

/

```
#Hello, this is notebook 1. - Q
# Hey!233
# It's peter
```

SELECT *
FROM 'train.csv'

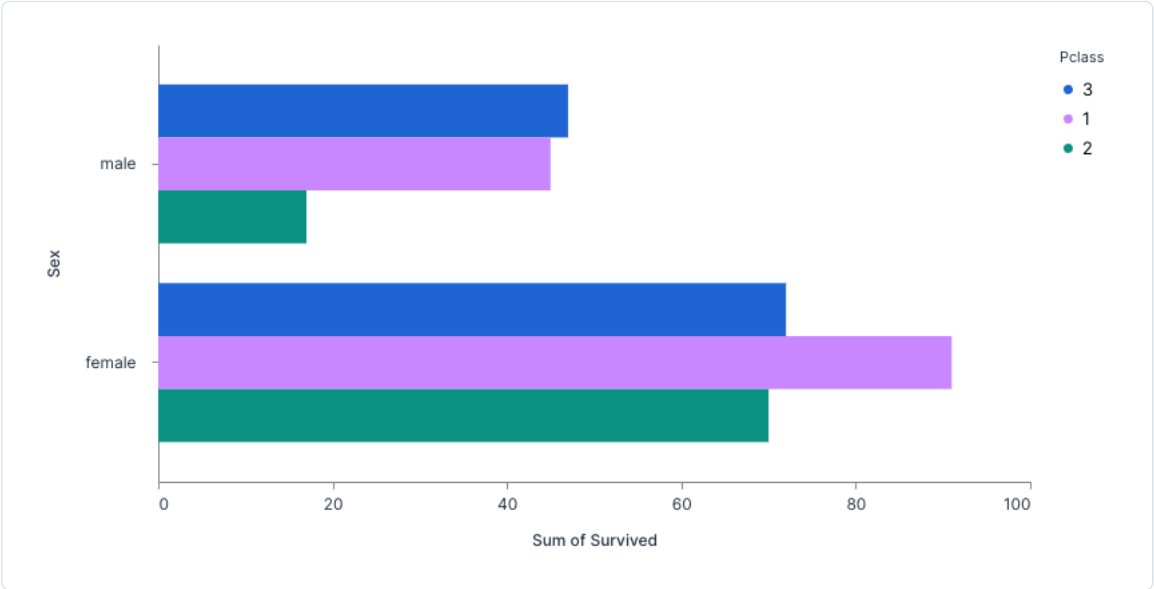
	PassengerId int64 1 - 891	Survived int64 0 - 1	Pclass int64 1 - 3	Name object Braund, Mr. ... 0.1% Cumings, Mr... 0.1% 889 others 99.8%	Sex object male 64.8% female 35.2%	Age float64 0.42 - 80.0	SibSp 0 - 8
0	1	0	3	Braund, Mr. Owen...	male	22	
1	2	1	1	Cumings, Mrs. Jo...	female	38	
2	3	1	3	Heikkinen, Miss. ...	female	26	
3	4	1	1	Futrelle, Mrs. Jac...	female	35	
4	5	0	3	Allen, Mr. William ...	male	35	
5	6	0	3	Moran, Mr. James	male	nan	
6	7	0	1	McCarthy, Mr. Ti...	male	54	
7	8	0	3	Palsson, Master. ...	male	2	
8	9	1	3	Johnson, Mrs. Os...	female	27	
9	10	1	2	Nasser, Mrs. Nich...	female	14	

891 rows, 12 cols, showing 10 rows/page << < Page 1 of 90 > >>

SELECT Name FROM df_train WHERE Survived = 1;

	Name object Cumings, Mr... 0.3% Heikkinen, M... 0.3% 340 others 99.4%
0	Cumings, Mrs. Jo...
1	Heikkinen, Miss. ...
2	Futrelle, Mrs. Jac...
3	Johnson, Mrs. Os...
4	Nasser, Mrs. Nich...
5	Sandstrom, Miss....
6	Bonnell, Miss. Eliz...
7	Hewlett, Mrs. (Ma...
8	Williams, Mr. Char...
9	Masselmani, Mrs....

342 rows, 1 col, showing 10 rows/page << < Page 1 of 35 > >>



SELECT *
FROM 'test.csv'

	PassengerId int64 892 - 1309	Pclass int64 1 - 3	Name object Kelly, Mr. Ja... 0.2% Wilkes, Mrs. ... 0.2% 416 others ... 99.5%	Sex object male 63.6% female 36.4%	Age float64 0.17 - 76.0	SibSp int64 0 - 8	Parch 0 - 9
0	892	3	Kelly, Mr. James	male	34.5	0	0
1	893	3	Wilkes, Mrs. Jam...	female	47	1	1
2	894	2	Myles, Mr. Thoma...	male	62	0	0
3	895	3	Wirz, Mr. Albert	male	27	0	0
4	896	3	Hirvonen, Mrs. Al...	female	22	1	1
5	897	3	Svensson, Mr. Jo...	male	14	0	0
6	898	3	Connolly, Miss. K...	female	30	0	0
7	899	2	Caldwell, Mr. Alb...	male	26	1	1
8	900	3	Abraham, Mrs. Jo...	female	18	0	0
9	901	3	Davies, Mr. John ...	male	21	2	2

418 rows, 11 cols, showing 10 rows/page << < Page 1 of 42 > >> [↓](#)

我想知道船舱等级跟存活率的关系，用图表示

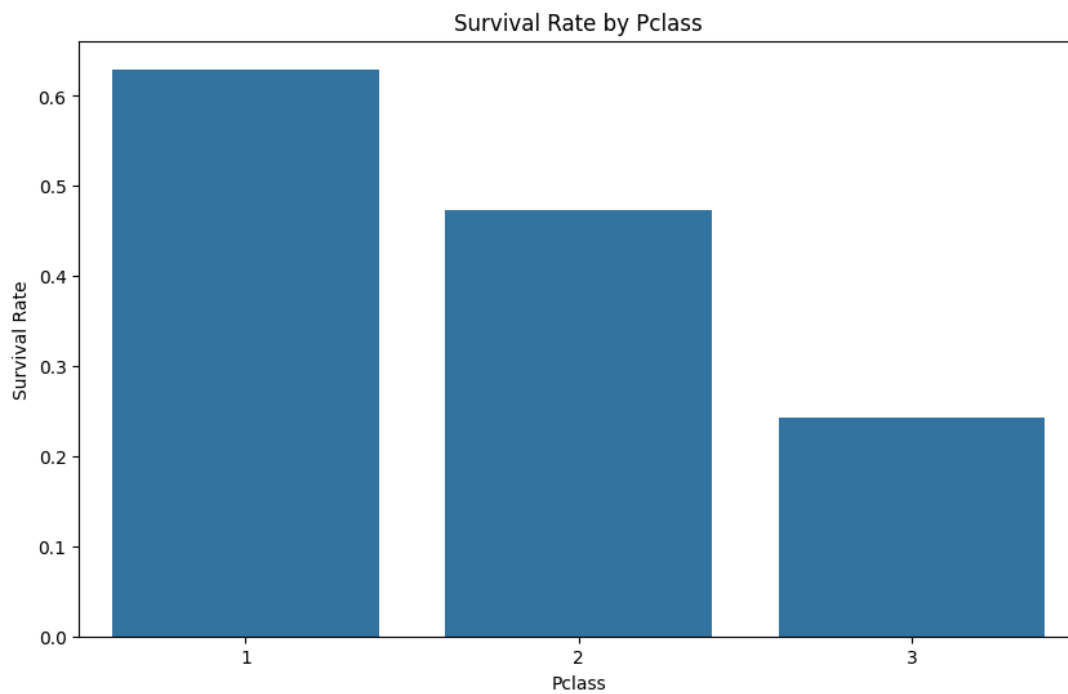
```
import matplotlib.pyplot as plt
import seaborn as sns

# Plotting the survival rate by Pclass
plt.figure(figsize=(10, 6))
sns.barplot(x='Pclass', y='Survived', data=df_train, ci=None)
plt.title('Survival Rate by Pclass')
plt.xlabel('Pclass')
plt.ylabel('Survival Rate')
plt.show()
```

/tmp/ipykernel_612/4259607381.py:6: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

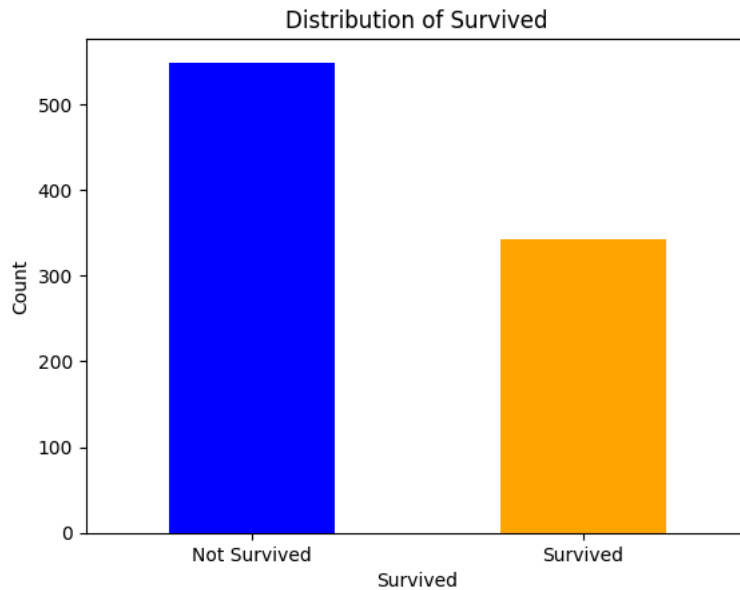
```
sns.barplot(x='Pclass', y='Survived', data=df_train, ci=None)
```



Visualize the distribution of the target variable to understand its class balance

```
import matplotlib.pyplot as plt

# Visualize the distribution of the target variable 'Survived'
df_train['Survived'].value_counts().plot(kind='bar', color=['blue', 'orange'])
plt.title('Distribution of Survived')
plt.xlabel('Survived')
plt.ylabel('Count')
plt.xticks(ticks=[0, 1], labels=['Not Survived', 'Survived'], rotation=0)
plt.show()
```



The bar chart above shows the distribution of the target variable 'Survived'. This helps us understand the class balance between those who survived and those who did not.

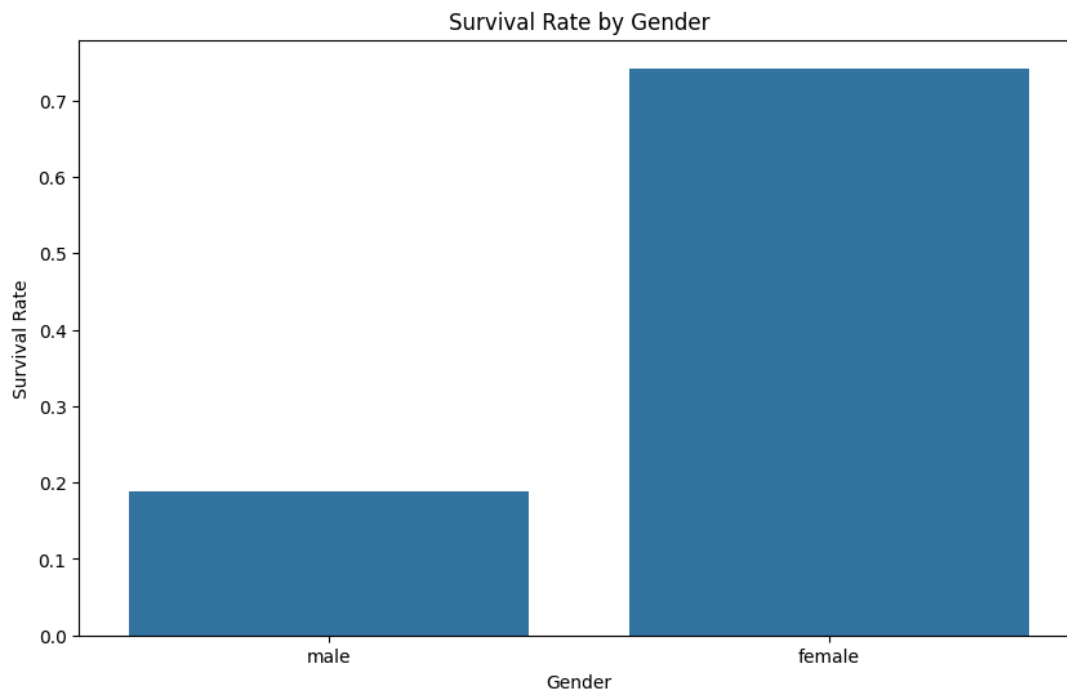
visualize the Gender vs Survival Rate

```
plt.figure(figsize=(10, 6))
sns.barplot(x='Sex', y='Survived', data=df_train, ci=None)
plt.title('Survival Rate by Gender')
plt.xlabel('Gender')
plt.ylabel('Survival Rate')
plt.show()
```

/tmp/ipykernel_612/3043830905.py:2: FutureWarning:

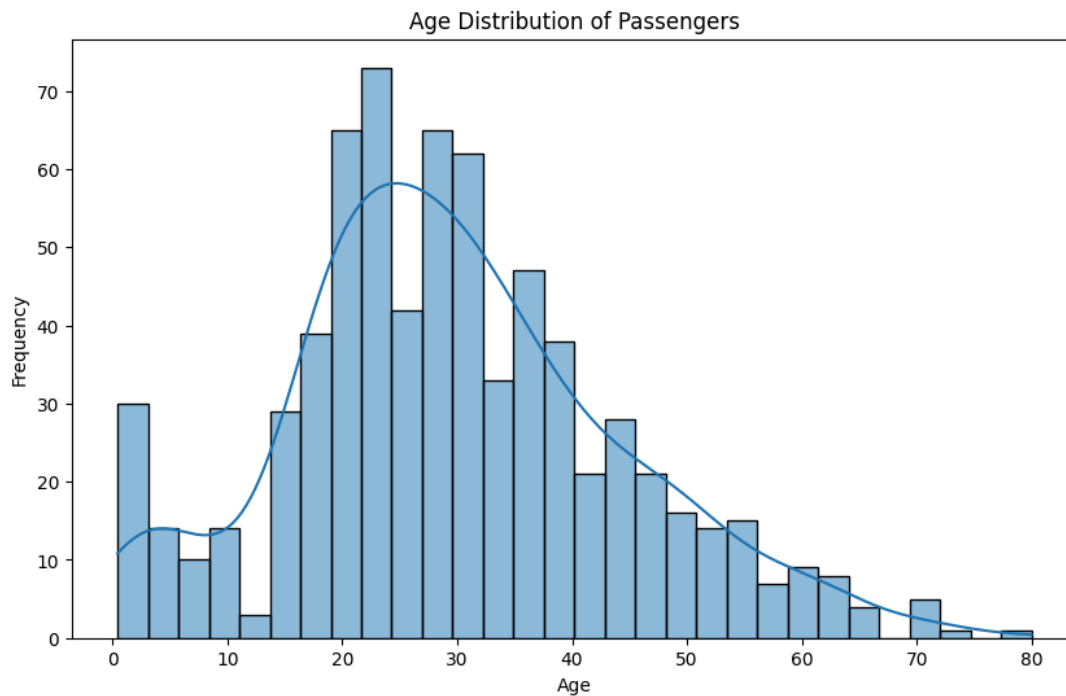
The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

```
sns.barplot(x='Sex', y='Survived', data=df_train, ci=None)
```



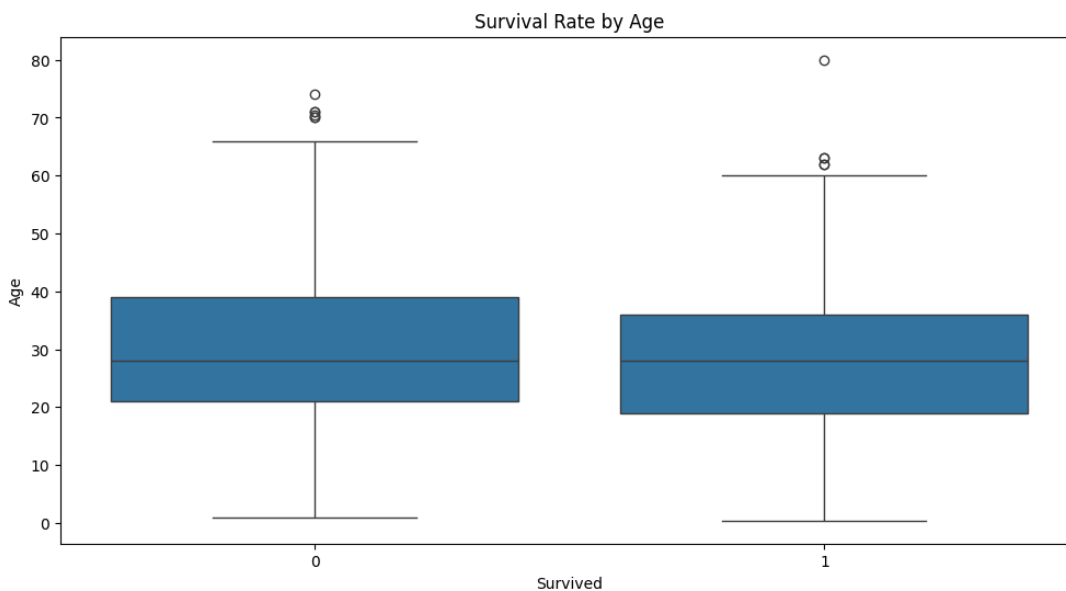
Use a histogram to show the age distribution of passengers

```
plt.figure(figsize=(10, 6))
sns.histplot(df_train['Age'].dropna(), bins=30, kde=True)
plt.title('Age Distribution of Passengers')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



Use a box plot to see the survival rate of passengers of different ages, and tell me what you find

```
plt.figure(figsize=(12, 6))
sns.boxplot(x='Survived', y='Age', data=df_train)
plt.title('Survival Rate by Age')
plt.xlabel('Survived')
plt.ylabel('Age')
plt.show()
```

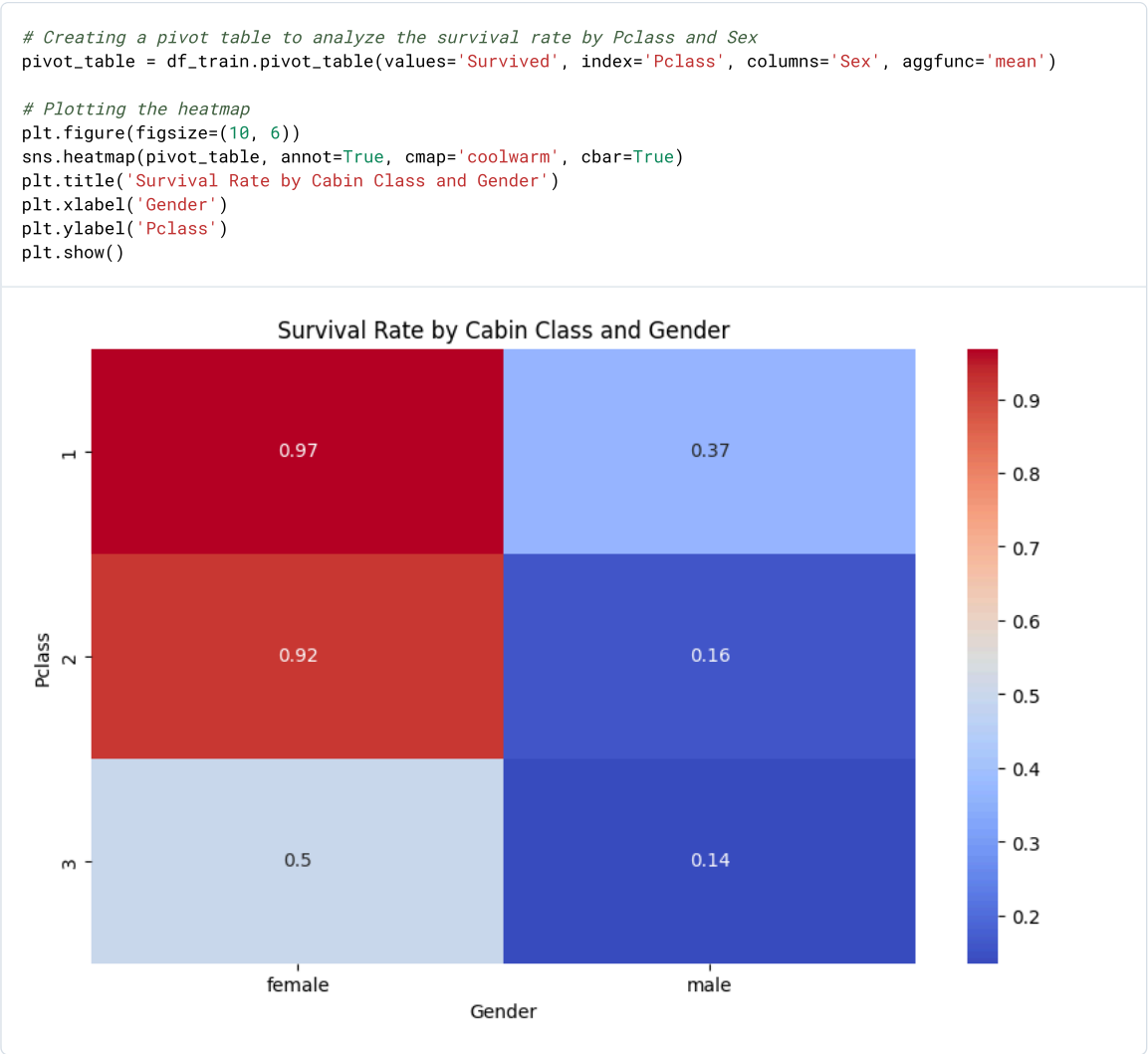


The box plot shows the distribution of ages for passengers who survived and those who did not. Here are some observations:

- 1. The median age of survivors is lower than that of non-survivors.
- 2. There is a wider age range among non-survivors compared to survivors.
- 3. Younger passengers (children) have a higher survival rate, as indicated by the lower median age and the presence of younger ages in the survivor group.

This suggests that younger passengers had a better chance of survival.

Use heatmap to analyze the survival rate of cabin class and gender combination and tell me the conclusion



The heatmap shows the survival rates for different combinations of cabin class (Pclass) and gender (Sex). Here are the conclusions:

- 1. Females had a higher survival rate across all cabin classes compared to males.
- 2. The highest survival rate was observed for females in the first class (Pclass 1).
- 3. Males in the third class (Pclass 3) had the lowest survival rate.
- 4. The survival rate for males increased with higher cabin classes, but it was still lower than that of females in the same class.

Overall, gender and cabin class both played significant roles in determining the survival rate, with females and higher cabin classes having better chances of survival.