


```
from google.colab import drive
drive.mount('/content/drive')
```

 Mounted at /content/drive

```
import pandas as pd
```


```
# Path to your folder in Google Drive
folder_path = '/content/drive/MyDrive/inten/'
```



```
# Read the CSVs
train_df = pd.read_csv(folder_path + 'train.csv')
test_df = pd.read_csv(folder_path + 'test.csv')
gender_submission_df = pd.read_csv(folder_path + 'gender_submission.csv')
```

```
print("Train Data:")
display(train_df.head())
```


```
print("Test Data:")
display(test_df.head())
```

```
print("Gender Submission Data:")
display(gender_submission_df.head())
```


 Train Data:

	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	
3	4	1	1	Futelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S	

Test Data:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S	


Gender Submission Data:

	PassengerId	Survived	
0	892	0	
1	893	1	
2	894	0	
3	895	0	

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

```
# Show plots in notebook
%matplotlib inline
```

```
print("Survived (Train):\n", train_df['Survived'].value_counts())
print("\nPclass (Train):\n", train_df['Pclass'].value_counts())
print("\nSex (Train):\n", train_df['Sex'].value_counts())
print("\nEmbarked (Train):\n", train_df['Embarked'].value_counts())
```

 Survived (Train):

```
Survived
0    549
1    342
Name: count, dtype: int64
```

Pclass (Train):

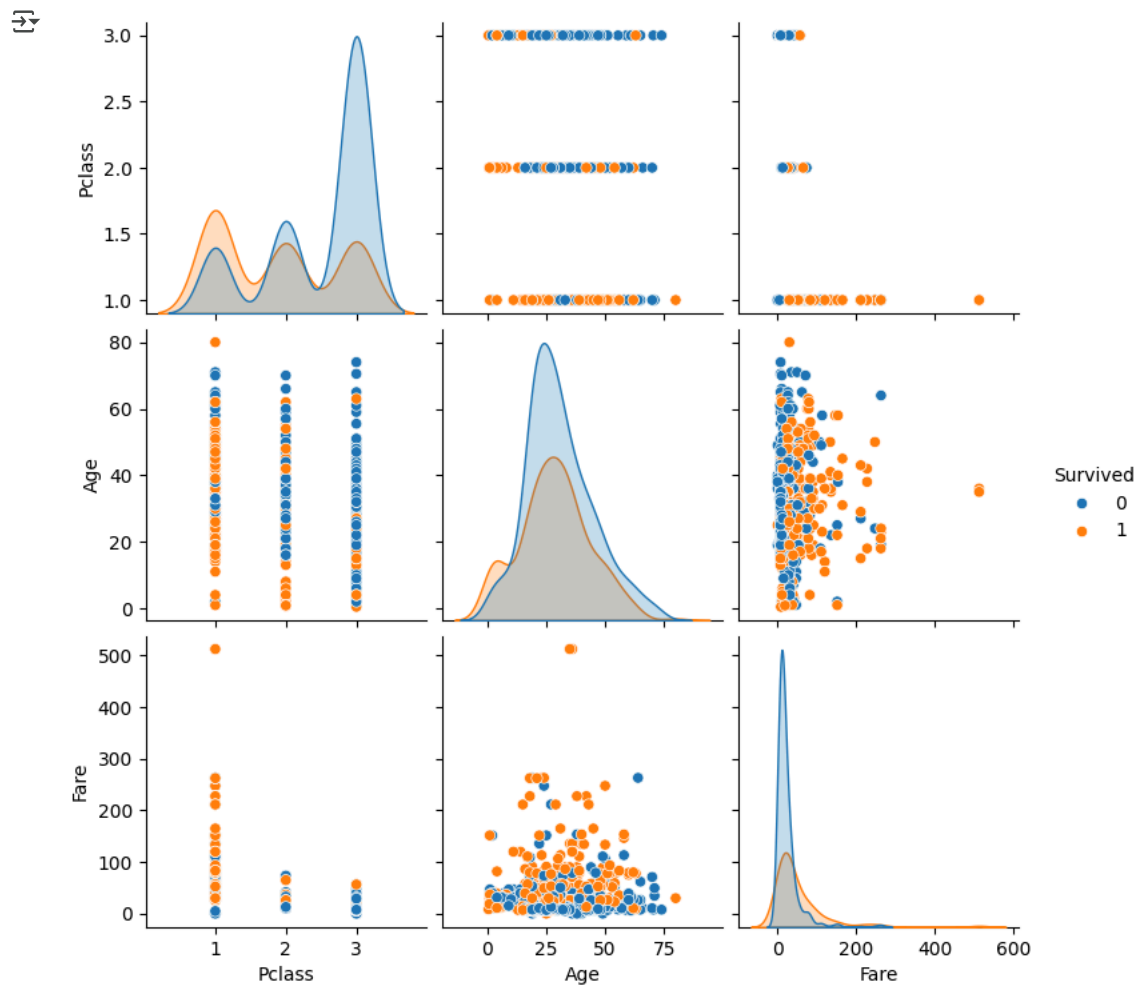
```
Pclass
3    491
1    216
2    184
Name: count, dtype: int64

Sex (Train):
Sex
male    577
female  314
Name: count, dtype: int64

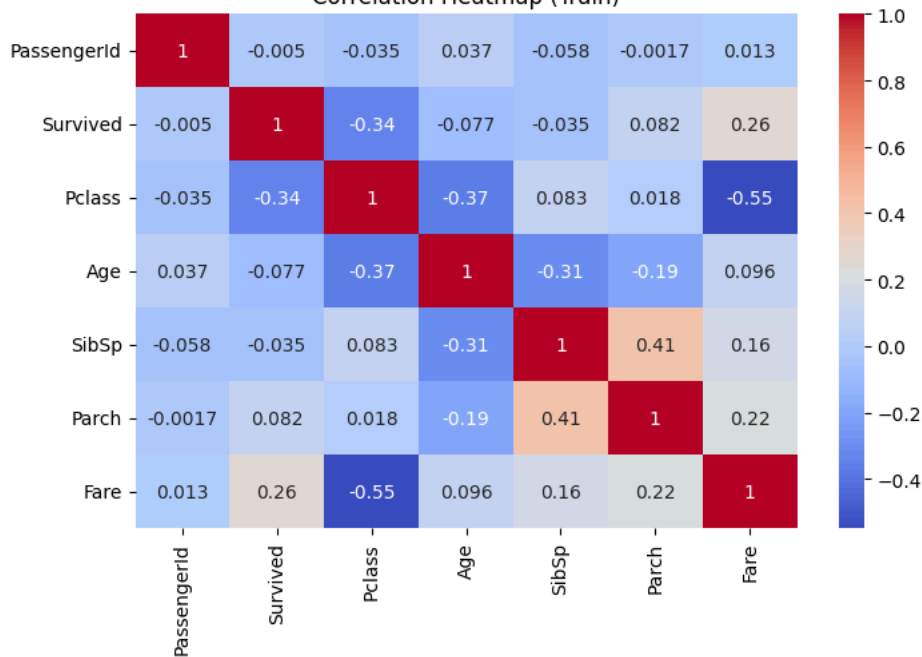
Embarked (Train):
Embarked
S    644
C    168
Q     77
Name: count, dtype: int64

# Pairplot on selected features (train)
sns.pairplot(train_df[['Survived', 'Pclass', 'Age', 'Fare']].dropna(), hue='Survived')
plt.show()

# Correlation heatmap
plt.figure(figsize=(8, 5))
# Select only numerical columns before calculating correlation
sns.heatmap(train_df.select_dtypes(include=np.number).corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap (Train)')
plt.show()
```

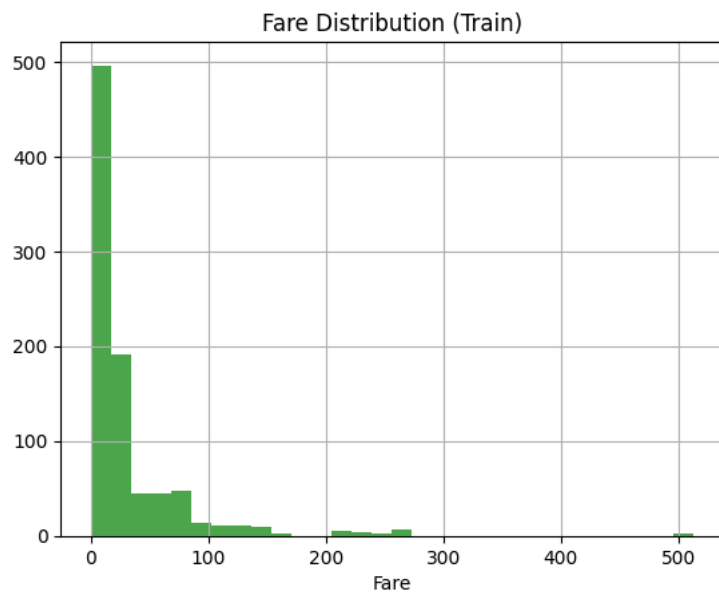
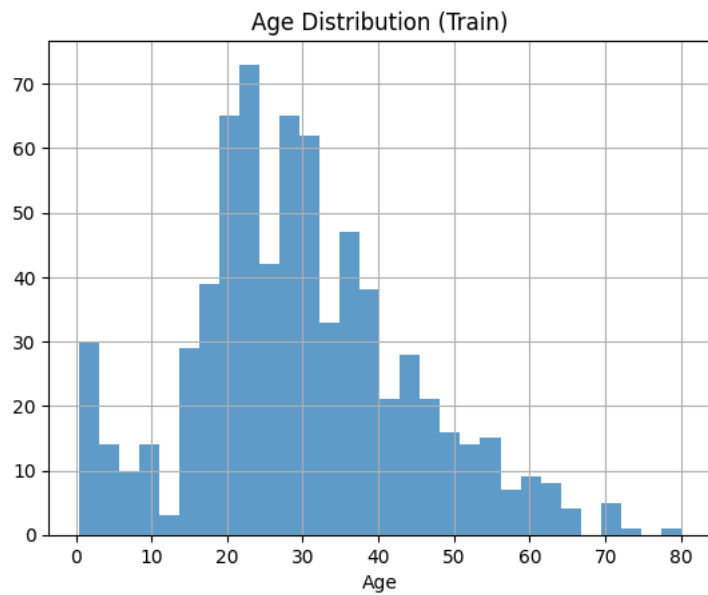


Correlation Heatmap (Train)



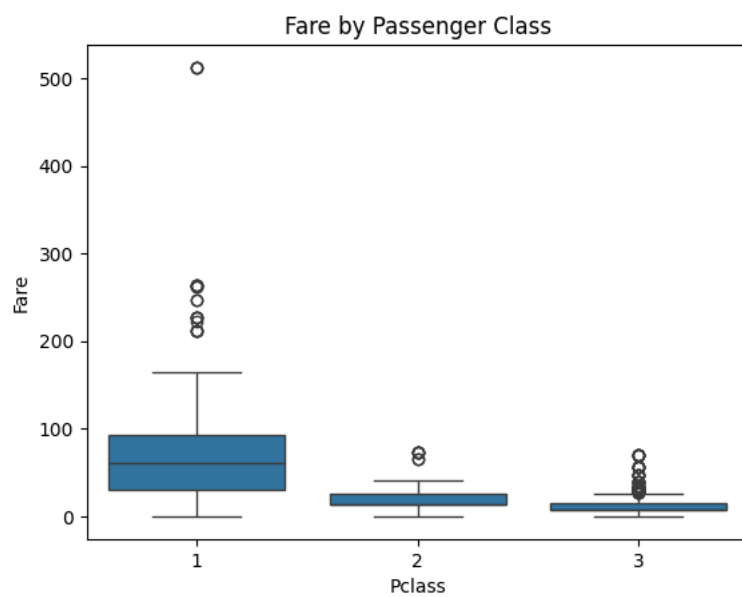
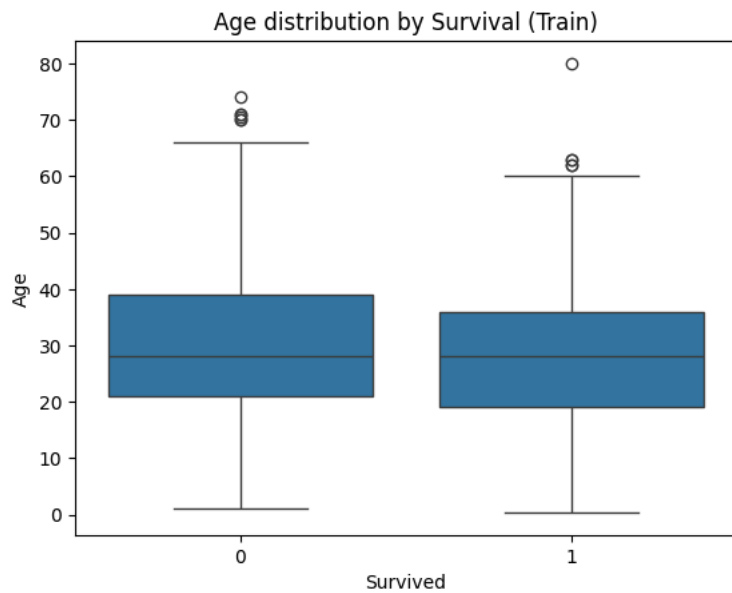
```
train_df['Age'].hist(bins=30, alpha=0.7)
plt.xlabel("Age")
plt.title("Age Distribution (Train)")
plt.show()
```

```
train_df['Fare'].hist(bins=30, alpha=0.7, color='g')
plt.xlabel("Fare")
plt.title("Fare Distribution (Train)")
plt.show()
```



```
sns.boxplot(data=train_df, x='Survived', y='Age')  
plt.title("Age distribution by Survival (Train)")  
plt.show()
```

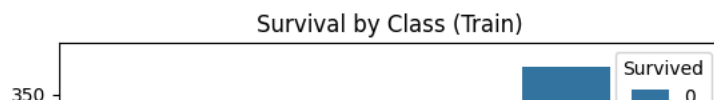
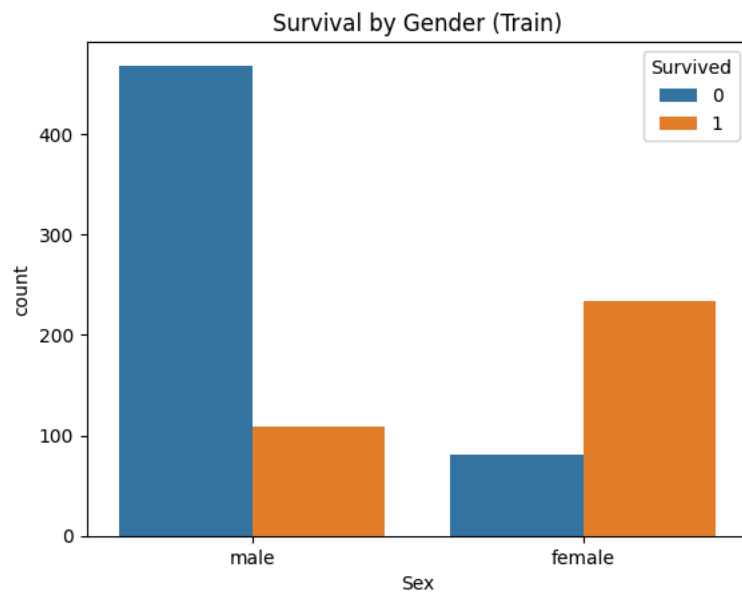
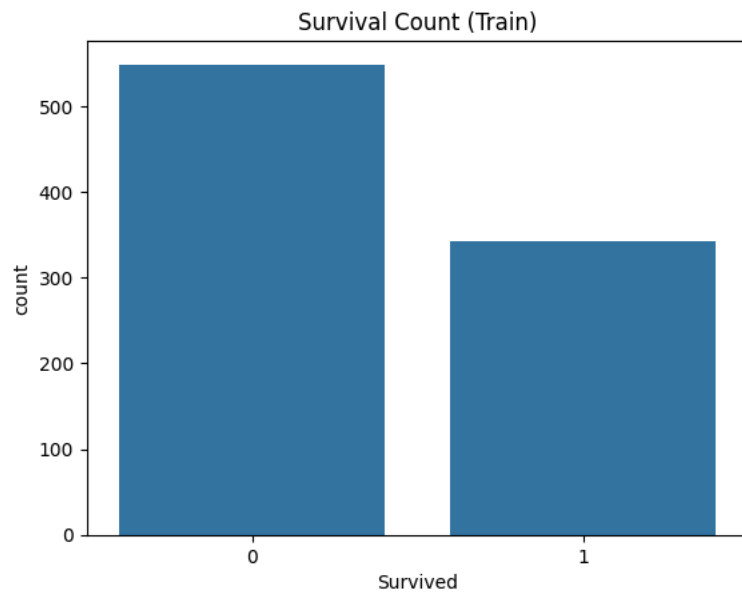
```
sns.boxplot(data=train_df, x='Pclass', y='Fare')  
plt.title("Fare by Passenger Class")  
plt.show()
```



```
sns.countplot(x='Survived', data=train_df)
plt.title("Survival Count (Train)")
plt.show()
```

```
sns.countplot(x='Sex', hue='Survived', data=train_df)
plt.title("Survival by Gender (Train)")
plt.show()
```

```
sns.countplot(x='Pclass', hue='Survived', data=train_df)
plt.title("Survival by Class (Train)")
plt.show()
```



```
test_df.info()
display(test_df.describe(include='all'))
```

```
gender_submission_df.head()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
```