# TIME SERIES DATA CLEANING, LOADING, AND PRE-PROCESSING TECHNIQUES

## AIM:

To implement programs for time series data cleaning, loading, handling, and pre-processing techniques.

## PROCEDURE:

• Load the data - Read the dataset using Pandas.

• Clean the column names - Remove unwanted characters and spaces.

• Handle missing values - Convert non-numeric values to NaN and drop missing data.

• Reshape data for time series - Convert wide format data into long format.

• Convert data types - Ensure numerical columns are properly formatted.

• Visualize the data - Plot time series data using Matplotlib.

## CODE:

import pandas as pd  
import matplotlib.pyplot as plt  
  
# Load the dataset  
file\_path = r"C:\Users\admin\Downloads\API\_SP.DYN.TFRT.IN\_DS2\_EN\_csv\_v2\_162\API\_SP.DYN.TFRT.IN\_DS2\_EN\_csv\_v2\_162.csv"  
df = pd.read\_csv(file\_path, delimiter=',', skiprows=4, on\_bad\_lines='skip')  
  
# Clean column names  
df.columns = df.columns.str.replace('"', '').str.strip()  
print("Cleaned Column Names:")  
print(df.columns)  
  
# Reshape data  
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  
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 (Line Plot)')  
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 successfully loads, cleans, processes, and visualizes time series data, demonstrating data handling techniques for better insights.