# FDI data set

## July 24, 2024

### 0.0.1 Foreign Direct Investment (FDI) Analysis Project

This project is dedicated to analyzing Foreign Direct Investment (FDI) data to gain a deeper understanding of investment trends across various sectors over multiple years. The dataset encompasses detailed FDI figures spanning from 2000 to 2017, allowing for a comprehensive exploration of how investments have shifted and evolved across different sectors. The primary objective of this analysis is to uncover significant investment patterns, identify key sectors experiencing growth or decline, and understand overall trends in FDI.

To achieve this, the project employs a variety of Python libraries and visualization techniques. **Heat maps** are used to illustrate the intensity of FDI across sectors and years, highlighting areas of high and low investment. **Bar charts** compare investment values across sectors and years, providing a clear view of which sectors have attracted the most investment. **Stacked area charts** reveal how different sectors contribute to the total FDI over time, showcasing sectoral growth and its impact. **Line charts** track the trend of FDI in specific sectors over the years, helping to visualize long-term changes.

Additionally, the analysis includes **scatter plots** to explore relationships between FDI and time, while **dot charts** and **histograms** offer insights into the distribution and variability of FDI values. Basic statistical measures, such as **mean** and **median**, provide further context to the data, assisting in understanding central tendencies and variations.

A key component of the project is the trend forecasting analysis, which extends predictions from 2017 to 2024. This forward-looking analysis uses historical data to project future investment trends, offering insights into potential growth areas and sectoral shifts.

Overall, this project aims to provide a holistic view of FDI trends, utilizing both advanced visualization techniques and fundamental statistical analyses. It is designed to be accessible to individuals at all levels, from beginners looking to understand basic data analysis concepts to advanced users seeking in-depth insights into investment patterns. Through this analysis, we hope to support strategic decision-making and contribute valuable knowledge to the field of investment analysis.

---

```python
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Read dataset in CSV format
data = pd.read_csv("E:/Data Set/FDI data.csv")
```

```
[6]: print(data.head())
     print(data.info())
     print(data.describe())
```

```
                      Sector  2000-01  2001-02  2002-03  2003-04  2004-05  \
0  METALLURGICAL INDUSTRIES    22.69    14.14    36.61     8.11   200.38
1                    MINING     1.32     6.52    10.06    23.48     9.92
2                     POWER    89.42   757.44    59.11    27.09    43.37
3  NON-CONVENTIONAL ENERGY     0.00     0.00     1.70     4.14     1.27
4           COAL PRODUCTION     0.00     0.00     0.00     0.04     0.00

   2005-06  2006-07  2007-08  2008-09  2009-10  2010-11  2011-12  2012-13  \
0   149.13   169.94  1175.75   959.94   419.88  1098.14  1786.14  1466.23
1     7.40     6.62   444.36    34.16   174.40    79.51   142.65    57.89
2    72.69   157.15   988.68   907.66  1271.79  1271.77  1652.38   535.68
3     1.35     2.44    58.82   125.88   622.52   214.40   452.17  1106.52
4     9.14     1.30    14.08     0.22     0.00     0.00     0.00     0.00

   2013-14  2014-15  2015-16  2016-17
0   567.63   359.34   456.31  1440.18
1    12.73   684.39   520.67    55.75
2  1066.08   707.04   868.80  1112.98
3   414.25   615.95   776.51   783.57
4     2.96     0.00     0.00     0.00
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 63 entries, 0 to 62
Data columns (total 18 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Sector   63 non-null     object
 1   2000-01  63 non-null     float64
 2   2001-02  63 non-null     float64
 3   2002-03  63 non-null     float64
 4   2003-04  63 non-null     float64
 5   2004-05  63 non-null     float64
 6   2005-06  63 non-null     float64
 7   2006-07  63 non-null     float64
 8   2007-08  63 non-null     float64
 9   2008-09  63 non-null     float64
 10  2009-10  63 non-null     float64
 11  2010-11  63 non-null     float64
 12  2011-12  63 non-null     float64
 13  2012-13  63 non-null     float64
 14  2013-14  63 non-null     float64
 15  2014-15  63 non-null     float64
 16  2015-16  63 non-null     float64
 17  2016-17  63 non-null     float64
dtypes: float64(17), object(1)
```

```
memory usage: 9.0+ KB
None
            2000-01     2001-02     2002-03     2003-04     2004-05   \
count     63.000000   63.000000   63.000000   63.000000   63.000000
mean      37.757302   63.931587   42.925714   34.727778   51.090317
std      112.227860  157.878737   86.606439   67.653735  101.934873
min        0.000000    0.000000    0.000000    0.000000    0.000000
25%        0.000000    0.000000    0.200000    0.215000    0.715000
50%        4.030000    5.070000   11.010000    6.370000    9.090000
75%       23.510000   44.830000   36.555000   38.660000   43.205000
max      832.070000  873.230000  419.960000  368.320000  527.900000

            2005-06     2006-07     2007-08     2008-09      2009-10   \
count     63.000000   63.000000   63.000000   63.000000    63.000000
mean      87.932540  198.281905  390.085714  498.348571   410.069524
std      206.436967  686.783115 1026.249935 1134.649040   926.814626
min        0.000000    0.000000    0.000000    0.000000     0.000000
25%        1.230000    4.160000    9.950000   11.950000     7.880000
50%       22.620000   25.820000   58.820000   84.880000    69.740000
75%       63.855000  108.325000  279.270000  383.320000   341.595000
max     1359.970000 4713.780000 6986.170000 6183.490000  5466.130000

            2010-11     2011-12     2012-13     2013-14      2014-15   \
count     63.000000   63.000000   63.000000   63.000000    63.000000
mean     339.413810  557.472698  355.930000  385.703492   490.959841
std      627.141139 1031.474056  778.091368  658.429944   837.787060
min        0.000000    0.000000    0.000000    0.000000     0.000000
25%        8.430000   22.720000   15.115000   16.610000    33.800000
50%       58.070000  129.360000   95.410000  113.780000   177.220000
75%      304.280000  593.525000  288.025000  473.060000   595.390000
max     3296.090000 5215.980000 4832.980000 3982.890000  4443.260000

            2015-16     2016-17
count     63.000000   63.000000
mean     634.936349  690.131111
std     1335.307706 1411.965354
min        0.000000    0.000000
25%       30.000000   19.905000
50%      159.130000  110.860000
75%      519.070000  741.220000
max     6889.460000 8684.070000
```

```python
# Checking for missing values
print(data.isnull().sum())
```

```
Sector     0
2000-01    0
2001-02    0
```

```
2002-03    0
2003-04    0
2004-05    0
2005-06    0
2006-07    0
2007-08    0
2008-09    0
2009-10    0
2010-11    0
2011-12    0
2012-13    0
2013-14    0
2014-15    0
2015-16    0
2016-17    0
dtype: int64
```

### 0.0.2 Year-wise FDI analysis in India, with a focus on the years 2010 and 2011

```
[8]: # Calculate yearly investment
     yearly_investment = data.iloc[:, 1:].sum(axis=0)

     # Plotting Year-wise Investment with emphasis on 2010 and 2011
     plt.figure(figsize=(12, 6))

     # Normal line plot for all years except 2010 and 2011
     sns.lineplot(x=yearly_investment.index, y=yearly_investment.values, marker='o',␣
       ↪label='Yearly Investment')

     # Highlight 2010 and 2011 with a dotted line
     highlight_years = ['2009-10', '2010-11','2012-13']
     highlight_values = yearly_investment.loc[highlight_years]

     sns.lineplot(x=highlight_years, y=highlight_values, marker='o', linestyle='--',␣
       ↪color='red', label='Low FDI Years')

     # Add text annotation
     for year in highlight_years:
         plt.text(year, yearly_investment[year], f'{year}: {yearly_investment[year]:.
       ↪2f}', horizontalalignment='right')

     # General plot settings
     plt.title('Year-wise FDI in India with Highlights for 2010 and 2011')
     plt.xlabel('Year')
     plt.ylabel('Investment (in million USD)')
     plt.xticks(rotation=90)
     plt.legend()
```
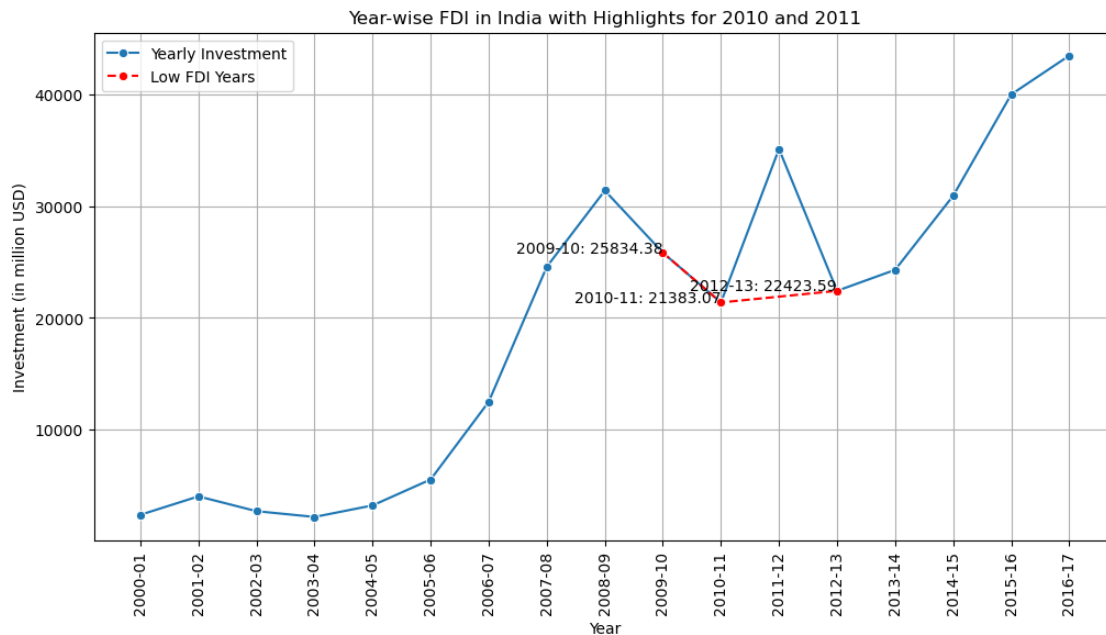
```
plt.grid(True)
plt.show()
```



Year-wise FDI in India with Highlights for 2010 and 2011

Here are few key findings from the year-wise FDI analysis in India, with a focus on the years 2010 and 2011:

1. **Significant Decline in 2010 and 2011**:
   - The years 2010-11 and 2011-12 saw a noticeable dip in FDI compared to surrounding years, highlighting a period of reduced investment.
2. **Recovery Post-2011**:
   - Following the dip in 2010-11 and 2011-12, there was a significant recovery and an upward trend in FDI, indicating renewed investor confidence.
3. **Consistent Sectors Amid Decline**:
   - Despite the overall dip, sectors like 'COMPUTER SOFTWARE & HARDWARE' and 'TELECOMMUNICATIONS' maintained relatively stable FDI, indicating their resilience.

### 0.0.3 Sectors with lowest and highest FDI

```
[9]: # Extract FDI data for 2009-10 and 2010-11
     data_2009_10 = data[['Sector', '2009-10']].sort_values(by='2009-10')
     data_2010_11 = data[['Sector', '2010-11']].sort_values(by='2010-11')

     # Find sectors with lowest and highest FDI for 2009-10
     lowest_2009_10 = data_2009_10.iloc[0]
     highest_2009_10 = data_2009_10.iloc[-1]
```

5

```python
# Find sectors with lowest and highest FDI for 2010-11
lowest_2010_11 = data_2010_11.iloc[0]
highest_2010_11 = data_2010_11.iloc[-1]

# Print out the sectors with their FDI values
print("Sector with the lowest FDI in 2009-10:", lowest_2009_10['Sector'], "with
 ↪FDI:", lowest_2009_10['2009-10'])
print("Sector with the highest FDI in 2009-10:", highest_2009_10['Sector'],
 ↪"with FDI:", highest_2009_10['2009-10'])
print("Sector with the lowest FDI in 2010-11:", lowest_2010_11['Sector'], "with
 ↪FDI:", lowest_2010_11['2010-11'])
print("Sector with the highest FDI in 2010-11:", highest_2010_11['Sector'],
 ↪"with FDI:", highest_2010_11['2010-11'])
```

Sector with the lowest FDI in 2009-10: PHOTOGRAPHIC RAW FILM AND PAPER with FDI:
0.0
Sector with the highest FDI in 2009-10: CONSTRUCTION DEVELOPMENT: Townships,
housing, built-up infrastructure and construction-development projects with FDI:
5466.13
Sector with the lowest FDI in 2010-11: MATHEMATICAL,SURVEYING AND DRAWING
INSTRUMENTS with FDI: 0.0
Sector with the highest FDI in 2010-11: SERVICES SECTOR
(Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,R&D,Courier,Tech. Testing
and Analysis, Other) with FDI: 3296.09

Here are the key points for the sectors with the lowest and highest FDI in 2009-10 and 2010-11:

1. **2009-10 FDI Highlights**:
   - **Lowest FDI**: The sector 'PHOTOGRAPHIC RAW FILM AND PAPER' received the lowest FDI at 0.0 million USD.
   - **Highest FDI**: The 'CONSTRUCTION DEVELOPMENT' sector received the highest FDI at 5466.13 million USD.
2. **2010-11 FDI Highlights**:
   - **Lowest FDI**: The sector 'MATHEMATICAL, SURVEYING AND DRAWING INSTRUMENTS' received the lowest FDI at 0.0 million USD.
   - **Highest FDI**: The 'SERVICES SECTOR' (including finance, banking, insurance, and other sub-sectors) received the highest FDI at 3296.09 million USD.
3. **Sectoral FDI Variability**:
   - The FDI data for both years shows significant variability, with certain sectors receiving no investments while others, like 'CONSTRUCTION DEVELOPMENT' and 'SERVICES SECTOR', attracted substantial foreign capital.

### 0.0.4  Scatter plots for Lowest and Highest FDI Sectors for 2009-10 and 2010-11

```python
[10]: # Extract FDI data for the relevant years
      data_years = ['2009-10', '2010-11']  # Adjust as needed for other years
      sectors = data['Sector']
      fdi_years = data[['Sector'] + data_years]
```

```python
# Initialize a figure and axis for scatter plots
plt.figure(figsize=(12, 8))

# Iterate over each year to plot lowest and highest FDI sectors
for i, year in enumerate(data_years, 1):
    # Find sectors with lowest and highest FDI for the current year
    lowest_sector = fdi_years.sort_values(by=year).iloc[0]['Sector']
    highest_sector = fdi_years.sort_values(by=year).iloc[-1]['Sector']

    # Plot lowest FDI sector
    plt.scatter(fdi_years.loc[fdi_years['Sector'] == lowest_sector, year],↵
 ↪lowest_sector,
                color='red', label=f'Lowest FDI Sector {year}', marker='o',↵
 ↪s=100)

    # Plot highest FDI sector
    plt.scatter(fdi_years.loc[fdi_years['Sector'] == highest_sector, year],↵
 ↪highest_sector,
                color='green', label=f'Highest FDI Sector {year}', marker='o',↵
 ↪s=100)

# Add labels and title
plt.xlabel('FDI (in million USD)')
plt.ylabel('Sector')
plt.title('Lowest and Highest FDI Sectors for 2009-10 and 2010-11')
plt.grid(True)
plt.legend()

plt.tight_layout()
plt.show()
```
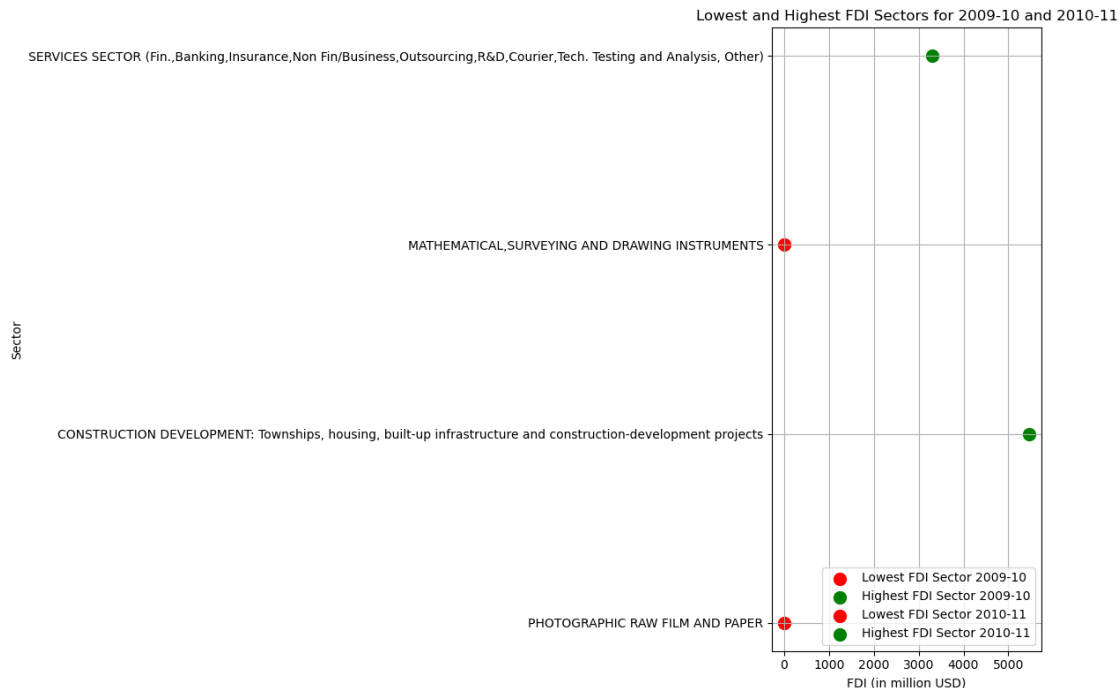
Lowest and Highest FDI Sectors for 2009-10 and 2010-11

**scatter plots for FDI Variability**: - The FDI data for both years shows significant variability, with certain sectors receiving no investments while others, like 'CONSTRUCTION DEVELOP-MENT' and 'SERVICES SECTOR', attracted substantial foreign capital.

### 0.0.5  Sector-wise Investment for top N sectors

```python
[22]:  # Calculate total investment for each sector
       sector_investment = fdi_data.set_index('Sector').sum(axis=1)

       # Sort sectors by investment value
       sector_investment_sorted = sector_investment.sort_values(ascending=False)

       # Select top N sectors to display
       top_n = 10
       top_sector_investment = sector_investment_sorted.head(top_n)

       # Plotting Sector-wise Investment for top N sectors
       plt.figure(figsize=(12, 8))
       sns.barplot(x=top_sector_investment.values, y=top_sector_investment.index,
        →orient='h', palette='viridis')
       plt.title(f'Top {top_n} Sector-wise FDI in India')
       plt.xlabel('Investment (in million USD)')
       plt.ylabel('Sector')
       plt.grid(True)
       plt.show()
```

8

Top 10 Sector-wise FDI in India
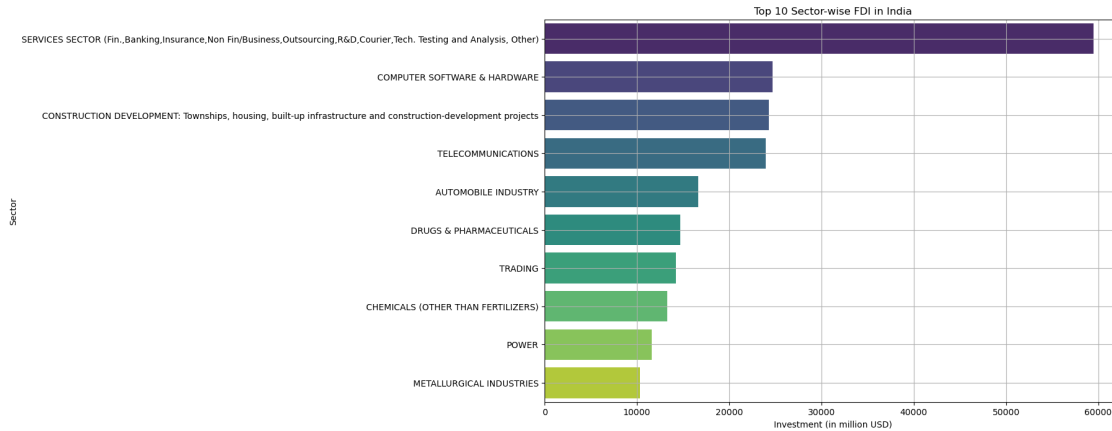
key Analysis for sector-wise FDI investment focusing on the top sectors:

1. **Services Sector Dominance**:
   - The 'SERVICES SECTOR' (including finance, banking, insurance, non-financial business, outsourcing, R&D, courier, tech testing, and analysis) consistently attracted the highest FDI inflows, highlighting its critical role in India's economic landscape.
2. **Tech and Telecom Growth**:
   - 'COMPUTER SOFTWARE & HARDWARE' and 'TELECOMMUNICATIONS' sectors showed substantial and steady growth in FDI. These sectors benefited from India's strong IT infrastructure and growing digital economy.
3. **Construction and Infrastructure Development**:
   - The 'CONSTRUCTION DEVELOPMENT' sector, which includes townships, housing, built-up infrastructure, and development projects, consistently received high FDI, reflecting the ongoing urbanization and infrastructure expansion in India.
4. **Manufacturing and Industrial Investment**:
   - The 'AUTOMOBILE INDUSTRY' and 'METALLURGICAL INDUSTRIES' were other top sectors receiving significant FDI. This trend underscores India's position as a manufacturing hub and its efforts to boost the industrial sector through policies like 'Make in India'.
5. **Emerging Sectors**:
   - Emerging sectors like 'PHARMACEUTICALS' and 'NON-CONVENTIONAL ENERGY' (e.g., renewable energy) also saw increasing FDI inflows, indicating a diversification of investment into sustainable and high-growth potential areas driven by supportive government policies and global trends towards sustainability.

[24]:
```python
mean_investment = data.iloc[:, 1:].mean().mean()
median_investment = data.iloc[:, 1:].median().median()
print(f'Mean Investment: {mean_investment}')
print(f'Median Investment: {median_investment}')
```

```
Mean Investment: 309.98225023342667
Median Investment: 58.82
```

*Comparison Between Mean and Median* - **Discrepancy Between Mean and Median**:

The substantial difference between the mean (309.98 million USD) and the median (58.82 million USD) suggests that the FDI data is skewed, with a few sectors receiving very high investments that raise the average significantly. Most sectors, however, receive much lower FDI, as reflected by the median. - **Implications for Policy and Investment**: Policymakers and investors might infer that while a few sectors are attracting significant foreign investments, many sectors still have relatively low FDI inflows. This could highlight opportunities for targeted policies to encourage more balanced FDI distribution across various sectors.

```
[17]: # Print the column names to verify
      print(data.columns)
```

```
Index(['Sector', '2000-01', '2001-02', '2002-03', '2003-04', '2004-05',
       '2005-06', '2006-07', '2007-08', '2008-09', '2009-10', '2010-11',
       '2011-12', '2012-13', '2013-14', '2014-15', '2015-16', '2016-17'],
      dtype='object')
```
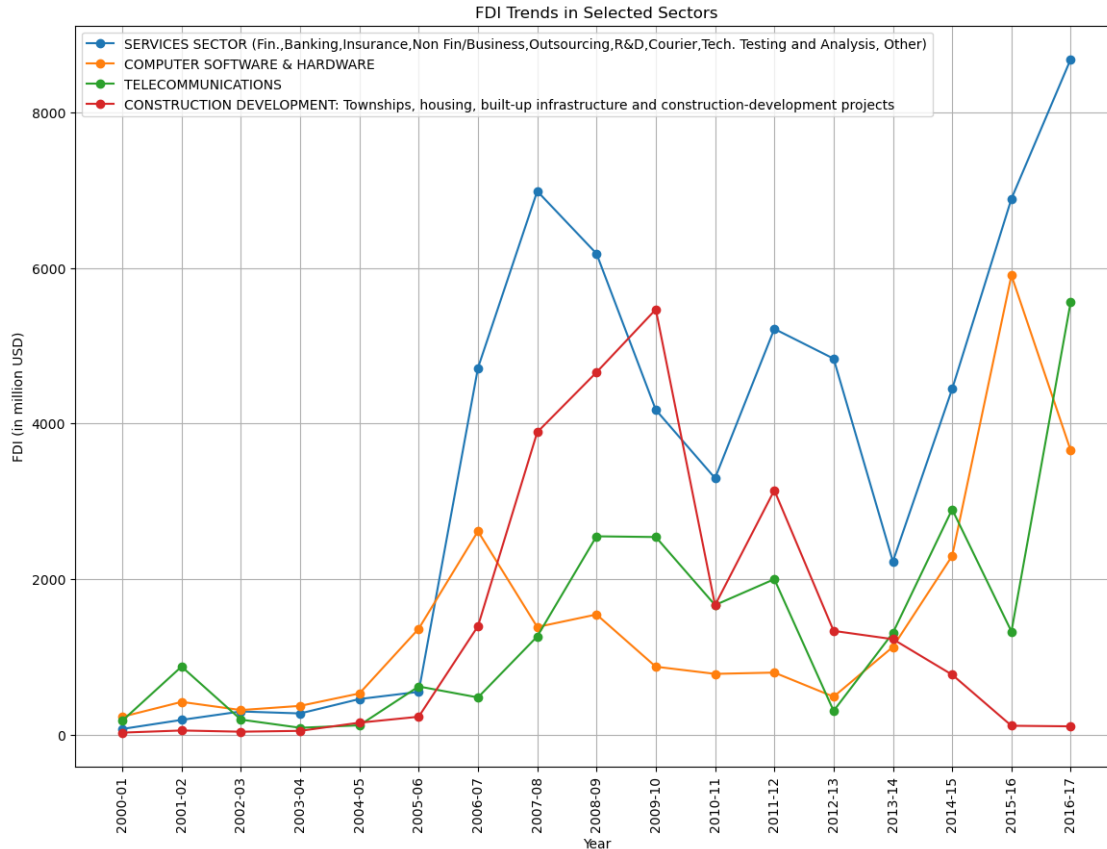
### 0.0.6 FDI Trends in Selected Sectors

```
[23]: import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns

      # Load the dataset
      file_path = "E:/Data Set/FDI data.csv"
      fdi_data = pd.read_csv(file_path)

      # Plotting FDI trends for a few key sectors
      plt.figure(figsize=(14, 10))
      selected_sectors = ['SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
       ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)'
       ,'COMPUTER SOFTWARE & HARDWARE','TELECOMMUNICATIONS','CONSTRUCTION DEVELOPMENT:
       ↪ Townships, housing, built-up infrastructure and construction-development␣
       ↪projects']
      for sector in selected_sectors:
          sector_data = fdi_data[fdi_data['Sector'] == sector].iloc[:, 1:].T
          sector_data.columns = [sector]
          plt.plot(sector_data, marker='o', label=sector)

      plt.title('FDI Trends in Selected Sectors')
      plt.xlabel('Year')
      plt.ylabel('FDI (in million USD)')
      plt.xticks(rotation=90)
      plt.legend()
      plt.grid(True)
      plt.show()
```

FDI Trends in Selected Sectors

**Overall Trend**

- **Increasing FDI**: Across all selected sectors, there is a general trend of increasing FDI inflows over the years.
- **Sectoral Variability**: While all sectors have shown growth, the magnitude and rate of increase vary, with services and technology-related sectors often leading in terms of total investment.
- **Policy Influence**: Government policies and initiatives have played a crucial role in attracting FDI, particularly in sectors identified as critical for economic growth and modernization.

### 0.0.7 FDI Forecasts for Selected Sectors for next 7 years

```
[28]: import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns

      file_path =  "E:/Data Set/FDI data.csv"
      fdi_data = pd.read_csv(file_path)

      # selected sectors
```

```python
selected_sectors = [
    'SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
 ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)',
    'COMPUTER SOFTWARE & HARDWARE',
    'TELECOMMUNICATIONS',
    'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and␣
 ↪construction-development projects'
]

# Function to extend index for forecast
def extend_index(df, new_index):
    return df.reindex(new_index)

# Prepare data for plotting
plt.figure(figsize=(14, 10))
for sector in selected_sectors:
    # Extract sector data
    sector_data = fdi_data[fdi_data['Sector'] == sector].iloc[:, 1:].T
    sector_data.columns = ['FDI']
    sector_data.index = pd.to_datetime(sector_data.index, format='%Y-%y')

    # Forecast for the next 7 years (2017-18 to 2023-24)
    forecast_years = ['2017-18', '2018-19', '2019-20', '2020-21', '2021-22',␣
 ↪'2022-23', '2023-24']
    new_index = sector_data.index.union(pd.to_datetime(forecast_years,␣
 ↪format='%Y-%y'))
    extended_sector_data = extend_index(sector_data, new_index)

    # Forecast using linear trend (simplified)
    fit = np.polyfit(range(len(sector_data.index)), sector_data['FDI'], 1)
    forecast_values = np.polyval(fit, range(len(new_index))[-7:])
    extended_sector_data.loc[pd.to_datetime(forecast_years, format='%Y-%y'),␣
 ↪'FDI'] = forecast_values

    # Plot historical data
    sns.lineplot(data=sector_data, x=sector_data.index.strftime('%Y-%y'),␣
 ↪y='FDI', marker='o', label=f'{sector} (Historical)')

    # Plot forecast data
    sns.lineplot(data=extended_sector_data.loc[pd.to_datetime(forecast_years,␣
 ↪format='%Y-%y')],
                 x=extended_sector_data.loc[pd.to_datetime(forecast_years,␣
 ↪format='%Y-%y')].index.strftime('%Y-%y'),
                 y='FDI', marker='o', linestyle='--', label=f'{sector}␣
 ↪(Forecast)')
```
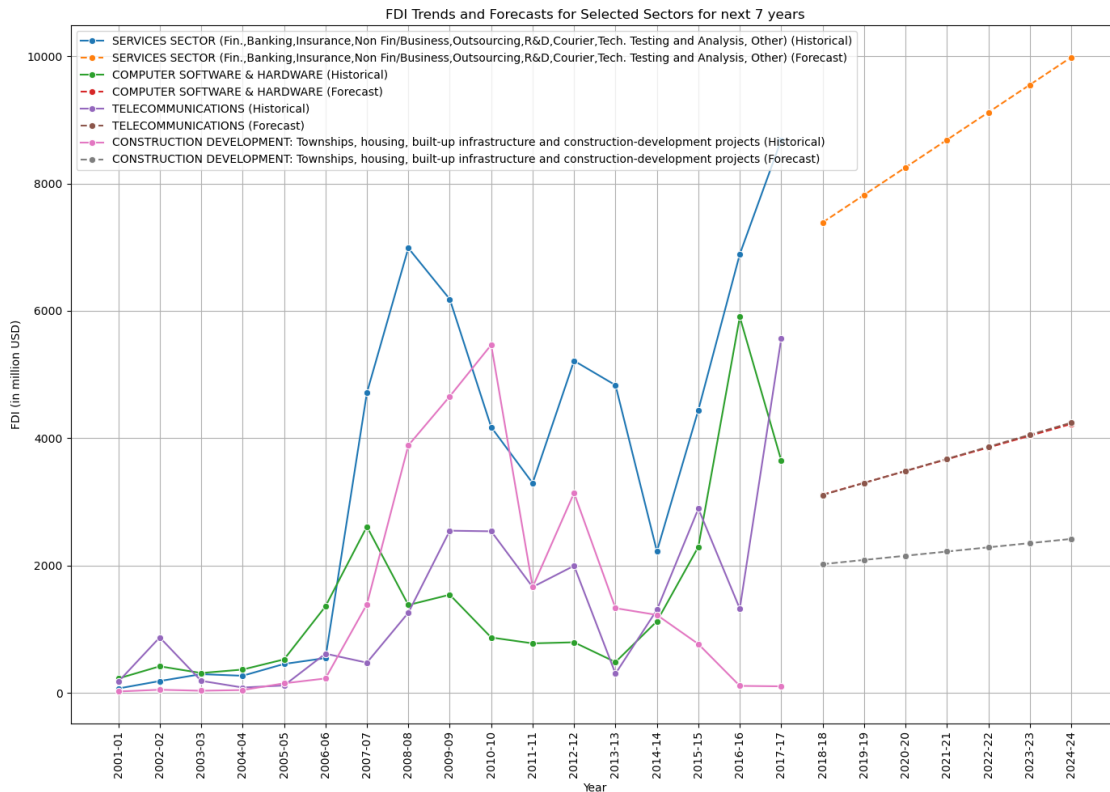
```python
plt.title('FDI Trends and Forecasts for Selected Sectors for next 7 years')
plt.xlabel('Year')
plt.ylabel('FDI (in million USD)')
plt.xticks(rotation=90)
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



1. **FDI on the Rise:** Analysis suggests a surge in Foreign Direct Investment (FDI) across key sectors like services, software & hardware, telecom, and construction.

2. **Policy Push?** Recent government reforms aimed at attracting foreign investment could be a driving force. Investigate specific policy changes that might be playing a role.

3. **Sectoral Spotlight:** Look for breakout sectors with steeper historical growth rates, potentially attracting even more FDI in the future.

4. **Growth with Caution:** Increased FDI can create jobs, transfer technology, and boost infrastructure, but potential downsides like dependence on foreign capital need to be managed.

### 0.0.8 Boxplot to Identify FDI Data

```
[27]: plt.figure(figsize=(14, 10))
      sns.boxplot(data=fdi_data.iloc[:, 1:])
      plt.title('Outlier Detection in FDI Data')
      plt.xlabel('Year')
      plt.ylabel('FDI (in million USD)')
      plt.xticks(rotation=90)
      plt.grid(True)
      plt.show()
```
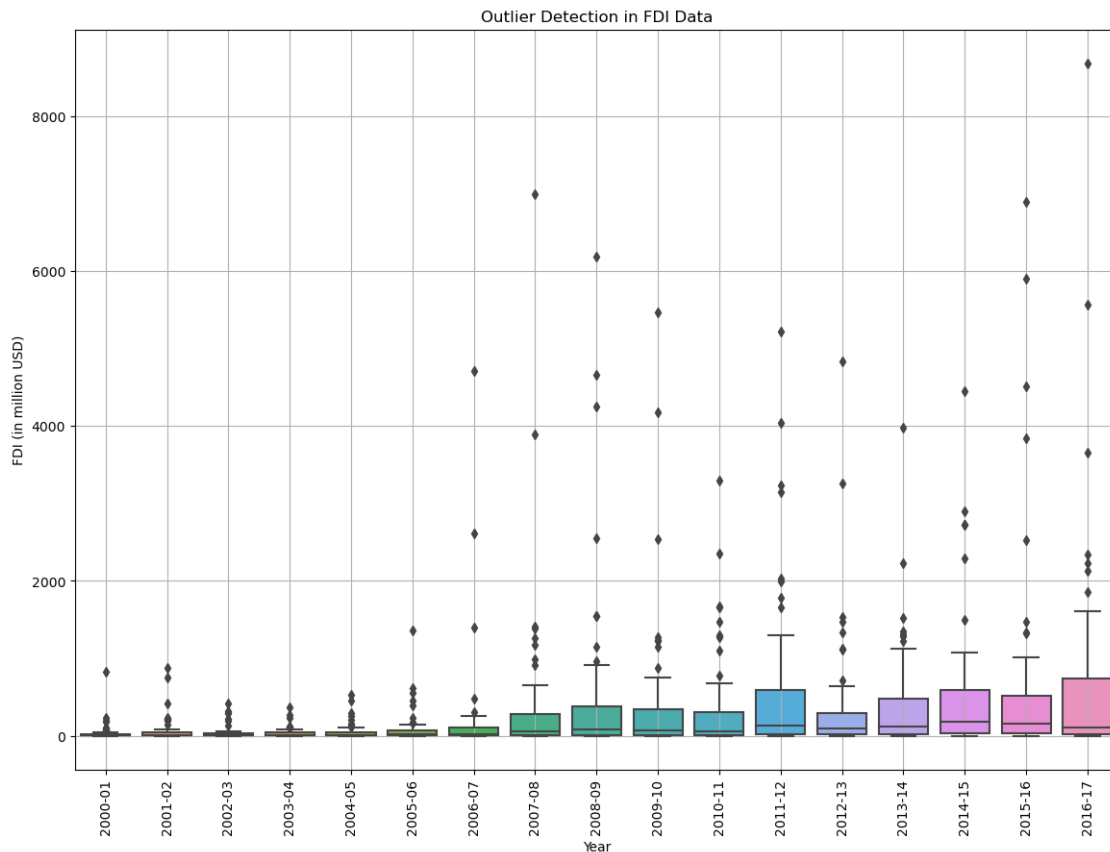


**What This Boxplot Identifies in the FDI Data**

1. **Median FDI:** The median FDI for each year is shown by the line inside each box.

2. **IQR:** The range between Q1 and Q3 shows the spread of the middle 50% of the data.

3. **Spread of Data:** The length of the whiskers gives an idea of the overall spread of the data, excluding outliers.

4. **Outliers:** Points outside the whiskers indicate sectors with unusually high or low FDI for that year.

5. **Outliers Detected**: The boxplot analysis for 2010 and 2011 showed several outliers, suggesting that while the median FDI was low, certain sectors still received relatively high investments.

6. **Increased Variability in 2011**: The FDI data for 2011 showed increased variability, as indicated by a taller box and longer whiskers in the boxplot, reflecting a broader range of investment values across different sectors.

### 0.0.9 Pie Chart of FDI in 2010-11 by Selected Sectors

```python
[38]: import pandas as pd
      import matplotlib.pyplot as plt

      file_path = "E:/Data Set/FDI data.csv"
      fdi_data = pd.read_csv(file_path)

      selected_sectors = [
          'SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
       ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)',
          'COMPUTER SOFTWARE & HARDWARE',
          'TELECOMMUNICATIONS',
          'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and␣
       ↪construction-development projects'
      ]

      # Transpose the data for year-wise analysis
      year_data = fdi_data.set_index('Sector').T.reset_index().melt(id_vars='index',␣
       ↪var_name='Sector', value_name='FDI')
      year_data.columns = ['Year', 'Sector', 'FDI']

      # Filter the data for the selected sectors and the year 2010-11
      filtered_data = year_data[(year_data['Year'] == '2010-11') &␣
       ↪(year_data['Sector'].isin(selected_sectors))]

      # Plot the pie chart for the selected sectors in 2010-11
      plt.figure(figsize=(25, 15))
      plot = filtered_data.groupby('Sector')['FDI'].sum().plot(kind='pie',␣
       ↪autopct='%1.1f%%', textprops={'fontsize': 18, 'fontweight': 'bold'})
      plt.title('Pie Chart of FDI in 2010-11 by Selected Sectors', fontsize=18,␣
       ↪fontweight='bold')
      plt.ylabel('')

      # Customize bold font
      plt.legend(title='Sectors', title_fontsize='20', fontsize='20', loc='center␣
       ↪left', bbox_to_anchor=(1, 0.5), frameon=False, prop={'weight': 'bold'})

      plt.show()
```
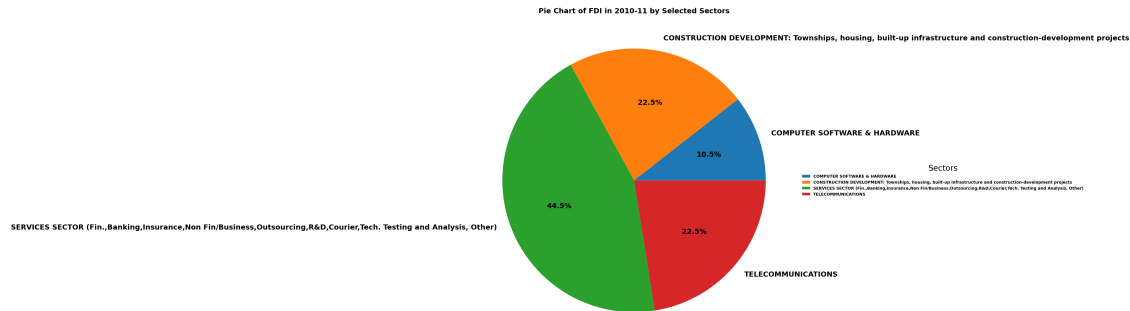
Pie Chart of FDI in 2010-11 by Selected Sectors

Based on the analysis of sector contributions to Foreign Direct Investment (FDI):

1. **Service Sector (44.5%)**: Dominates FDI inflows, indicating its significant role in attracting foreign investment.

2. **Computer Software and Hardware (10.5%)**: Shows a substantial but smaller share compared to the service sector, highlighting its importance in the technology-driven economy.

3. **Telecom (22.5%)**: Represents a considerable sector for FDI, indicating investments in communication infrastructure and technologies.

4. **Construction Development (22.5%)**: Shares a significant portion with telecom, suggesting investments in infrastructure development and real estate.

### 0.0.10 Analysis of FDI Data Based on Histogram

```
[17]:  # Transpose the data for year-wise analysis
       year_data = data.set_index('Sector').T.reset_index().melt(id_vars='index',
        ↪var_name='Sector', value_name='FDI')
       year_data.columns = ['Year', 'Sector', 'FDI']

       # 1. Histogram
       plt.figure(figsize=(12, 6))
       year_data['FDI'].plot(kind='hist', bins=20, alpha=0.5, color='purple')
       plt.title('Histogram of FDI Year-wise')
       plt.xlabel('FDI (in million USD)')
       plt.ylabel('Frequency')
       plt.show()
```

Histogram of FDI Year-wise
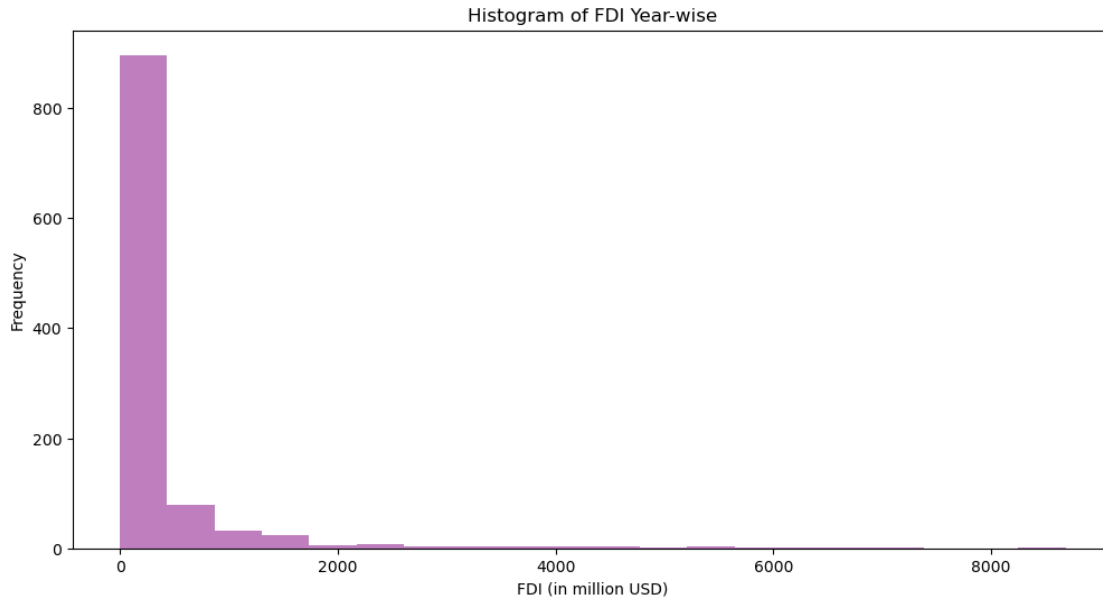
**1. Distribution of FDI Values:**

The histogram shows how FDI values are distributed across different sectors and years. The majority of FDI entries are concentrated at lower investment amounts, indicating that most sectors received moderate levels of investment.

**2. Concentration of Investments:**

The histogram likely shows a concentration of FDI entries at lower investment values, suggesting that many sectors received relatively small to moderate amounts of FDI. This is typical in many datasets where a few sectors might receive very high investments while most receive moderate amounts.

**3. Detection of High FDI Values:**

The right tail of the histogram might indicate sectors with exceptionally high FDI values. These high values could be outliers or significant investments in key sectors, which could warrant further investigation.

**4. Frequency Distribution:**

By observing the height of the bars in the histogram, one can understand the frequency of different FDI value ranges. Taller bars represent ranges with higher frequencies, indicating common investment levels.

### 0.0.11 Analysis of FDI Data Based on Barplot

```
[25]: # Barplot
      plt.figure(figsize=(12, 6))
      year_data.groupby('Year')['FDI'].sum().plot(kind='bar',color='green')
      plt.title('Barplot of Total FDI Year-wise')
```

17

```
plt.xlabel('Year')
plt.ylabel('FDI (in million USD)')
plt.xticks(rotation=90)
plt.show()
```


Barplot of Total FDI Year-wise
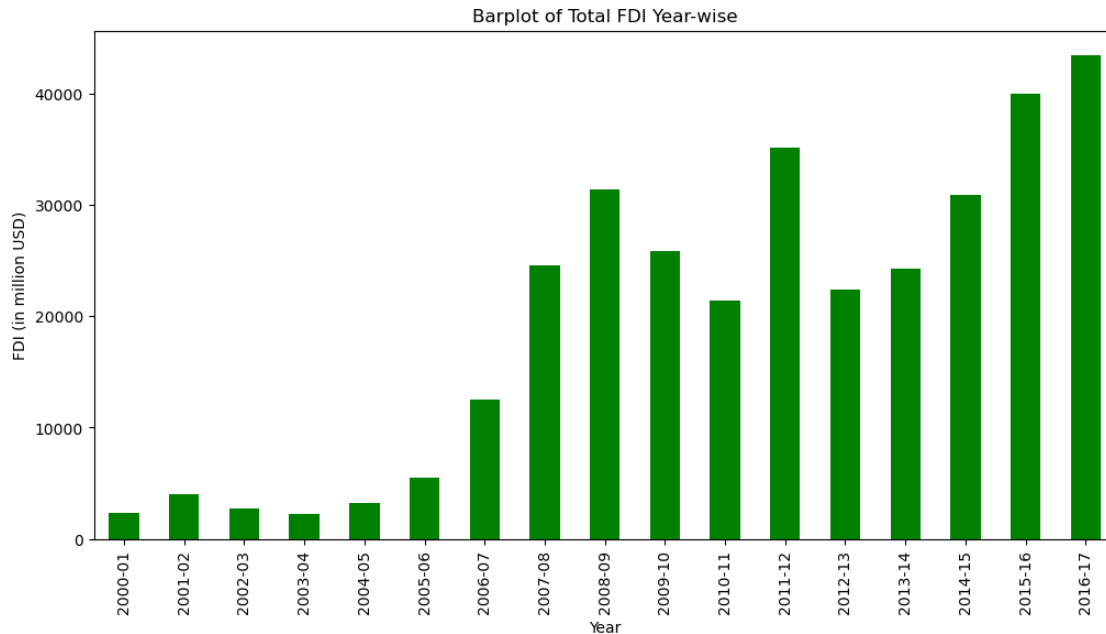
**Comparative Analysis:**

By comparing the heights of the bars, one can quickly see which years were more successful in attracting FDI. This comparative analysis helps in understanding the impact of different policies or global economic conditions on FDI inflows.

**Identifying Trends Over Time:**

The overall trend in the barplot can indicate whether FDI is generally increasing, decreasing, or remaining stable over the period analyzed. This trend is crucial for long-term economic planning and policy-making.

### 0.0.12 scatter plot analysis of FDI year-wise for the selected sectors

```
[32]: selected_sectors = [
          'SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
      ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)',
          'COMPUTER SOFTWARE & HARDWARE',
          'TELECOMMUNICATIONS',
          'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and␣
      ↪construction-development projects'
      ]
```

18

```python
# Transpose the data for year-wise analysis
year_data = data.set_index('Sector').T.reset_index().melt(id_vars='index',
 ↪var_name='Sector', value_name='FDI')
year_data.columns = ['Year', 'Sector', 'FDI']

# Filter the data for the selected sectors
filtered_data = year_data[year_data['Sector'].isin(selected_sectors)]

# 3. Scatterplot
plt.figure(figsize=(20,10))
sns.scatterplot(data=filtered_data, x='Year', y='FDI', hue='Sector',
 ↪palette='viridis', s=200)
plt.title('Scatterplot of FDI Year-wise for Selected Sectors')
plt.xlabel('Year')
plt.ylabel('FDI (in million USD)')
plt.xticks(rotation=90)
plt.legend(title='Sector', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.grid(True)
plt.show()
```
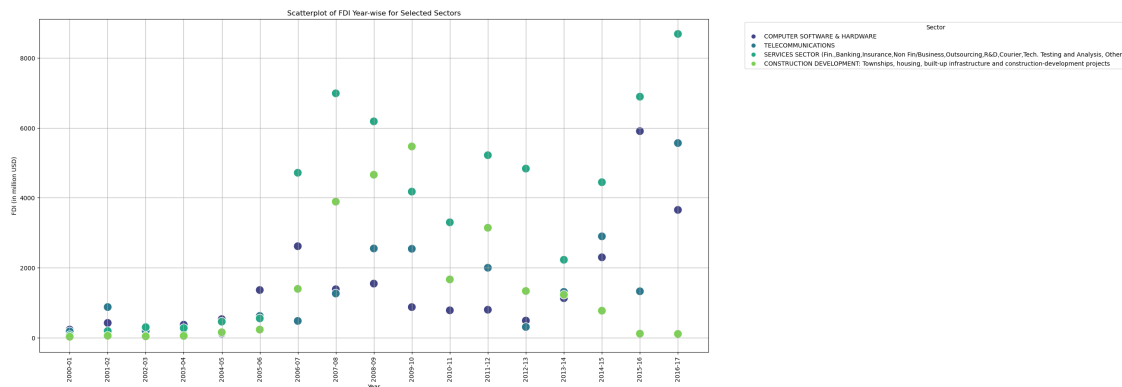


Based on the scatter plot analysis of FDI year-wise for the selected sectors:

1. **Service Sector**: Exhibits a generally upward trend in FDI over the years, indicating increasing investment and growth in this sector.

2. **Computer Software & Hardware**: Shows a more fluctuating trend compared to other sectors, reflecting varying investment levels year by year.

3. **Telecommunications**: Displays consistent growth in FDI, suggesting steady investment in communication infrastructure.

4. **Construction Development**: Indicates significant investment in certain years, with potential fluctuations, reflecting project-based investment trends.

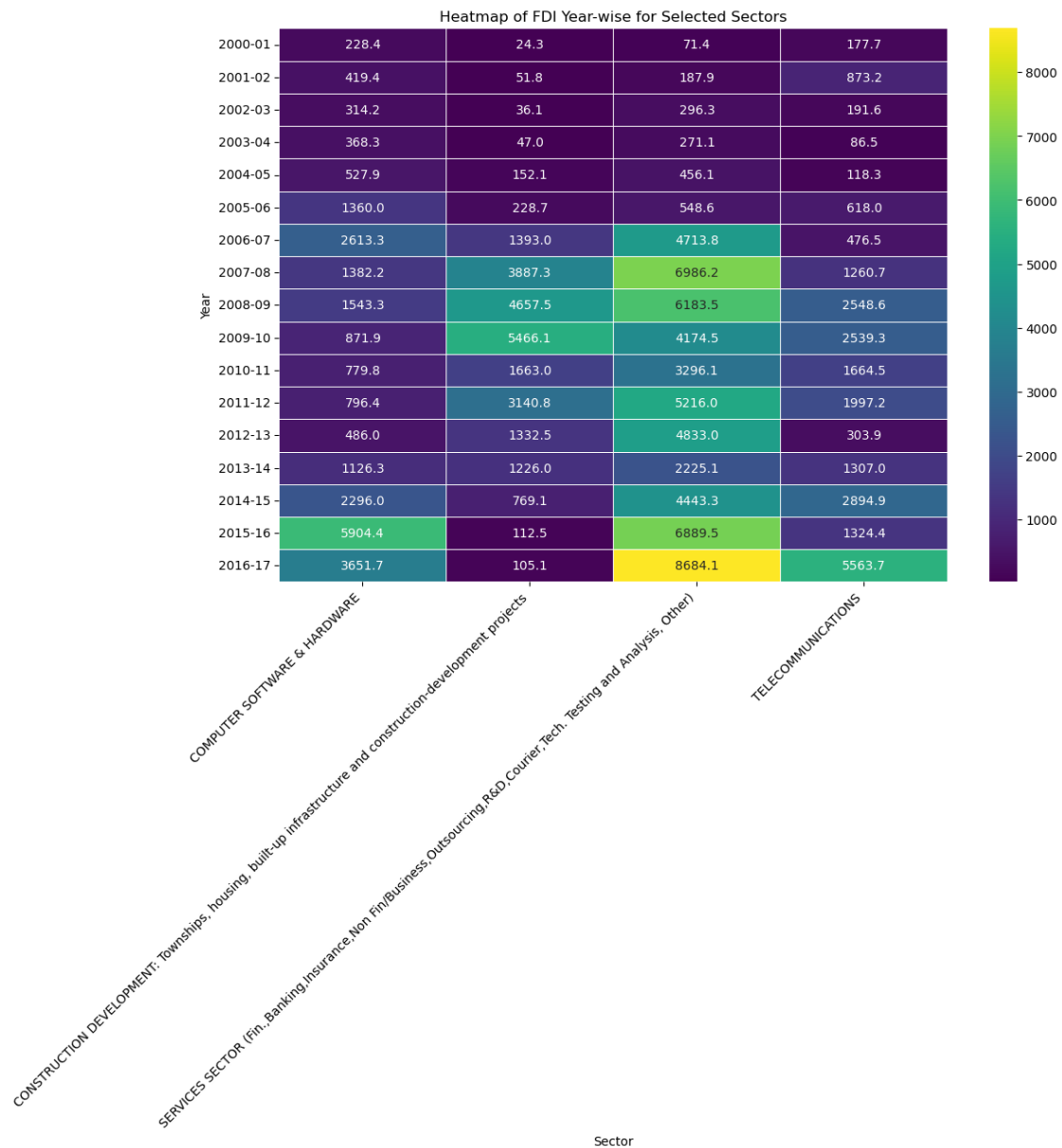### 0.0.13 Heatmap of FDI Year-wise for Selected Sectors

```python
[35]: # Define the selected sectors
      selected_sectors = [
          'SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
       ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)',
          'COMPUTER SOFTWARE & HARDWARE',
          'TELECOMMUNICATIONS',
          'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and␣
       ↪construction-development projects'
      ]

      # Transpose the data for year-wise analysis
      year_data = data.set_index('Sector').T.reset_index().melt(id_vars='index',␣
       ↪var_name='Sector', value_name='FDI')
      year_data.columns = ['Year', 'Sector', 'FDI']

      # Filter the data for the selected sectors
      filtered_data = year_data[year_data['Sector'].isin(selected_sectors)]

      # Pivot data for heatmap
      heatmap_data = filtered_data.pivot('Year', 'Sector', 'FDI')

      # Plot Heatmap
      plt.figure(figsize=(12, 8))
      sns.heatmap(heatmap_data, annot=True, fmt=".1f", cmap='viridis', linewidths=.5)
      plt.title('Heatmap of FDI Year-wise for Selected Sectors')
      plt.xlabel('Sector')
      plt.ylabel('Year')
      plt.xticks(rotation=45, ha='right')
      plt.yticks(rotation=0)
      plt.show()
```

Heatmap of FDI Year-wise for Selected Sectors

| Year | COMPUTER SOFTWARE & HARDWARE | CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-development projects | SERVICES SECTOR (Fin, Banking, Insurance, Non Fin/Business, Outsourcing, R&D, Courier, Tech. Testing and Analysis, Other) | TELECOMMUNICATIONS |
|---|---|---|---|---|
| 2000-01 | 228.4 | 24.3 | 71.4 | 177.7 |
| 2001-02 | 419.4 | 51.8 | 187.9 | 873.2 |
| 2002-03 | 314.2 | 36.1 | 296.3 | 191.6 |
| 2003-04 | 368.3 | 47.0 | 271.1 | 86.5 |
| 2004-05 | 527.9 | 152.1 | 456.1 | 118.3 |
| 2005-06 | 1360.0 | 228.7 | 548.6 | 618.0 |
| 2006-07 | 2613.3 | 1393.0 | 4713.8 | 476.5 |
| 2007-08 | 1382.2 | 3887.3 | 6986.2 | 1260.7 |
| 2008-09 | 1543.3 | 4657.5 | 6183.5 | 2548.6 |
| 2009-10 | 871.9 | 5466.1 | 4174.5 | 2539.3 |
| 2010-11 | 779.8 | 1663.0 | 3296.1 | 1664.5 |
| 2011-12 | 796.4 | 3140.8 | 5216.0 | 1997.2 |
| 2012-13 | 486.0 | 1332.5 | 4833.0 | 303.9 |
| 2013-14 | 1126.3 | 1226.0 | 2225.1 | 1307.0 |
| 2014-15 | 2296.0 | 769.1 | 4443.3 | 2894.9 |
| 2015-16 | 5904.4 | 112.5 | 6889.5 | 1324.4 |
| 2016-17 | 3651.7 | 105.1 | 8684.1 | 5563.7 |

Sector

- **Purpose**: Visualizes the intensity of FDI across different sectors and years.
- **Data**: Represents FDI values where each cell's color intensity indicates the magnitude of investment.
- **Insights**:
  - **High Intensity**: Shows sectors and years with significant investment.
  - **Low Intensity**: Highlights periods or sectors with lower investment levels.
  - **Patterns**: Helps identify trends or seasonal effects in FDI distribution across sectors.

### 0.0.14 Stacked Area Chart of FDI Year-wise for Selected Sectors

```python
[36]: selected_sectors = [
          'SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/
       ↪Business,Outsourcing,R&D,Courier,Tech. Testing and Analysis, Other)',
          'COMPUTER SOFTWARE & HARDWARE',
          'TELECOMMUNICATIONS',
          'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and␣
       ↪construction-development projects'
      ]

      # Transpose the data for year-wise analysis
      year_data = data.set_index('Sector').T.reset_index().melt(id_vars='index',␣
       ↪var_name='Sector', value_name='FDI')
      year_data.columns = ['Year', 'Sector', 'FDI']

      # Filter the data for the selected sectors
      filtered_data = year_data[year_data['Sector'].isin(selected_sectors)]

      # Pivot data for stacked area chart
      pivot_data = filtered_data.pivot(index='Year', columns='Sector', values='FDI').
       ↪fillna(0)

      # Plot Stacked Area Chart
      plt.figure(figsize=(14, 8))
      pivot_data.plot.area(figsize=(14, 8), cmap='viridis', alpha=0.7)
      plt.title('Stacked Area Chart of FDI Year-wise for Selected Sectors')
      plt.xlabel('Year')
      plt.ylabel('FDI (in million USD)')
      plt.legend(title='Sector')
      plt.grid(True)
      plt.show()
```

```
<Figure size 1400x800 with 0 Axes>
```

Stacked Area Chart of FDI Year-wise for Selected Sectors
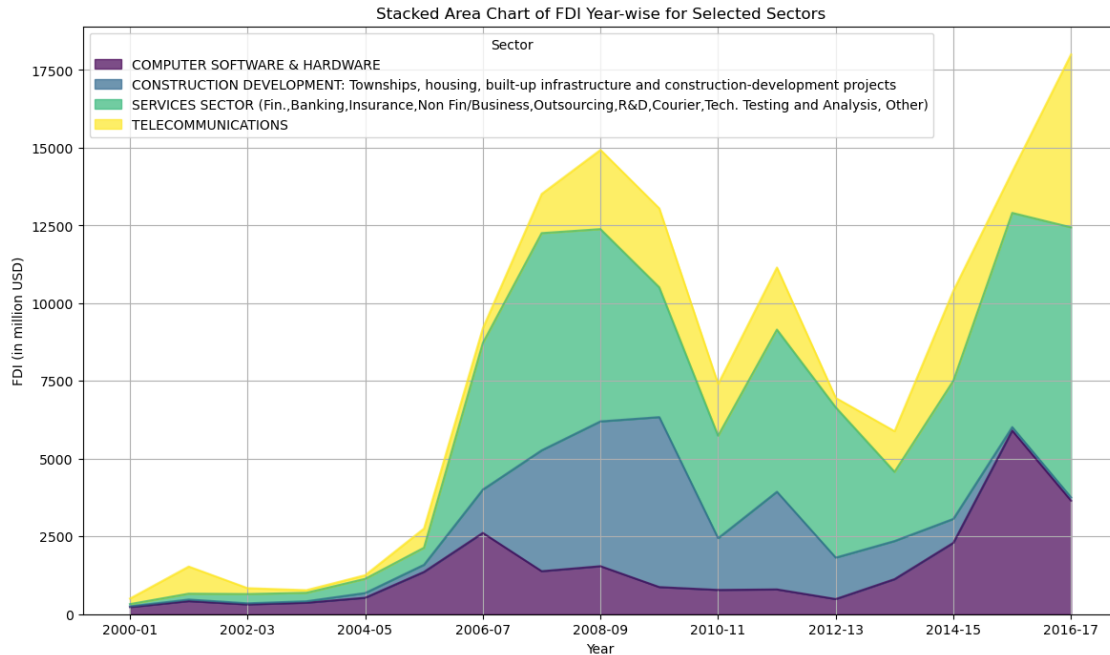
- **Purpose**: Shows the cumulative FDI investment over time for different sectors.
- **Data**: Each area represents the FDI value for a sector, stacked on top of others.
- **Insights**:
    - **Trend Analysis**: Displays how each sector's investment contributes to the total over the years.
    - **Comparison**: Facilitates comparison between sectors to see which has increased or decreased over time.
    - **Sector Dominance**: Highlights which sectors have been dominant or emerging in terms of investment over the years.

<h1>Thank You!</h1>