**EX:No.3 221501028**

**07/01/25**

**IMPLEMENTING PROGRAM FOR TIME SERIES DATA CLEANING, LOADING TIME SERIES DATA AND PREPROCESSING TECHNIQUE**

**AIM:**  
To preprocess and visualize time series data for climate change indicators.

**ALGORITHM:**

1. Load the time series dataset and set the appropriate column as the index for easy time-based operations.
2. Handle missing values using forward fill to maintain data continuity.
3. Remove duplicate entries to ensure data integrity.
4. Generate lag features to analyze past trends and their influence on current values.
5. Visualize the cleaned dataset using line plots for better understanding of trends over time.

**PROCESS**:

*# Load Data and Set 'climate' as the Index*

**import** pandas **as** pd

**def** load\_timeseries\_data(file\_path: str):

df **=** pd**.**read\_csv(file\_path)

df['Year'] **=** pd**.**to\_datetime(df['Year'], format**=**'%Y')

df**.**set\_index('Year', inplace**=True**)

**return** df

df **=** load\_timeseries\_data('C:/Users/Lenovo/Downloads/Climate\_Change\_Indicators.csv')

print(df**.**head())

Global Average Temperature (°C) CO2 Concentration (ppm) \

Year

1948-01-01 13.17 397.04

1996-01-01 13.10 313.17

2015-01-01 14.67 311.95

1966-01-01 14.79 304.25

1992-01-01 13.15 354.52

Sea Level Rise (mm) Arctic Ice Area (million km²)

Year

1948-01-01 116.25 5.97

1996-01-01 277.92 9.66

2015-01-01 290.32 8.40

1966-01-01 189.71 11.83

1992-01-01 14.84 11.23

In [7]:

*# Clean and Preprocess the Data*

*# Clean missing values (e.g., forward fill missing values)*

**def** clean\_missing\_values(df):

df **=** df**.**fillna(method**=**'ffill') *# Forward fill missing values*

**return** df

*# Remove duplicates if any*

**def** remove\_duplicates(df):

df **=** df**.**drop\_duplicates()

**return** df

*# Example usage*

df\_cleaned **=** clean\_missing\_values(df)

df\_cleaned **=** remove\_duplicates(df\_cleaned)

*# Visualize the time series data*

**import** matplotlib.pyplot **as** plt

**def** plot\_timeseries(df):

df**.**plot(figsize**=**(10, 6))

plt**.**title("Climate Change Indicators")

plt**.**xlabel('Year')

plt**.**ylabel('Values')

plt**.**show()

*# Plot the cleaned data*

plot\_timeseries(df\_cleaned)

In [8]:

*# Create lag features (e.g., lag of 1 year for temperature)*

**def** create\_lag\_features(df, lag**=**1):

df['lag\_1'] **=** df['Global Average Temperature (°C)']**.**shift(lag)

**return** df

*# Example usage*

df\_with\_lags **=** create\_lag\_features(df\_cleaned)

plot\_timeseries(df\_with\_lags)

In [9]:

**def** preprocess\_timeseries(file\_path):

*# Step 1: Load the data*

df **=** load\_timeseries\_data(file\_path)

*# Step 2: Clean the data*

df **=** clean\_missing\_values(df)

df **=** remove\_duplicates(df)

*# Step 3: Feature Engineering (Create lag features, etc.)*

df **=** create\_lag\_features(df)

*# Step 4: Visualize the cleaned data*

plot\_timeseries(df)

**return** df

*# Example usage*

df\_preprocessed **=** preprocess\_timeseries('C:/Users/Lenovo/Downloads/Climate\_Change\_Indicators.csv')

Global Average Temperature (°C) CO2 Concentration (ppm) \

Year

1948-01-01 13.17 397.04

1996-01-01 13.10 313.17

2015-01-01 14.67 311.95

1966-01-01 14.79 304.25

1992-01-01 13.15 354.52

Sea Level Rise (mm) Arctic Ice Area (million km²)

Year

1948-01-01 116.25 5.97

1996-01-01 277.92 9.66

2015-01-01 290.32 8.40

1966-01-01 189.71 11.83

1992-01-01 14.84 11.23



**RESULT:**  
The preprocessing and visualization of time series data for climate change indicators were successfully implemented.