TIME SERIES ANALYSIS AND FORECASTING

EX:1

AIM:TO IMPLEMENT A PYTHON PROGRAMFOR TIME SERIES ,DATA CLEANING,LOADING,HANDLING,PREPROCESSING TECHNIQUEOF BIRTH RATE EACH COUNTRY

import pandas as pd  
import matplotlib.pyplot as plt  
  
# Load the dataset with correct delimiter and skip bad lines  
file\_path = r"C:\Users\Lenovo\Downloads\API\_SP.DYN.TFRT.IN\_DS2\_EN\_csv\_v2\_162\birthrate.csv"  
df = pd.read\_csv(file\_path, delimiter=',', skiprows=4, on\_bad\_lines='skip')  
  
# Clean the column names by removing extra quotes and spaces  
df.columns = df.columns.str.replace('"', '').str.strip()  
  
# Preview the column names to ensure they've been cleaned  
print("Cleaned Column Names:")  
print(df.columns)  
  
# Filter only the relevant columns (Country Name + the years 1960–2023)  
df\_clean = df[['Country Name', '1960', '1961', '1962', '1963', '1964', '1965', '1966', '1967', '1968', '1969',   
               '1970', '1971', '1972', '1973', '1974', '1975', '1976', '1977', '1978', '1979', '1980', '1981',   
               '1982', '1983', '1984', '1985', '1986', '1987', '1988', '1989', '1990', '1991', '1992', '1993',   
               '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003', '2004', '2005',   
               '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017',   
               '2018', '2019', '2020', '2021', '2022', '2023']]  
  
# Reshape the dataset to long format (years in 'Year' and birth rates in 'Birth Rate')  
df\_long = pd.melt(df\_clean, id\_vars=["Country Name"], var\_name="Year", value\_name="Birth Rate")  
  
# Convert 'Year' to integer and 'Birth Rate' to numeric, handling errors gracefully  
df\_long['Year'] = pd.to\_numeric(df\_long['Year'], errors='coerce')  
df\_long['Birth Rate'] = pd.to\_numeric(df\_long['Birth Rate'], errors='coerce')  
  
# Drop rows with missing birth rate data  
df\_long = df\_long.dropna(subset=['Birth Rate'])  
  
# Preview the reshaped data  
print(df\_long.head())  
  
# Choose a country to visualize (e.g., 'United States')  
country\_name = 'United States'  # Change this to any country present in your dataset  
df\_country = df\_long[df\_long['Country Name'] == country\_name]  
  
# Plotting the birth rate over time for the selected country  
plt.figure(figsize=(10, 6))  
plt.plot(df\_country['Year'], df\_country['Birth Rate'], marker='o', color='b', label=country\_name)  
plt.title(f'Birth Rate Over Time for {country\_name}')  
plt.xlabel('Year')  
plt.ylabel('Birth Rate (per 1000 people)')  
plt.grid(True)  
plt.legend()  
plt.xticks(rotation=45)  
plt.tight\_layout()  
  
# Show the plot  
plt.show()

#data cleaning and loading  
  
import pandas as pd  
file\_path = r"C:\Users\Lenovo\Downloads\API\_SP.DYN.TFRT.IN\_DS2\_EN\_csv\_v2\_162\birthrate.csv"  
df = pd.read\_csv(file\_path, delimiter=',', skiprows=4, on\_bad\_lines='skip')  
df.columns = df.columns.str.replace('"', '').str.strip()  
print("Cleaned Column Names:")  
print(df.columns)

#Handling & Preprocessing  
  
df\_long = pd.melt(df[['Country Name'] + [str(year) for year in range(1960, 2024)]],   
                  id\_vars=["Country Name"], var\_name="Year", value\_name="Birth Rate")  
df\_long['Year'] = pd.to\_numeric(df\_long['Year'])  
df\_long['Birth Rate'] = pd.to\_numeric(df\_long['Birth Rate'], errors='coerce')  
df\_long.dropna(subset=['Birth Rate'], inplace=True)  
print(df\_long.head())

#Visualization  
  
import matplotlib.pyplot as plt  
df\_country = df\_long[df\_long['Country Name'] == 'India']  
plt.figure(figsize=(10, 6))  
plt.plot(df\_country['Year'], df\_country['Birth Rate'], marker='o', color='b')  
plt.title('Birth Rate Over Time for India')  
plt.xlabel('Year')  
plt.ylabel('Birth Rate (per 1000 people)')  
plt.grid(True)  
plt.xticks(rotation=45)  
plt.tight\_layout()  
plt.show()

RESULT:THE PROGRAM HAS BEEN SUCESSFULLY EXECUTED