```python
# Step 1: Upload the CSV File
from google.colab import files
uploaded = files.upload()  # Select and upload 'train.csv'


# Step 2: Import Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


# Step 3: Load Dataset (After Uploading)
data = pd.read_csv('/content/train.csv')


# Step 4: Basic Information
print("✅ Shape of the dataset:", data.shape)
print("\n✅ Data Types:\n", data.dtypes)
print("\n✅ First 5 Rows:\n", data.head())
print("\n✅ Statistical Summary:\n", data.describe())


# Step 5: Check Missing Values
print("\n✅ Missing Values:\n", data.isnull().sum())


# Step 6: Histograms for All Numerical Features
data.hist(figsize=(15, 10))
plt.suptitle('Feature Distributions', fontsize=20)
plt.show()


# Step 7: Boxplots for Numerical Columns
for column in data.select_dtypes(include=['float64', 'int64']).columns:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=data[column])
```

```
    plt.title(f'Boxplot of {column}', fontsize=16)

    plt.show()
```

```
# Step 8: Correlation Heatmap (Corrected)

# Only select numeric columns

numeric_data = data.select_dtypes(include=['float64', 'int64'])


# Calculate correlation matrix

correlation_matrix = numeric_data.corr()


# Plot heatmap

plt.figure(figsize=(12, 10))

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")

plt.title('Correlation Heatmap', fontsize=20)

plt.show()
```

<center>"**OUT PUT**"</center>

⬛ **train.csv**(text/csv) - 61194 bytes, last modified: 28/4/2025 - 100% done

Saving train.csv to train (1).csv

✅ Shape of the dataset: (891, 12)

✅ Data Types:

```
 PassengerId     int64

Survived        int64

Pclass          int64

Name            object

Sex             object

Age             float64

SibSp           int64

Parch           int64

Ticket          object

Fare            float64
```

Cabin   object

Embarked   object

dtype: object

✅ First 5 Rows:

     PassengerId Survived Pclass \

0   1   0   3

1   2   1   1

2   3   1   3

3   4   1   1

4   5   0   3

                    Name   Sex   Age SibSp \

0   Braund, Mr. Owen Harris   male 22.0   1

1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0   1

2   Heikkinen, Miss. Laina female 26.0   0

3   Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0   1

4   Allen, Mr. William Henry   male 35.0   0

   Parch   Ticket   Fare Cabin Embarked

0   0   A/5 21171 7.2500 NaN   S

1   0   PC 17599 71.2833 C85   C

2   0 STON/O2. 3101282 7.9250 NaN   S

3   0   113803 53.1000 C123   S

4   0   373450 8.0500 NaN   S

✅ Statistical Summary:

     PassengerId   Survived   Pclass   Age   SibSp \

count 891.000000 891.000000 891.000000 714.000000 891.000000

mean   446.000000   0.383838   2.308642 29.699118   0.523008

std   257.353842   0.486592   0.836071 14.526497   1.102743

min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

✅ Missing Values:

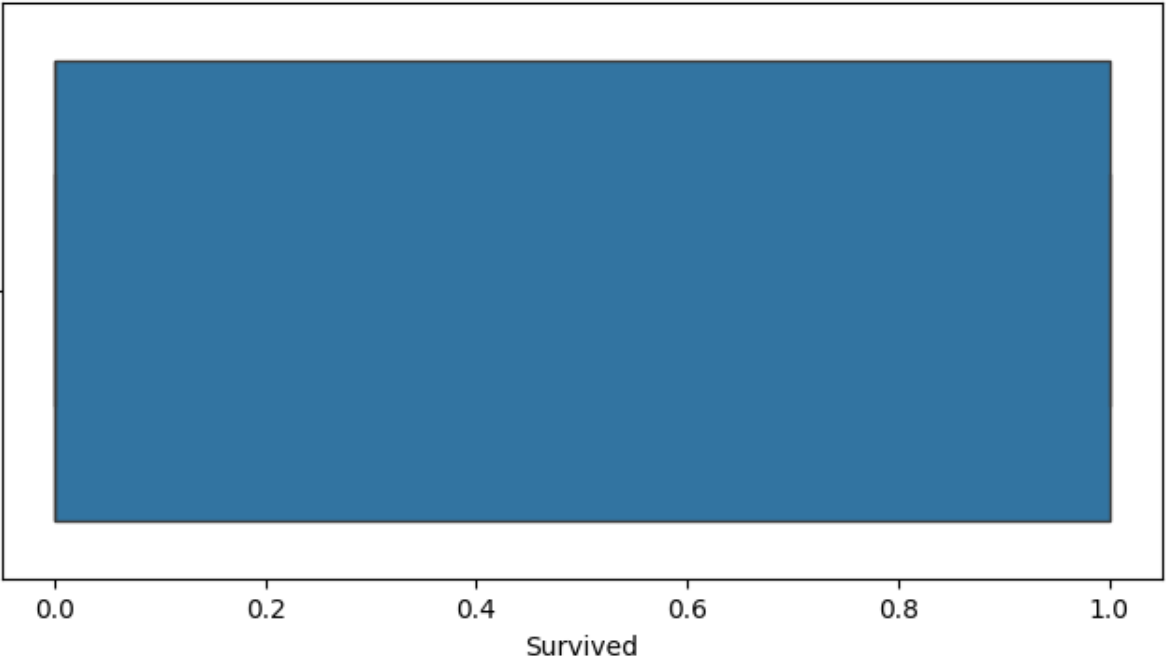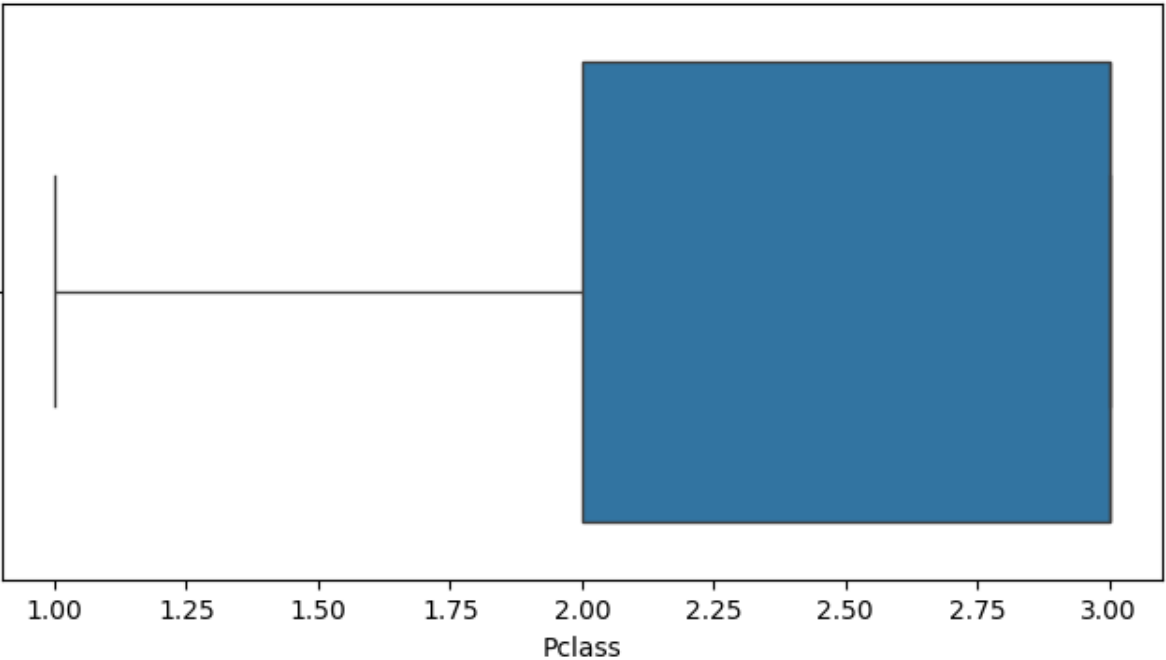|	
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2

dtype: int64

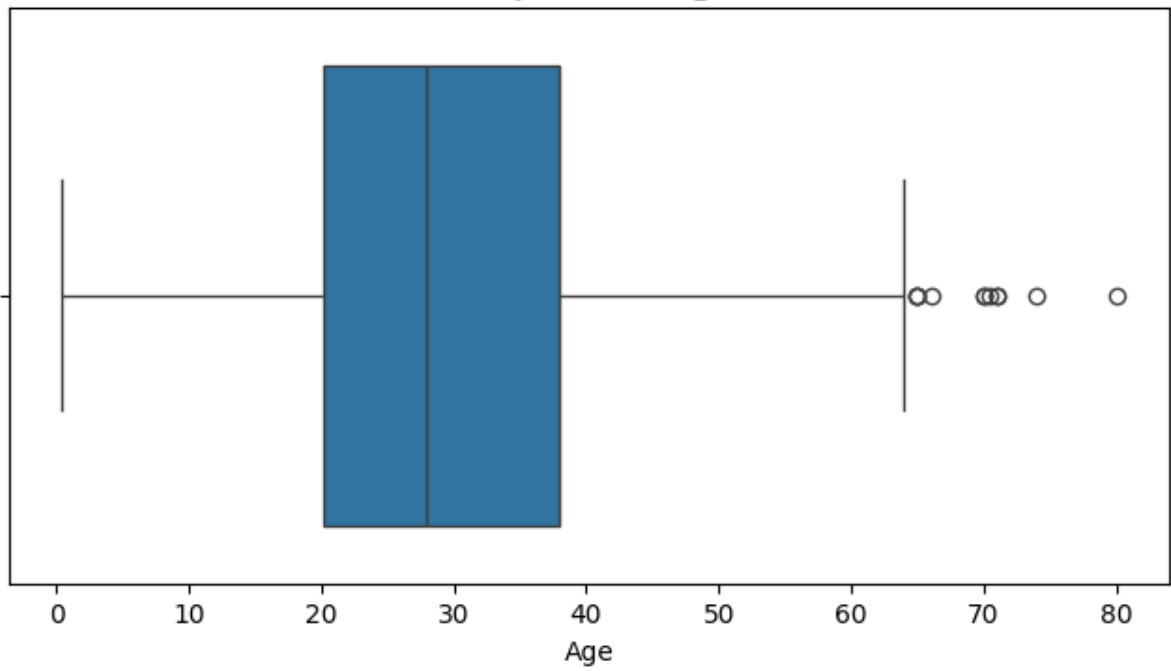# Feature Distributions
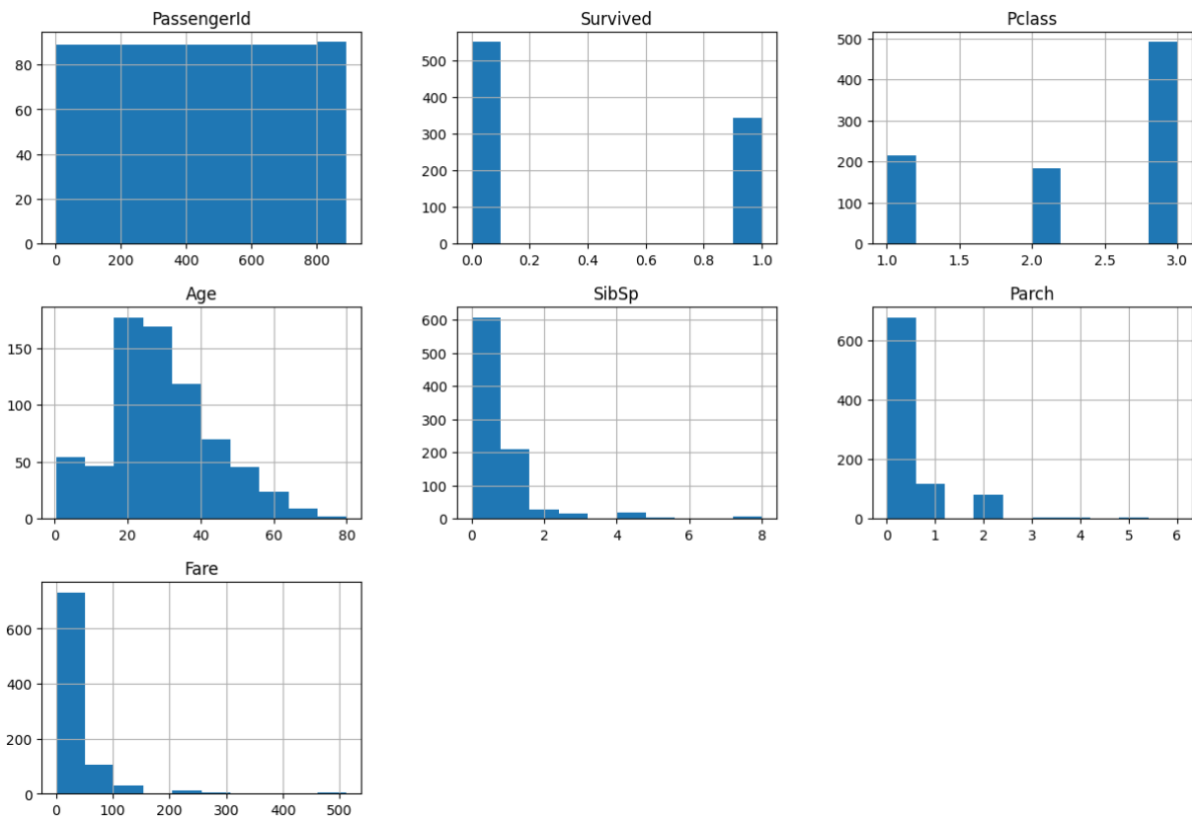
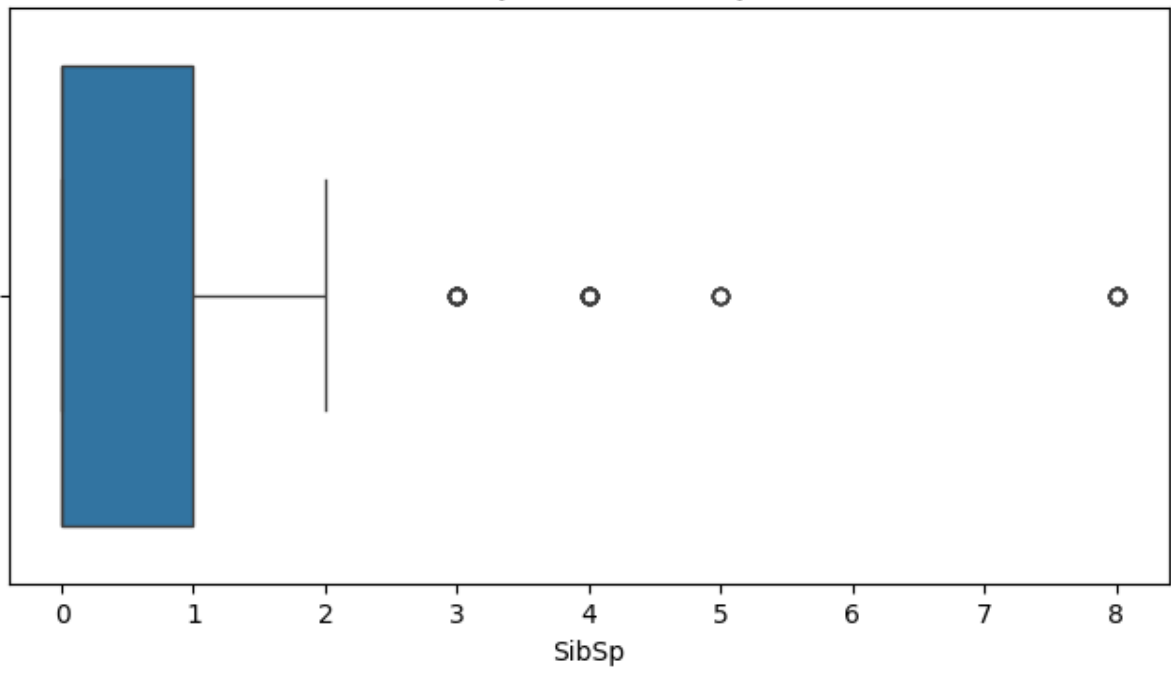# Boxplot of PassengerId
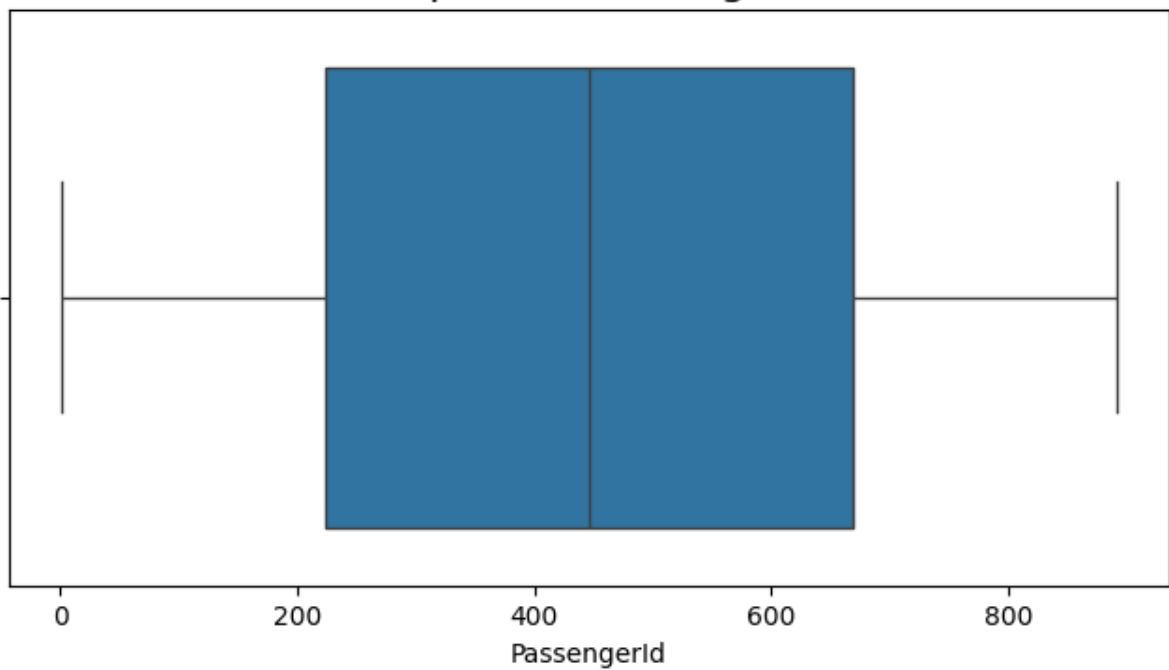
Boxplot of Survived
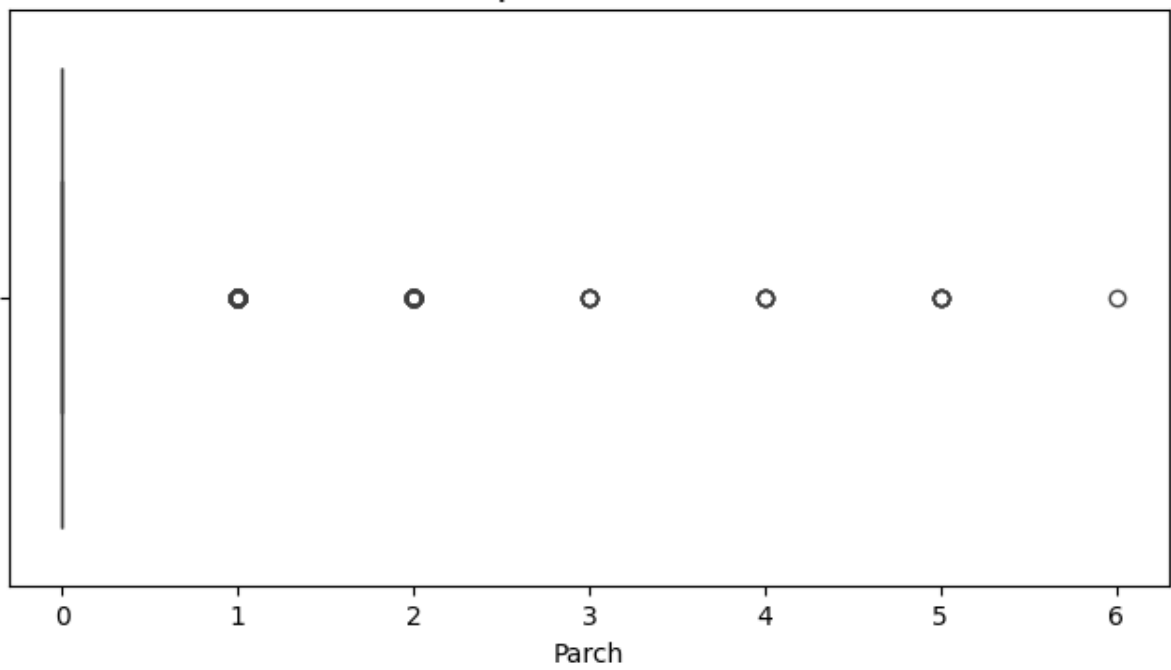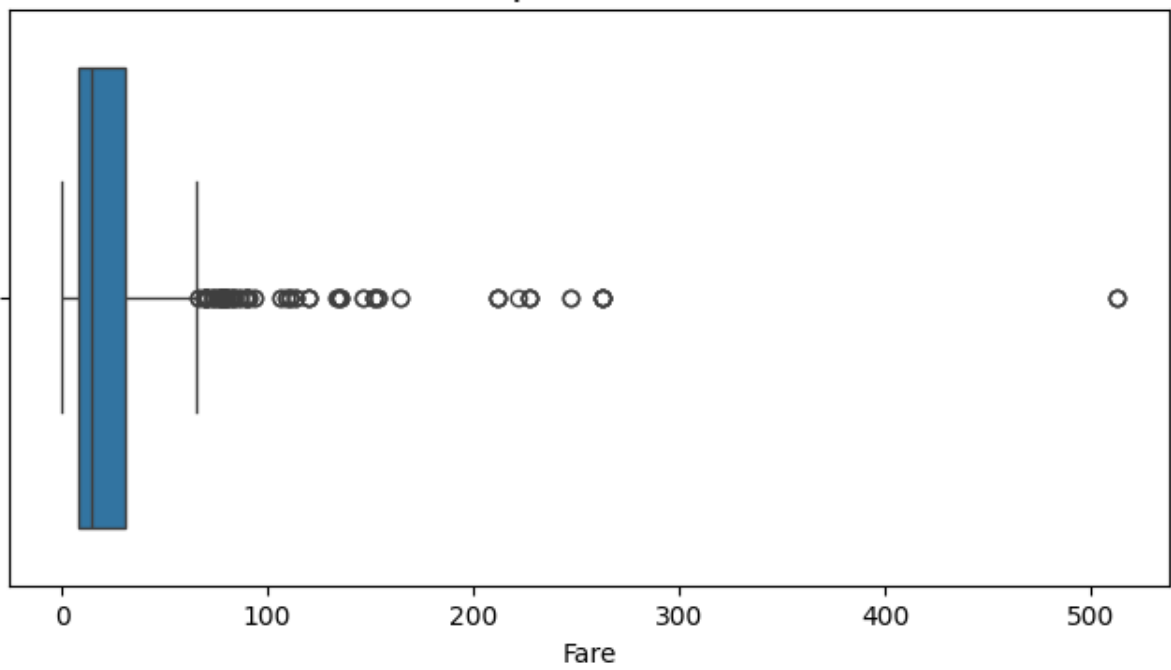

Boxplot of Pclass

Boxplot of Age

Boxplot of SibSp

Boxplot of Parch

Boxplot of Fare

Correlation Heatmap