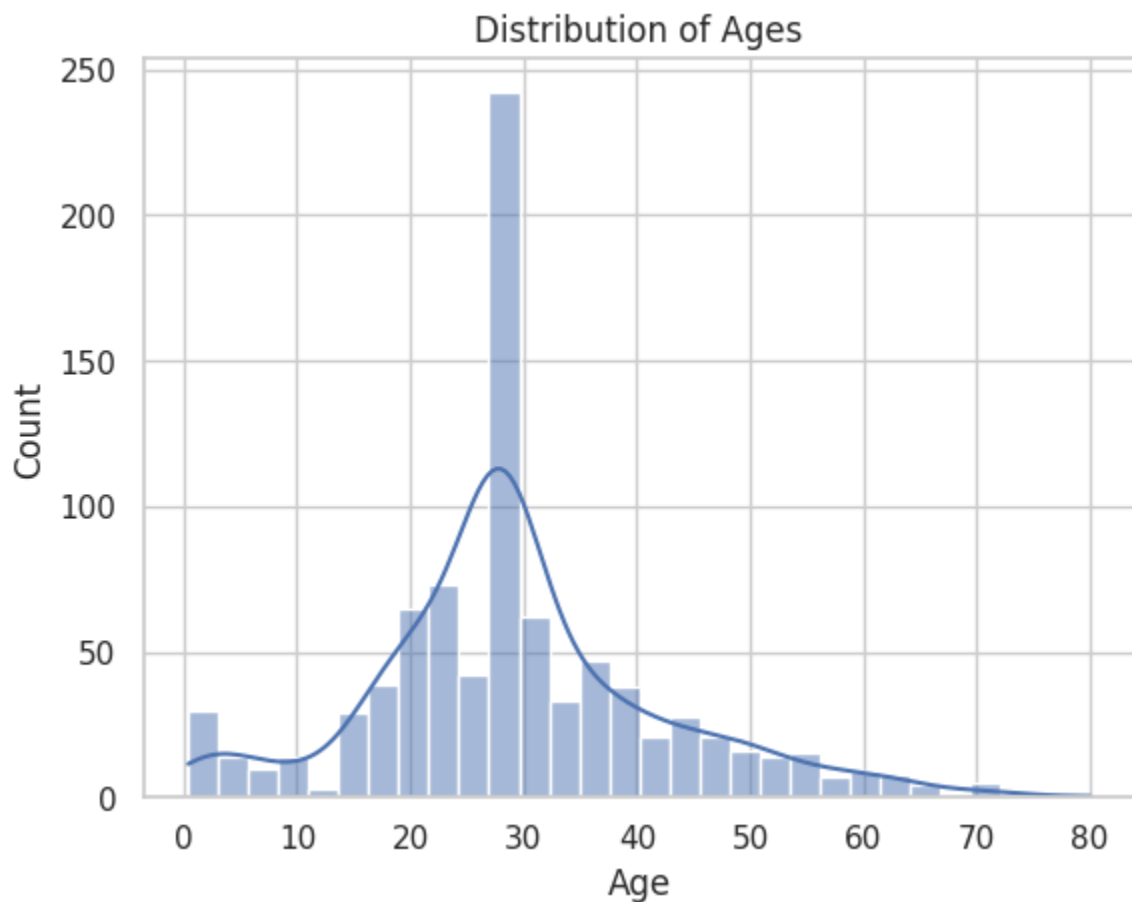
Out[19]:

	Name	Sex	Ticket	Embarked
count	891	891	891	889
unique	891	2	681	3
top	Dooley, Mr. Patrick	male	347082	S
freq	1	577	7	644

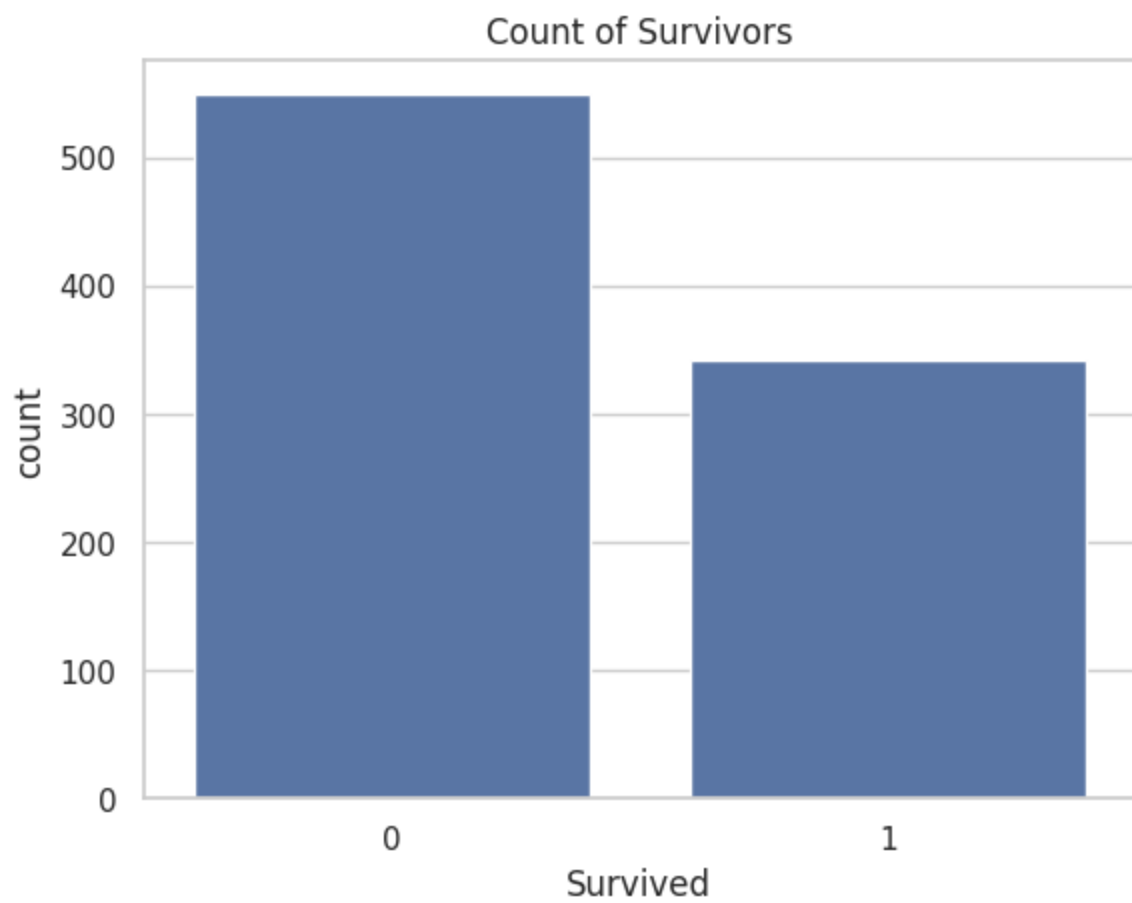
```
In [21]: import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [22]: sns.set(style="whitegrid")
```

```
In [23]: sns.histplot(df['Age'], kde=True, bins=30)
plt.title('Distribution of Ages')
plt.show()
```



```
In [24]: sns.countplot(x='Survived', data=df)
plt.title('Count of Survivors')
plt.show()
```



In [ ]: