import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

df = pd.read\_csv('dataset.csv')

df.head()

df.isna().sum()

df['FE Score'] = df["FE Score"].fillna(df["FE Score"].mean())

df['FE Score'].mean()

df['SE Score']=df['SE Score'].fillna(df['SE Score'].median())

df.isna().sum()

plt.figure(figsize=(8, 4))

plt.hist(df['SE Score'].dropna(), bins=15, color='skyblue', edgecolor='black')

plt.title("Original SE Score Distribution")

plt.xlabel("SE Score")

plt.ylabel("Frequency")

plt.show()

df['Log\_SE\_Score'] = df['SE Score'].apply(lambda x: np.log1p(x) if pd.notna(x) else x)

plt.figure(figsize=(8, 4))

plt.hist(df['Log\_SE\_Score'].dropna(), bins=15, color='salmon', edgecolor='black')

plt.title("Log Transformed SE Score Distribution")

plt.xlabel("Log(SE Score)")

plt.ylabel("Frequency")

plt.show()