**DATA SCIENCE TASK 2**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

# Load the dataset

data = pd.read\_csv("C:/Users/Ishita/Desktop/DSTASK2/MOVIE RATING DATASET.csv")

# Display first few rows to understand the structure of data

data.head()

# Clean the data

data.isnull().sum()

# Calculate summary statistics

mean\_rating = data['Movie Rating'].mean()

median\_rating = data['Movie Rating'].median()

mode\_rating = data['Movie Rating'].mode()[0]

print(f"Mean Rating: {mean\_rating}")

print(f"Median Rating: {median\_rating}")

print(f"Mode Rating: {mode\_rating}")

# Top 10 rated movies

top\_rated\_movies = data[['Movie Name', 'Movie Rating']].sort\_values(by='Movie Rating', ascending=False).head(10)

print("Top 10 Rated Movies:")

print(top\_rated\_movies)

# Plotting the histogram

plt.figure(figsize=(10,6))

sns.histplot(data['Movie Rating'], bins=20, kde=True, color='blue', edgecolor='black')

plt.title('Distribution of Movie Ratings')

plt.xlabel('Rating')

plt.ylabel('Frequency')

plt.show()

# Box Plot for movie rating distribution

plt.figure(figsize=(8,6))

sns.boxplot(x=data['Movie Rating'], color='green')

plt.title('Box Plot of Movie Ratings')

plt.xlabel('Rating')

plt.show()



