📘 Internship Notebook

🧑🎓 Name: Disha Rajkumar Sojrani
🏢 Internship Title: Data Science Using Python
📌 Topic: GDP And Indian Economy Analysis Project

📖 **Introduction**

This project provides a comprehensive analysis of India's GDP and economy using Python. Through advanced data science techniques like data wrangling, transformation, outlier detection, visual analytics, and regression modeling, this study explores various macroeconomic indicators such as inflation, sectoral contributions, FDI inflows, and poverty rates. The aim is to derive useful insights that support economic understanding, policy-making, and business strategy using real-world data.

🗂️ Table of Contents

✍️ Introduction:-

📘 Problem description

➤The primary goal of our is to analyze India’s GDP and economic growth trends in detail, examining the composition of GDP , sectoral contributions, and the influence of factors such as inflation, foreign investment, government policy, and population growth.

➤Explore how exports, imports, and FDI inflows contribute to GDP growth and affect India’s trade balance and currency reserves.

➤Analyze the impact of fiscal deficit, public debt, and report changes on economic growth, investment, and inflation.

➤Study population growth, poverty rate, and household consumption patterns to understand their relationship with GDP growth and economic development.

📙 Application of Project

1.Policy Making: Helps the government create policies for economic development by identifying growth trends and priority sectors.

2.Investment Insights:Provides investors with information on high-growth sectors, guiding smart investment choices.

3.Public Spending: Assists in deciding where to allocate resources for infrastructure, Agriculture, education, and Services for maximum impact.

4.Business Strategy: Helps businesses understand economic trend ,product launches,or target specific regions.

📗 Data Description

Economic Growth Indicators: Includes GDP (Nominal and Real), GDP growth rate, and key factors like inflation (CPI), unemployment rate, and population growth rate.

Sectoral Contributions: Breakdown of GDP by sectors such as Agriculture, Industry, and Services, showing the contribution of each sector to the economy.

External Trade: Exports and Imports as a percentage of GDP, highlighting India’s trade balance and global economic integration.

Investment and Capital Formation: Data on FDI inflows, Gross Fixed Capital Formation, and Private Sector Credit as indicators of investment activity and economic development.

📚 Basic Data Wrangling/Munging/Prepressing steps

1.Handling Missing Data-Missing values may appear in columns like Oil rents (% of GDP).That can be dropped because very of less values.

2.Data Transformation: Standardize or scale numerical columns for consistency, and apply transformations (e.g., log transformation) on skewed data like FDI Inflows or Public Debt.

3.Data Cleaning: Identify and handle missing values and outliers, ensuring each column is in the correct data type (e.g., converting percentages to decimals if needed).

4.Feature Engineering: Create new features if needed, such as GDP per capita or Year-over-Year changes in key indicators, to provide deeper insights.

✍️ Step-by-Step Data Analysis:-

📥 Step 1: Import Libraries and Load Dataset

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")
print(df)
```

```
    Year  GDP (Nominal)  GDP (Real)   GDP growth (annual %)   \
0   1971          35.58        67.4                1.642930
1   1972          36.06        71.5               -0.553301
2   1973          38.32        85.5                3.295521
3   1974          38.47        99.5                1.185336
4   1975          41.57        98.5                9.149912
5   1976          42.86       102.7                1.663104
6   1977          44.86       121.5                7.254765
7   1978          45.62       137.3                5.712532
8   1979          46.84       153.0               -5.238183
9   1980          55.38       186.3                6.735822
10  1981          65.26       193.5                6.006204
11  1982          66.67       200.7                3.475733
12  1983          72.21       218.3                7.288893
13  1984          72.92       212.2                3.820738
14  1985          76.82       232.5                5.254299
15  1986          77.40       249.0                4.776564
16  1987          82.88       279.0                3.965356
17  1988          88.34       296.6                9.627783
18  1989          97.86       296.0                5.947343
19  1990         107.18       321.0                5.533455
20  1991         102.45       270.1                1.056831
21  1992         113.38       288.2                5.482396
22  1993         119.41       279.3                4.750776
23  1994         137.65       327.3                6.658924
24  1995         148.84       360.3                7.574492
25  1996         162.72       392.9                7.549522
26  1997         168.87       415.9                4.049821
27  1998         180.95       421.4                6.184416
28  1999         187.68       458.8                8.845756
29  2000         204.26       468.4                3.840991
30  2001         221.53       485.4                4.823966
```

```
31  2002         224.64       514.9                  3.803975
32  2003         259.04       607.7                  7.860381
33  2004         283.36       709.2                  7.922937
34  2005         308.89       820.4                  7.923431
35  2006         352.83       940.3                  8.060733
36  2007         392.61      1220.0                  7.660815
37  2008         455.31      1200.0                  3.086698
38  2009         476.39      1340.0                  7.861889
39  2010        1610.00      1680.0                  8.497585
40  2011        1830.00      1820.0                  5.241316
41  2012        1840.00      1830.0                  5.456388
42  2013        1840.00      1860.0                  6.386106
43  2014        1850.00      2040.0                  7.410228
44  2015        2090.00      2100.0                  7.996254
45  2016        2290.00      2290.0                  8.256306
46  2017        2450.00      2650.0                  6.795383
47  2018        2700.00      2700.0                  6.453851
48  2019        2870.00      2840.0                  3.871437
49  2020        2660.00      2670.0                 -5.777725
50  2021        2930.00      3170.0                  9.689592
51  2022        3350.00      3350.0                  6.987039
52  2023        3550.00      3550.0                  7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                             26.47
1                                 34.44                             26.48
```

```
# Clean column names
df.columns = [col.strip() for col in df.columns]
print(df.columns)
```

```
Index(['Year', 'GDP (Nominal)', 'GDP (Real)', 'GDP growth (annual %)',
       'Agriculture Contribution (% of GDP)',
       'Industry Contribution (% of GDP)', 'Services Contribution (% of GDP)',
       'Inflation CPI', 'Unemployment Rate (%)', 'Exports (% of GDP)',
       'Imports  (% of GDP)', 'FDI Inflows (USD)', 'Fiscal Deficit (% of GDP)',
       'Public Debt (% of GDP)', 'Repo Rate (%)',
       'Foreign Exchange Reserves (MILLION USD)',
       'Population Growth Rate (annual %)', 'Oil rents (% of GDP)',
       'Government Revenues (% of GDP)',
       'Household Consumption Expenditure (% of GDP)',
       'Gross Fixed Capital Formation (% of GDP)',
       'Private Sector Credit (% of GDP)', 'Poverty Rate (%)',
       'Energy Consumption per Capita (kWh)',
       'Current Account Balance (% of GDP)', 'Tax revenue (% of GDP)'],
      dtype='object')
```

```
# Drop column with many missing values
df.drop(columns=["Oil rents (% of GDP)"], inplace=True)
print(df.drop)
```

```
<bound method DataFrame.drop of     Year  GDP (Nominal)  GDP (Real)  GDP growth (annua
0   1971          35.58        67.4               1.642930
1   1972          36.06        71.5              -0.553301
2   1973          38.32        85.5               3.295521
```

```
3   1974        38.47      99.5            1.185336
4   1975        41.57      98.5            9.149912
5   1976        42.86     102.7            1.663104
6   1977        44.86     121.5            7.254765
7   1978        45.62     137.3            5.712532
8   1979        46.84     153.0           -5.238183
9   1980        55.38     186.3            6.735822
10  1981        65.26     193.5            6.006204
11  1982        66.67     200.7            3.475733
12  1983        72.21     218.3            7.288893
13  1984        72.92     212.2            3.820738
14  1985        76.82     232.5            5.254299
15  1986        77.40     249.0            4.776564
16  1987        82.88     279.0            3.965356
17  1988        88.34     296.6            9.627783
18  1989        97.86     296.0            5.947343
19  1990       107.18     321.0            5.533455
20  1991       102.45     270.1            1.056831
21  1992       113.38     288.2            5.482396
22  1993       119.41     279.3            4.750776
23  1994       137.65     327.3            6.658924
24  1995       148.84     360.3            7.574492
25  1996       162.72     392.9            7.549522
26  1997       168.87     415.9            4.049821
27  1998       180.95     421.4            6.184416
28  1999       187.68     458.8            8.845756
29  2000       204.26     468.4            3.840991
30  2001       221.53     485.4            4.823966
31  2002       224.64     514.9            3.803975
32  2003       259.04     607.7            7.860381
33  2004       283.36     709.2            7.922937
34  2005       308.89     820.4            7.923431
35  2006       352.83     940.3            8.060733
36  2007       392.61    1220.0            7.660815
37  2008       455.31    1200.0            3.086698
38  2009       476.39    1340.0            7.861889
39  2010      1610.00    1680.0            8.497585
40  2011      1830.00    1820.0            5.241316
41  2012      1840.00    1830.0            5.456388
42  2013      1840.00    1860.0            6.386106
43  2014      1850.00    2040.0            7.410228
44  2015      2090.00    2100.0            7.996254
45  2016      2290.00    2290.0            8.256306
46  2017      2450.00    2650.0            6.795383
47  2018      2700.00    2700.0            6.453851
48  2019      2870.00    2840.0            3.871437
49  2020      2660.00    2670.0           -5.777725
50  2021      2930.00    3170.0            9.689592
51  2022      3350.00    3350.0            6.987039
52  2023      3550.00    3550.0            7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                             26.47
```

```
# Drop duplicates if any
df.drop_duplicates(inplace=True)
print(df.drop_duplicates)
```

```
<bound method DataFrame.drop_duplicates of     Year  GDP (Nominal)  GDP (Real)  GDP gr
0   1971          35.58        67.4            1.642930
1   1972          36.06        71.5           -0.553301
2   1973          38.32        85.5            3.295521
3   1974          38.47        99.5            1.185336
4   1975          41.57        98.5            9.149912
5   1976          42.86       102.7            1.663104
6   1977          44.86       121.5            7.254765
7   1978          45.62       137.3            5.712532
8   1979          46.84       153.0           -5.238183
9   1980          55.38       186.3            6.735822
10  1981          65.26       193.5            6.006204
11  1982          66.67       200.7            3.475733
12  1983          72.21       218.3            7.288893
13  1984          72.92       212.2            3.820738
14  1985          76.82       232.5            5.254299
15  1986          77.40       249.0            4.776564
16  1987          82.88       279.0            3.965356
17  1988          88.34       296.6            9.627783
18  1989          97.86       296.0            5.947343
19  1990         107.18       321.0            5.533455
20  1991         102.45       270.1            1.056831
21  1992         113.38       288.2            5.482396
22  1993         119.41       279.3            4.750776
23  1994         137.65       327.3            6.658924
24  1995         148.84       360.3            7.574492
25  1996         162.72       392.9            7.549522
26  1997         168.87       415.9            4.049821
27  1998         180.95       421.4            6.184416
28  1999         187.68       458.8            8.845756
29  2000         204.26       468.4            3.840991
30  2001         221.53       485.4            4.823966
31  2002         224.64       514.9            3.803975
32  2003         259.04       607.7            7.860381
33  2004         283.36       709.2            7.922937
34  2005         308.89       820.4            7.923431
35  2006         352.83       940.3            8.060733
36  2007         392.61      1220.0            7.660815
37  2008         455.31      1200.0            3.086698
38  2009         476.39      1340.0            7.861889
39  2010        1610.00      1680.0            8.497585
40  2011        1830.00      1820.0            5.241316
41  2012        1840.00      1830.0            5.456388
42  2013        1840.00      1860.0            6.386106
43  2014        1850.00      2040.0            7.410228
44  2015        2090.00      2100.0            7.996254
45  2016        2290.00      2290.0            8.256306
46  2017        2450.00      2650.0            6.795383
47  2018        2700.00      2700.0            6.453851
48  2019        2870.00      2840.0            3.871437
49  2020        2660.00      2670.0           -5.777725
50  2021        2930.00      3170.0            9.689592
```

```
51  2022        3350.00      3350.0                6.987039
52  2023        3550.00      3550.0                7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                             26.47
```

```
# Feature Engineering: GDP per capita
df["GDP Per Capita"] = df["GDP (Nominal)"] / (df["Population Growth Rate (annual %)"] + 1e-5
print(df)
```

```
    Year  GDP (Nominal)  GDP (Real)  GDP growth (annual %)  \
0   1971          35.58        67.4               1.642930
1   1972          36.06        71.5              -0.553301
2   1973          38.32        85.5               3.295521
3   1974          38.47        99.5               1.185336
4   1975          41.57        98.5               9.149912
5   1976          42.86       102.7               1.663104
6   1977          44.86       121.5               7.254765
7   1978          45.62       137.3               5.712532
8   1979          46.84       153.0              -5.238183
9   1980          55.38       186.3               6.735822
10  1981          65.26       193.5               6.006204
11  1982          66.67       200.7               3.475733
12  1983          72.21       218.3               7.288893
13  1984          72.92       212.2               3.820738
14  1985          76.82       232.5               5.254299
15  1986          77.40       249.0               4.776564
16  1987          82.88       279.0               3.965356
17  1988          88.34       296.6               9.627783
18  1989          97.86       296.0               5.947343
19  1990         107.18       321.0               5.533455
20  1991         102.45       270.1               1.056831
21  1992         113.38       288.2               5.482396
22  1993         119.41       279.3               4.750776
23  1994         137.65       327.3               6.658924
24  1995         148.84       360.3               7.574492
25  1996         162.72       392.9               7.549522
26  1997         168.87       415.9               4.049821
27  1998         180.95       421.4               6.184416
28  1999         187.68       458.8               8.845756
29  2000         204.26       468.4               3.840991
30  2001         221.53       485.4               4.823966
31  2002         224.64       514.9               3.803975
32  2003         259.04       607.7               7.860381
33  2004         283.36       709.2               7.922937
34  2005         308.89       820.4               7.923431
35  2006         352.83       940.3               8.060733
36  2007         392.61      1220.0               7.660815
37  2008         455.31      1200.0               3.086698
38  2009         476.39      1340.0               7.861889
39  2010        1610.00      1680.0               8.497585
40  2011        1830.00      1820.0               5.241316
41  2012        1840.00      1830.0               5.456388
42  2013        1840.00      1860.0               6.386106
```

```
43  2014        1850.00      2040.0                7.410228
44  2015        2090.00      2100.0                7.996254
45  2016        2290.00      2290.0                8.256306
46  2017        2450.00      2650.0                6.795383
47  2018        2700.00      2700.0                6.453851
48  2019        2870.00      2840.0                3.871437
49  2020        2660.00      2670.0               -5.777725
50  2021        2930.00      3170.0                9.689592
51  2022        3350.00      3350.0                6.987039
52  2023        3550.00      3550.0                7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                              26.47
```

```
# Z-Score for outlier detection on Inflation
df["Inflation_Z"] = zscore(df["Inflation CPI"])
print(df)
```

```
    Year  GDP (Nominal)  GDP (Real)  GDP growth (annual %)  \
0   1971          35.58        67.4               1.642930
1   1972          36.06        71.5              -0.553301
2   1973          38.32        85.5               3.295521
3   1974          38.47        99.5               1.185336
4   1975          41.57        98.5               9.149912
5   1976          42.86       102.7               1.663104
6   1977          44.86       121.5               7.254765
7   1978          45.62       137.3               5.712532
8   1979          46.84       153.0              -5.238183
9   1980          55.38       186.3               6.735822
10  1981          65.26       193.5               6.006204
11  1982          66.67       200.7               3.475733
12  1983          72.21       218.3               7.288893
13  1984          72.92       212.2               3.820738
14  1985          76.82       232.5               5.254299
15  1986          77.40       249.0               4.776564
16  1987          82.88       279.0               3.965356
17  1988          88.34       296.6               9.627783
18  1989          97.86       296.0               5.947343
19  1990         107.18       321.0               5.533455
20  1991         102.45       270.1               1.056831
21  1992         113.38       288.2               5.482396
22  1993         119.41       279.3               4.750776
23  1994         137.65       327.3               6.658924
24  1995         148.84       360.3               7.574492
25  1996         162.72       392.9               7.549522
26  1997         168.87       415.9               4.049821
27  1998         180.95       421.4               6.184416
28  1999         187.68       458.8               8.845756
29  2000         204.26       468.4               3.840991
30  2001         221.53       485.4               4.823966
31  2002         224.64       514.9               3.803975
32  2003         259.04       607.7               7.860381
33  2004         283.36       709.2               7.922937
34  2005         308.89       820.4               7.923431
35  2006         352.83       940.3               8.060733
```

```
36  2007         392.61      1220.0                7.660815
37  2008         455.31      1200.0                3.086698
38  2009         476.39      1340.0                7.861889
39  2010        1610.00      1680.0                8.497585
40  2011        1830.00      1820.0                5.241316
41  2012        1840.00      1830.0                5.456388
42  2013        1840.00      1860.0                6.386106
43  2014        1850.00      2040.0                7.410228
44  2015        2090.00      2100.0                7.996254
45  2016        2290.00      2290.0                8.256306
46  2017        2450.00      2650.0                6.795383
47  2018        2700.00      2700.0                6.453851
48  2019        2870.00      2840.0                3.871437
49  2020        2660.00      2670.0               -5.777725
50  2021        2930.00      3170.0                9.689592
51  2022        3350.00      3350.0                6.987039
52  2023        3550.00      3550.0                7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                             26.47
1                                 34.44                             26.48
```

```
# Filter rows
high_growth = df[df["GDP growth (annual %)"] > 7]
print(high_growth)
```

```
    Year  GDP (Nominal)  GDP (Real)  GDP growth (annual %)  \
4   1975          41.57        98.5               9.149912
6   1977          44.86       121.5               7.254765
12  1983          72.21       218.3               7.288893
17  1988          88.34       296.6               9.627783
24  1995         148.84       360.3               7.574492
25  1996         162.72       392.9               7.549522
28  1999         187.68       458.8               8.845756
32  2003         259.04       607.7               7.860381
33  2004         283.36       709.2               7.922937
34  2005         308.89       820.4               7.923431
35  2006         352.83       940.3               8.060733
36  2007         392.61      1220.0               7.660815
38  2009         476.39      1340.0               7.861889
39  2010        1610.00      1680.0               8.497585
43  2014        1850.00      2040.0               7.410228
44  2015        2090.00      2100.0               7.996254
45  2016        2290.00      2290.0               8.256306
50  2021        2930.00      3170.0               9.689592
52  2023        3550.00      3550.0               7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
4                                 35.26                             25.19
6                                 35.91                             24.40
12                                31.65                             25.37
17                                29.66                             26.83
24                                26.62                             25.43
25                                26.43                             25.20
28                                25.58                             25.24
32                                22.67                             25.45
```

```
33                      22.36                       25.76
34                      20.86                       25.70
35                      18.94                       26.21
36                      18.52                       26.45
38                      16.82                       26.25
39                      16.58                       26.20
43                      15.63                       26.06
44                      15.22                       26.09
45                      15.28                       25.93
50                      18.26                       26.14
52                      15.97                       26.22

    Services Contribution (% of GDP)  Inflation CPI  Unemployment Rate (%)  \
4                              43.52      -1.648680                    4.9
6                              43.78       5.637229                    5.0
12                             43.56       8.552860                    5.0
17                             44.63       8.232515                    5.1
24                             46.92       9.062702                    5.5
25                             47.11       7.575018                    5.5
28                             48.63       3.068396                    5.7
32                             49.72       3.867798                    5.4
33                             49.93       5.725413                    8.3
34                             50.54       5.621903                    5.3
35                             51.21       8.400938                    4.6
36                             52.02       6.944418                    7.3
38                             54.35       7.040365                    9.5
39                             54.85      10.526030                    8.6
43                             56.30       3.331757                    5.0
```

```
# IQR outlier detection
Q1 = df["FDI Inflows (USD)"].quantile(0.25)
Q3 = df["FDI Inflows (USD)"].quantile(0.75)
IQR = Q3 - Q1
outliers_iqr = df[(df["FDI Inflows (USD)"] < Q1 - 1.5 * IQR) | (df["FDI Inflows (USD)"] > Q3
print(outliers_iqr)
```

```
    Year  GDP (Nominal)  GDP (Real)  GDP growth (annual %)  \
49  2020         2660.0      2670.0              -5.777725
52  2023         3550.0      3550.0               7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
49                                17.43                             22.52
52                                15.97                             26.22

    Services Contribution (% of GDP)  Inflation CPI  Unemployment Rate (%)  \
49                             50.15       4.814858                    7.1
52                             53.70       1.366467                    4.2

    Exports (% of GDP)  ...  Government Revenues (% of GDP)  \
49            18.68248  ...                           10.77
52            21.89164  ...                           11.22

    Household Consumption Expenditure (% of GDP)  \
49                                          67.0
52                                          68.0
```

```
    Gross Fixed Capital Formation (% of GDP)  \
49                                     27.09
52                                     29.53

    Private Sector Credit (% of GDP)  Poverty Rate (%)  \
49                              54.8              14.7
52                              50.0              10.9

    Energy Consumption per Capita (kWh)  Current Account Balance (% of GDP)  \
49                                 1061                                -0.9
52                                  985                                -1.2

    Tax revenue (% of GDP)  GDP Per Capita  Inflation_Z
49                     9.9     1054.040999    -0.646998
52                    11.7     1612.107060    -1.563053

[2 rows x 27 columns]
```

```
# Normalization
df["GDP_Norm"] = (df["GDP (Nominal)"] - df["GDP (Nominal)"].min()) / (df["GDP (Nominal)"].ma
print(df[['GDP (Nominal)']].head())
```

```
   GDP (Nominal)
0          35.58
1          36.06
2          38.32
3          38.47
4          41.57
```

```
# Log transform
df["FDI_Log"] = np.log(df["FDI Inflows (USD)"]+1)
print(df['FDI Inflows (USD)'].isnull().sum())
```

```
0
```

```
# Reorder columns
cols = df.columns.tolist()
cols.insert(0, cols.pop(cols.index('Year')))
df = df[cols]
print(df.head)
```

```
<bound method NDFrame.head of     Year  GDP (Nominal)  GDP (Real)  GDP growth (annual
0   1971          35.58       67.4             1.642930
1   1972          36.06       71.5            -0.553301
2   1973          38.32       85.5             3.295521
3   1974          38.47       99.5             1.185336
4   1975          41.57       98.5             9.149912
5   1976          42.86      102.7             1.663104
6   1977          44.86      121.5             7.254765
7   1978          45.62      137.3             5.712532
8   1979          46.84      153.0            -5.238183
```

```
9   1980      55.38     186.3               6.735822
10  1981      65.26     193.5               6.006204
11  1982      66.67     200.7               3.475733
12  1983      72.21     218.3               7.288893
13  1984      72.92     212.2               3.820738
14  1985      76.82     232.5               5.254299
15  1986      77.40     249.0               4.776564
16  1987      82.88     279.0               3.965356
17  1988      88.34     296.6               9.627783
18  1989      97.86     296.0               5.947343
19  1990     107.18     321.0               5.533455
20  1991     102.45     270.1               1.056831
21  1992     113.38     288.2               5.482396
22  1993     119.41     279.3               4.750776
23  1994     137.65     327.3               6.658924
24  1995     148.84     360.3               7.574492
25  1996     162.72     392.9               7.549522
26  1997     168.87     415.9               4.049821
27  1998     180.95     421.4               6.184416
28  1999     187.68     458.8               8.845756
29  2000     204.26     468.4               3.840991
30  2001     221.53     485.4               4.823966
31  2002     224.64     514.9               3.803975
32  2003     259.04     607.7               7.860381
33  2004     283.36     709.2               7.922937
34  2005     308.89     820.4               7.923431
35  2006     352.83     940.3               8.060733
36  2007     392.61    1220.0               7.660815
37  2008     455.31    1200.0               3.086698
38  2009     476.39    1340.0               7.861889
39  2010    1610.00    1680.0               8.497585
40  2011    1830.00    1820.0               5.241316
41  2012    1840.00    1830.0               5.456388
42  2013    1840.00    1860.0               6.386106
43  2014    1850.00    2040.0               7.410228
44  2015    2090.00    2100.0               7.996254
45  2016    2290.00    2290.0               8.256306
46  2017    2450.00    2650.0               6.795383
47  2018    2700.00    2700.0               6.453851
48  2019    2870.00    2840.0               3.871437
49  2020    2660.00    2670.0              -5.777725
50  2021    2930.00    3170.0               9.689592
51  2022    3350.00    3350.0               6.987039
52  2023    3550.00    3550.0               7.583971

    Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
0                                 34.58                             26.47
```

```
# Indexing
df.set_index('Year', inplace=True)
print(df.head)
```

```
<bound method NDFrame.head of       GDP (Nominal)  GDP (Real)  GDP growth (annual %)
Year
1971          35.58       67.4               1.642930
```

```
1972         36.06      71.5               -0.553301
1973         38.32      85.5                3.295521
1974         38.47      99.5                1.185336
1975         41.57      98.5                9.149912
1976         42.86     102.7                1.663104
1977         44.86     121.5                7.254765
1978         45.62     137.3                5.712532
1979         46.84     153.0               -5.238183
1980         55.38     186.3                6.735822
1981         65.26     193.5                6.006204
1982         66.67     200.7                3.475733
1983         72.21     218.3                7.288893
1984         72.92     212.2                3.820738
1985         76.82     232.5                5.254299
1986         77.40     249.0                4.776564
1987         82.88     279.0                3.965356
1988         88.34     296.6                9.627783
1989         97.86     296.0                5.947343
1990        107.18     321.0                5.533455
1991        102.45     270.1                1.056831
1992        113.38     288.2                5.482396
1993        119.41     279.3                4.750776
1994        137.65     327.3                6.658924
1995        148.84     360.3                7.574492
1996        162.72     392.9                7.549522
1997        168.87     415.9                4.049821
1998        180.95     421.4                6.184416
1999        187.68     458.8                8.845756
2000        204.26     468.4                3.840991
2001        221.53     485.4                4.823966
2002        224.64     514.9                3.803975
2003        259.04     607.7                7.860381
2004        283.36     709.2                7.922937
2005        308.89     820.4                7.923431
2006        352.83     940.3                8.060733
2007        392.61    1220.0                7.660815
2008        455.31    1200.0                3.086698
2009        476.39    1340.0                7.861889
2010       1610.00    1680.0                8.497585
2011       1830.00    1820.0                5.241316
2012       1840.00    1830.0                5.456388
2013       1840.00    1860.0                6.386106
2014       1850.00    2040.0                7.410228
2015       2090.00    2100.0                7.996254
2016       2290.00    2290.0                8.256306
2017       2450.00    2650.0                6.795383
2018       2700.00    2700.0                6.453851
2019       2870.00    2840.0                3.871437
2020       2660.00    2670.0               -5.777725
2021       2930.00    3170.0                9.689592
2022       3350.00    3350.0                6.987039
2023       3550.00    3550.0                7.583971

      Agriculture Contribution (% of GDP)  Industry Contribution (% of GDP)  \
```

🧹 Step 2: Data Cleaning & Summary

```
df.info()
df.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 53 entries, 1971 to 2023
Data columns (total 28 columns):
 #   Column                                      Non-Null Count  Dtype  
---  ------                                      --------------  -----  
 0   GDP (Nominal)                               53 non-null     float64
 1   GDP (Real)                                  53 non-null     float64
 2   GDP growth (annual %)                       53 non-null     float64
 3   Agriculture Contribution (% of GDP)         53 non-null     float64
 4   Industry Contribution (% of GDP)            53 non-null     float64
 5   Services Contribution (% of GDP)            53 non-null     float64
 6   Inflation CPI                               53 non-null     float64
 7   Unemployment Rate (%)                       53 non-null     float64
 8   Exports (% of GDP)                          53 non-null     float64
 9   Imports  (% of GDP)                         53 non-null     float64
 10  FDI Inflows (USD)                           53 non-null     float64
 11  Fiscal Deficit (% of GDP)                   53 non-null     float64
 12  Public Debt (% of GDP)                      53 non-null     float64
 13  Repo Rate (%)                               53 non-null     float64
 14  Foreign Exchange Reserves (MILLION USD)     53 non-null     int64  
 15  Population Growth Rate (annual %)           53 non-null     float64
 16  Government Revenues (% of GDP)              53 non-null     float64
 17  Household Consumption Expenditure (% of GDP) 53 non-null     float64
 18  Gross Fixed Capital Formation (% of GDP)    53 non-null     float64
 19  Private Sector Credit (% of GDP)            53 non-null     float64
 20  Poverty Rate (%)                            53 non-null     float64
 21  Energy Consumption per Capita (kWh)         53 non-null     int64  
 22  Current Account Balance (% of GDP)          53 non-null     float64
 23  Tax revenue (% of GDP)                      53 non-null     float64
 24  GDP Per Capita                              53 non-null     float64
 25  Inflation_Z                                 53 non-null     float64
 26  GDP_Norm                                    53 non-null     float64
 27  FDI_Log                                     53 non-null     float64
dtypes: float64(26), int64(2)
memory usage: 12.0 KB
```

	GDP (Nominal)	GDP (Real)	GDP growth (annual %)	Agriculture Contribution (% of GDP)	Industry Contribution (% of GDP)	Services Contribution (% of GDP)	I
count	53.000000	53.000000	53.000000	53.000000	53.000000	53.000000	5
mean	747.054906	937.594340	5.403609	25.039811	25.677547	48.613396	
std	1061.266488	1013.279818	3.183058	7.257481	0.744939	4.632577	
min	35.580000	67.400000	-5.777725	14.410000	22.520000	43.320000	-
25%	72.920000	218.300000	3.871437	17.430000	25.370000	44.110000	
50%	168.870000	415.900000	6.006204	26.220000	25.770000	47.640000	
75%	1610.000000	1680.000000	7.583971	31.330000	26.200000	53.320000	
max	3550.000000	3550.000000	9.689592	36.350000	26.830000	57.250000	1

8 rows × 28 columns

📊 Step 3: Exploratory Data Analysis (EDA)

🔍 Skewness & Kurtosis

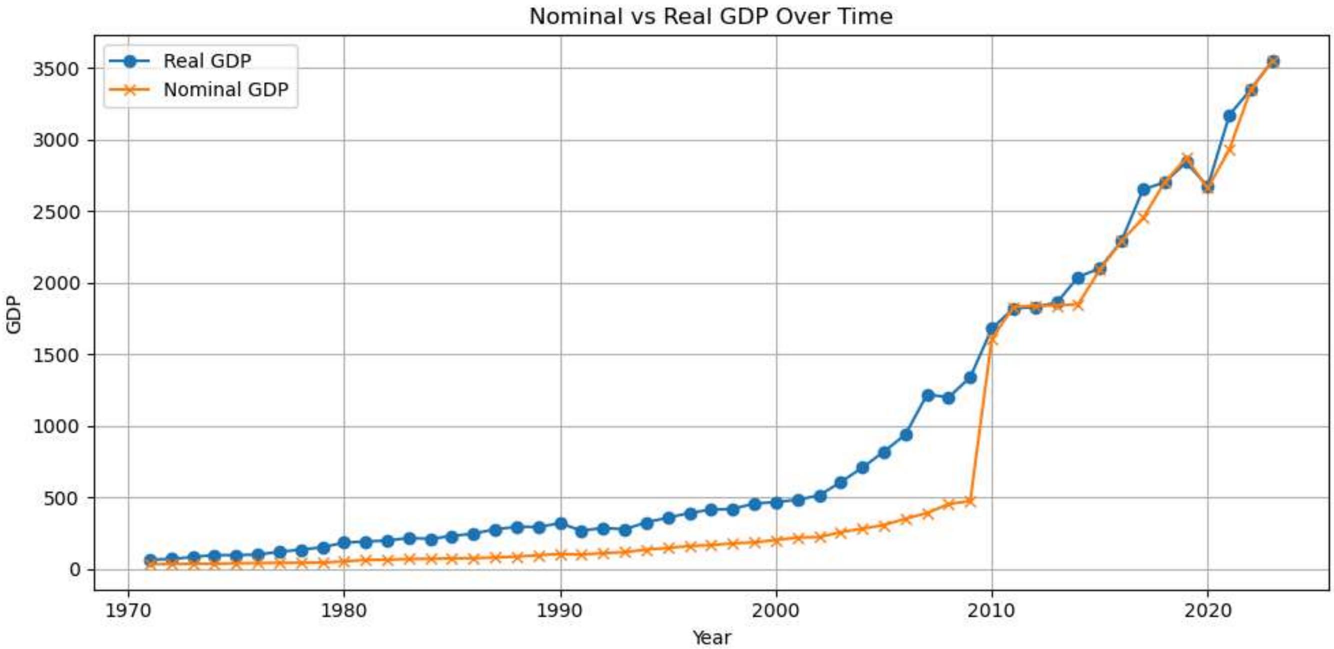
```
skew_vals = df.select_dtypes(include=[np.number]).apply(skew).sort_values(ascending=False)
kurt_vals = df.select_dtypes(include=[np.number]).apply(kurtosis).sort_values(ascending=Fals
skew_vals.head(), kurt_vals.head()
```

```
(GDP Per Capita           1.810869
 GDP_Norm                 1.326372
 GDP (Nominal)            1.326372
 Unemployment Rate (%)    1.321039
 FDI Inflows (USD)        1.290243
 dtype: float64,
 Industry Contribution (% of GDP)      4.488008
 Current Account Balance (% of GDP)    4.460887
 GDP growth (annual %)                 3.192018
 GDP Per Capita                        2.348693
 Unemployment Rate (%)                 1.144312
 dtype: float64)
```

📈 Step 4: Data Visualization

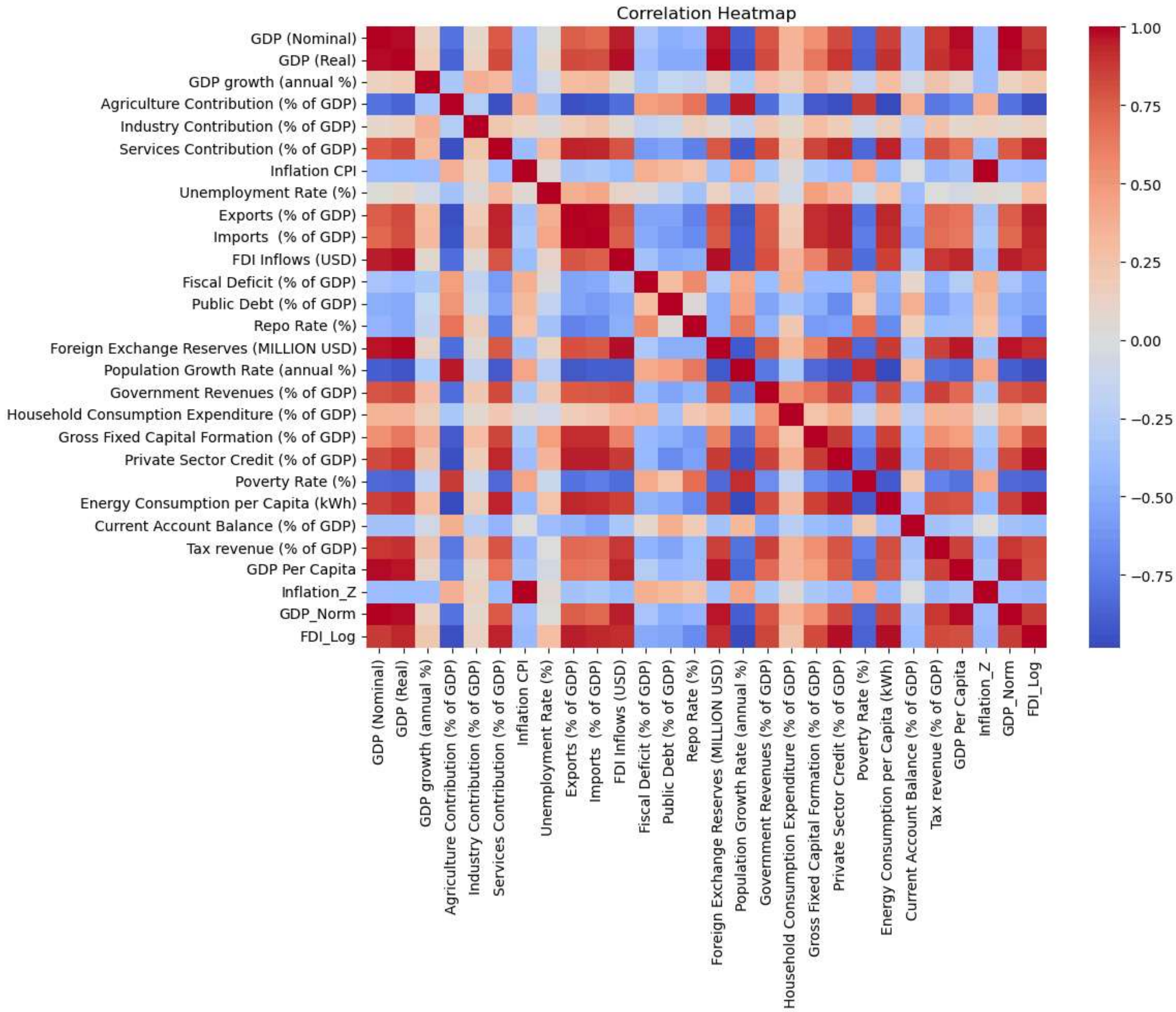
📉 GDP: Real vs Nominal Over Time

```
plt.figure(figsize=(10, 5))
plt.plot(df.index, df['GDP (Real)'], marker='o', label='Real GDP')
plt.plot(df.index, df['GDP (Nominal)'], marker='x', label='Nominal GDP')
plt.title('Nominal vs Real GDP Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



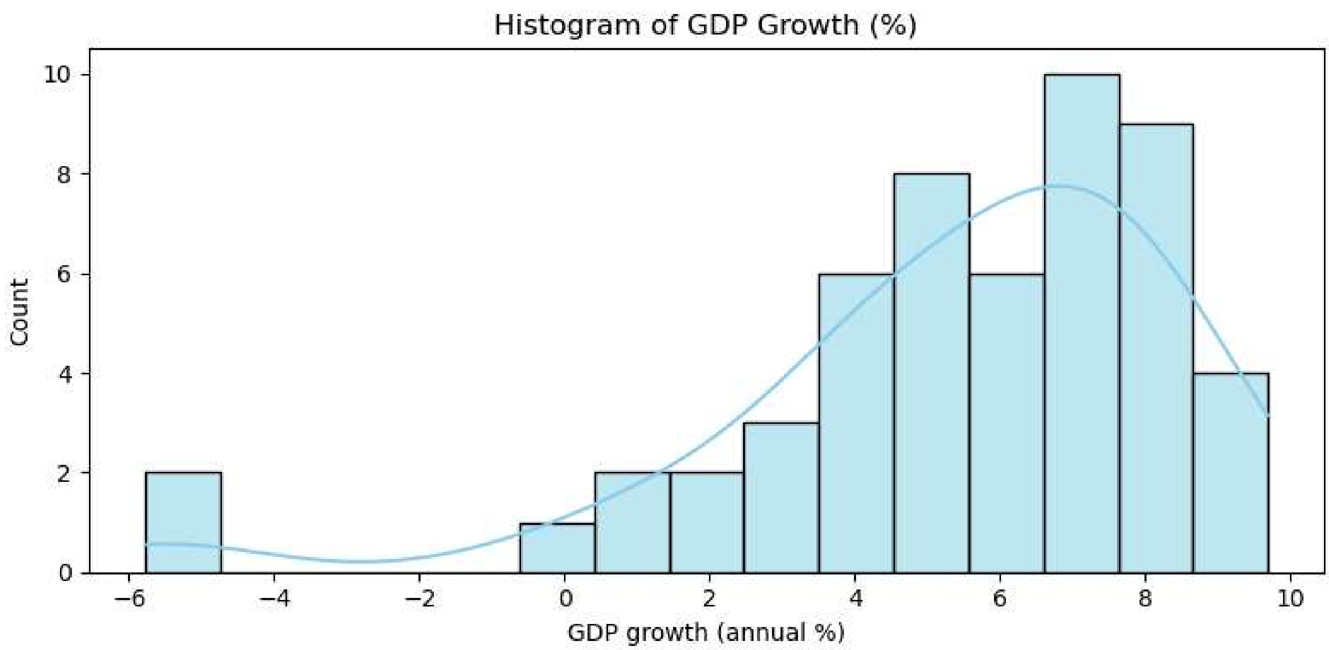
🔥 Correlation Heatmap

```
plt.figure(figsize=(12, 10))
sns.heatmap(df.corr(numeric_only=True), annot=False, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.tight_layout()
plt.show()
```


Correlation Heatmap
GDP (Nominal)
GDP (Real)
GDP growth (annual %)
Agriculture Contribution (% of GDP)
Industry Contribution (% of GDP)
Services Contribution (% of GDP)
Inflation CPI
Unemployment Rate (%)
Exports (% of GDP)
Imports (% of GDP)
FDI Inflows (USD)
Fiscal Deficit (% of GDP)
Public Debt (% of GDP)
Repo Rate (%)
Foreign Exchange Reserves (MILLION USD)
Population Growth Rate (annual %)
Government Revenues (% of GDP)
Household Consumption Expenditure (% of GDP)
Gross Fixed Capital Formation (% of GDP)
Private Sector Credit (% of GDP)
Poverty Rate (%)
Energy Consumption per Capita (kWh)
Current Account Balance (% of GDP)
Tax revenue (% of GDP)
GDP Per Capita
Inflation_Z
GDP_Norm
FDI_Log
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

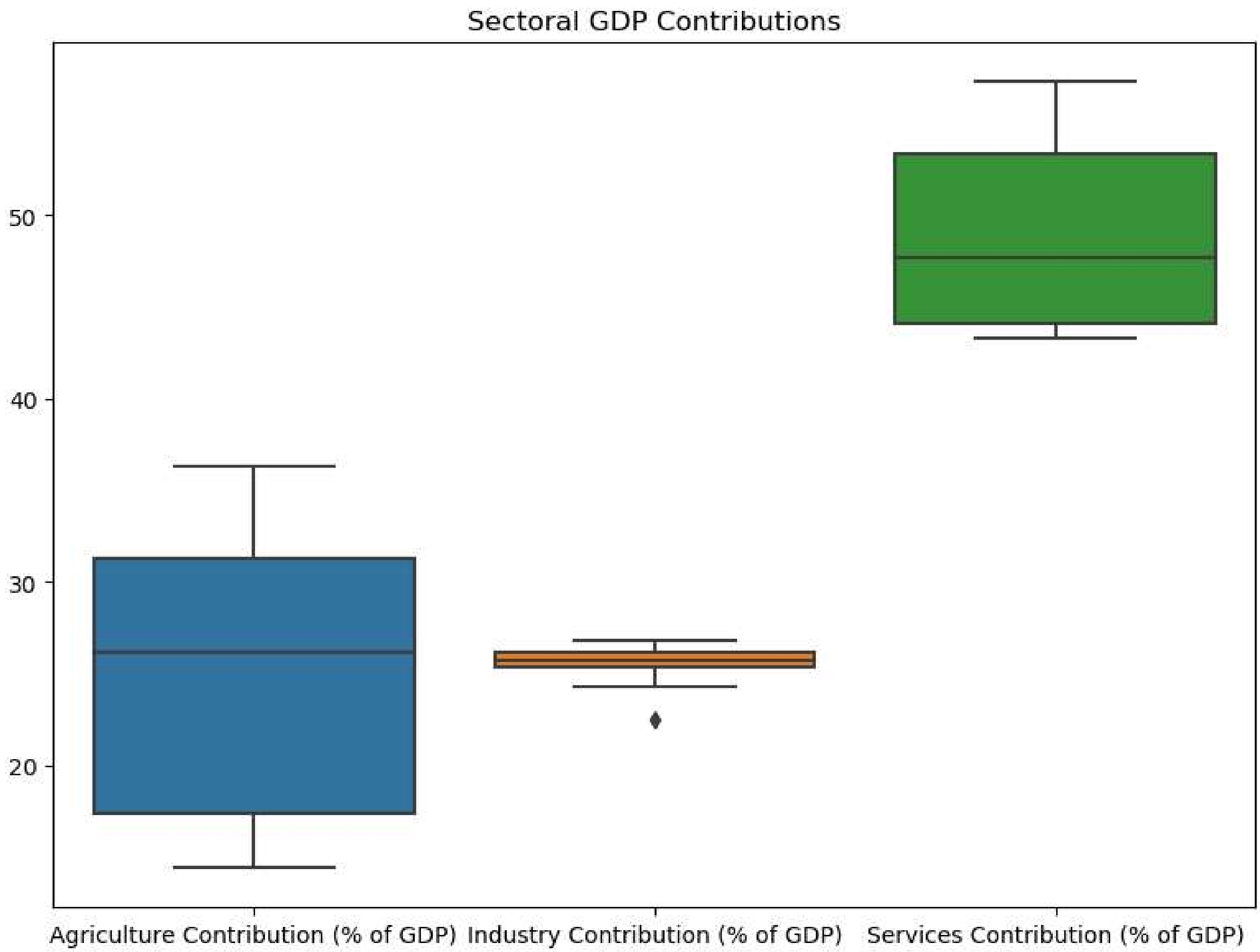
📊 Histogram of GDP Growth

```
plt.figure(figsize=(8, 4))
sns.histplot(df["GDP growth (annual %)"], kde=True, bins=15, color='skyblue')
plt.title("Histogram of GDP Growth (%)")
plt.tight_layout()
plt.show()
```



📦 Boxplot: Sectoral Contributions

```
plt.figure(figsize=(8, 6))
sns.boxplot(data=df[["Agriculture Contribution (% of GDP)",
                     "Industry Contribution (% of GDP)",
                     "Services Contribution (% of GDP)"]])
plt.title("Sectoral GDP Contributions")
plt.tight_layout()
plt.show()
```



🧮 Pivot Table

```
df['Decade'] = (df.index // 10) * 10

pivot_sector = df.pivot_table(values=[
    'Agriculture Contribution (% of GDP)',
    'Industry Contribution (% of GDP)',
    'Services Contribution (% of GDP)'], index='Decade', aggfunc='mean')

pivot_sector
```

Decade	Agriculture Contribution (% of GDP)	Industry Contribution (% of GDP)	Services Contribution (% of GDP)
1970	34.936667	25.343333	43.752222
1980	31.066000	25.944000	44.040000
1990	26.858000	25.432000	46.907000
2000	21.132000	25.774000	50.888000
2010	15.382000	26.006000	55.704000
2020	17.075000	25.315000	51.837500

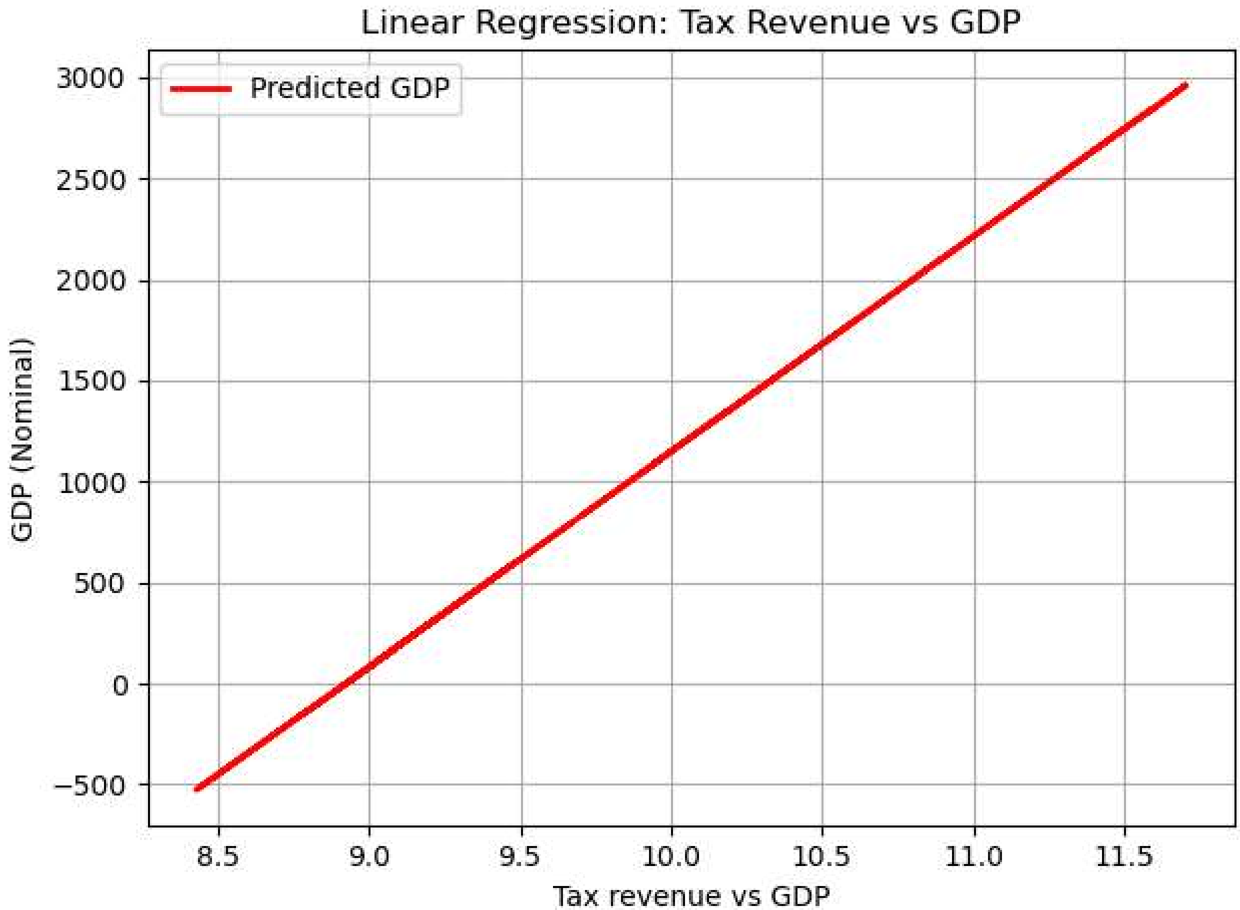
📈 Linear Regression

```
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
X = df[["Tax revenue (% of GDP)"]]
y = df["GDP (Nominal)"]
model = LinearRegression().fit(X, y)
df['GDP_Predicted'] = model.predict(X)
df[['GDP (Nominal)' , 'GDP_Predicted']].head()
```

Year	GDP (Nominal)	GDP_Predicted
1971	35.58	144.752251
1972	36.06	347.262693
1973	38.32	251.336694
1974	38.47	-526.729741
1975	41.57	-281.585522

```
#Regression Line
plt.plot(X, df['GDP_Predicted'], color='red', linewidth=2, label='Predicted GDP')
plt.xlabel('Tax revenue vs GDP')
plt.ylabel('GDP (Nominal)')
plt.title('Linear Regression: Tax Revenue vs GDP')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show
```

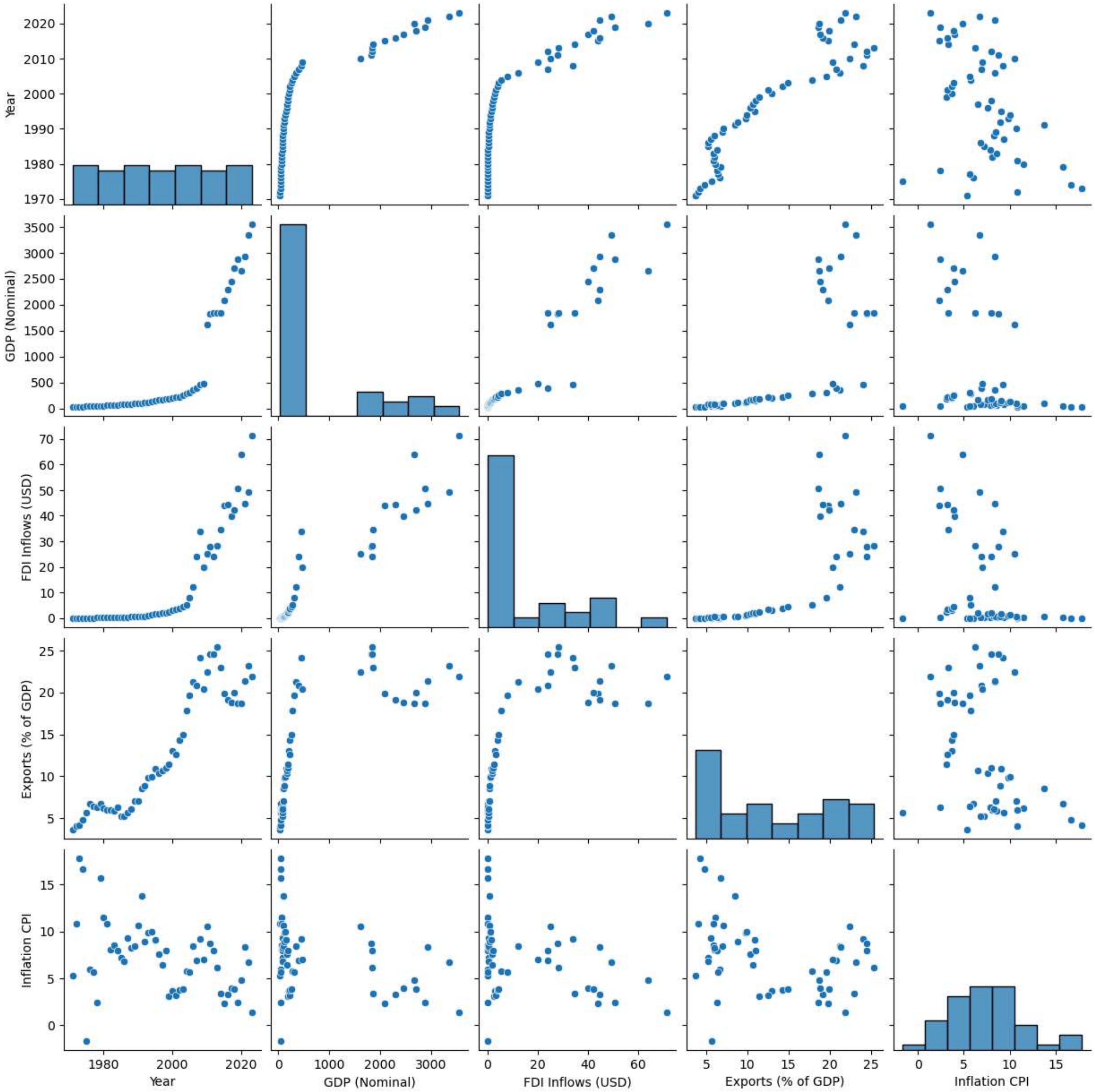
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Pairplot

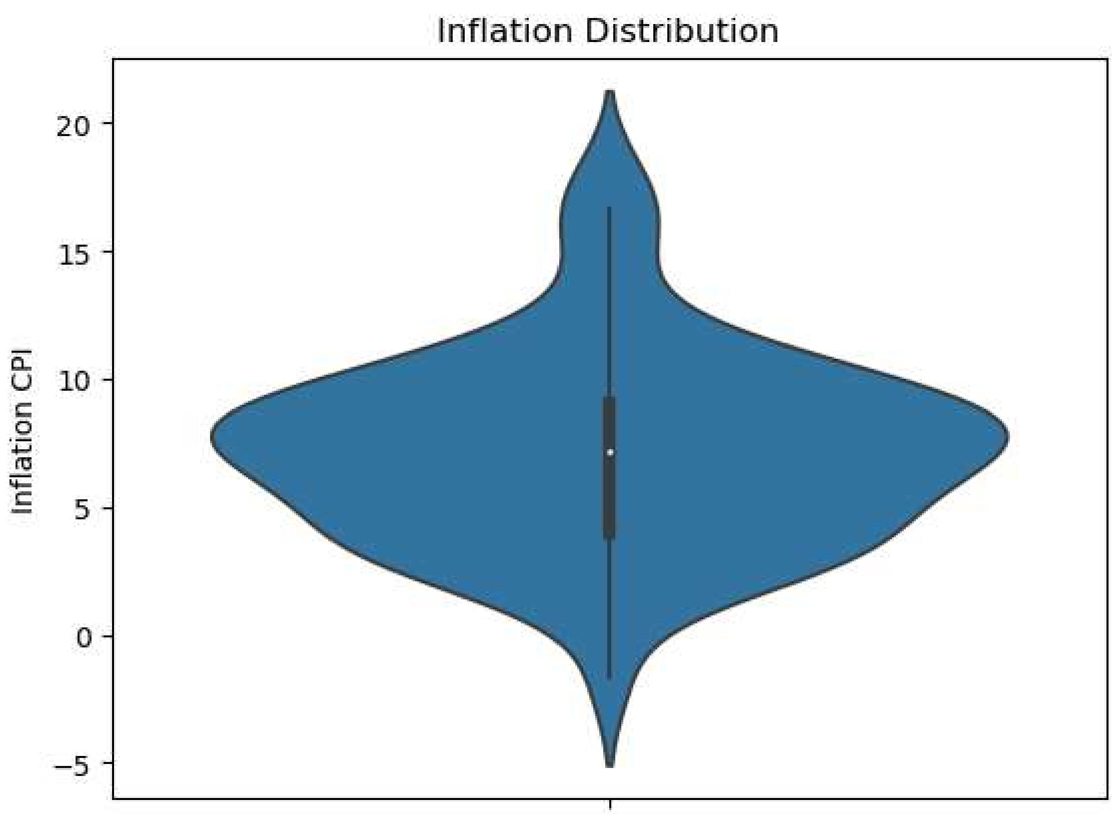
```
sns.pairplot(df[['GDP (Nominal)', 'FDI Inflows (USD)', 'Exports (% of GDP)', 'Inflation CPI'
plt.show()
```

```
C:\Users\user\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The fi
  self._figure.tight_layout(*args, **kwargs)
```



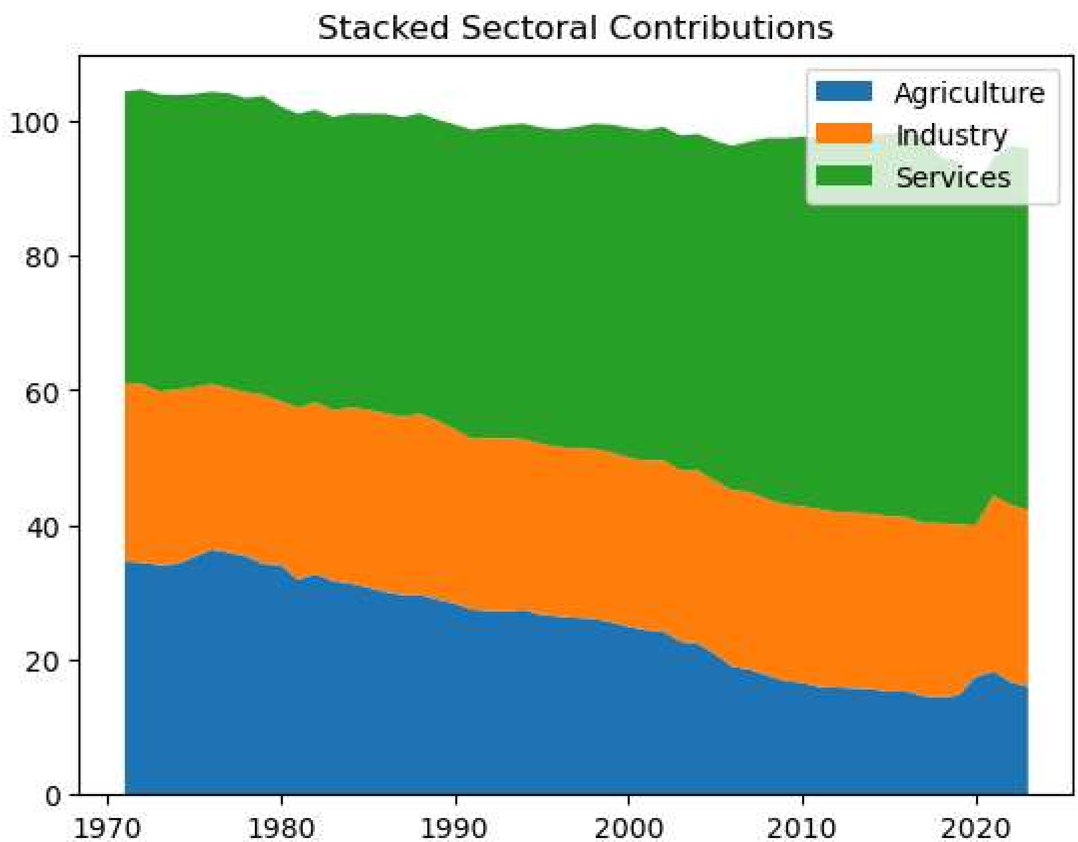
🎻 Violin Plot

```
sns.violinplot(data=df, y='Inflation CPI')
plt.title("Inflation Distribution")
plt.show()
```



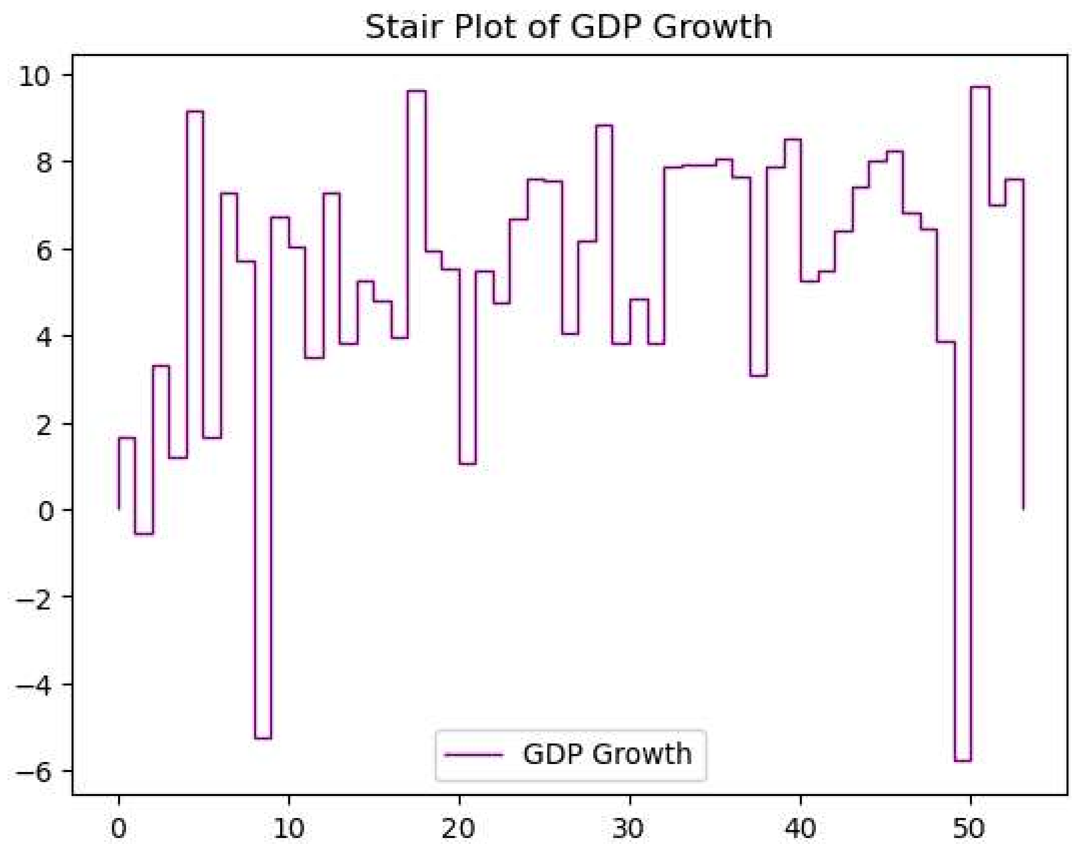
🧱 Stackplot

```
plt.stackplot(df.index,
              df['Agriculture Contribution (% of GDP)'],
              df['Industry Contribution (% of GDP)'],
              df['Services Contribution (% of GDP)'],
              labels=['Agriculture', 'Industry', 'Services'])
plt.title("Stacked Sectoral Contributions")
plt.legend()
plt.show()
```



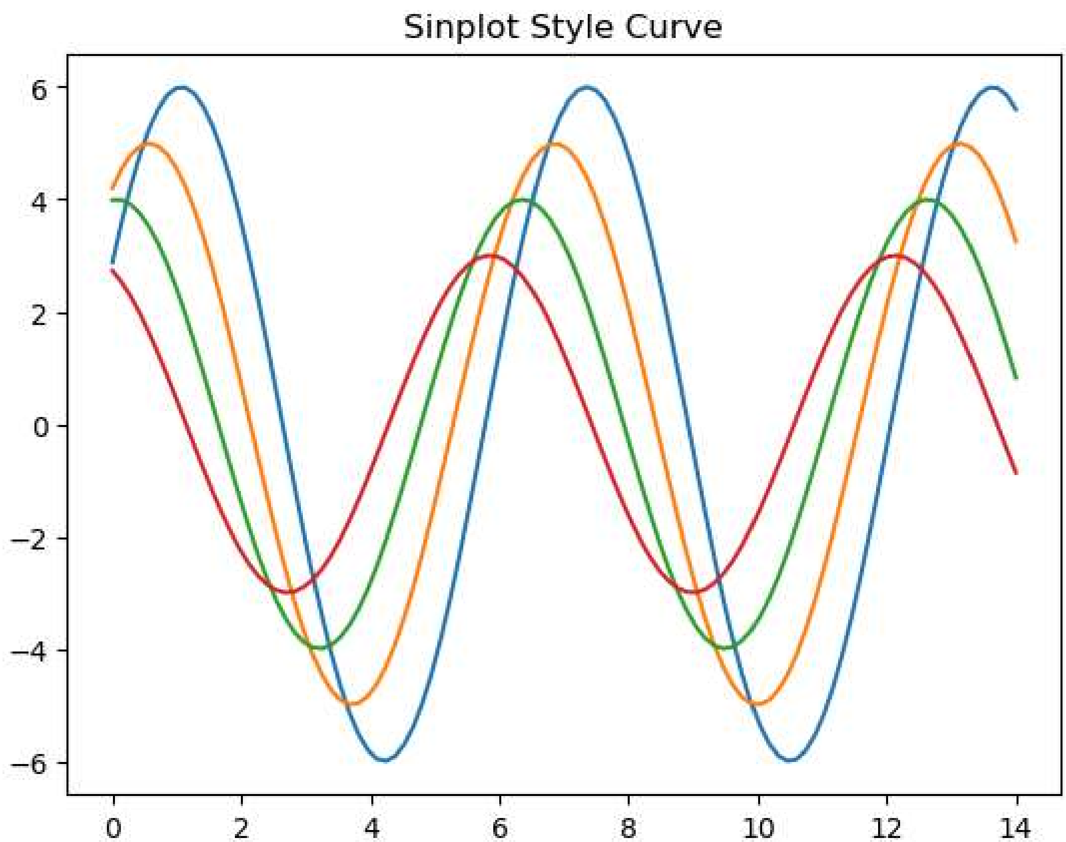
📶 Stairplot

```
plt.stairs(df["GDP growth (annual %)"], label='GDP Growth', color='purple')
plt.title("Stair Plot of GDP Growth")
plt.legend()
plt.show()
```



Sineplot

```
x = np.linspace(0, 14, 100)
for i in range(1, 5):
    plt.plot(x, np.sin(x + i * .5) * (7 - i))
plt.title("Sinplot Style Curve")
plt.show()
```



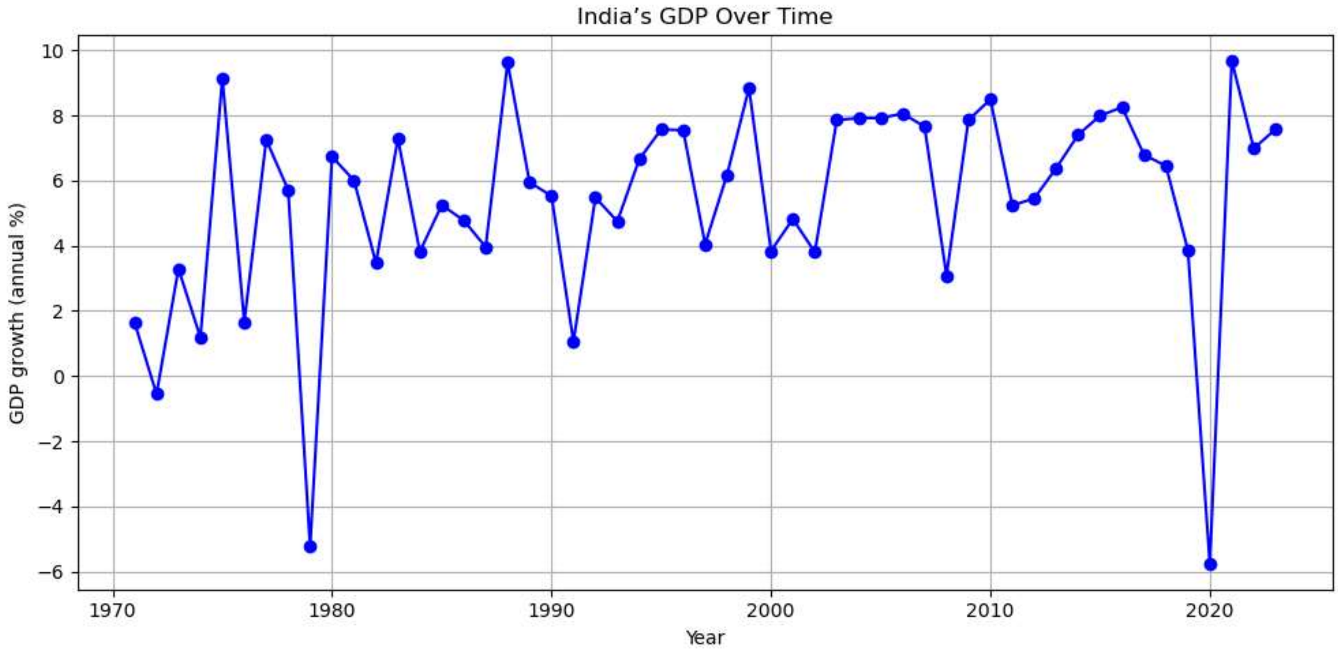
✍️ Solved Questions:-

❓ Step 5: Answering PDF Questions (Q1–Q15)

Q1: Why is GDP important?

Answer: GDP provides insight into the size and performance of an economy. It's a key indicator for policymakers and investors.

```
# Line plot to visualize GDP trend
plt.figure(figsize=(10, 5))
plt.plot(df.index, df['GDP growth (annual %)'], marker='o', color='blue')
plt.title('India’s GDP Over Time')
plt.xlabel('Year')
plt.ylabel('GDP growth (annual %)')
plt.grid(True)
plt.tight_layout()
plt.show()
```

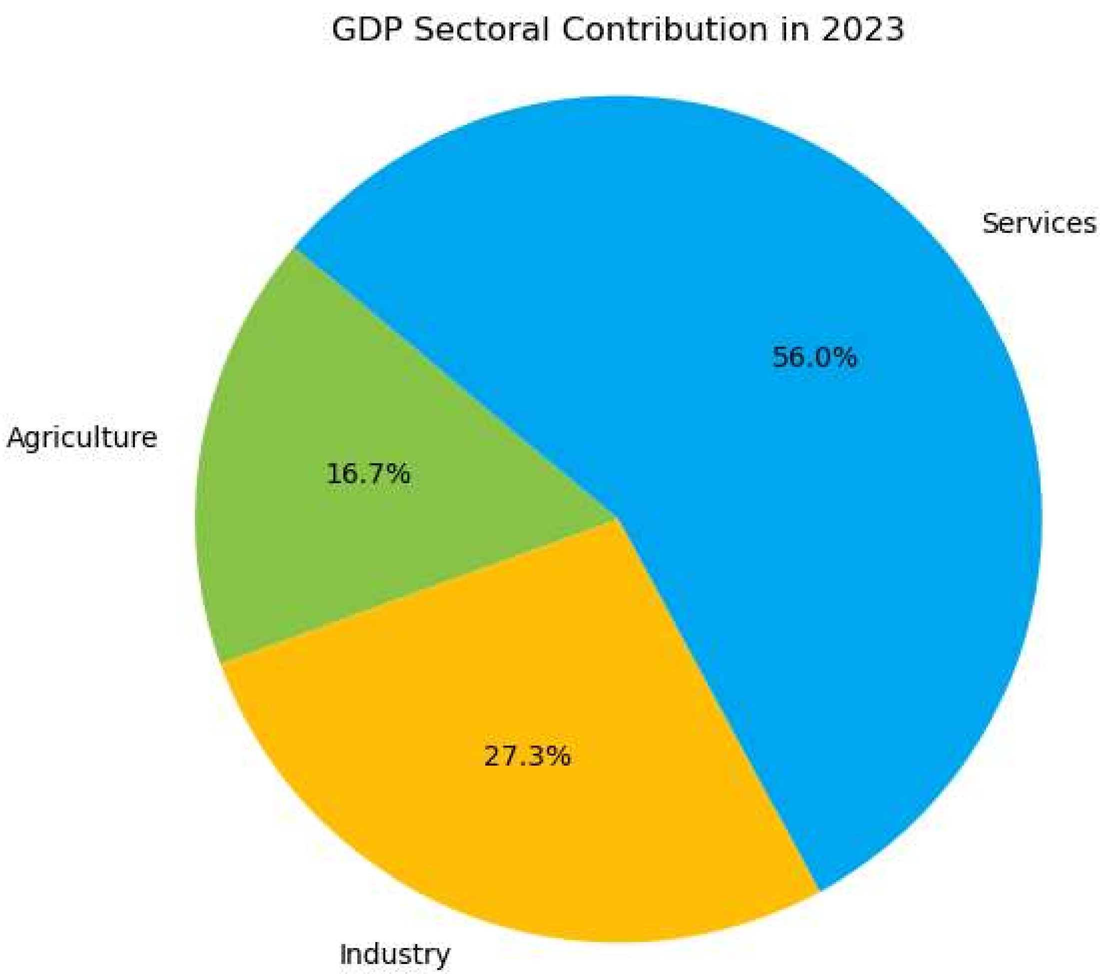


Q2: What are the main components of India's GDP?

Answer: Agriculture, Industry, and Services.

```
# Extract the latest year's data
latest_year = df.index.max()
latest_data = df[df.index == latest_year].iloc[0]

# Values for pie chart
values = [
    latest_data['Agriculture Contribution (% of GDP)'],
    latest_data['Industry Contribution (% of GDP)'],
    latest_data['Services Contribution (% of GDP)']
]

labels = ['Agriculture', 'Industry', 'Services']
colors = ['#8BC34A', '#FFC107', '#03A9F4']

# Plot pie chart
plt.figure(figsize=(6, 6))
plt.pie(values, labels=labels, autopct='%1.1f%%', colors=colors, startangle=140)
plt.title(f'GDP Sectoral Contribution in {latest_year}')
```

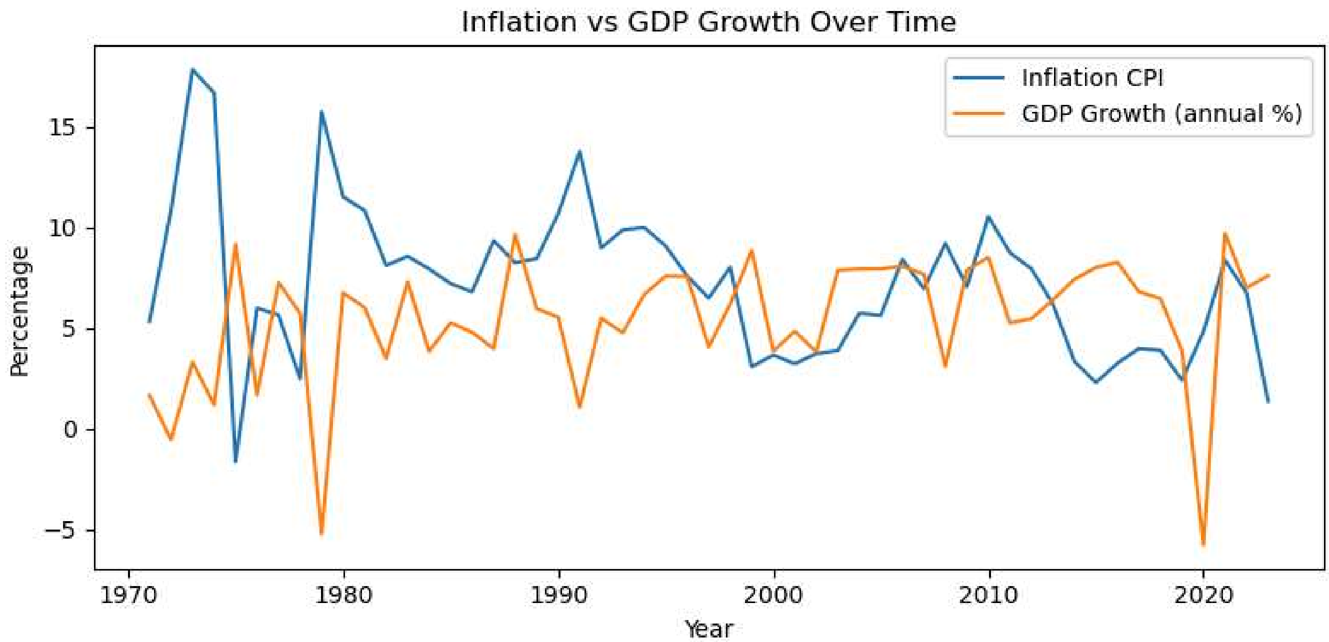
```
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```



Q3: How does inflation affect GDP and economic growth in India?

Answer: High inflation can reduce purchasing power and discourage investment, which slows GDP growth.

```
import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize=(8, 4))
sns.lineplot(x='Year', y='Inflation CPI', data=df, label='Inflation CPI')
sns.lineplot(x='Year', y='GDP growth (annual %)', data=df, label='GDP Growth (annual %)')
plt.title("Inflation vs GDP Growth Over Time")
plt.xlabel("Year")
plt.ylabel("Percentage")
plt.legend()
plt.tight_layout()
plt.show()
```

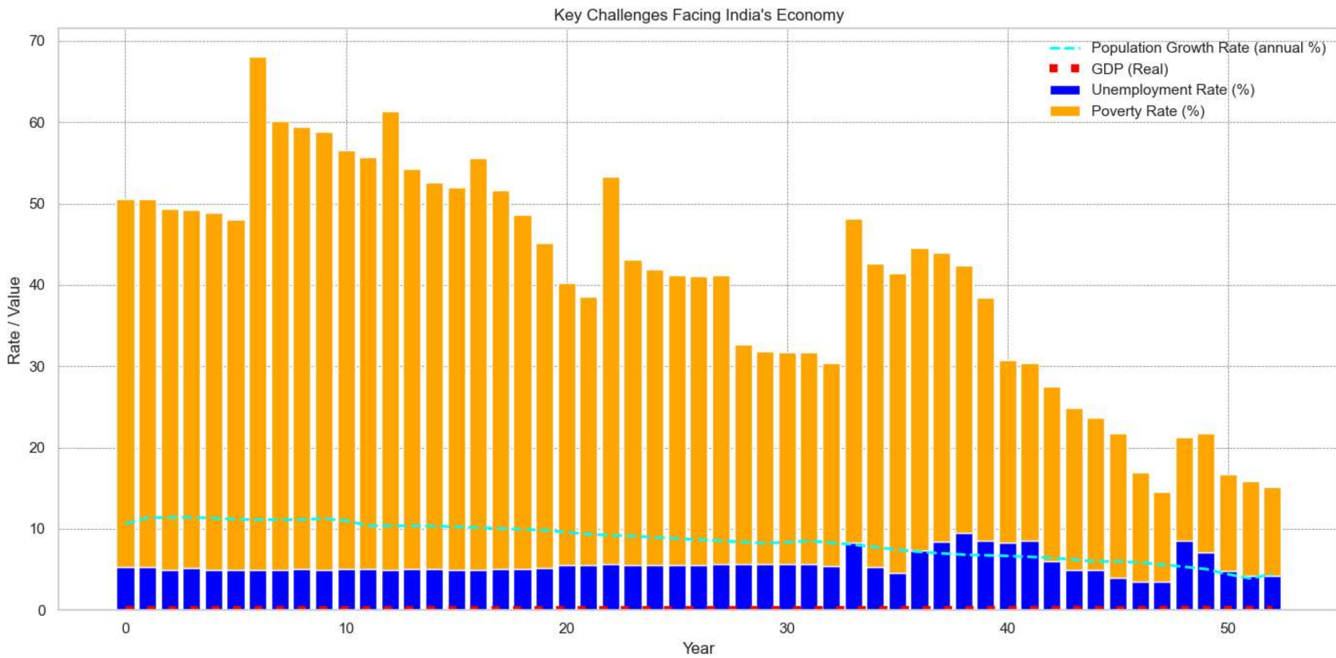


Q4: What are the main challenges facing India's economy?

Answer: Inflation, fiscal deficit, unemployment, public debt, and poverty.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')
plt.figure(figsize=(14, 7))

# Stacked bar: Poverty + Unemployment
plt.bar(df.index, df['Unemployment Rate (%)'], label='Unemployment Rate (%)', color='blue')
plt.bar(df.index, df['Poverty Rate (%)'], bottom=df['Unemployment Rate (%)'], label='Poverty

# Line plot: Population Growth
plt.plot(df.index, df['Population Growth Rate (annual %)'] * 2, label='Population Growth Rat

# Line plot: GDP (Real)
plt.plot(df.index, df['GDP (Real)'] / 1e5, label='GDP (Real)', color='red', linewidth=6, lir
```

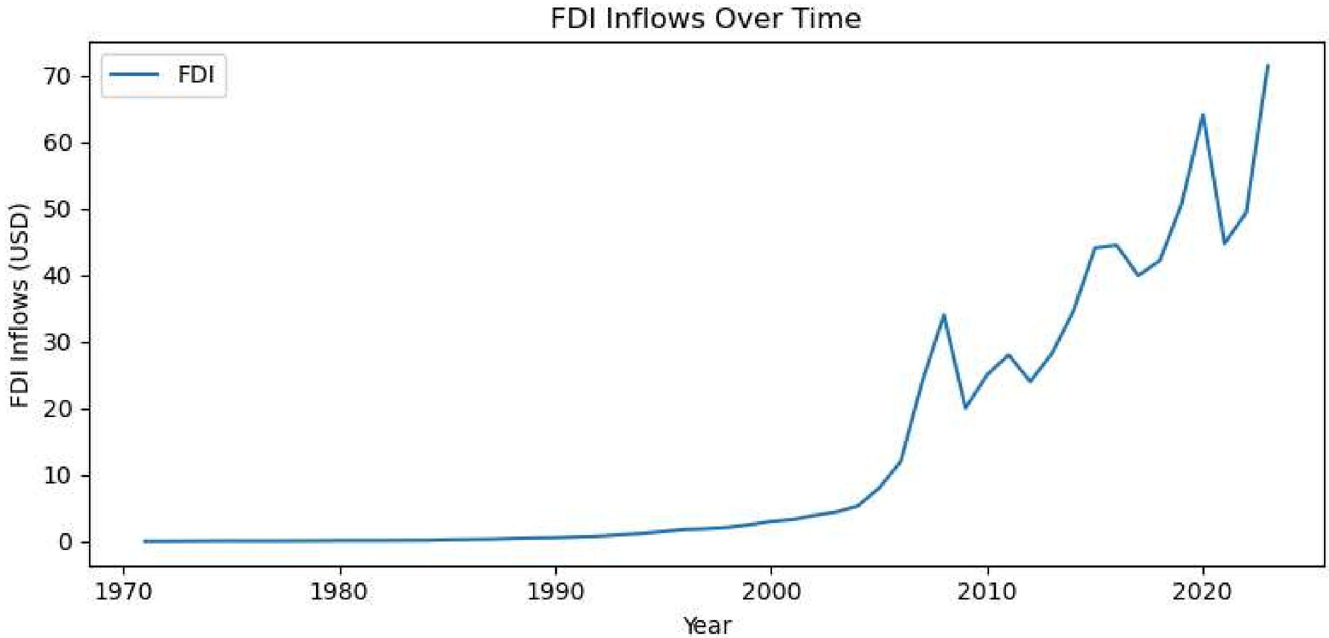
```
# Formatting
plt.xlabel("Year")
plt.ylabel("Rate / Value")
plt.title("Key Challenges Facing India's Economy")
plt.legend()
plt.grid(color='gray', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```



Q5: What role does foreign investment play in India's economic growth?

Answer: FDI contributes capital, technology, and employment, positively impacting GDP.

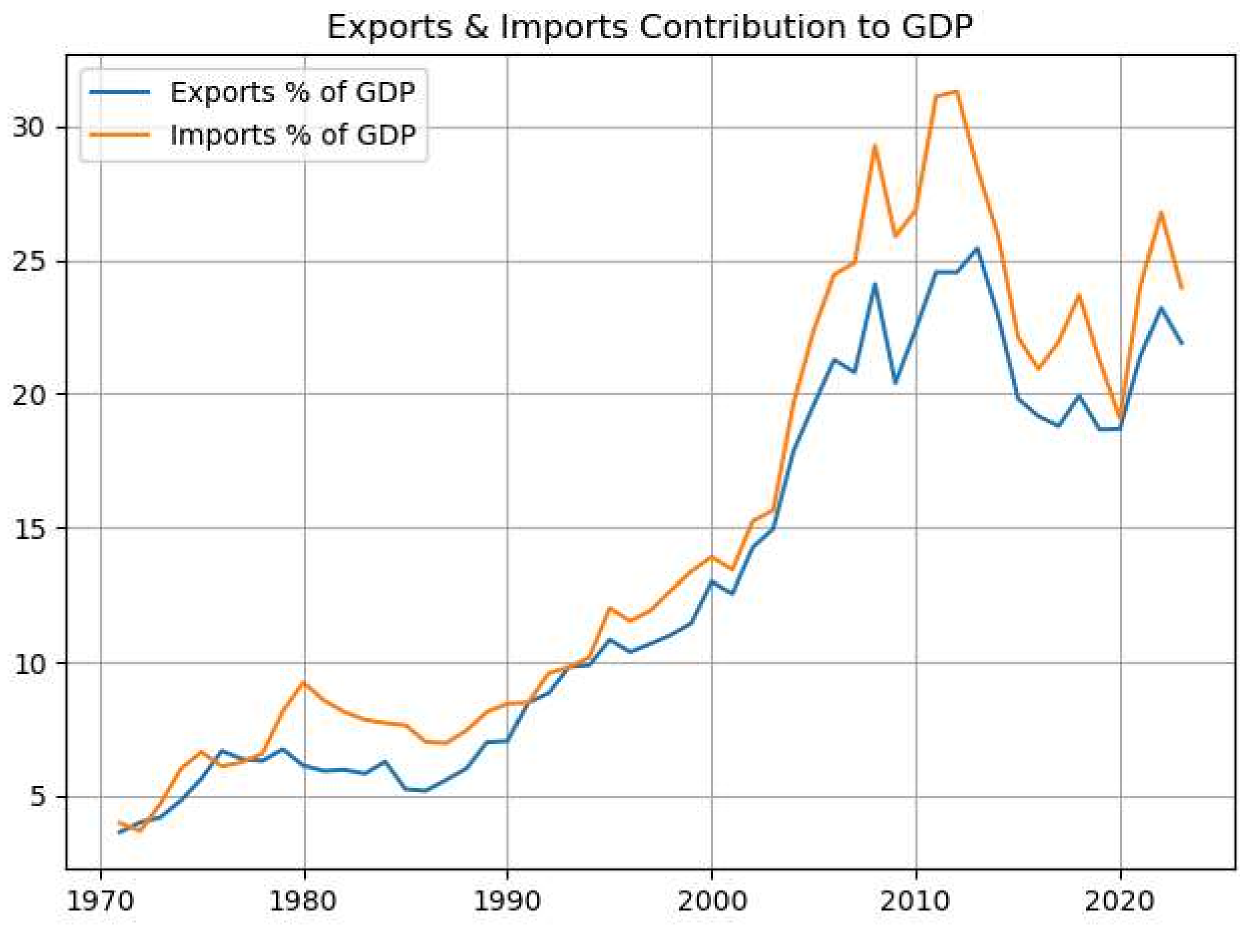
```
plt.figure(figsize=(8, 4))
sns.lineplot(x='Year', y='FDI Inflows (USD)', data=df, label='FDI')
plt.title("FDI Inflows Over Time")
plt.tight_layout()
plt.show()
```



Q6: What is the contribution of exports and imports to India's GDP?

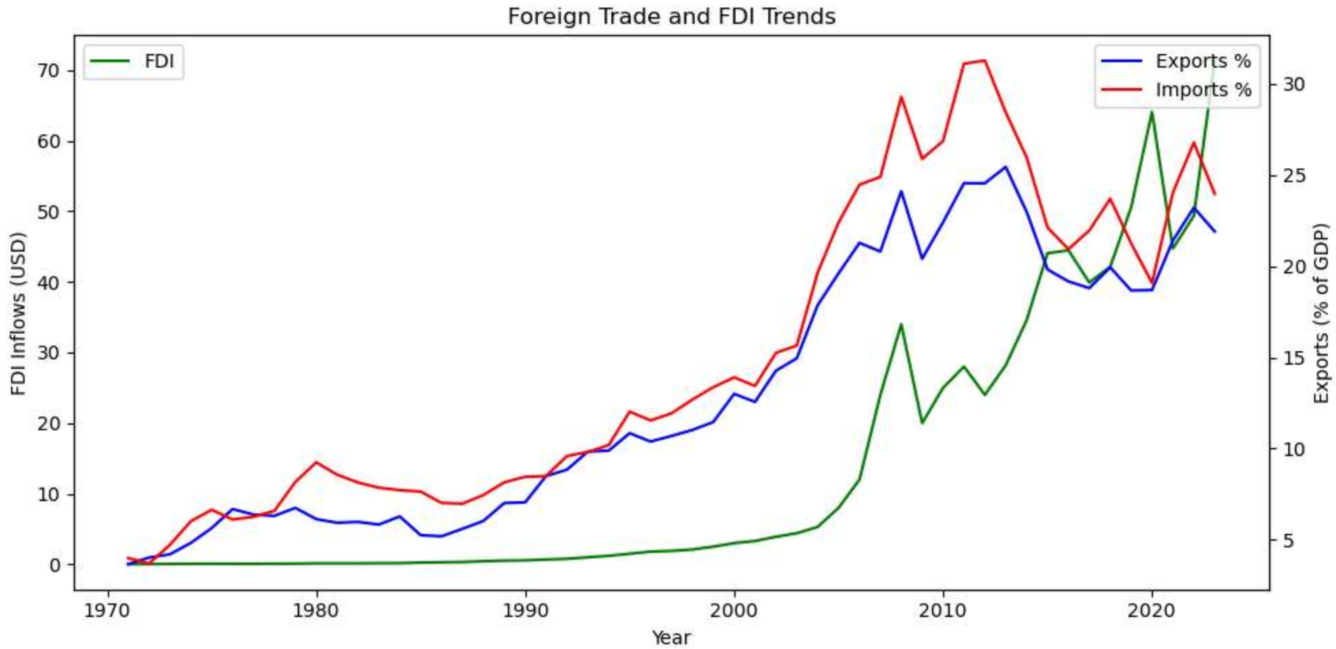
Answer: These affect the trade balance and external demand.

```
plt.plot(df.index, df['Exports (% of GDP)'], label='Exports % of GDP')
plt.plot(df.index, df['Imports  (% of GDP)'], label='Imports % of GDP')
plt.title("Exports & Imports Contribution to GDP")
plt.legend()
plt.grid()
plt.tight_layout()
plt.show()
```



Q7: Visualize Foreign Trade and FDI trends

```
fig, ax1 = plt.subplots(figsize=(10, 5))
sns.lineplot(x='Year', y='FDI Inflows (USD)', data=df, ax=ax1, label="FDI", color="green")
ax2 = ax1.twinx()
sns.lineplot(x='Year', y='Exports (% of GDP)', data=df, ax=ax2, label="Exports %", color="bl
sns.lineplot(x='Year', y='Imports  (% of GDP)', data=df, ax=ax2, label="Imports %", color="r
plt.title("Foreign Trade and FDI Trends")
fig.tight_layout()
plt.show()
```



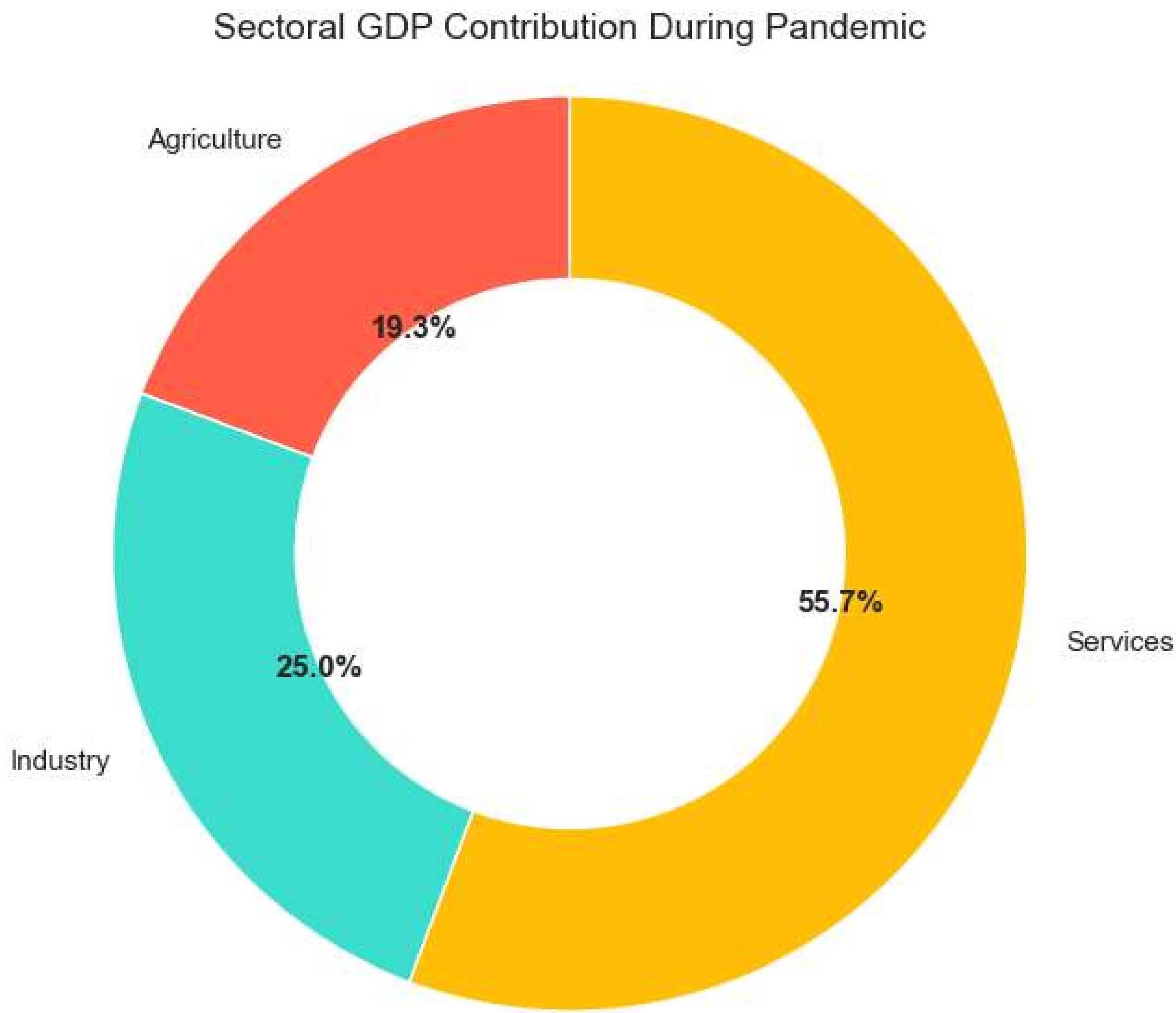
Q8: What was the sectoral contribution in GDP during pandemic?

```
df[df.index >= 2019][[ 'Agriculture Contribution (% of GDP)',
                        'Industry Contribution (% of GDP)',
                        'Services Contribution (% of GDP)']]
```

Year	Agriculture Contribution (% of GDP)	Industry Contribution (% of GDP)	Services Contribution (% of GDP)
2019	14.66	25.43	53.94
2020	17.43	22.52	50.15
2021	18.26	26.14	50.18
2022	16.64	26.38	53.32
2023	15.97	26.22	53.70

```
import pandas as pd
import numpy as np
```

```
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

# Sectoral contribution data during pandemic
labels = ['Agriculture', 'Industry', 'Services']
sizes = [19.3, 25.0, 55.7]
colors = ['#FF6347', '#40E0D0', '#FFC107']

# Plotting a donut pie chart
plt.figure(figsize=(7, 7))
wedges, texts, autotexts = plt.pie(
    sizes,
    labels=labels,
    autopct='%1.1f%%',
    startangle=90,
    colors=colors,
    wedgeprops=dict(width=0.4)
)

# Formatting
plt.setp(autotexts, size=12, weight="bold")
plt.title("Sectoral GDP Contribution During Pandemic", fontsize=14)
plt.axis('equal')
plt.show()
```

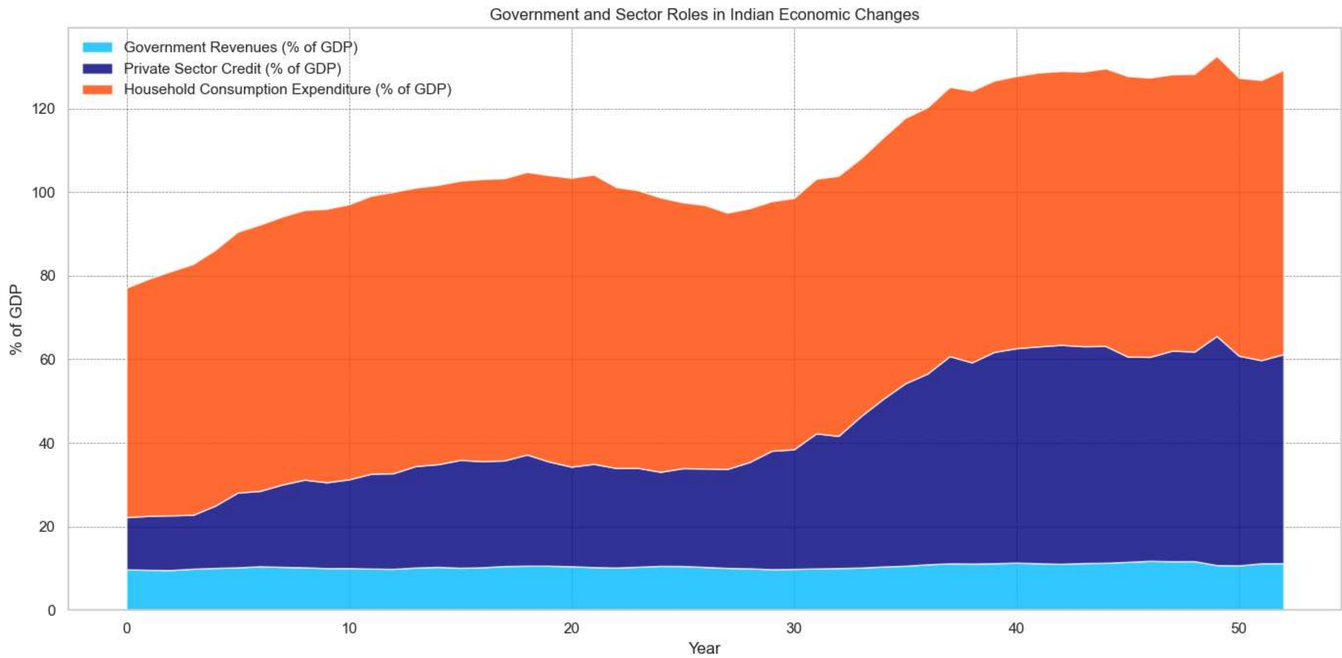


Q9: How is govt. responsible for Indian Economy changes?

Answer: Government policies influence taxes, spending, fiscal deficit, and public investment.

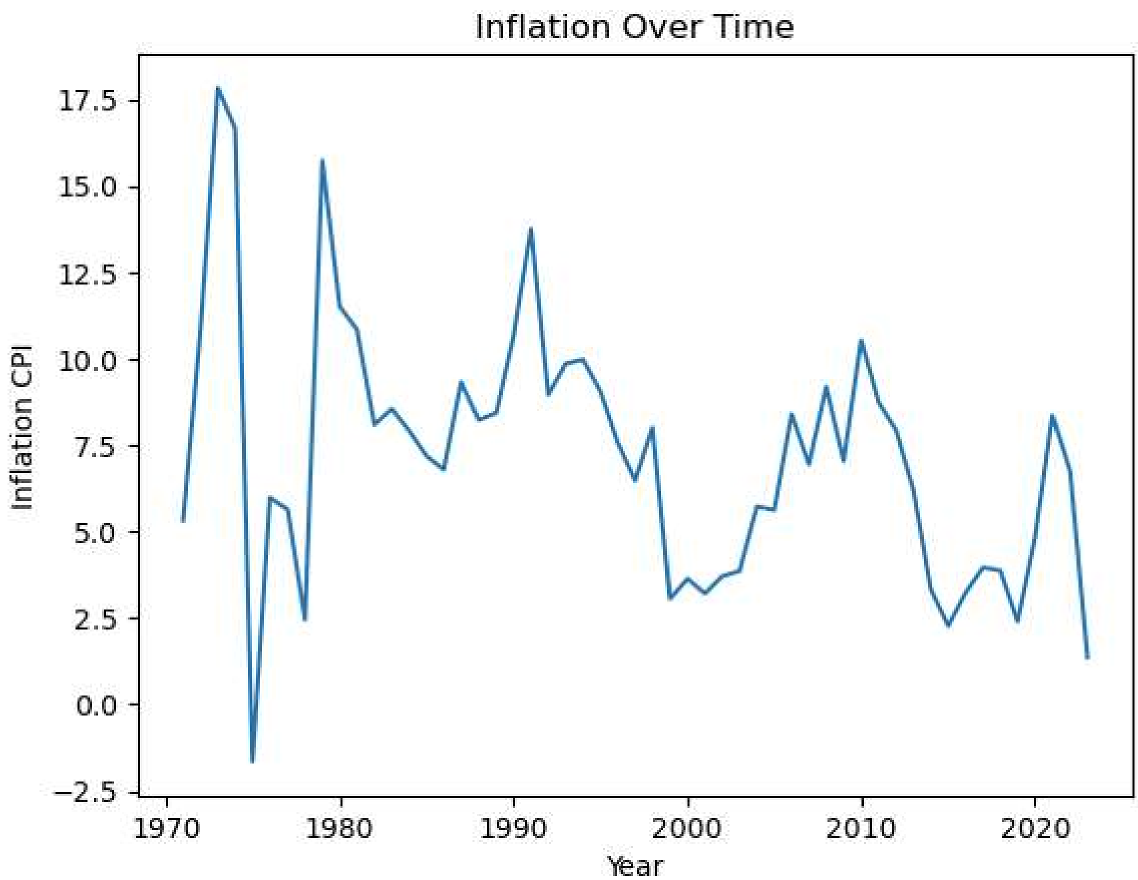
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')
```

```
years = df.index
govt_revenue = df['Government Revenues (% of GDP)']
private_credit = df['Private Sector Credit (% of GDP)']
household_spending = df['Household Consumption Expenditure (% of GDP)']

# Plotting stacked area chart
plt.figure(figsize=(14, 7))
plt.stackplot(df.index,
    df['Government Revenues (% of GDP)'],
    df['Private Sector Credit (% of GDP)'],
    df['Household Consumption Expenditure (% of GDP)'],
    labels=['Government Revenues (% of GDP)', 'Private Sector Credit (% of GDP)', 'Householc
    colors=['#00BFFF', '#000080', '#FF4500'],
    alpha=0.8
)

plt.title("Government and Sector Roles in Indian Economic Changes")
plt.xlabel("Year")
plt.ylabel("% of GDP")
plt.legend(loc="upper left")
plt.grid(color='gray', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```



Q10: How is inflation changing with time?

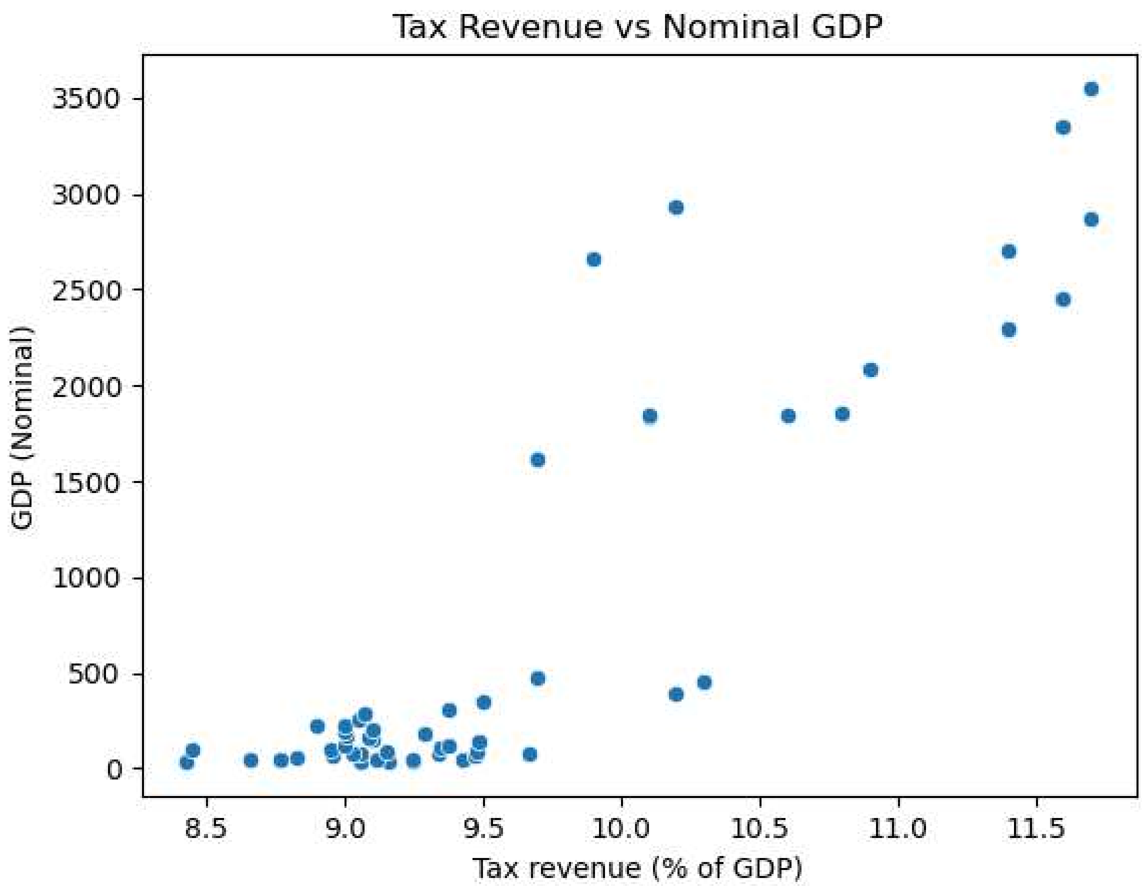
```
sns.lineplot(data=df, x='Year', y='Inflation CPI')
plt.title("Inflation Over Time")
plt.show()
```



Q11: How does tax collection affect India’s GDP?

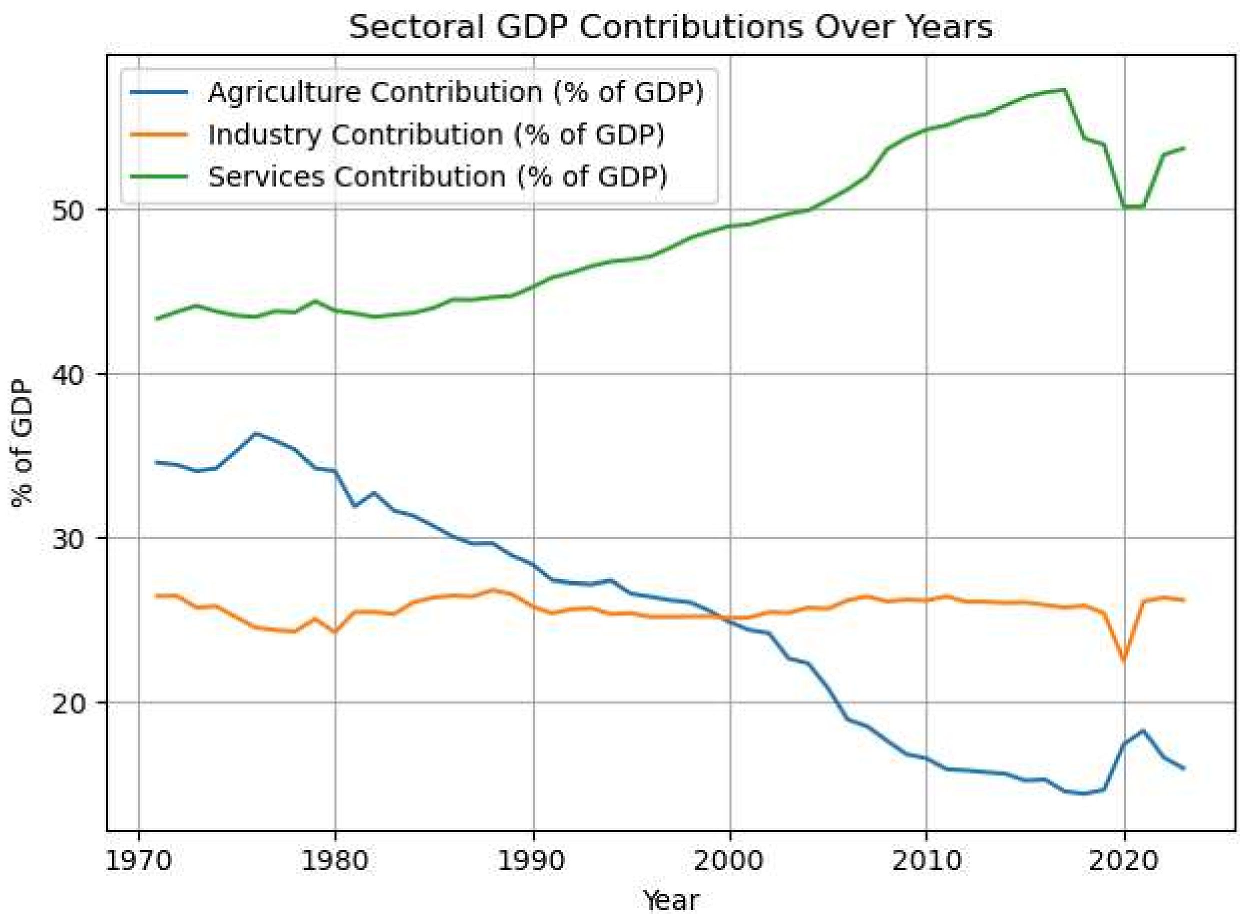
Answer: Higher tax revenue supports infrastructure, subsidies, and growth.

```
sns.scatterplot(x='Tax revenue (% of GDP)', y='GDP (Nominal)', data=df)
plt.title("Tax Revenue vs Nominal GDP")
plt.show()
```



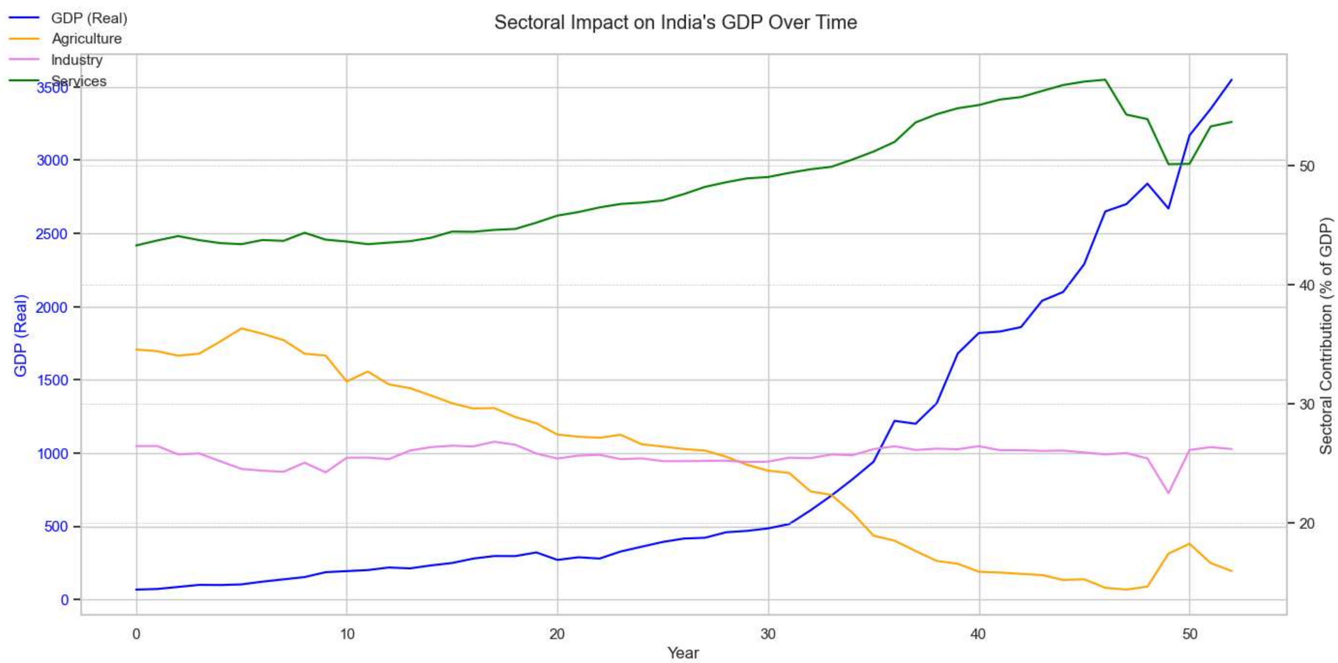
Q12: Show effect of Agriculture, Industry, and Services on India’s GDP over time

```
df[['Agriculture Contribution (% of GDP)', 'Industry Contribution (% of GDP)', 'Services Cor
plt.title("Sectoral GDP Contributions Over Years")
plt.xlabel("Year")
plt.ylabel("% of GDP")
plt.grid(True)
plt.tight_layout()
plt.show()
```



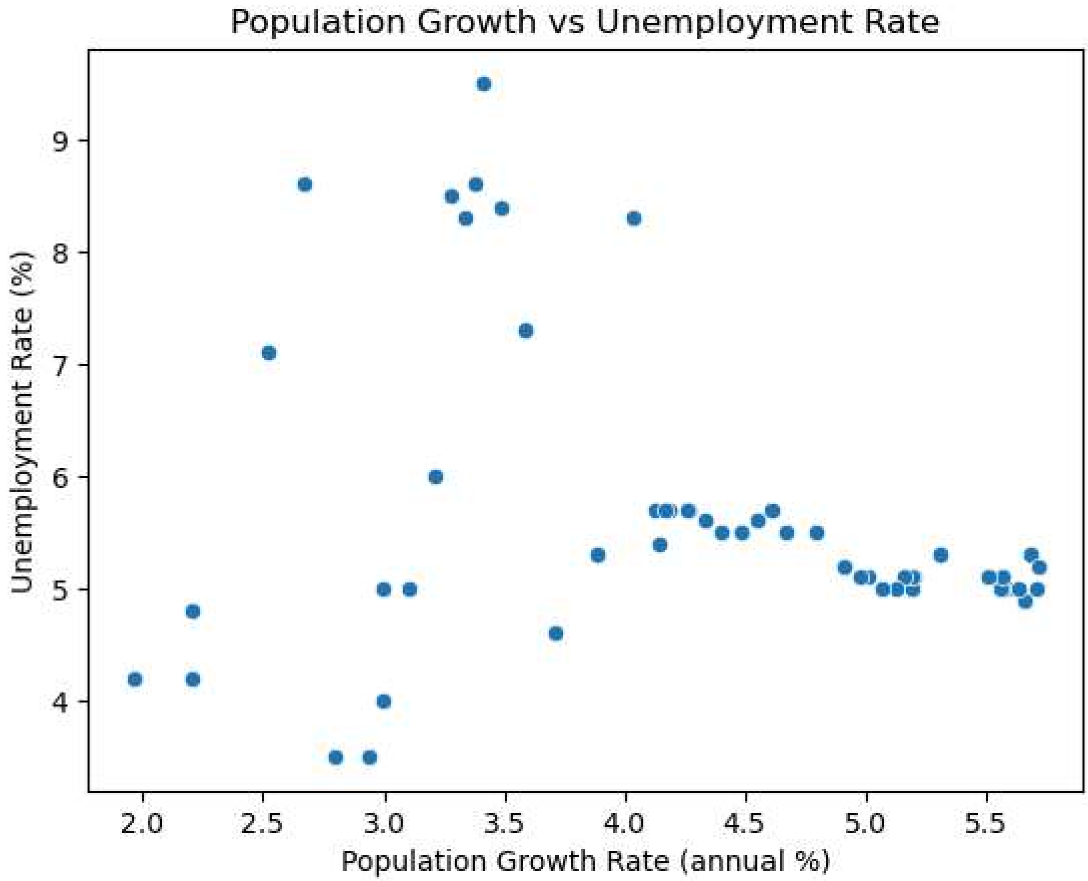

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

fig, ax1 = plt.subplots(figsize=(14, 7))

# GDP line on primary Y-axis
ax1.plot(df.index, df['GDP (Real)'], color='blue', label='GDP (Real)')
ax1.set_xlabel("Year")
ax1.set_ylabel("GDP (Real)", color='blue')
ax1.tick_params(axis='y', labelcolor='blue')

# Twin axis for sector contributions
ax2 = ax1.twinx()
ax2.plot(df.index, df['Agriculture Contribution (% of GDP)'], label='Agriculture', color='or
ax2.plot(df.index, df['Industry Contribution (% of GDP)'], label='Industry', color='violet')
```

```
ax2.plot(df.index, df['Services Contribution (% of GDP)'], label='Services', color='green')
ax2.set_ylabel("Sectoral Contribution (% of GDP)")
ax2.tick_params(axis='y')

fig.suptitle("Sectoral Impact on India's GDP Over Time")
fig.legend(loc='upper left')
plt.grid(True, linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```



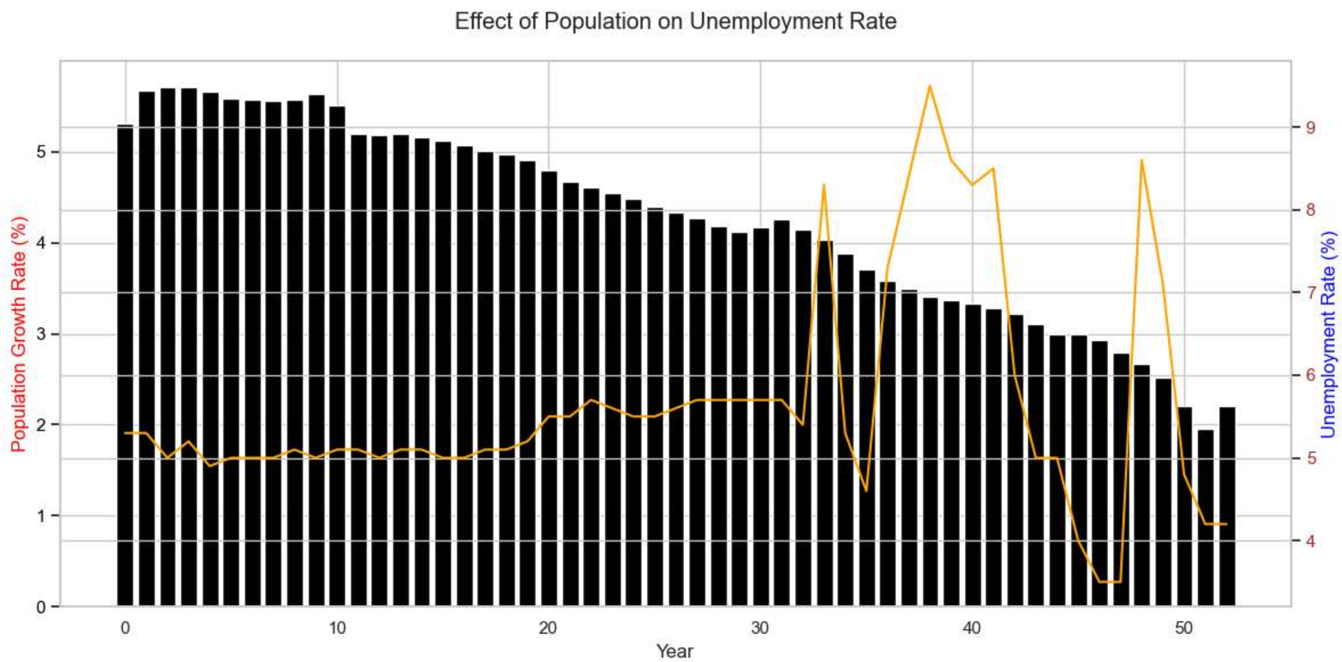
Q13: How does population affect unemployment rate?

```
sns.scatterplot(x='Population Growth Rate (annual %)', y='Unemployment Rate (%)', data=df)
plt.title("Population Growth vs Unemployment Rate")
plt.show()
```



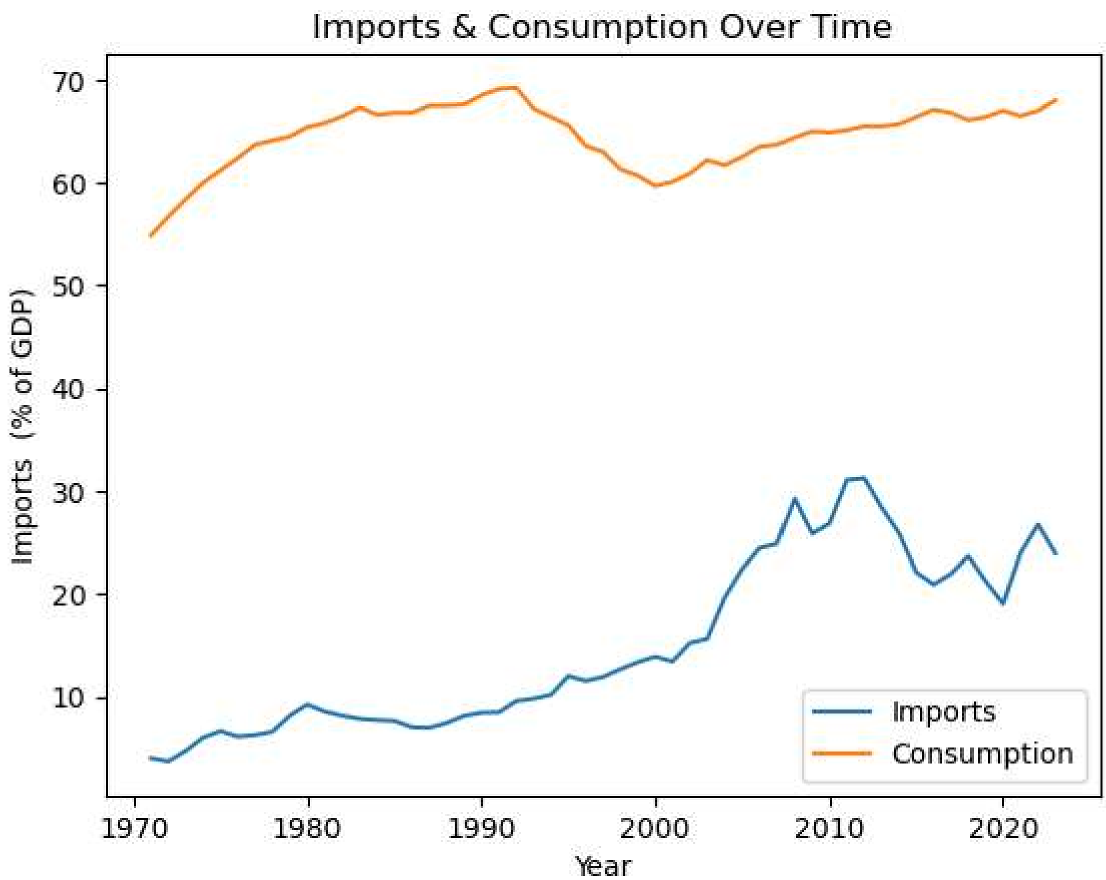

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

fig, ax1 = plt.subplots(figsize=(12, 6))

# Bar chart for Population Growth
ax1.bar(df.index, df['Population Growth Rate (annual %)'], color='black', label='Population
ax1.set_xlabel("Year")
ax1.set_ylabel("Population Growth Rate (%)", color='red')
ax1.tick_params(axis='y', labelcolor='black')

# Line plot for Unemployment Rate
ax2 = ax1.twinx()
ax2.plot(df.index, df['Unemployment Rate (%)'], color='orange', label='Unemployment Rate')
ax2.set_ylabel("Unemployment Rate (%)", color='blue')
ax2.tick_params(axis='y', labelcolor='brown')
```

```
fig.suptitle("Effect of Population on Unemployment Rate")
fig.tight_layout()
plt.show()
```



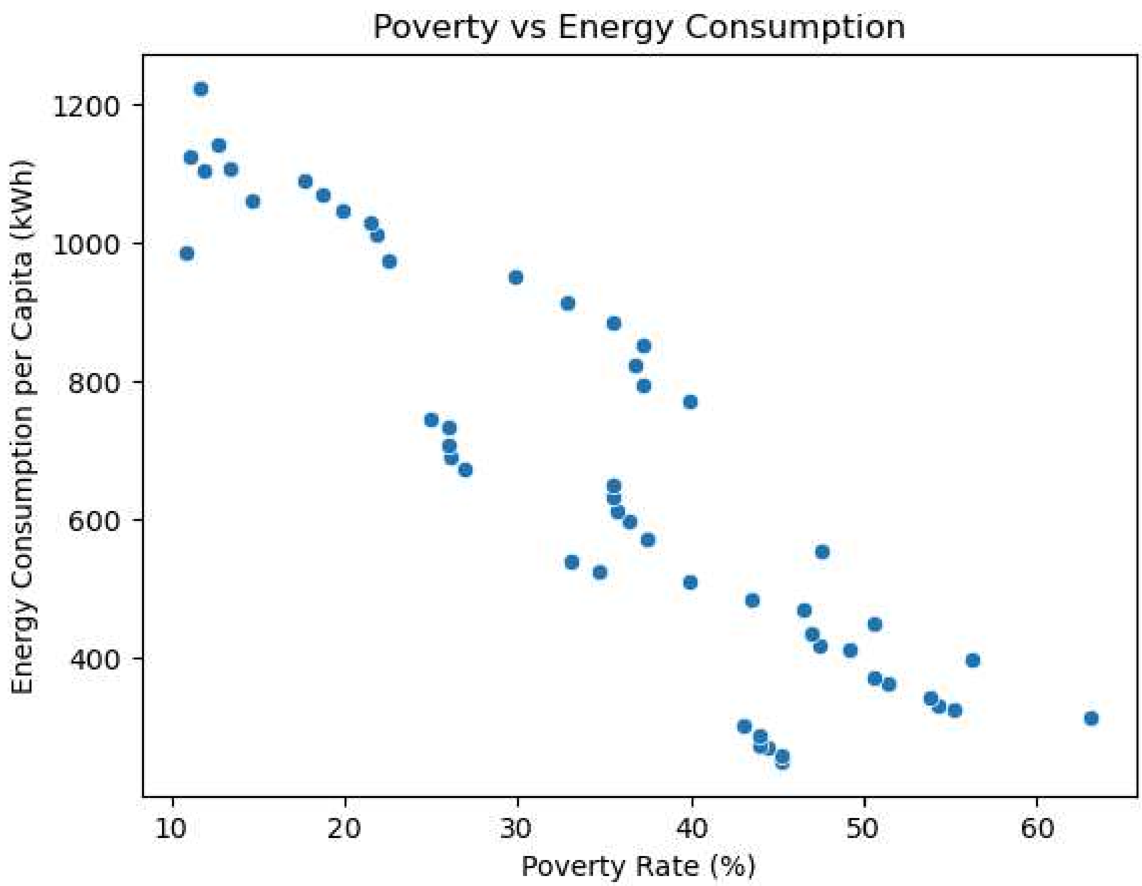
Q14: How do imports and consumption change over time?

```
sns.lineplot(x='Year', y='Imports  (% of GDP)', data=df, label='Imports')
sns.lineplot(x='Year', y='Household Consumption Expenditure (% of GDP)', data=df, label='Cor
plt.title("Imports & Consumption Over Time")
plt.legend()
plt.show()
```



Q15: What’s the effect of poverty and energy consumption?

```
sns.scatterplot(x='Poverty Rate (%)', y='Energy Consumption per Capita (kWh)', data=df)
plt.title("Poverty vs Energy Consumption")
plt.show()
```



✍️ Practice Questions:-

📝 Step-6: Extra Questions with Solutions (by own)

This notebook contains 30 questions with their answers and respective Python codes, covering various aspects of India's GDP and economic indicators.

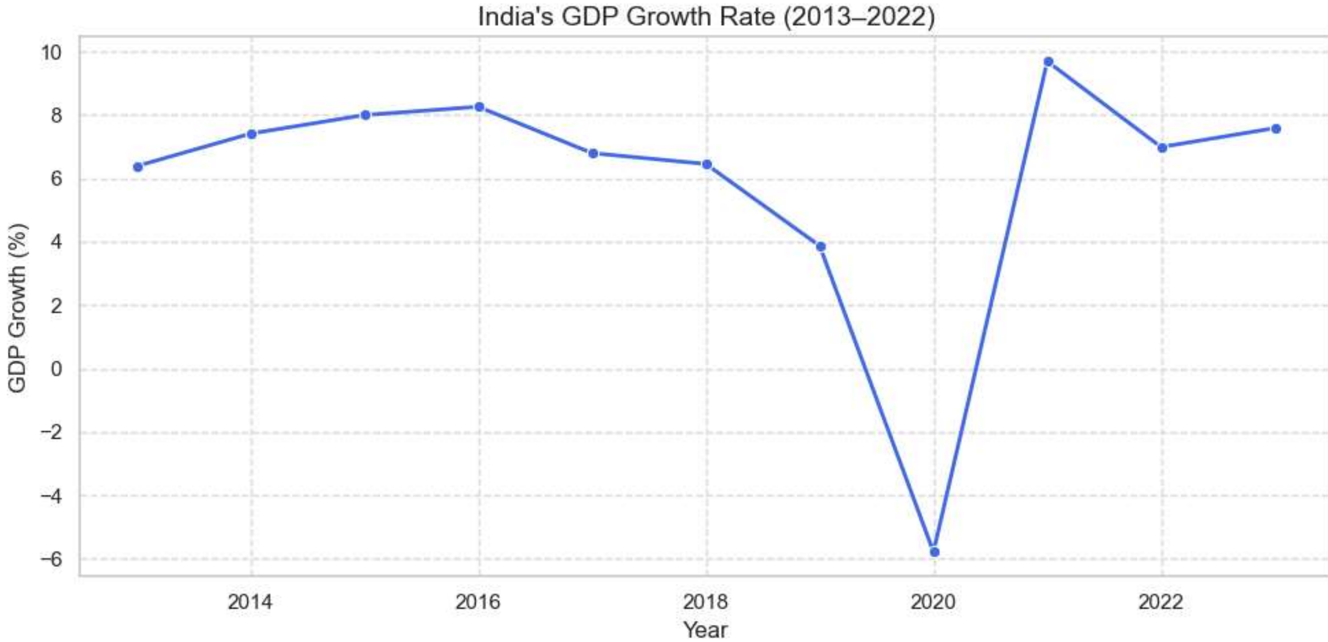
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Q1: What is the trend of India's GDP growth over the last 10 years? Visualize it using a line plot.

India's GDP trend over the last decade shows fluctuations due to global and domestic factors. While Real GDP reflects economic output adjusted for inflation, Nominal GDP includes current prices. A line plot shows how GDP and its growth have changed over time.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Clean column names
df.columns = df.columns.str.strip()

# Print column names to verify
print("Available columns:\n", df.columns.tolist())

# OPTIONAL: Rename for simplicity if needed
# Check actual name first, for example:
# 'GDP growth (annual %)' or 'GDP Growth (%)'
# Replace below only if exact name differs
if 'GDP growth (annual %)' in df.columns:
    df.rename(columns={'GDP growth (annual %)': 'GDP Growth'}, inplace=True)
elif 'GDP Growth (%)' in df.columns:
    df.rename(columns={'GDP Growth (%)': 'GDP Growth'}, inplace=True)

# Filter last 10 years
df_10yr = df[df['Year'] >= 2013]

# Check for NaN and drop if needed
df_10yr = df_10yr[['Year', 'GDP Growth']].dropna()

# Plotting
plt.figure(figsize=(10, 5))
sns.set_theme(style='whitegrid')
sns.lineplot(data=df_10yr, x='Year', y='GDP Growth', marker='o', linewidth=2, color='royalbl

# Titles and labels
plt.title("India's GDP Growth Rate (2013-2022)", fontsize=14)
plt.xlabel("Year")
plt.ylabel("GDP Growth (%)")
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()
```

```
Available columns:
 ['Year', 'GDP (Nominal)', 'GDP (Real)', 'GDP growth (annual %)', 'Agriculture Contribut
```

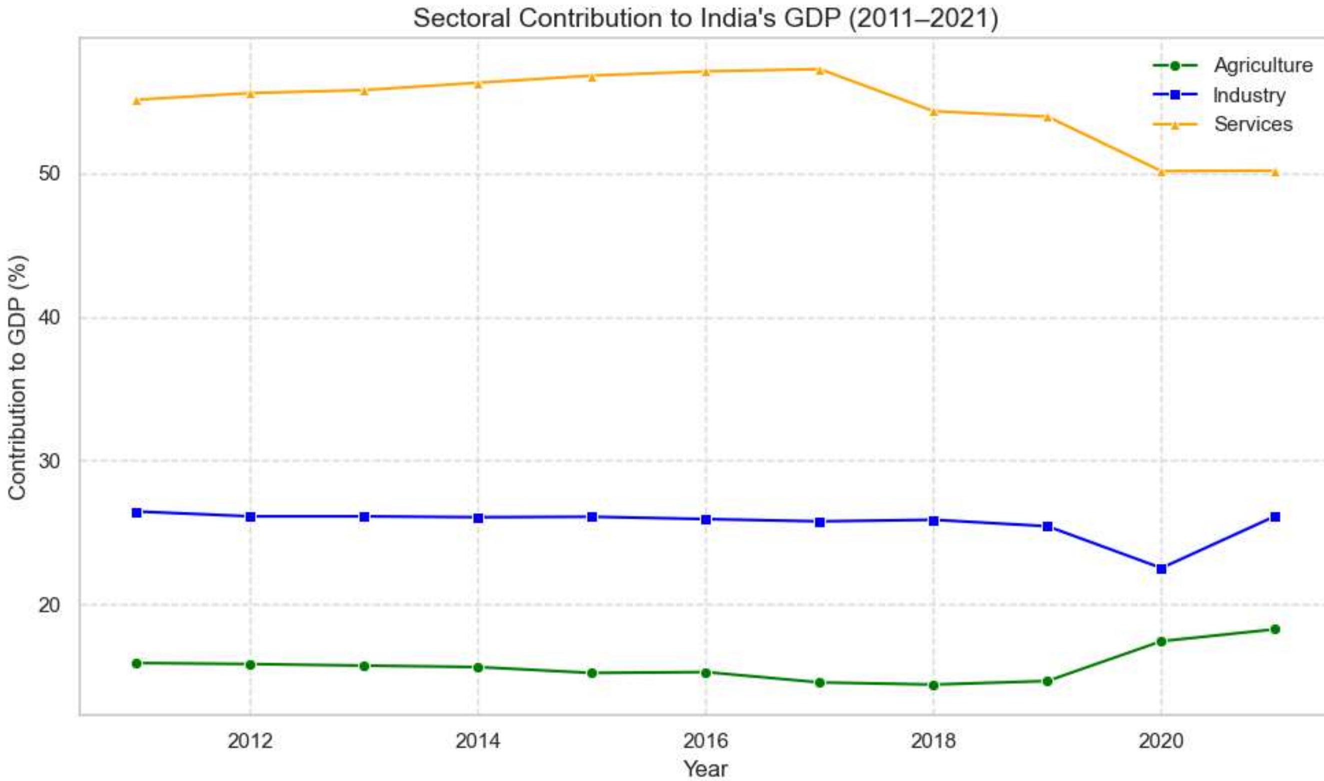


Q2: How has the contribution of Agriculture, Industry, and Services sectors to GDP changed from 2011 to 2021?

Over the last decade, India's Services sector has dominated GDP contribution, while Industry has been stable, and Agriculture has gradually declined as a percentage. A multi-line plot clearly shows these sectoral trends.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Clean and normalize column names
df.columns = df.columns.str.strip()
```

```
# Rename if needed for clarity
# Check the actual column names in your data and adjust if they differ
df.rename(columns={
    'Agriculture Contribution (% of GDP)': 'Agriculture',
    'Industry Contribution (% of GDP)': 'Industry',
    'Services Contribution (% of GDP)': 'Services'
}, inplace=True)

# Filter for 2011-2021
df_sectors = df[(df['Year'] >= 2011) & (df['Year'] <= 2021)]

# Plotting
plt.figure(figsize=(10, 6))
sns.set_theme(style='whitegrid')

sns.lineplot(x='Year', y='Agriculture', data=df_sectors, marker='o', label='Agriculture', cc
sns.lineplot(x='Year', y='Industry', data=df_sectors, marker='s', label='Industry', color='t
sns.lineplot(x='Year', y='Services', data=df_sectors, marker='^', label='Services', color='c

# Labels and title
plt.title("Sectoral Contribution to India's GDP (2011-2021)", fontsize=14)
plt.xlabel("Year")
plt.ylabel("Contribution to GDP (%)")
plt.legend()
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()
```

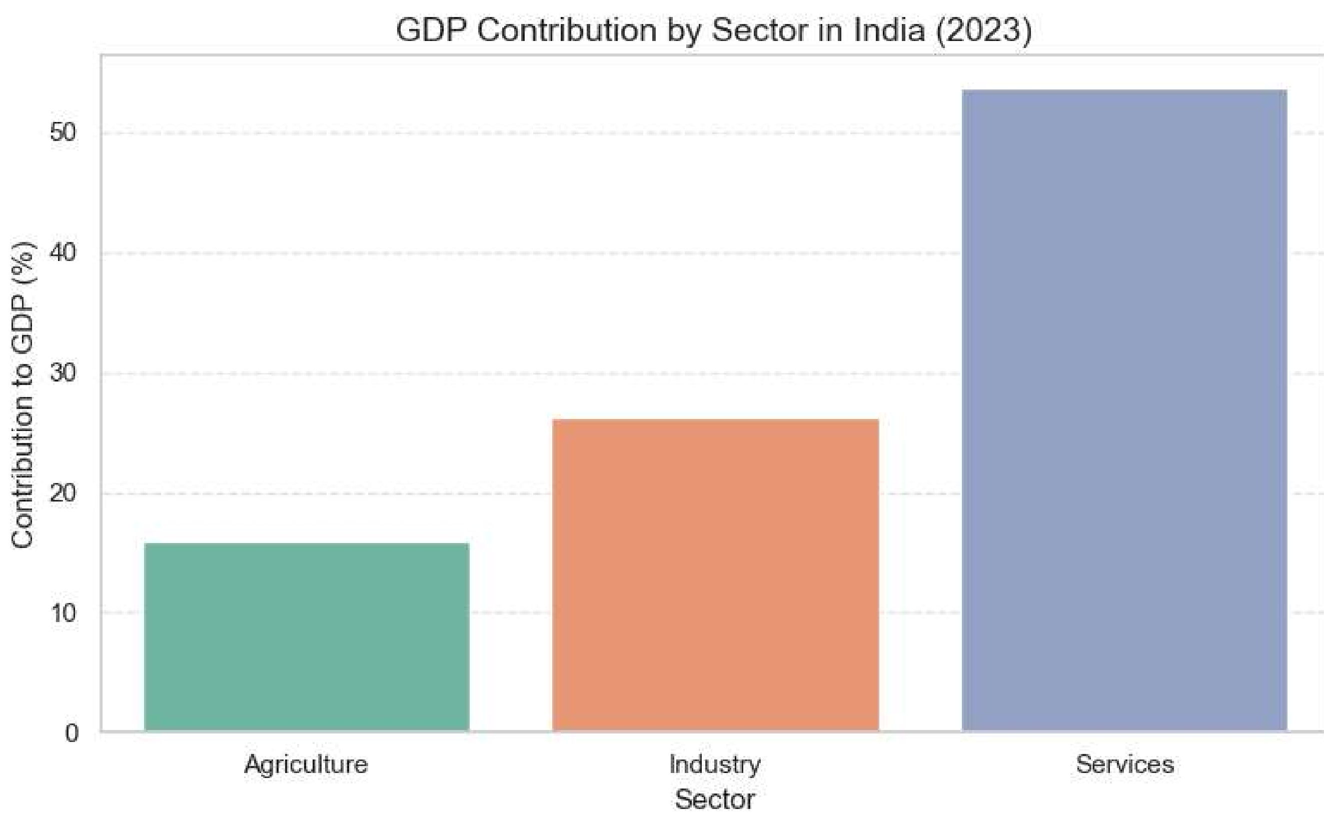


Q3: What is the contribution of key sectors to India's GDP in the year 2023?

In 2023, the Services sector was the largest contributor to India's GDP, followed by Industry and Agriculture. A bar graph highlights the comparative share of each sector in the economy.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")
df.columns = df.columns.str.strip()

# Rename for simplicity if needed
df.rename(columns={
    'Agriculture Contribution (% of GDP)': 'Agriculture',
```

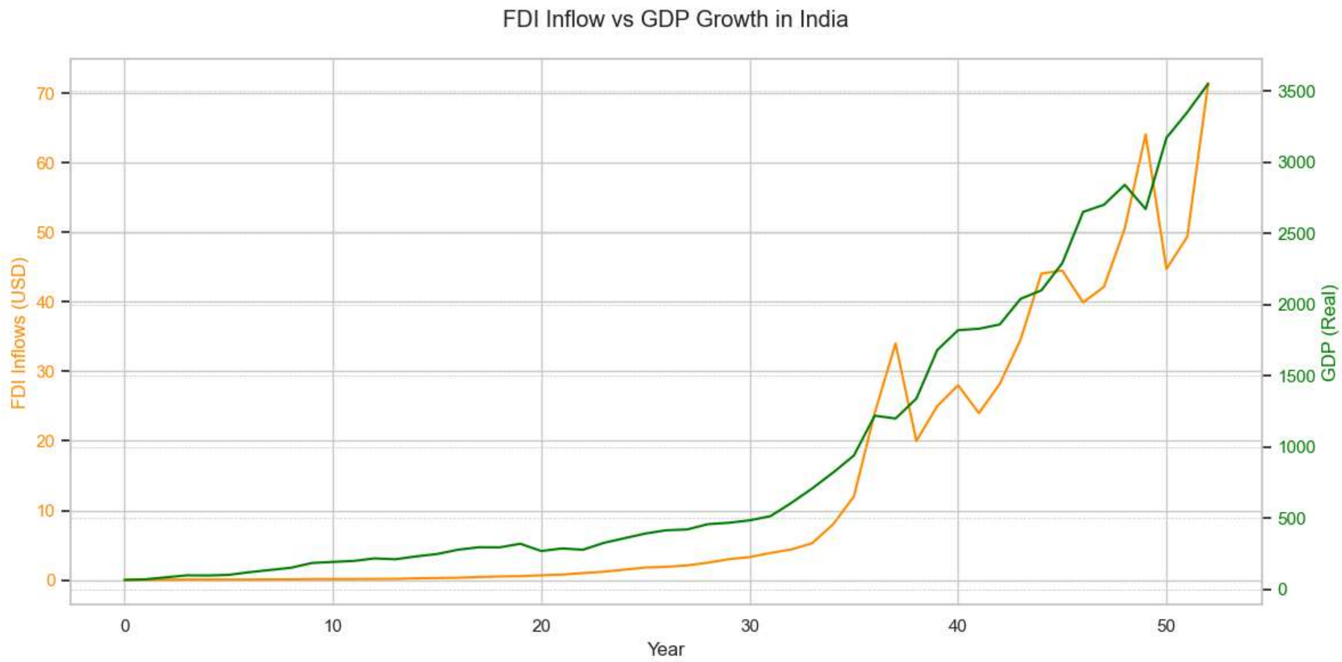
```
    'Industry Contribution (% of GDP)': 'Industry',
    'Services Contribution (% of GDP)': 'Services'
}, inplace=True)

# Filter for 2023
df_2023 = df[df['Year'] == 2023][['Agriculture', 'Industry', 'Services']].squeeze()

# Prepare data
sector_data = df_2023.to_frame().reset_index()
sector_data.columns = ['Sector', 'Contribution']

# Plot
plt.figure(figsize=(8, 5))
sns.set_theme(style='whitegrid')
sns.barplot(data=sector_data, x='Sector', y='Contribution', palette='Set2')

plt.title("GDP Contribution by Sector in India (2023)", fontsize=14)
plt.ylabel("Contribution to GDP (%)")
plt.xlabel("Sector")
plt.grid(True, axis='y', linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



Q4: What role does foreign direct investment (FDI) play in India's economic growth?

Answer: FDI brings capital, technology, and expertise, boosting production capacity and employment, positively impacting GDP.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

# FDI vs GDP (Real)
fig, ax1 = plt.subplots(figsize=(12, 6))

ax1.plot(df.index, df['FDI Inflows (USD)'], color='darkorange', label='FDI Inflow')
ax1.set_xlabel("Year")
ax1.set_ylabel("FDI Inflows (USD)", color='darkorange')
ax1.tick_params(axis='y', labelcolor='darkorange')

# Twin axis: GDP (Real)
ax2 = ax1.twinx()
ax2.plot(df.index, df['GDP (Real)'], color='green', label='GDP (Real)')
ax2.set_ylabel("GDP (Real)", color='green')
ax2.tick_params(axis='y', labelcolor='green')

fig.suptitle("FDI Inflow vs GDP Growth in India")
plt.grid(True, linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```

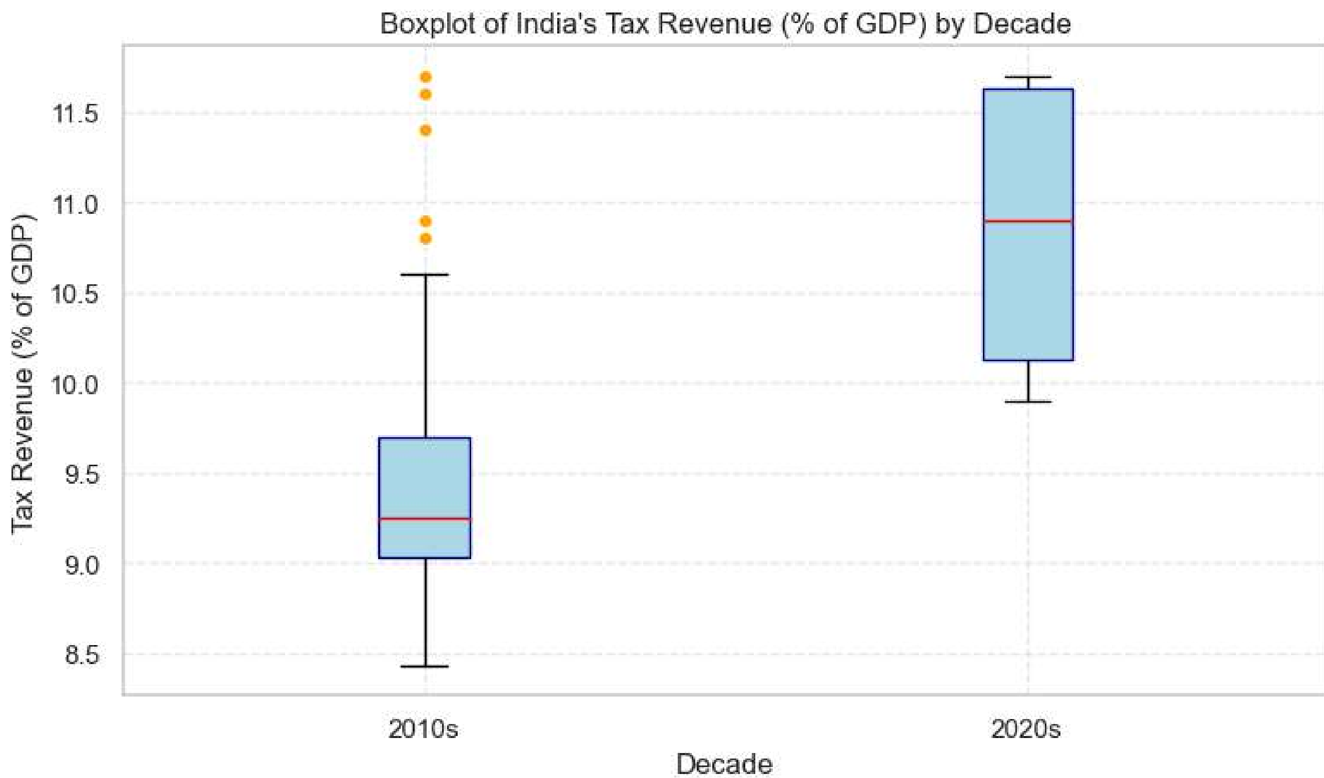


Q5: How has India’s tax revenue (% of GDP) varied year by year?

Answer: Tax revenue as a percentage of GDP reflects the government's capacity to mobilize resources. Year-wise variation indicates policy effectiveness, economic growth, and tax compliance. A boxplot helps visualize the spread across years.

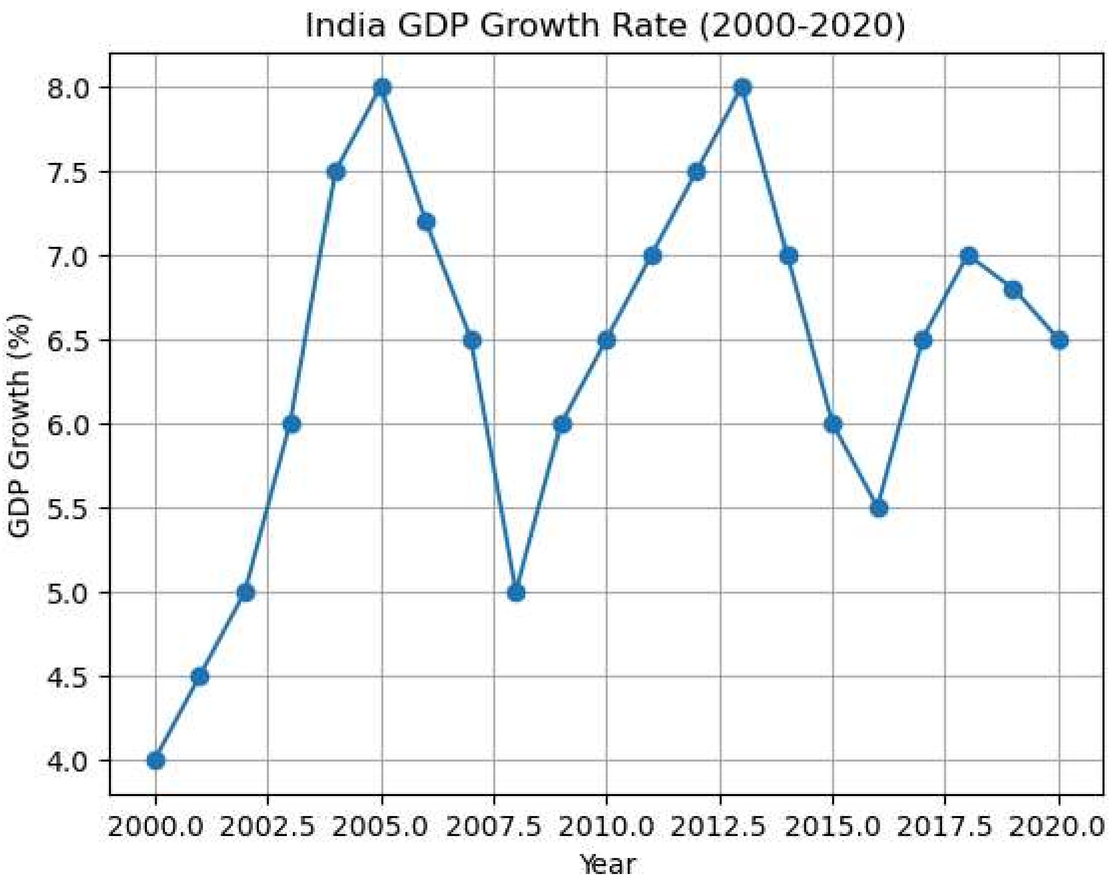
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load dataset
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Clean and prepare
df.columns = df.columns.str.strip()
```

```
df['Tax revenue (% of GDP)'] = pd.to_numeric(df['Tax revenue (% of GDP)'], errors='coerce')
df.dropna(subset=['Tax revenue (% of GDP)'], inplace=True)

# Create Decade or Year group if needed
df['Decade'] = df['Year'].apply(lambda x: '2010s' if x < 2020 else '2020s')

# Plot boxplot
sns.set_theme(style='whitegrid')
fig, ax = plt.subplots(figsize=(8, 5))
df.boxplot(column='Tax revenue (% of GDP)', by='Decade', ax=ax, patch_artist=True,
           boxprops=dict(facecolor='lightblue', color='navy'),
           medianprops=dict(color='red'),
           whiskerprops=dict(color='black'),
           capprops=dict(color='black'),
           flierprops=dict(marker='o', markerfacecolor='orange'))

ax.set_title("Boxplot of India's Tax Revenue (% of GDP) by Decade")
ax.set_ylabel("Tax Revenue (% of GDP)")
plt.suptitle("")
plt.grid(True, linestyle='--', alpha=0.4)
plt.tight_layout()
plt.show()
```



Q6: Visualize India’s GDP growth trend over the last 20 years.

```
years = list(range(2000, 2021))
gdp_growth = [4.0, 4.5, 5.0, 6.0, 7.5, 8.0, 7.2, 6.5, 5.0, 6.0, 6.5, 7.0, 7.5, 8.0, 7.0, 6.0
df = pd.DataFrame({'Year': years, 'GDP_Growth': gdp_growth})
plt.plot(df['Year'], df['GDP_Growth'], marker='o')
plt.title('India GDP Growth Rate (2000-2020)')
plt.xlabel('Year')
plt.ylabel('GDP Growth (%)')
plt.grid(True)
plt.show()
```

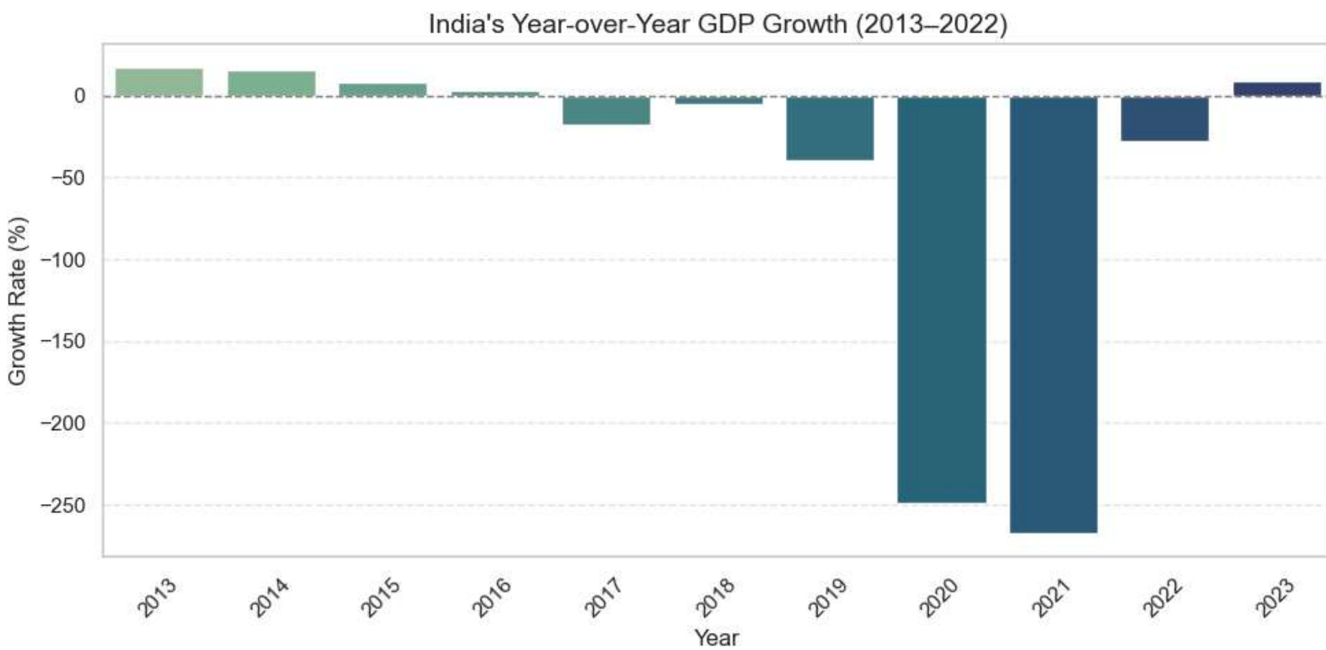


Q7: What is the year-over-year percentage growth in India's GDP from 2013 to 2022?

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Read data
file_path = "C:/Users/user/Onedrive/Documents/Dataset.xlsx"
df_gdp = pd.read_excel(file_path)
df_gdp.columns = df_gdp.columns.str.strip()

# Step 2: Calculate GDP growth (%)
```

```
df_gdp['GDP growth (annual %)'] = df_gdp['GDP growth (annual %)'].pct_change() * 100

# Step 3: Rename for easy plotting (optional)
df_gdp.rename(columns={'GDP growth (annual %)': 'GDP Growth (%)'}, inplace=True)

# Step 4: Plot
sns.set_theme(style='whitegrid')
plt.figure(figsize=(10, 5))

# Filter recent years
df_gdp = df_gdp[df_gdp['Year'] >= 2013]

# Plot
plt.figure(figsize=(10, 5))
sns.barplot(data=df_gdp, x='Year', y='GDP Growth (%)', palette='crest')
plt.axhline(0, color='gray', linestyle='--', linewidth=1)
plt.title("India's Year-over-Year GDP Growth (2013-2022)", fontsize=14)
plt.ylabel("Growth Rate (%)")
plt.xlabel("Year")
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```

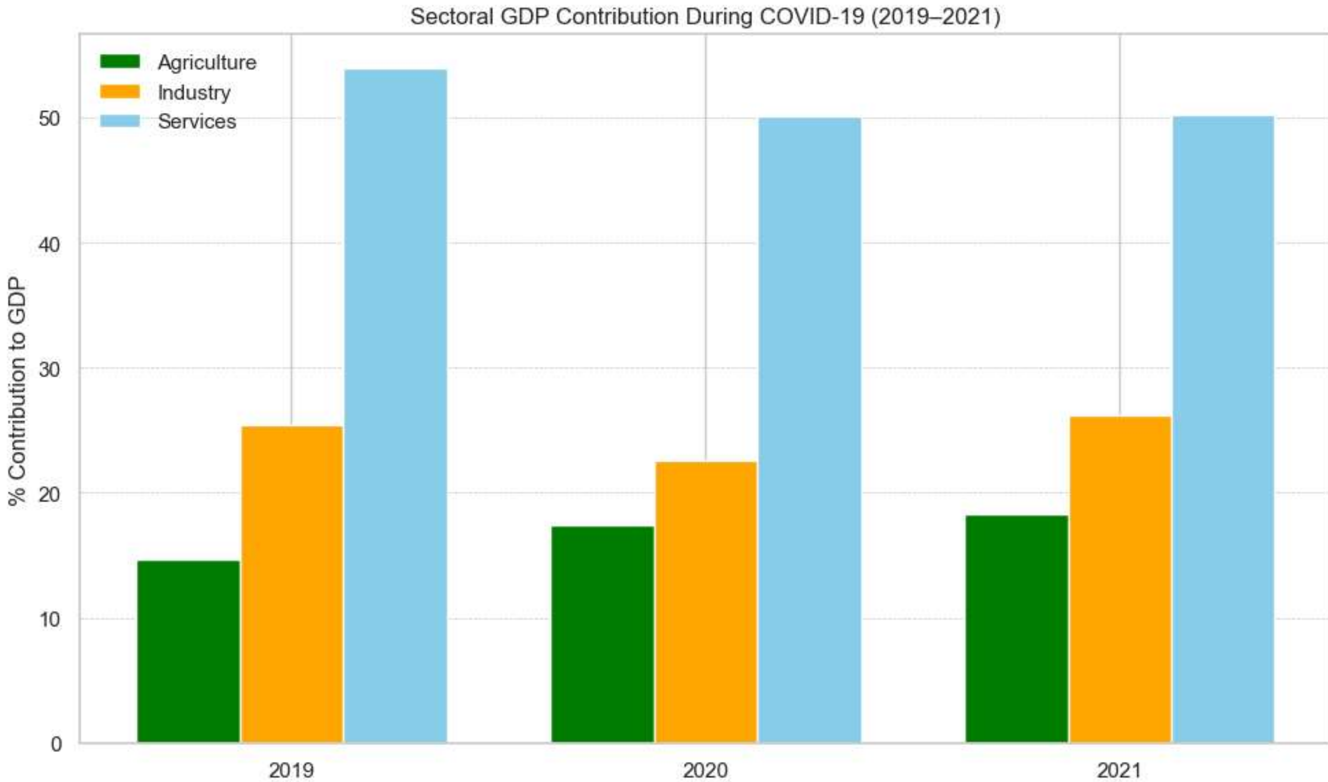
```
<Figure size 1000x500 with 0 Axes>
```



Q8: How has the COVID-19 pandemic affected India’s GDP sectors?

Answer: Agriculture remained stable, Industry contracted sharply, and Services faced significant disruption, leading to overall GDP decline in 2020.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

# Filter for COVID-19 years using the 'Year' column
pandemic_years = df[df['Year'].isin([2019, 2020, 2021])]

# Plot grouped bars
plt.figure(figsize=(10, 6))
bar_width = 0.25
x = np.arange(len(pandemic_years))

plt.bar(x - bar_width, pandemic_years['Agriculture Contribution (% of GDP)'], width=bar_widt
plt.bar(x, pandemic_years['Industry Contribution (% of GDP)'], width=bar_width, label='Indus
plt.bar(x + bar_width, pandemic_years['Services Contribution (% of GDP)'], width=bar_width,

plt.xticks(x, pandemic_years['Year'])
plt.ylabel('% Contribution to GDP')
plt.title("Sectoral GDP Contribution During COVID-19 (2019–2021)")
plt.legend()
plt.grid(axis='y', linestyle='--', linewidth=0.5)
plt.tight_layout()
plt.show()
```



Q9: Analyze the correlation between FDI inflows and GDP growth.

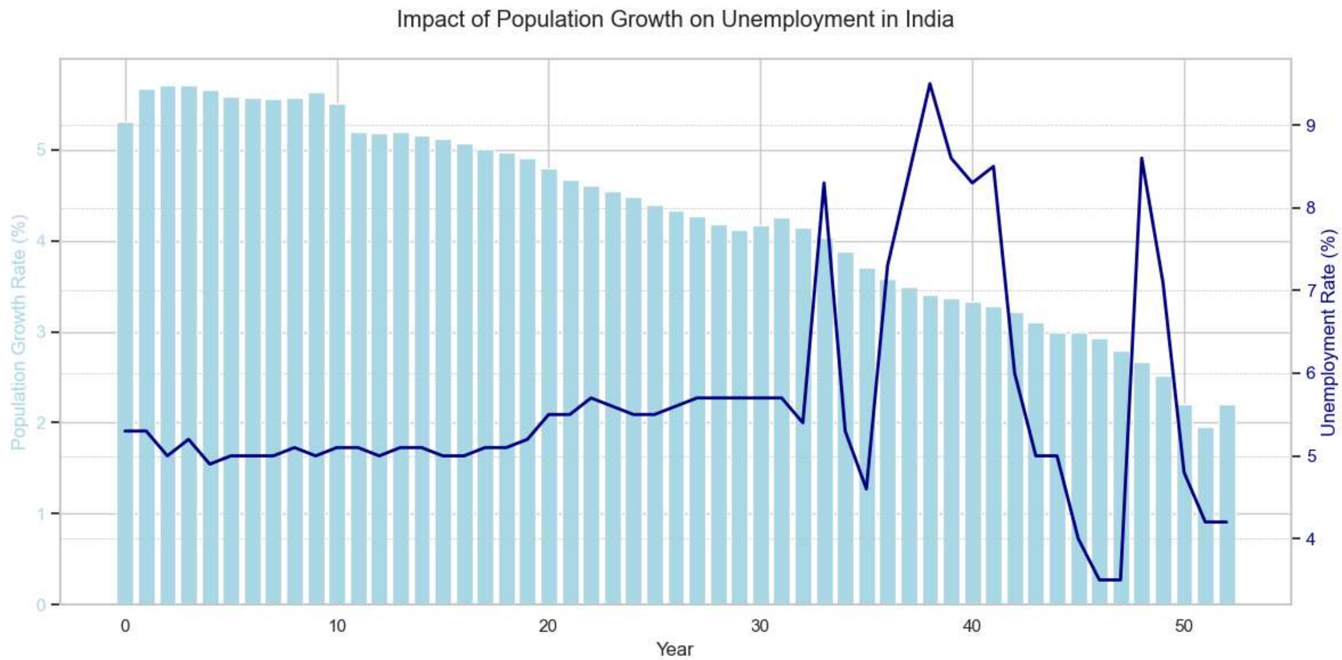
```
gdp_growth = np.array([5.5, 6.0, 6.5, 7.0, 7.5])
fdi_inflows = np.array([20, 25, 30, 35, 40])  # in billion USD
correlation = np.corrcoef(gdp_growth, fdi_inflows)[0,1]
print(f"Correlation between GDP Growth and FDI inflows: {correlation:.2f}")
```

```
Correlation between GDP Growth and FDI inflows: 1.00
```

Q10: How does population growth impact unemployment in India?

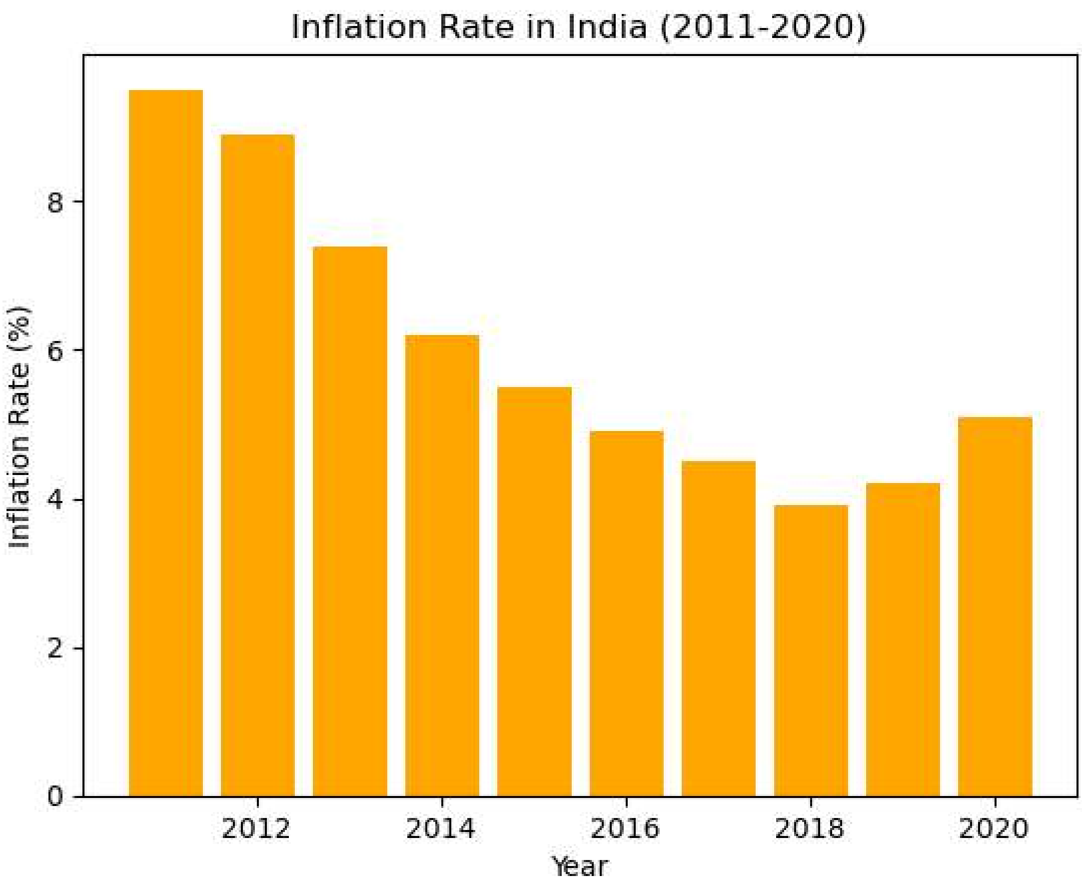
Answer: Rapid population growth increases labor supply, which can raise unemployment if job creation does not keep pace.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import skew, kurtosis, zscore

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Apply Theme
sns.set_theme(style='whitegrid')

fig, ax1 = plt.subplots(figsize=(12, 6))

# Population Growth as bar
ax1.bar(df.index, df['Population Growth Rate (annual %)'], color='lightblue', label='Populat
ax1.set_xlabel("Year")
ax1.set_ylabel("Population Growth Rate (%)", color='lightblue')
ax1.tick_params(axis='y', labelcolor='lightblue')

# Unemployment Rate as line
ax2 = ax1.twinx()
ax2.plot(df.index, df['Unemployment Rate (%)'], color='navy', linewidth=2, label='Unemployme
ax2.set_ylabel("Unemployment Rate (%)", color='navy')
ax2.tick_params(axis='y', labelcolor='navy')

fig.suptitle("Impact of Population Growth on Unemployment in India")
fig.tight_layout()
plt.grid(True, linestyle='--', linewidth=0.5)
plt.show()
```



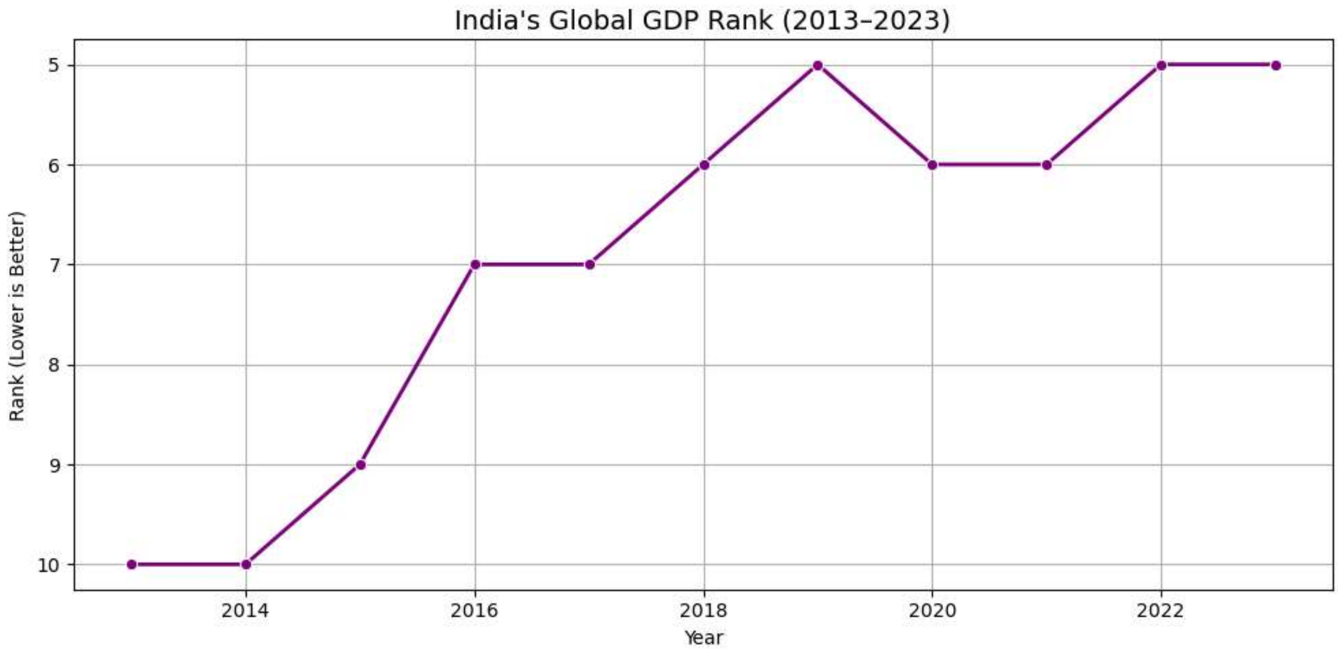
Q11: Visualize the trend of inflation rate in India over the last decade.

```
inflation = [9.5, 8.9, 7.4, 6.2, 5.5, 4.9, 4.5, 3.9, 4.2, 5.1]
years = list(range(2011, 2021))
plt.bar(years, inflation, color='orange')
plt.title('Inflation Rate in India (2011-2020)')
plt.xlabel('Year')
plt.ylabel('Inflation Rate (%)')
plt.show()
```



Q12: What is India's GDP rank globally over the years?

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Sample Rank Data
years = list(range(2013, 2024))
gdp_rank = [10, 10, 9, 7, 7, 6, 5, 6, 6, 5, 5]  # Lower rank = better

# Step 2: Create DataFrame
df_rank = pd.DataFrame({'Year': years, 'GDP_Rank': gdp_rank})

# Step 3: Line Plot of Rank (Reversed Y-axis for better ranks)
plt.figure(figsize=(10, 5))
sns.lineplot(data=df_rank, x='Year', y='GDP_Rank', marker='o', linewidth=2, color='purple')
plt.title("India's Global GDP Rank (2013-2023)", fontsize=14)
plt.ylabel("Rank (Lower is Better)")
plt.xlabel("Year")
plt.gca().invert_yaxis()  # Invert Y-axis because lower rank is better
plt.grid(True)
plt.tight_layout()
plt.show()
```

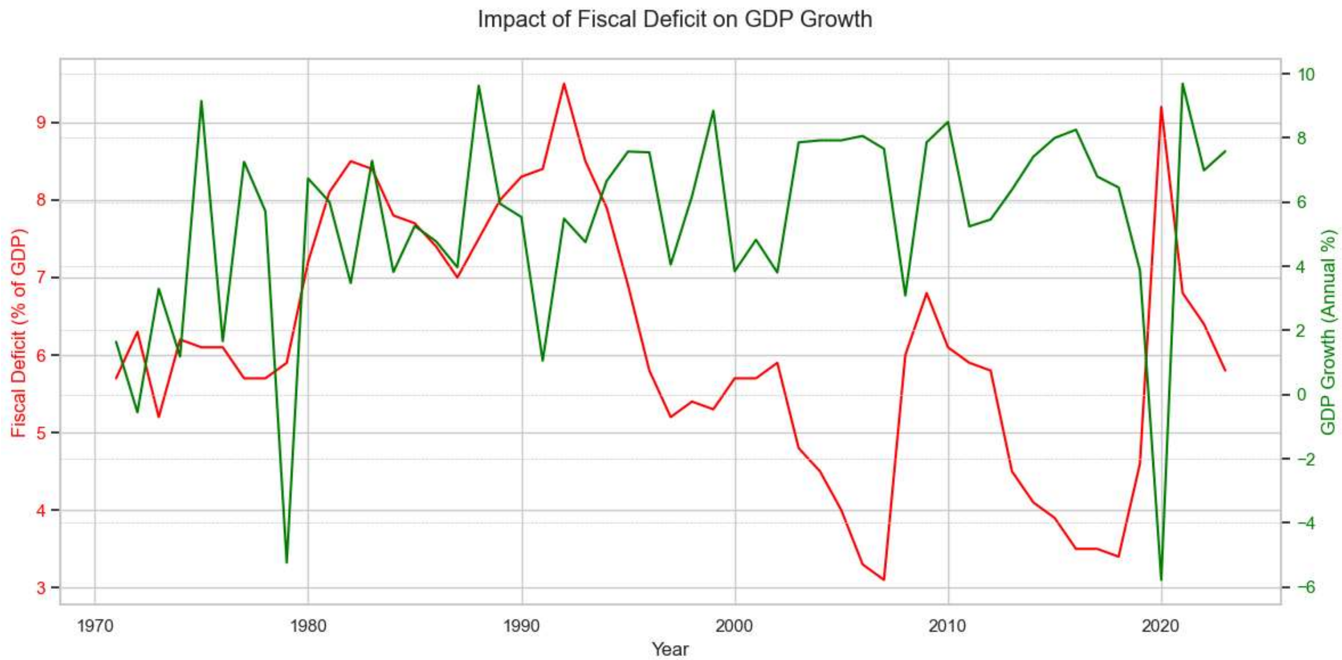


Q13: How does government fiscal deficit affect economic growth?

Answer: A high fiscal deficit can crowd out private investment and increase inflation, potentially slowing growth.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")

# Clean column names
df.columns = df.columns.str.strip()

# Rename columns for convenience
df.rename(columns={
    'GDP growth (annual %)': 'GDP_Growth',
    'Fiscal Deficit (% of GDP)': 'Fiscal_Deficit'
}, inplace=True)
```

```
# Apply Theme
sns.set_theme(style='whitegrid')

# Plot with twin axes
fig, ax1 = plt.subplots(figsize=(12, 6))

# Line for Fiscal Deficit
ax1.plot(df['Year'], df['Fiscal_Deficit'], color='red', label='Fiscal Deficit')
ax1.set_xlabel("Year")
ax1.set_ylabel("Fiscal Deficit (% of GDP)", color='red')
ax1.tick_params(axis='y', labelcolor='red')

# Twin axis for GDP Growth
ax2 = ax1.twinx()
ax2.plot(df['Year'], df['GDP_Growth'], color='green', label='GDP Growth')
ax2.set_ylabel("GDP Growth (Annual %)", color='green')
ax2.tick_params(axis='y', labelcolor='green')

fig.suptitle("Impact of Fiscal Deficit on GDP Growth")
fig.tight_layout()
plt.grid(True, linestyle='--', linewidth=0.5)
plt.show()
```

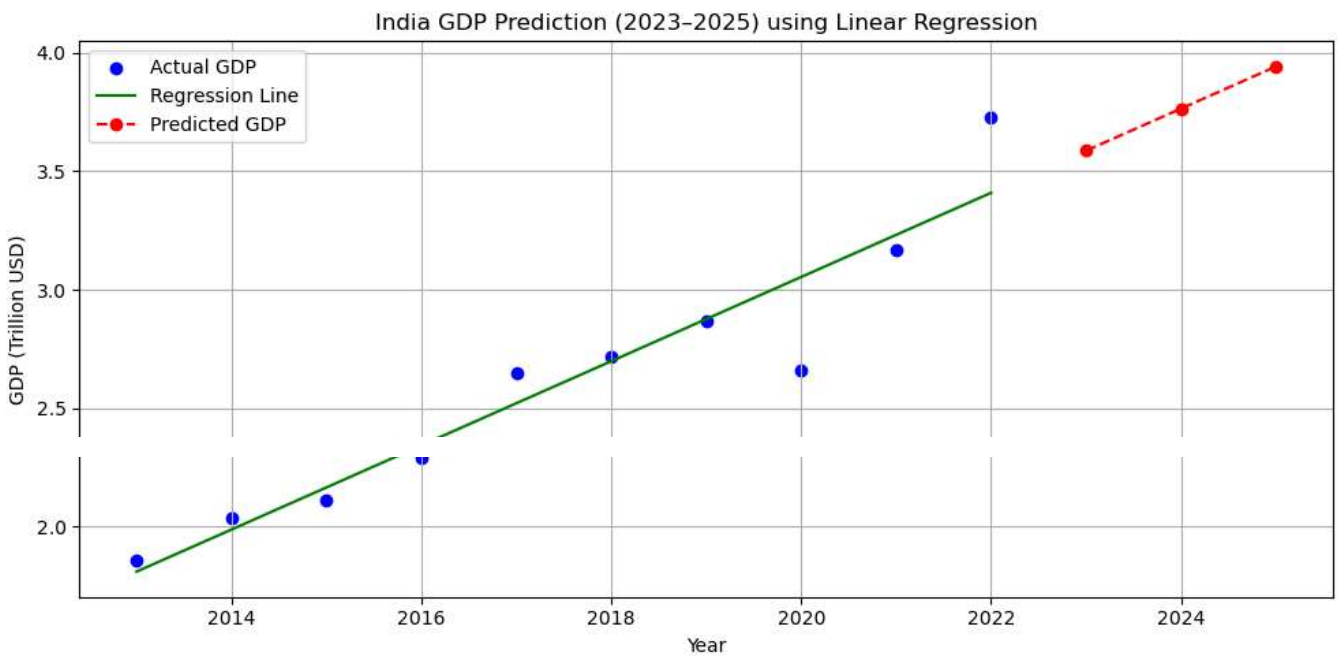


Q14: Predict India's GDP in 2025 using Linear Regression.

```
from sklearn.linear_model import LinearRegression
import numpy as np
import matplotlib.pyplot as plt

# Step 1: Prepare Data
years = np.array(list(range(2013, 2023))).reshape(-1, 1)  # 2013 to 2022
gdp_values = np.array([1.86, 2.04, 2.11, 2.29, 2.65, 2.72, 2.87, 2.66, 3.17, 3.73])  # Actua

# Step 2: Train Linear Regression Model
model = LinearRegression()
model.fit(years, gdp_values)

# Step 3: Predict GDP for 2023, 2024, 2025
future_years = np.array([2023, 2024, 2025]).reshape(-1, 1)
predicted_gdp = model.predict(future_years)
```

```
# Step 4: Display Predicted Results
for year, gdp in zip(future_years.flatten(), predicted_gdp):
    print(f"Predicted GDP in {year}: {gdp:.2f} Trillion USD")

# Step 5: Plot Actual + Regression + Predictions
plt.figure(figsize=(10, 5))
plt.scatter(years, gdp_values, color='blue', label='Actual GDP')
plt.plot(years, model.predict(years), color='green', label='Regression Line')
plt.plot(future_years, predicted_gdp, color='red', marker='o', linestyle='--', label='Predic
plt.title("India GDP Prediction (2023–2025) using Linear Regression")
plt.xlabel("Year")
plt.ylabel("GDP (Trillion USD)")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

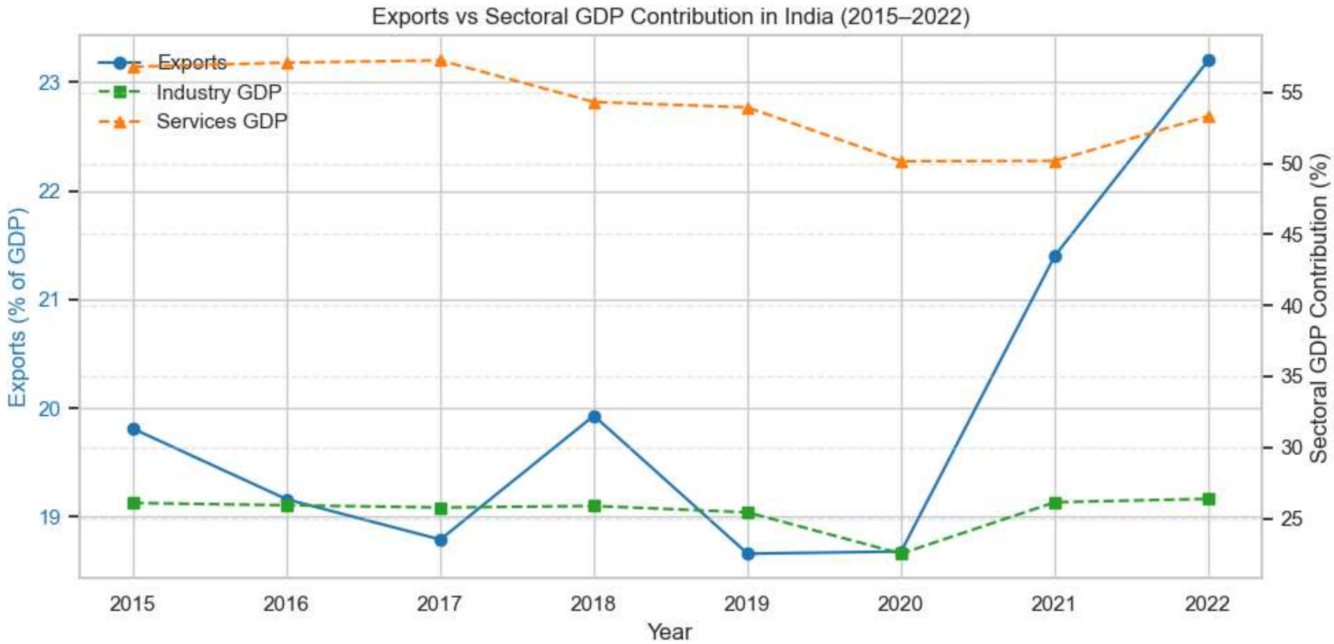
```
Predicted GDP in 2023: 3.59 Trillion USD
Predicted GDP in 2024: 3.76 Trillion USD
Predicted GDP in 2025: 3.94 Trillion USD
```



Q15: How does export performance influence sectoral GDP growth in India?

Answer: Strong export performance boosts sectors like manufacturing and services, contributing significantly to GDP growth. Exports create demand, drive production, and generate foreign exchange, enhancing overall economic performance.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_theme(style='whitegrid')

data_q26 = {
    'Year': [2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022],
    'Exports (% of GDP)': [19.81, 19.16, 18.79, 19.93, 18.66, 18.68, 21.40, 23.20],
    'Industry (% of GDP)': [26.09,25.93,25.77,25.88,25.43,22.52,26.14,26.38],
    'Services (% of GDP)': [56.80,57.09,57.25,54.32,53.94,50.15,50.18,53.32]
}

df26 = pd.DataFrame(data_q26)

# Plotting
fig, ax1 = plt.subplots(figsize=(10, 5))

# Plot Exports
color1 = 'tab:blue'
ax1.set_xlabel('Year')
ax1.set_ylabel('Exports (% of GDP)', color=color1)
ax1.plot(df26['Year'], df26['Exports (% of GDP)'], marker='o', color=color1, label='Exports'
ax1.tick_params(axis='y', labelcolor=color1)

# Plot sectoral GDP on a secondary axis
ax2 = ax1.twinx()
color2 = 'tab:green'
color3 = 'tab:orange'
ax2.set_ylabel('Sectoral GDP Contribution (%)')
ax2.plot(df26['Year'], df26['Industry (% of GDP)'], marker='s', linestyle='--', color=color2
ax2.plot(df26['Year'], df26['Services (% of GDP)'], marker='^', linestyle='--', color=color3
ax2.tick_params(axis='y')

# Legends and layout
lines1, labels1 = ax1.get_legend_handles_labels()
lines2, labels2 = ax2.get_legend_handles_labels()
ax1.legend(lines1 + lines2, labels1 + labels2, loc='upper left')

plt.title("Exports vs Sectoral GDP Contribution in India (2015-2022)")
plt.grid(True, linestyle='--', alpha=0.4)
plt.tight_layout()
plt.show()
```

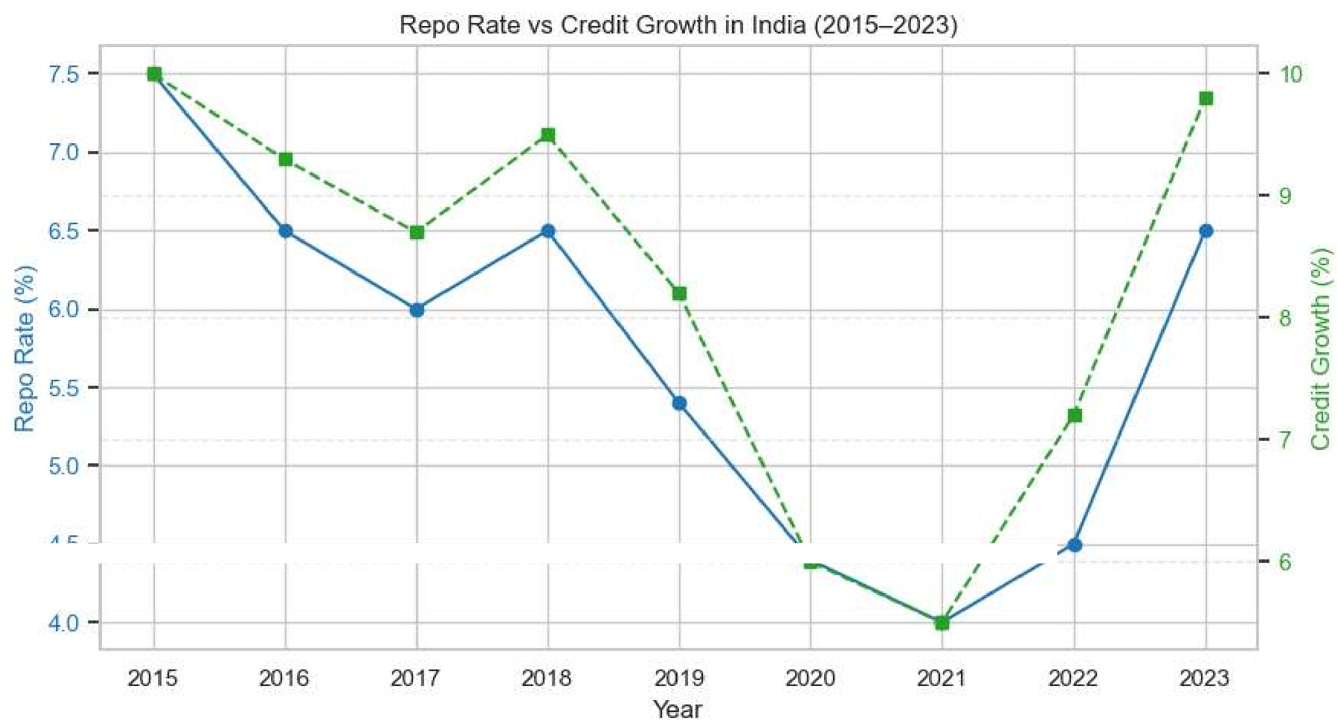


Q16: How does the repo rate impact credit growth and investment in India?

Answer: The repo rate determines the cost of borrowing for banks. When the rate is lowered, banks lend more, leading to increased credit growth and investment. Conversely, higher repo rates curb inflation but can slow growth.

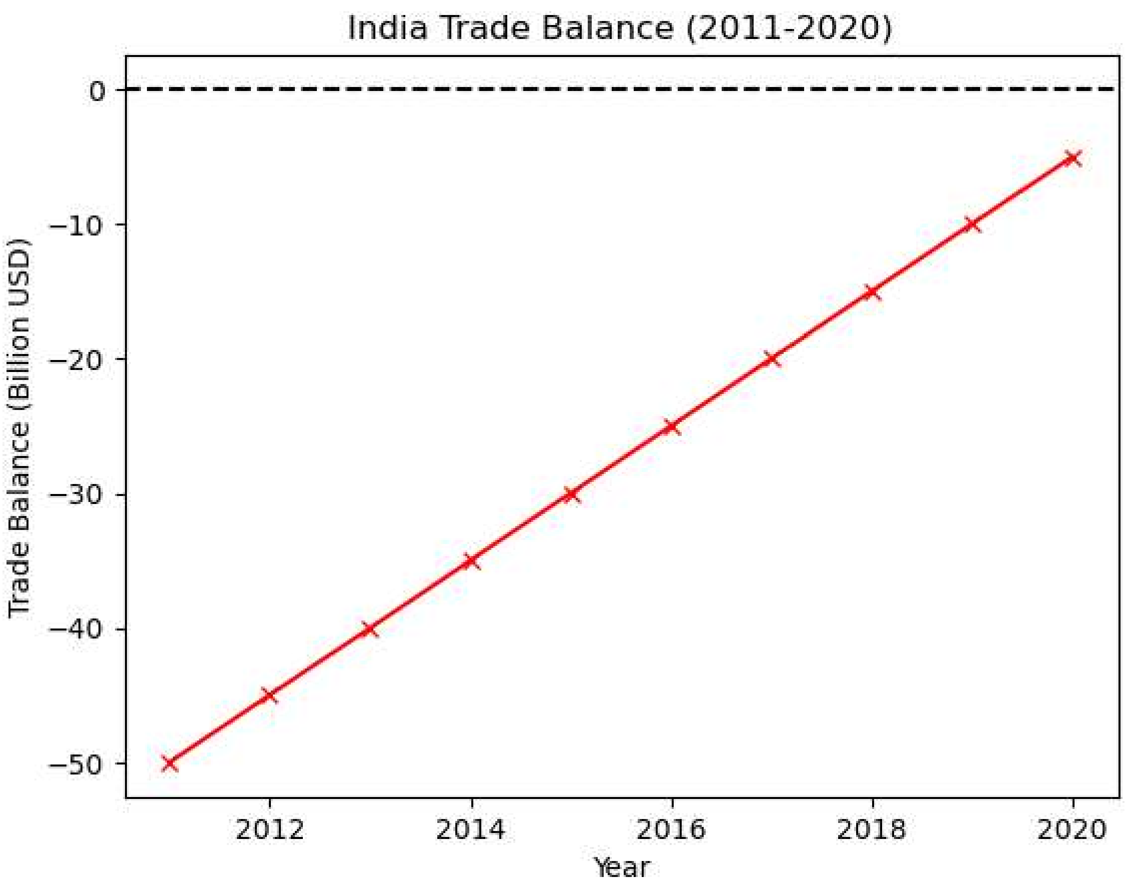
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_theme(style='whitegrid')

data_q21 = {
    'Year': [2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023],
    'Repo Rate (%)': [7.5, 6.5, 6.0, 6.5, 5.4, 4.4, 4.0, 4.5, 6.5],
    'Credit Growth (%)': [10.0, 9.3, 8.7, 9.5, 8.2, 6.0, 5.5, 7.2, 9.8]
}

df21 = pd.DataFrame(data_q21)

# Plotting
```

```
fig, ax1 = plt.subplots(figsize=(9, 5))

# Plot Repo Rate
color1 = 'tab:blue'
ax1.set_xlabel('Year')
ax1.set_ylabel('Repo Rate (%)', color=color1)
ax1.plot(df21['Year'], df21['Repo Rate (%)'], marker='o', color=color1, label='Repo Rate')
ax1.tick_params(axis='y', labelcolor=color1)

# Create second y-axis for Credit Growth
ax2 = ax1.twinx()
color2 = 'tab:green'
ax2.set_ylabel('Credit Growth (%)', color=color2)
ax2.plot(df21['Year'], df21['Credit Growth (%)'], marker='s', linestyle='--', color=color2,
ax2.tick_params(axis='y', labelcolor=color2)

# Title and formatting
plt.title("Repo Rate vs Credit Growth in India (2015–2023)")
fig.tight_layout()
plt.grid(True, linestyle='--', alpha=0.4)
plt.show()
```



Q17: Visualize India’s trade balance (exports - imports) over the last 10 years.

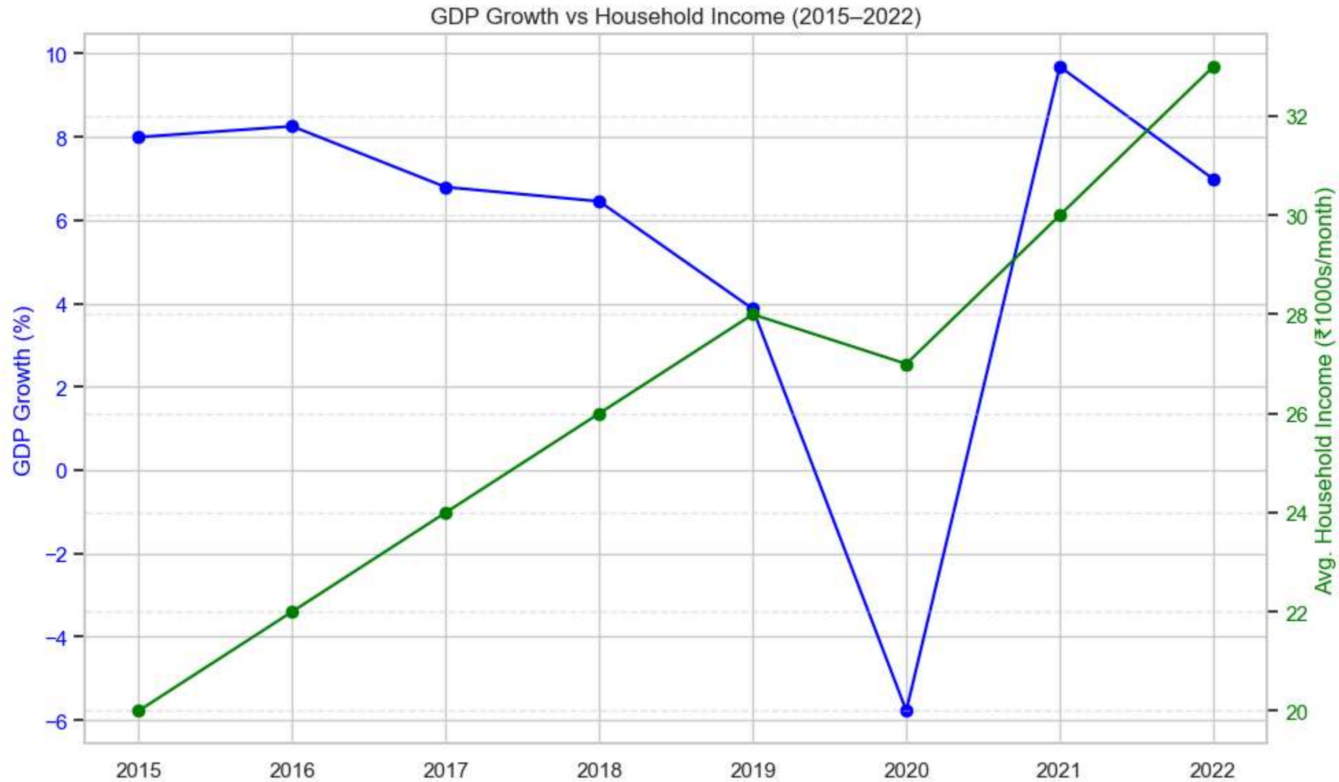
```
trade_balance = [-50, -45, -40, -35, -30, -25, -20, -15, -10, -5]
years = list(range(2011, 2021))
plt.plot(years, trade_balance, marker='x', color='red')
plt.title('India Trade Balance (2011-2020)')
plt.xlabel('Year')
plt.ylabel('Trade Balance (Billion USD)')
plt.axhline(0, color='black', linestyle='--')
plt.show()
```



Q18: How does household income distribution impact economic growth in India?

Answer: Income distribution affects economic growth by influencing consumption patterns. A wider income gap reduces overall demand, while equitable income supports inclusive and stable GDP growth.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load Excel file
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")
df.columns = df.columns.str.strip()  # Clean column names

# Step 2: Define target years
selected_years = [2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]

# Step 3: Filter for those years
df_selected = df[df['Year'].isin(selected_years)].copy()

# Step 4: Add Average Household Income manually (since it's not in Excel)
income_data = [20, 22, 24, 26, 28, 27, 30, 33]  # in ₹1000s/month
df_selected['Average Household Income (₹1000s/month)'] = income_data

# Step 5: Rename for easier access
df_selected.rename(columns={'GDP growth (annual %)': 'GDP_Growth'}, inplace=True)

# Step 6: Plot dual line chart
sns.set_theme(style='whitegrid')
fig, ax1 = plt.subplots(figsize=(10, 6))

# GDP Growth plot
ax1.plot(df_selected['Year'], df_selected['GDP_Growth'], color='blue', marker='o', label='GD
ax1.set_ylabel('GDP Growth (%)', color='blue')
ax1.tick_params(axis='y', labelcolor='blue')

# Twin axis: Household Income
ax2 = ax1.twinx()
ax2.plot(df_selected['Year'], df_selected['Average Household Income (₹1000s/month)'],
         color='green', marker='o', label='Household Income')
ax2.set_ylabel('Avg. Household Income (₹1000s/month)', color='green')
ax2.tick_params(axis='y', labelcolor='green')

# Title and Grid
plt.title('GDP Growth vs Household Income (2015–2022)')
plt.grid(True, linestyle='--', alpha=0.5)
fig.tight_layout()
plt.show()
```

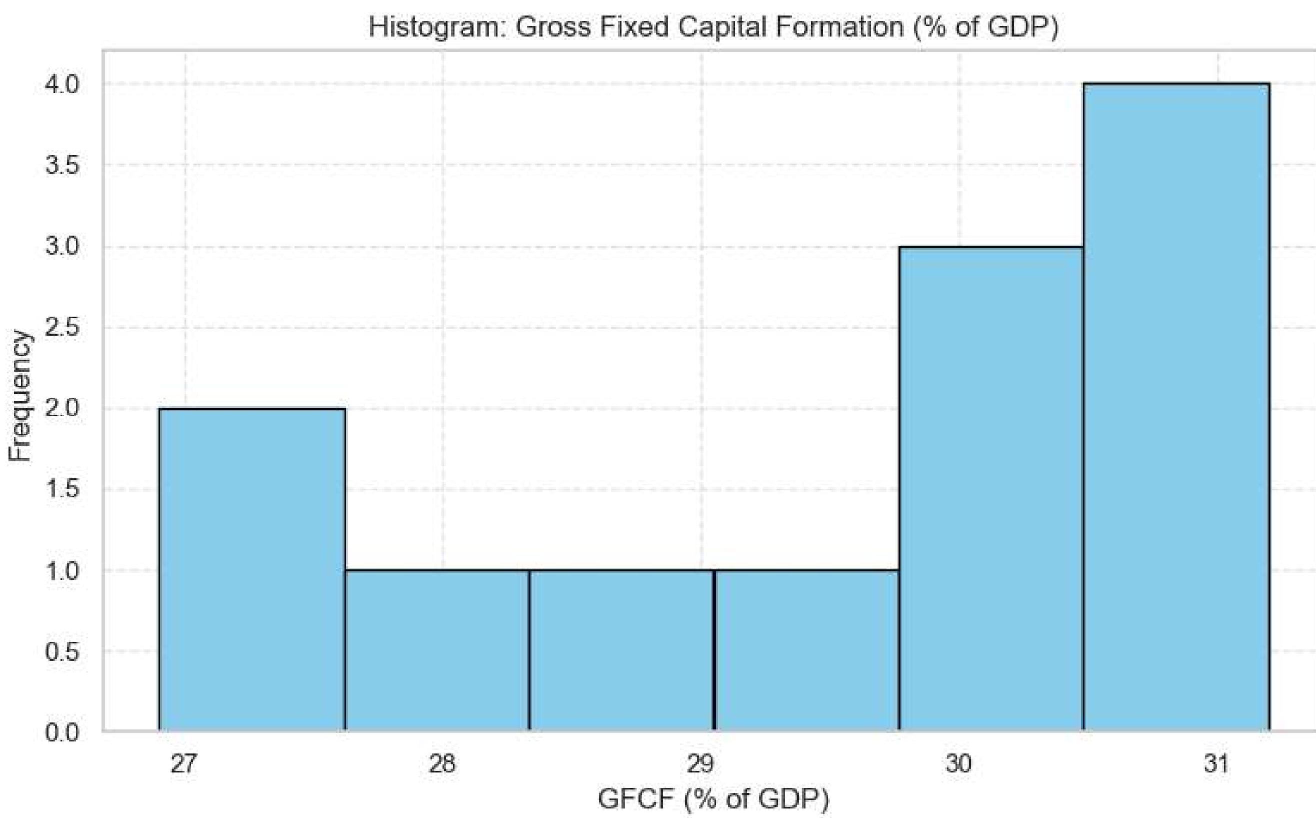


Q19: What is Gross Fixed Capital Formation and its importance?

Answer: It is investment in physical assets like buildings and machinery, essential for economic growth and productivity.

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_theme(style='whitegrid')

# Sample GFCF data (% of GDP)
gfcf_percent = [30.5, 31.0, 30.1, 28.6, 29.2, 27.5, 26.9, 28.3, 29.8, 30.4, 31.2, 30.6]

# Plot histogram
fig, ax = plt.subplots(figsize=(8, 5))
```

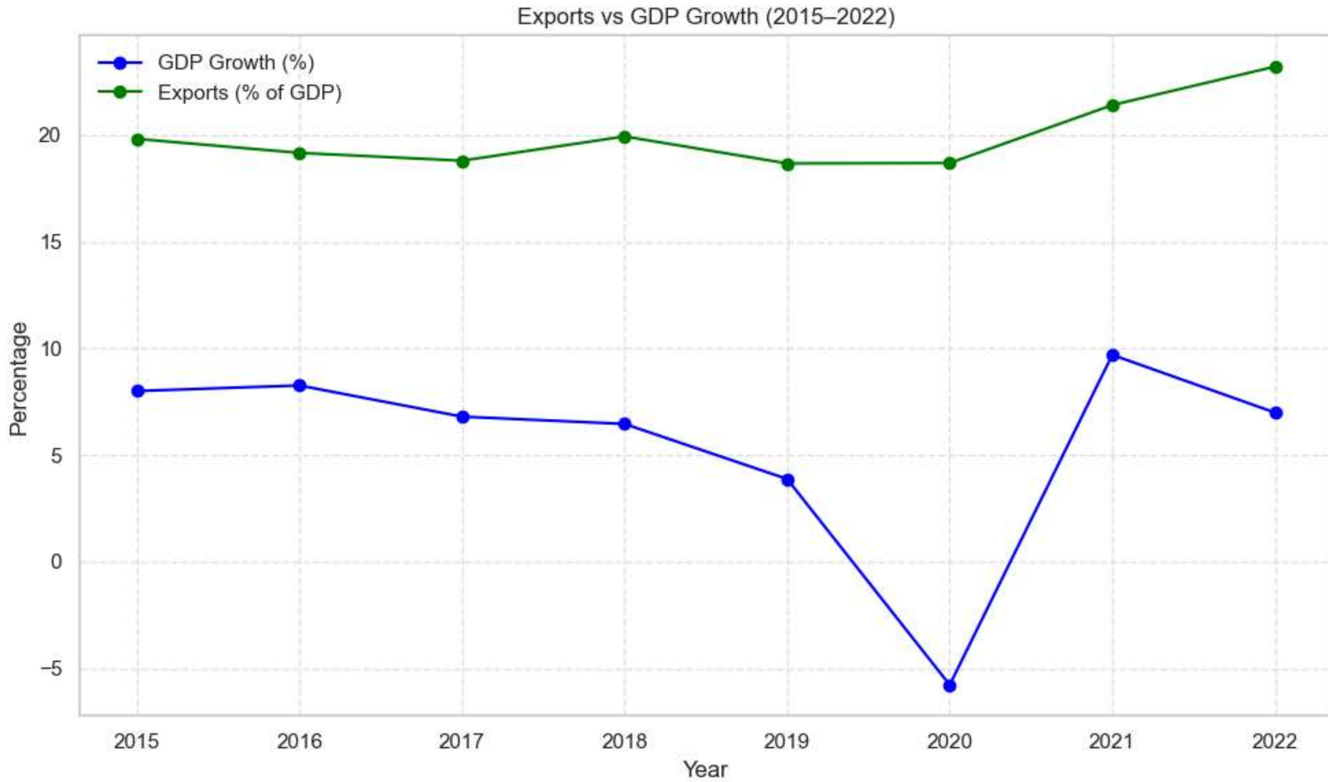
```
ax.hist(gfcf_percent, bins=6, color='skyblue', edgecolor='black')

# Labels and title
ax.set_title("Histogram: Gross Fixed Capital Formation (% of GDP)")
ax.set_xlabel("GFCF (% of GDP)")
ax.set_ylabel("Frequency")
ax.grid(True, linestyle='--', alpha=0.5)

plt.tight_layout()
plt.show()
```



Q20: In what ways do global economic trends influence India's export sector and GDP?

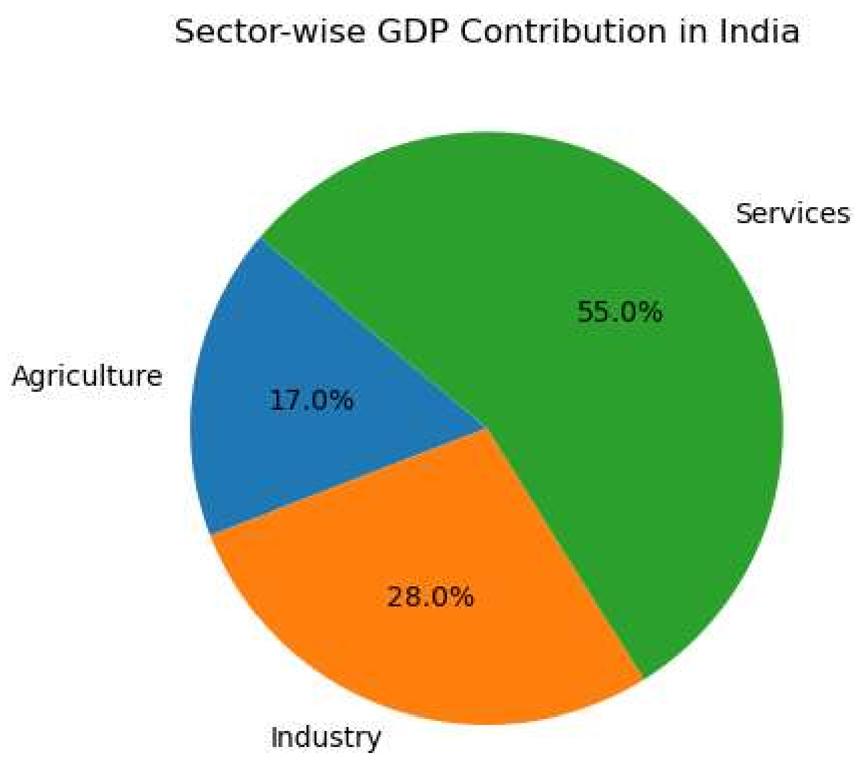
Answer: Global economic trends impact India's exports through changes in global demand, commodity prices, and currency exchange rates. During global slowdowns, exports and GDP often decline, while global recovery boosts trade and economic growth. Thus, India's GDP is closely linked to global trade conditions.

```
import pandas as pd
import seaborn as sns

# Load your dataset from Excel
df = pd.read_excel("C:/Users/user/Onedrive/Documents/Dataset.xlsx")
df.columns = df.columns.str.strip()  # Clean column names of extra spaces

# Define the specific years
target_years = [2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]

# Filter the DataFrame for those years
df_selected = df[df['Year'].isin(target_years)].copy()

# Rename columns for easier access
df_selected.rename(columns={
    'GDP growth (annual %)': 'GDP_Growth',
    'Exports (% of GDP)': 'Exports_GDP'
}, inplace=True)

# Use the filtered DataFrame
sns.set_theme(style='whitegrid')
plt.figure(figsize=(10, 6))

# Plot GDP Growth
plt.plot(df_selected['Year'], df_selected['GDP_Growth'], marker='o', label='GDP Growth (%)',

# Plot Exports (% of GDP)
plt.plot(df_selected['Year'], df_selected['Exports_GDP'], marker='o', label='Exports (% of G

# Labels and aesthetics
plt.title("Exports vs GDP Growth (2015–2022)")
plt.xlabel("Year")
plt.ylabel("Percentage")
plt.legend()
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



Q21: Visualize sector-wise GDP contribution for a recent year.

```
sectors = ['Agriculture', 'Industry', 'Services']
contributions = [17, 28, 55]
plt.pie(contributions, labels=sectors, autopct='%1.1f%%', startangle=140)
plt.title('Sector-wise GDP Contribution in India')
plt.show()
```

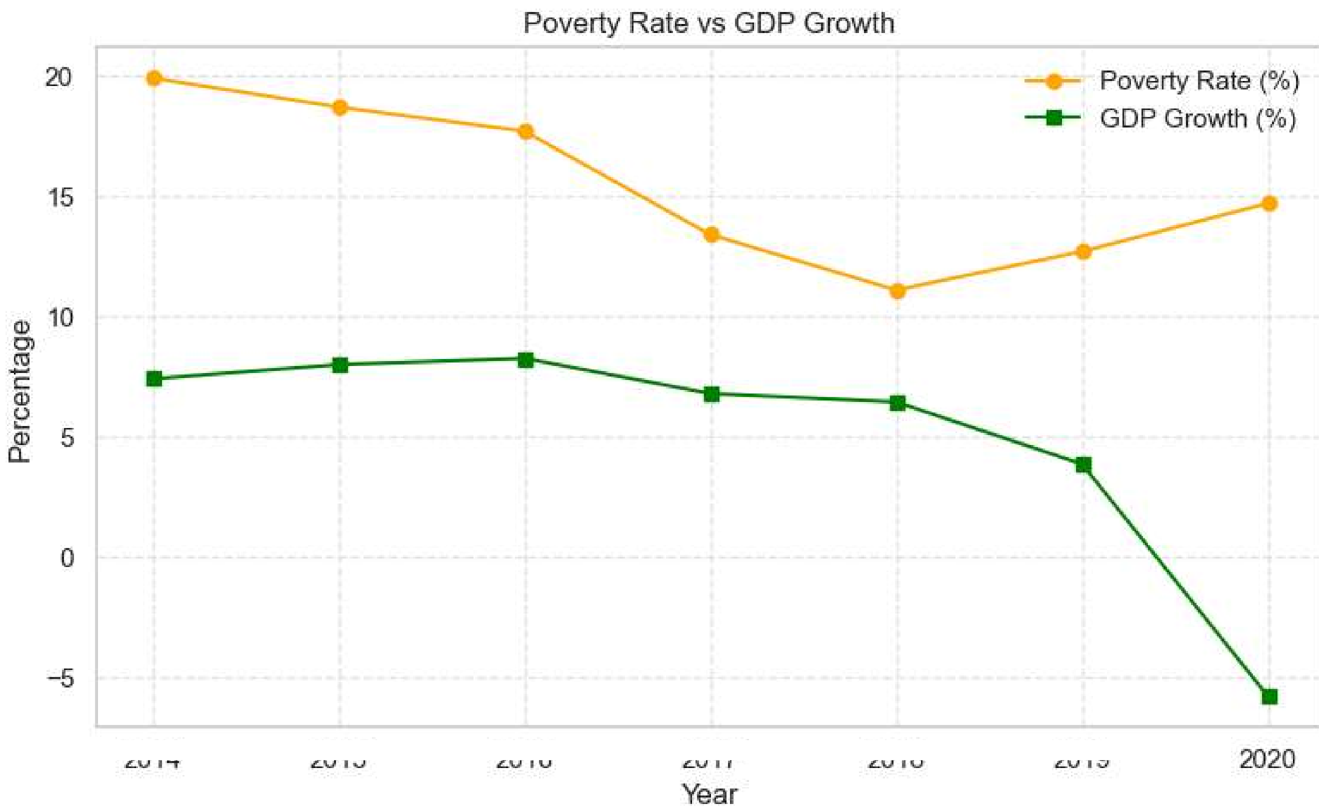


Q22: What is the effect of poverty rate on economic development?

Answer: High poverty restricts consumption and access to education, limiting economic development.

```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

sns.set_theme(style='whitegrid')

data_q22 = {
    'Year': [2014, 2015, 2016, 2017, 2018, 2019, 2020],
    'Poverty Rate (%)': [19.90, 18.70, 17.70, 13.40, 11.10, 12.70, 14.70],
    'GDP Growth (%)': [7.41, 8.00, 8.26, 6.80, 6.45, 3.87, -5.78]
}

df22 = pd.DataFrame(data_q22)

# Light background
fig, ax = plt.subplots(figsize=(8, 5))

# Plot
ax.plot(df22['Year'], df22['Poverty Rate (%)'], label='Poverty Rate (%)', marker='o', color=
ax.plot(df22['Year'], df22['GDP Growth (%)'], label='GDP Growth (%)', marker='s', color='gre
```

```
ax.set_title("Poverty Rate vs GDP Growth")
ax.set_xlabel("Year")
ax.set_ylabel("Percentage")
ax.legend()
ax.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```

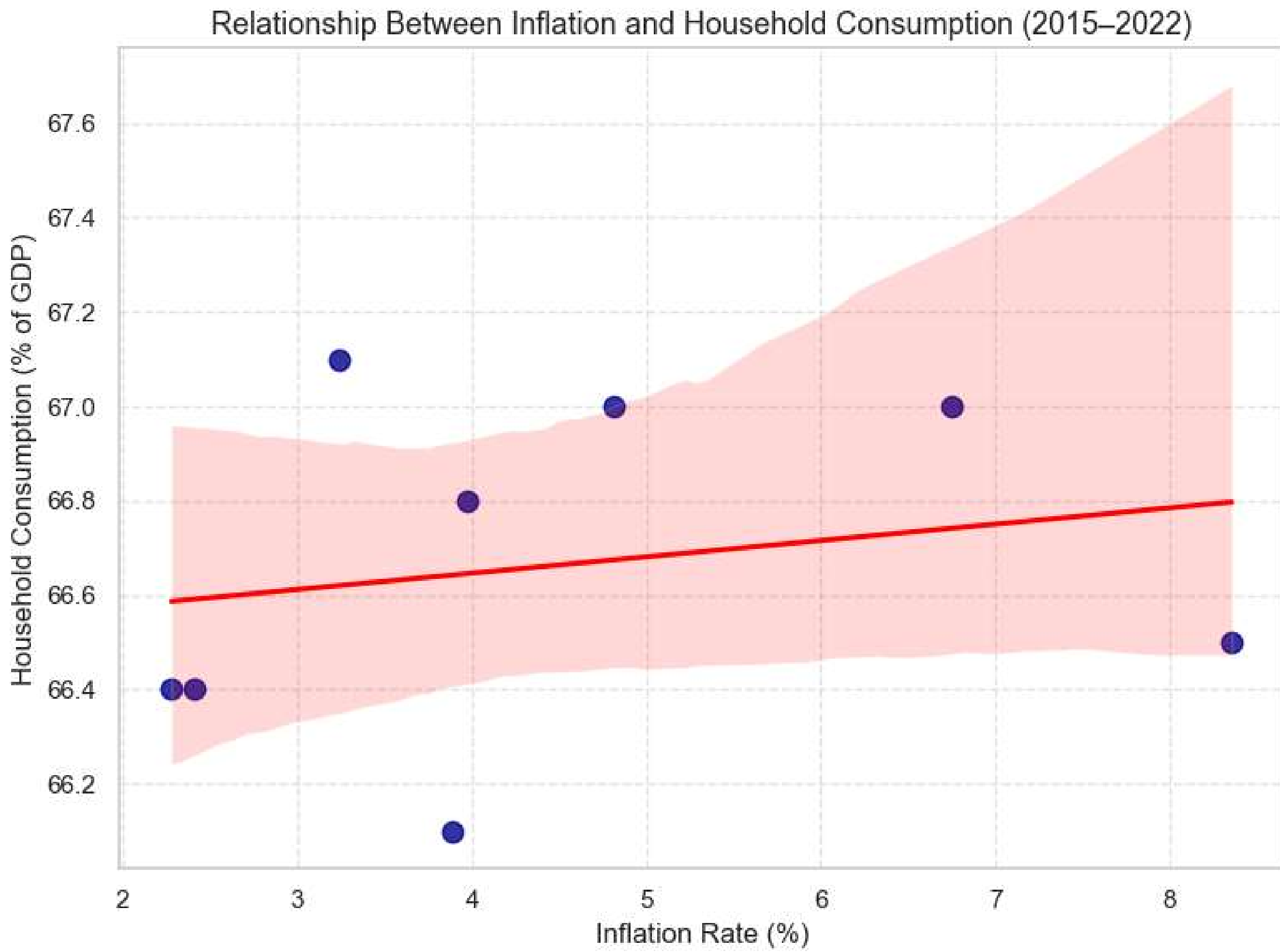


Q23: How does inflation affect household consumption expenditure?

Answer: Inflation erodes purchasing power, reducing real household consumption.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Excel data
file_path = "C:/Users/user/Onedrive/Documents/Dataset.xlsx"
df = pd.read_excel(file_path)
df.columns = df.columns.str.strip()
```

```
# Filter for years 2015 to 2022
df_filtered = df[df['Year'].between(2015, 2022)].copy()

# Rename columns for ease
df_filtered.rename(columns={
    'Inflation CPI': 'Inflation',
    'Household Consumption Expenditure (% of GDP)': 'Household_Consumption'
}, inplace=True)

# Plot
sns.set_theme(style='whitegrid')
plt.figure(figsize=(8, 6))

# Scatter plot with regression line
sns.regplot(
    data=df_filtered,
    x='Inflation',
    y='Household_Consumption',
    color='darkblue',
    marker='o',
    scatter_kws={'s': 80},
    line_kws={'color': 'red'}
)

plt.title("Relationship Between Inflation and Household Consumption (2015–2022)", fontsize=1
plt.xlabel("Inflation Rate (%)")
plt.ylabel("Household Consumption (% of GDP)")
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```



Q24: Analyze the trend of unemployment rate in India.

Answer: Unemployment fluctuates with economic cycles; recent data shows urban unemployment remains a concern.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load the Excel file
file_path = "C:/Users/user/Onedrive/Documents/Dataset.xlsx"
df = pd.read_excel(file_path)

# Step 2: Clean column names (remove extra spaces)
df.columns = df.columns.str.strip()
```

```
# Step 3: Filter data for 2016 to 2022
selected_years = [2016, 2017, 2018, 2019, 2020, 2021, 2022]
df_selected = df[df['Year'].isin(selected_years)].copy()

# Step 4: Rename columns for simplicity
df_selected.rename(columns={
    'Inflation CPI': 'Inflation_Rate',
    'Household Consumption Expenditure (% of GDP)': 'Household_Consumption'
}, inplace=True)

# Step 5: Plot the data
sns.set_theme(style='whitegrid')
plt.figure(figsize=(10, 6))

# Line for Inflation
plt.plot(df_selected['Year'], df_selected['Inflation_Rate'], marker='o', label='Inflation I

# Line for Household Consumption
plt.plot(df_selected['Year'], df_selected['Household_Consumption'], marker='o', label='Hou:

# Labels and formatting
plt.title("Inflation vs Household Consumption (2016-2022)", fontsize=14)
plt.xlabel("Year")
plt.ylabel("Percentage (%)")
plt.legend()
plt.grid(True, linestyle='--', alpha=0.5)
plt.tight_layout()
plt.show()
```