

```
# Import the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder, StandardScaler

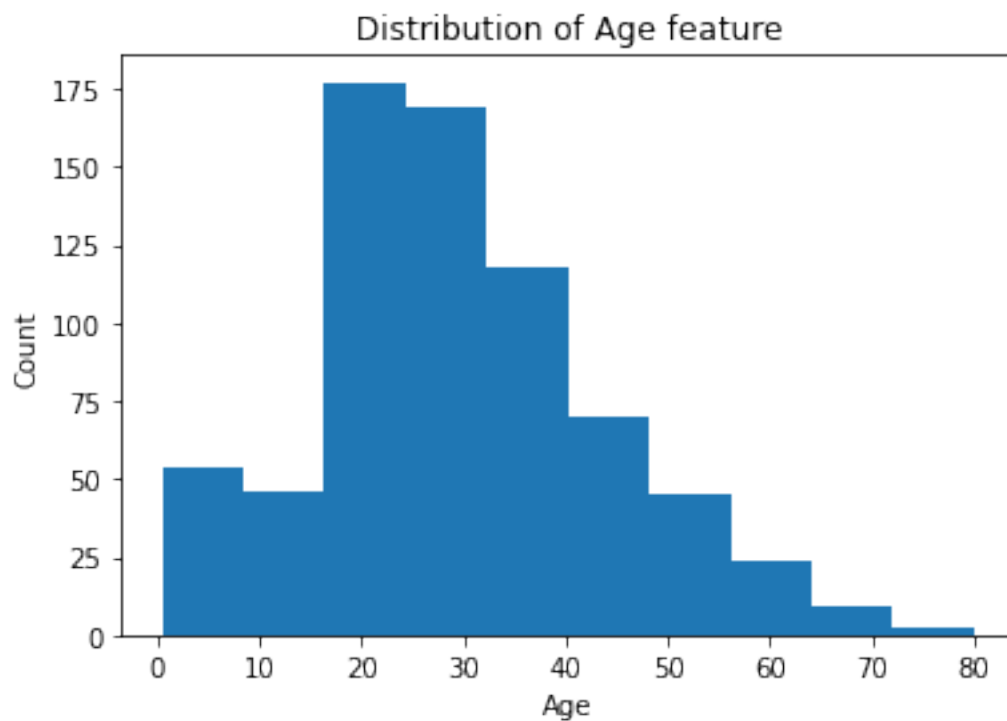
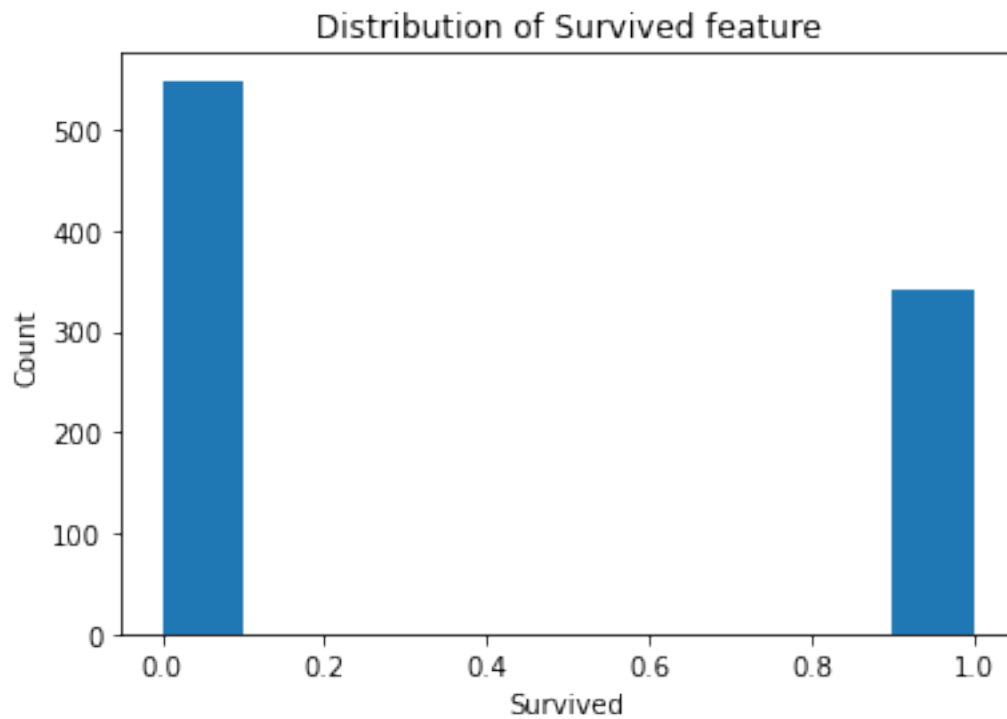
# Importing the dataset.
df = pd.read_csv('Titanic-Dataset.csv')

# Checking for Null Values
df.isnull().sum()

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

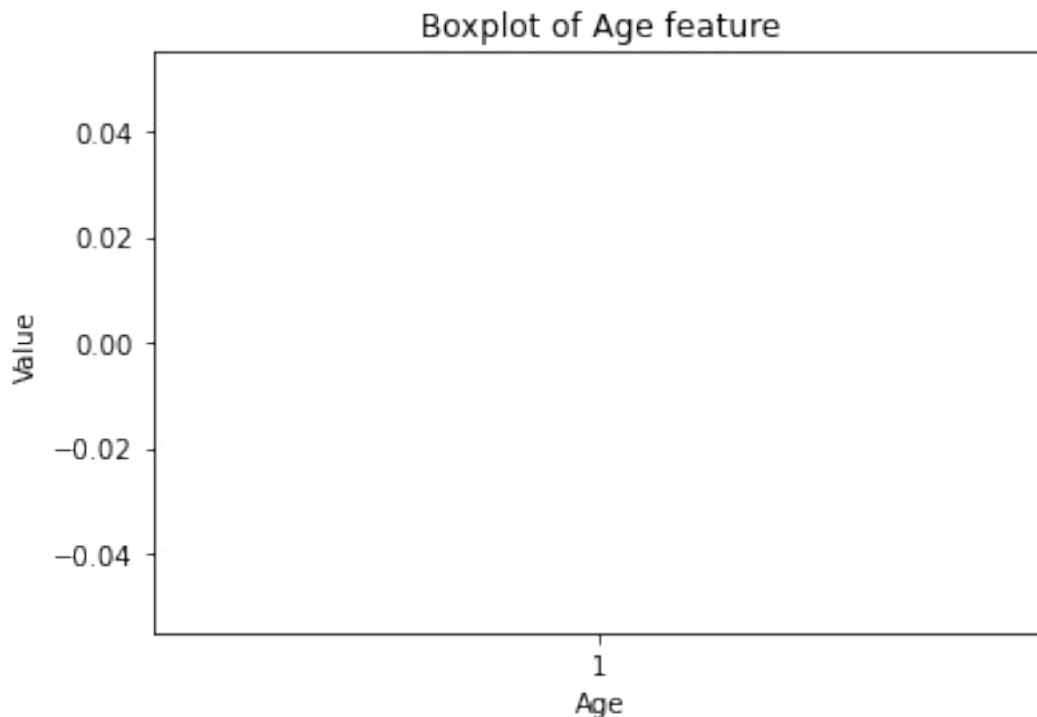
# Data Visualization
# Plot the distribution of the 'Survived' feature
plt.hist(df['Survived'])
plt.xlabel('Survived')
plt.ylabel('Count')
plt.title('Distribution of Survived feature')
plt.show()

# Plot the distribution of the 'Age' feature
plt.hist(df['Age'])
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Distribution of Age feature')
plt.show()
```



```
# Outlier Detection
plt.boxplot(df['Age'])
plt.xlabel('Age')
plt.ylabel('Value')
```

```
plt.title('Boxplot of Age feature')
plt.show()
```



Splitting the data into Dependent and Independent variables

```
X_train = df.drop('Survived', axis=1)
```

```
y_train = df['Survived']
```

```
print(X_train)
```

```
print(y_train)
```

| | PassengerId | Pclass | |
|--------|-------------|--------|---|
| Name \ | | | |
| 0 | 1 | 3 | Braund, Mr. Owen |
| Harris | | | |
| 1 | 2 | 1 | Cumings, Mrs. John Bradley (Florence Briggs |
| Th... | | | |
| 2 | 3 | 3 | Heikkinen, Miss. |
| Laina | | | |
| 3 | 4 | 1 | Futrelle, Mrs. Jacques Heath (Lily May |
| Peel) | | | |
| 4 | 5 | 3 | Allen, Mr. William |
| Henry | | | |
| .. | ... | ... | |
| ... | | | |
| 886 | 887 | 2 | Montvila, Rev. |
| Juozas | | | |
| 887 | 888 | 1 | Graham, Miss. Margaret |
| Edith | | | |

| | | | |
|----------|-----|---|---------------------------------|
| 888 | 889 | 3 | Johnston, Miss. Catherine Helen |
| "Carrie" | | | |
| 889 | 890 | 1 | Behr, Mr. Karl |
| Howell | | | |
| 890 | 891 | 3 | Dooley, Mr. |
| Patrick | | | |

| | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin |
|----------|--------|------|-------|-------|------------------|---------|-------|
| Embarked | | | | | | | |
| 0 | male | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN |
| S | | | | | | | |
| 1 | female | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 |
| C | | | | | | | |
| 2 | female | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN |
| S | | | | | | | |
| 3 | female | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 |
| S | | | | | | | |
| 4 | male | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN |
| S | | | | | | | |
| .. | ... | ... | ... | ... | ... | ... | ... |
| ... | | | | | | | |
| 886 | male | 27.0 | 0 | 0 | 211536 | 13.0000 | NaN |
| S | | | | | | | |
| 887 | female | 19.0 | 0 | 0 | 112053 | 30.0000 | B42 |
| S | | | | | | | |
| 888 | female | NaN | 1 | 2 | W./C. 6607 | 23.4500 | NaN |
| S | | | | | | | |
| 889 | male | 26.0 | 0 | 0 | 111369 | 30.0000 | C148 |
| C | | | | | | | |
| 890 | male | 32.0 | 0 | 0 | 370376 | 7.7500 | NaN |
| Q | | | | | | | |

[891 rows x 11 columns]

| | |
|-----|---|
| 0 | 0 |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 0 |
| .. | |
| 886 | 0 |
| 887 | 1 |
| 888 | 0 |
| 889 | 1 |
| 890 | 0 |

Name: Survived, Length: 891, dtype: int64

Perform Encoding

categorical_features = ['Pclass', 'Sex', 'Embarked']

for col in categorical_features:

```

encoder = LabelEncoder()
X_train[col] = encoder.fit_transform(X_train[col])
df[col] = encoder.transform(df[col])

```

```

print(encoder)
print(X_train)

```

```
LabelEncoder()
```

| | PassengerId | Pclass | |
|----------|-------------|--------|---|
| Name \ | | | |
| 0 | 1 | 2 | Braund, Mr. Owen |
| Harris | | | |
| 1 | 2 | 0 | Cumings, Mrs. John Bradley (Florence Briggs |
| Th... | | | |
| 2 | 3 | 2 | Heikkinen, Miss. |
| Laina | | | |
| 3 | 4 | 0 | Futrelle, Mrs. Jacques Heath (Lily May |
| Peel) | | | |
| 4 | 5 | 2 | Allen, Mr. William |
| Henry | | | |
| .. | ... | ... | |
| ... | | | |
| 886 | 887 | 1 | Montvila, Rev. |
| Juozas | | | |
| 887 | 888 | 0 | Graham, Miss. Margaret |
| Edith | | | |
| 888 | 889 | 2 | Johnston, Miss. Catherine Helen |
| "Carrie" | | | |
| 889 | 890 | 0 | Behr, Mr. Karl |
| Howell | | | |
| 890 | 891 | 2 | Dooley, Mr. |
| Patrick | | | |

| | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin |
|----------|-----|------|-------|-------|------------------|---------|-------|
| Embarked | | | | | | | |
| 0 | 1 | 22.0 | 1 | 0 | A/5 21171 | 7.2500 | NaN |
| 2 | | | | | | | |
| 1 | 0 | 38.0 | 1 | 0 | PC 17599 | 71.2833 | C85 |
| 0 | | | | | | | |
| 2 | 0 | 26.0 | 0 | 0 | STON/O2. 3101282 | 7.9250 | NaN |
| 2 | | | | | | | |
| 3 | 0 | 35.0 | 1 | 0 | 113803 | 53.1000 | C123 |
| 2 | | | | | | | |
| 4 | 1 | 35.0 | 0 | 0 | 373450 | 8.0500 | NaN |
| 2 | | | | | | | |
| .. | ... | ... | ... | ... | ... | ... | ... |
| . | | | | | | | |
| 886 | 1 | 27.0 | 0 | 0 | 211536 | 13.0000 | NaN |
| 2 | | | | | | | |
| 887 | 0 | 19.0 | 0 | 0 | 112053 | 30.0000 | B42 |

```

2
888    0    NaN    1    2    W./C. 6607    23.4500    NaN
2
889    1    26.0    0    0    111369    30.0000    C148
0
890    1    32.0    0    0    370376    7.7500    NaN
1

```

[891 rows x 11 columns]

Feature Scaling

```

scaler = StandardScaler()
X_train[['Age', 'Fare']] = scaler.fit_transform(X_train[['Age',
'Fare']])
print(X_train)

```

| | PassengerId | Pclass | Name | Sex | Age |
|---------|-------------|--------|------------------------------|-----|-----------|
| SibSp \ | | | | | |
| 405 | 406 | 1 | Gale, Mr. Shadrach | 1 | 0.315877 |
| 1 | | | | | |
| 547 | 548 | 1 | Padro y Manent, Mr. Julian | 1 | NaN |
| 0 | | | | | |
| 170 | 171 | 0 | Van der hoef, Mr. Wyckoff | 1 | 2.163216 |
| 0 | | | | | |
| 857 | 858 | 0 | Daly, Mr. Peter Denis | 1 | 1.479016 |
| 0 | | | | | |
| 295 | 296 | 0 | Lewy, Mr. Ervin G | 1 | NaN |
| 0 | | | | | |
| .. | ... | ... | ... | ... | ... |
| ... | | | | | |
| 173 | 174 | 2 | Sivola, Mr. Antti Wilhelm | 1 | -0.573583 |
| 0 | | | | | |
| 613 | 614 | 2 | Horgan, Mr. John | 1 | NaN |
| 0 | | | | | |
| 107 | 108 | 2 | Moss, Mr. Albert Johan | 1 | NaN |
| 0 | | | | | |
| 788 | 789 | 2 | Dean, Master. Bertram Vere | 1 | -1.941982 |
| 1 | | | | | |
| 816 | 817 | 2 | Heininen, Miss. Wendla Maria | 0 | -0.436743 |
| 0 | | | | | |

| | Parch | Ticket | Fare | Cabin | Embarked |
|-----|-------|-------------------|-----------|-------|----------|
| 405 | 0 | 28664 | -0.234297 | NaN | 2 |
| 547 | 0 | SC/PARIS 2146 | -0.362113 | NaN | 0 |
| 170 | 0 | 111240 | -0.010451 | B19 | 2 |
| 857 | 0 | 113055 | -0.134910 | E17 | 2 |
| 295 | 0 | PC 17612 | -0.113943 | NaN | 0 |
| .. | ... | ... | ... | ... | ... |
| 173 | 0 | STON/O 2. 3101280 | -0.468440 | NaN | 2 |
| 613 | 0 | 370377 | -0.471574 | NaN | 1 |

| | | | | | | |
|-----|---|----------|-----------|-----------|-----|---|
| 107 | 0 | | 312991 | -0.471126 | NaN | 2 |
| 788 | 2 | | C.A. 2315 | -0.241908 | NaN | 2 |
| 816 | 0 | STON/O2. | 3101290 | -0.468440 | NaN | 2 |

[501 rows x 11 columns]

Splitting Data into Train and Test

```
X_train, X_test, y_train, y_test = train_test_split(X_train, y_train,
test_size=0.25, random_state=42)
```

```
print(X_train)
```

```
print(X_test)
```

```
print(y_train)
```

```
print(y_test)
```

| | PassengerId | Pclass | Name | Sex | Age |
|---------|-------------|--------|------------------------------|-----|------|
| SibSp \ | | | | | |
| 405 | 406 | 1 | Gale, Mr. Shadrach | 1 | 34.0 |
| 1 | | | | | |
| 547 | 548 | 1 | Padro y Manent, Mr. Julian | 1 | NaN |
| 0 | | | | | |
| 170 | 171 | 0 | Van der hoef, Mr. Wyckoff | 1 | 61.0 |
| 0 | | | | | |
| 857 | 858 | 0 | Daly, Mr. Peter Denis | 1 | 51.0 |
| 0 | | | | | |
| 295 | 296 | 0 | Lewy, Mr. Ervin G | 1 | NaN |
| 0 | | | | | |
| .. | ... | ... | ... | ... | ... |
| .. | | | | | |
| 173 | 174 | 2 | Sivola, Mr. Antti Wilhelm | 1 | 21.0 |
| 0 | | | | | |
| 613 | 614 | 2 | Horgan, Mr. John | 1 | NaN |
| 0 | | | | | |
| 107 | 108 | 2 | Moss, Mr. Albert Johan | 1 | NaN |
| 0 | | | | | |
| 788 | 789 | 2 | Dean, Master. Bertram Vere | 1 | 1.0 |
| 1 | | | | | |
| 816 | 817 | 2 | Heininen, Miss. Wendla Maria | 0 | 23.0 |
| 0 | | | | | |

| | Parch | Ticket | Fare | Cabin | Embarked |
|-----|-------|-------------------|---------|-------|----------|
| 405 | 0 | 28664 | 21.0000 | NaN | 2 |
| 547 | 0 | SC/PARIS 2146 | 13.8625 | NaN | 0 |
| 170 | 0 | 111240 | 33.5000 | B19 | 2 |
| 857 | 0 | 113055 | 26.5500 | E17 | 2 |
| 295 | 0 | PC 17612 | 27.7208 | NaN | 0 |
| .. | ... | ... | ... | ... | ... |
| 173 | 0 | STON/O 2. 3101280 | 7.9250 | NaN | 2 |
| 613 | 0 | 370377 | 7.7500 | NaN | 1 |
| 107 | 0 | 312991 | 7.7750 | NaN | 2 |
| 788 | 2 | C.A. 2315 | 20.5750 | NaN | 2 |

```

816      0  STON/02. 3101290  7.9250  NaN      2

[501 rows x 11 columns]
  PassengerId  Pclass
Name \
651      652      1      Doling, Miss.
Elsie
167      168      2  Skoog, Mrs. William (Anna Bernhardina
Karlsson)
376      377      2      Landergren, Miss. Aurora
Adelia
392      393      2      Gustafsson, Mr. Johan
Birger
568      569      2      Doharr, Mr.
Tannous
..      ...      ...
...
272      273      1      Mellinger, Mrs. (Elizabeth Anne
Maidment)
366      367      0  Warren, Mrs. Frank Manley (Anna Sophia
Atkinson)
207      208      2      Albimona, Mr. Nassef
Cassem
676      677      2      Sawyer, Mr. Frederick
Charles
847      848      2      Markoff, Mr.
Marin

```

```

      Sex  Age  SibSp  Parch  Ticket  Fare  Cabin  Embarked
651    0  18.0    0      1  231919  23.0000  NaN      2
167    0  45.0    1      4  347088  27.9000  NaN      2
376    0  22.0    0      0  C 7077   7.2500  NaN      2
392    1  28.0    2      0  3101277  7.9250  NaN      2
568    1   NaN    0      0    2686   7.2292  NaN      0
..    ...    ...    ...    ...    ...    ...    ...
272    0  41.0    0      1  250644  19.5000  NaN      2
366    0  60.0    1      0  110813  75.2500  D37      0
207    1  26.0    0      0    2699  18.7875  NaN      0
676    1  24.5    0      0  342826   8.0500  NaN      2
847    1  35.0    0      0  349213   7.8958  NaN      0

```

```

[167 rows x 11 columns]
405    0
547    1
170    0
857    1
295    0
..
173    0
613    0

```



```
107      1
788      1
816      0
Name: Survived, Length: 501, dtype: int64
651      1
167      0
376      1
392      0
568      0
      ..
272      1
366      1
207      1
676      0
847      0
Name: Survived, Length: 167, dtype: int64
```