Practical 8: Data Visualization I

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the Seaborn library to see if we can find any patterns in the data.
2. Write a code to check how the price of the ticket (column name: 'fare') for each passenger is distributed by plotting a histogram.

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

# Load the Titanic dataset
titanic = sns.load_dataset('titanic')

# Display basic information about the dataset
print(titanic.info())

# Visualize the distribution of passenger ages
plt.figure(figsize=(10, 6))
sns.histplot(titanic['age'].dropna(), bins=30, kde=True) #Kernel Density Estimation, probability density function, Adds the KDE curve
#It plots the frequencies of values in the specified dataset column (titanic['age']).
plt.title('Age Distribution of Passengers')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()

# Visualize the survival rate by gender
plt.figure(figsize=(10, 6))
sns.countplot(data=titanic, x='sex', hue='survived') # count of males and females divided into survivors and non-survivors.
plt.title('Survival Count by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()

# Visualize the survival rate by passenger class
plt.figure(figsize=(10, 6))
sns.countplot(data=titanic, x='class', hue='survived')
```

```
plt.title('Survival Count by Passenger Class')
plt.xlabel('Passenger Class')
plt.ylabel('Count')
plt.show()

# Plot a histogram of the 'fare' column
plt.figure(figsize=(10, 6))
sns.histplot(titanic['fare'].dropna(), bins=50, kde=True)
plt.title('Distribution of Ticket Fares')
plt.xlabel('Fare')
plt.ylabel('Frequency')
plt.show()
```