```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline

df = pd.read_csv("train.csv")

df.shape

df.head()

df.info()

df.describe()

df.isna().sum()

df["Age"] = df["Age"].fillna(df["Age"].mean())

df.isna().sum()

def fun1(value):
    if (value == "male"):
        return 1
    else:
        return 0
```

```
def fun2(value):
    if (value == 'S'):
        return 0
    elif (value == 'C'):
        return 1
    elif (value == 'Q'):
        return 2
    else:
        return 0

df["Sex"] = df["Sex"].apply(fun1)

df["Embarked"] = df["Embarked"].apply(fun2)

df = df.drop("Cabin", axis=1)

df.shape

px.box(df["Sex"], df["Age"], color=df["Survived"])

plt.figure(figsize=(10,7))
box = sns.boxplot(df["Sex"], df["Age"], hue=df["Survived"])
plt.show()
```