# Developing a Linear Regression Model for Forecasting Time Series Data

## Aim:

To develop a linear regression model to forecast birth rates over time using historical data.

## Procedure:

• Load the Dataset: Import the dataset containing historical birth rate data.

• Data Cleaning: Remove unnecessary characters, rename columns, and handle missing values.

• Data Reshaping: Convert the dataset into a long format for analysis.

• Data Filtering: Select data specific to India.

• Model Training: Train a linear regression model using year as the independent variable and birth rate as the dependent variable.

• Forecasting: Predict future birth rates for the years 2024 to 2030.

• Visualization: Plot the actual data, regression line, and predicted values.

• Result Analysis: Print and interpret the predicted values.

## Code Implementation:

import pandas as pd  
import matplotlib.pyplot as plt  
from sklearn.linear\_model import LinearRegression  
import numpy as np  
  
# Load the dataset  
file\_path = r"C:\Users\Lenovo\Downloads\API\_SP.DYN.TFRT.IN\_DS2\_EN\_csv\_v2\_162 (2)\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()  
  
# 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)  
  
# Filter data for India  
df\_country = df\_long[df\_long['Country Name'] == 'India']  
  
# Prepare data for Linear Regression  
X = df\_country[['Year']].values # Independent variable (Years)  
y = df\_country['Birth Rate'].values # Dependent variable (Birth Rate)  
  
# Train Linear Regression model  
model = LinearRegression()  
model.fit(X, y)  
  
# Forecast future birth rates (2024-2030)  
future\_years = np.array(range(2024, 2031)).reshape(-1, 1)  
future\_predictions = model.predict(future\_years)  
  
# Plot historical data  
plt.figure(figsize=(10, 6))  
plt.scatter(df\_country['Year'], df\_country['Birth Rate'], color='blue', label='Actual Data')  
plt.plot(df\_country['Year'], model.predict(X), color='red', linestyle='--', label='Regression Line')  
  
# Plot predictions  
plt.scatter(future\_years, future\_predictions, color='green', label='Predicted Data')  
plt.plot(future\_years, future\_predictions, color='green', linestyle='--')  
  
# Labels & Title  
plt.title('Birth Rate Forecasting for India (Linear Regression)')  
plt.xlabel('Year')  
plt.ylabel('Birth Rate (per 1000 people)')  
plt.legend()  
plt.grid(True)  
plt.xticks(rotation=45)  
plt.tight\_layout()  
plt.show()  
  
# Print predictions  
for year, rate in zip(future\_years.flatten(), future\_predictions):  
 print(f"Predicted Birth Rate for {year}: {rate:.2f} per 1000 people")

## Results:

The program to develop a linear regression model to forecast birth rates over time using historical data has been executed successfully.