

**Project Title = Foreign Direct Investment Analytics**

**Technologies = Data Science**

**Domain = Finance**

**Project Difficulties level = Intermediate**

**By Nikhil Reddy**

## ▼ Problem Statement:

Investment is a game of understanding historic data of investment objects under different events but it is still a game of chances to minimize the risk we apply analytics to find the equilibrium investment. To understand the Foreign direct investment in India for the last 17 years from 2000-01 to 2016-17. This dataset contains sector and financial year-wise data of FDI in India Sector-wise investment analysis Year-wise investment analysis.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
from google.colab import files
uploaded = files.upload()
```

→ Choose files FDI data.csv  
 • **FDI data.csv**(text/csv) - 7992 bytes, last modified: 08/01/2025 - 100% done  
 Saving FDI data.csv to FDI data (1).csv

```
df = pd.read_csv('FDI data.csv')
df.head()
```

	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
0	METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94	419.88	1098.14	1786.14	1466.23	567.63	35
1	MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16	174.40	79.51	142.65	57.89	12.73	68
2	POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68	907.66	1271.79	1271.77	1652.38	535.68	1066.08	70
3	NON-CONVENTIONAL ENERGY	0.00	0.00	1.70	4.14	1.27	1.35	2.44	58.82	125.88	622.52	214.40	452.17	1106.52	414.25	61
4	COAL PRODUCTION	0.00	0.00	0.00	0.04	0.00	9.14	1.30	14.08	0.22	0.00	0.00	0.00	0.00	0.00	2.96

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
df.tail()
```

	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	201
58	PRINTING OF BOOKS (INCLUDING LITHO PRINTING IN...	0.00	0.00	6.30	0.00	0.06	9.90	20.04	35.54	31.61	70.51	36.63	47.39	14.34	113
59	COIR	0.00	0.00	0.00	0.00	0.47	0.59	0.04	0.01	0.00	0.25	0.10	0.55	0.15	0
60	CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES	0.00	0.00	0.00	0.00	0.00	0.93	64.06	182.92	172.70	324.56	675.07	386.28	283.89	485
61	CONSTRUCTION DEVELOPMENT: Townships, housing, ...	24.33	51.75	36.10	47.04	152.06	228.71	1392.95	3887.33	4657.51	5466.13	1663.03	3140.78	1332.49	1226
62	MISCELLANEOUS INDUSTRIES	832.07	221.37	218.76	235.48	121.83	164.76	304.87	528.42	1549.70	1147.56	1475.97	813.38	229.49	468

df.shape

(63, 18)

df.info()

```
↙ <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
```

## ✓ All the datatypes are correct

df.describe()

	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
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63
mean	37.757302	63.931587	42.925714	34.727778	51.090317	87.932540	198.281905	390.085714	498.348571	410.069524	339						
std	112.227860	157.878737	86.606439	67.653735	101.934873	206.436967	686.783115	1026.249935	1134.649040	926.814626	627						
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0						
25%	0.000000	0.000000	0.200000	0.215000	0.715000	1.230000	4.160000	9.950000	11.950000	7.880000	8						
50%	4.030000	5.070000	11.010000	6.370000	9.090000	22.620000	25.820000	58.820000	84.880000	69.740000	58						
75%	23.510000	44.830000	36.555000	38.660000	43.205000	63.855000	108.325000	279.270000	383.320000	341.595000	304						
max	832.070000	873.230000	419.960000	368.320000	527.900000	1359.970000	4713.780000	6986.170000	6183.490000	5466.130000	3296						

✓ Here from above we can understand that the data of FDI is largely skewed specifically right skewed as mean is greater than median(50%)

```
df.isnull().sum()
```

```
→ 0
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
```

✓ There are no null values

```
df.duplicated().sum()
```

```
→ 0
```

✓ From above we can conclude that there are no duplicates involved here!!!

```
df.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')
```

```
df1 = df.set_index('Sector') #####making Sector as index column
df1.head()
```

	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
Sector															
METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94	419.88	1098.14	1786.14	1466.23	567.63	359.3
MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16	174.40	79.51	142.65	57.89	12.73	684.3
POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68	907.66	1271.79	1271.77	1652.38	535.68	1066.08	707.0
NON-CONVENTIONAL ENERGY	0.00	0.00	1.70	4.14	1.27	1.35	2.44	58.82	125.88	622.52	214.40	452.17	1106.52	414.25	615.5
COAL PRODUCTION	0.00	0.00	0.00	0.04	0.00	9.14	1.30	14.08	0.22	0.00	0.00	0.00	0.00	2.96	0.0

Next steps: [Generate code with df1](#) [View recommended plots](#) [New interactive sheet](#)

```
df['Sector'].nunique()
```

63

## Univariate Plots

### Distribution of FDI For Each Year

```
def histplots(df1):
    columns = df1.columns
    n_cols = 3
    n_rows = -(-len(columns) // n_cols)

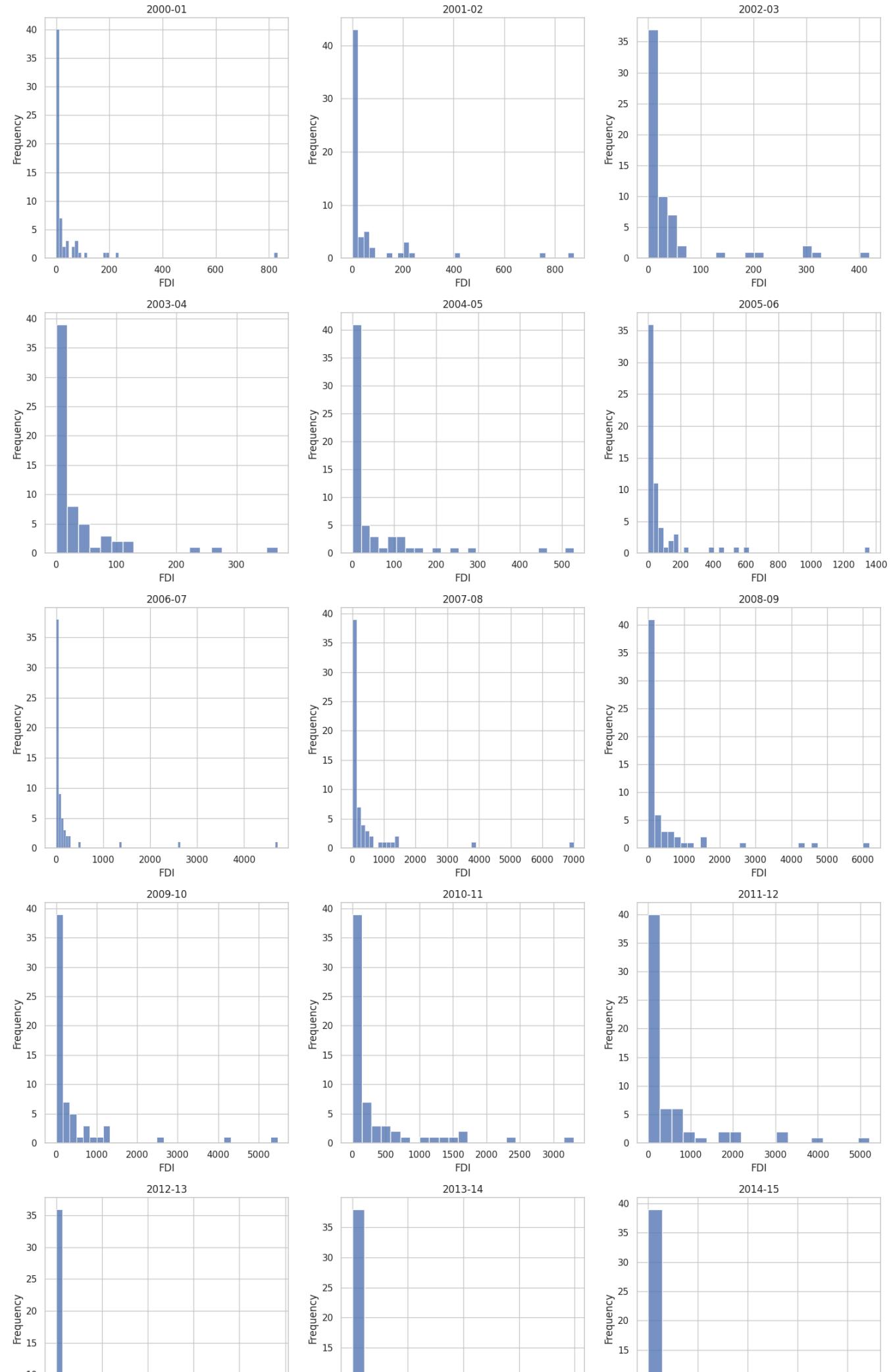
    fig, axes = plt.subplots(n_rows, n_cols, figsize=(15, 5*n_rows))
    axes = axes.flatten()

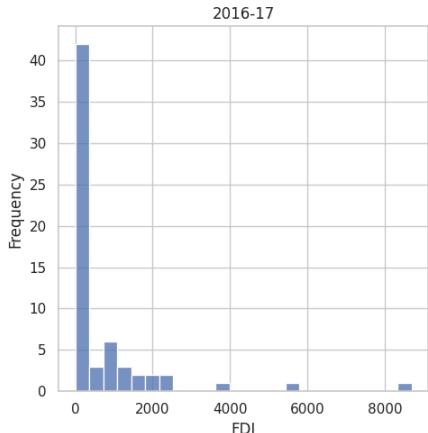
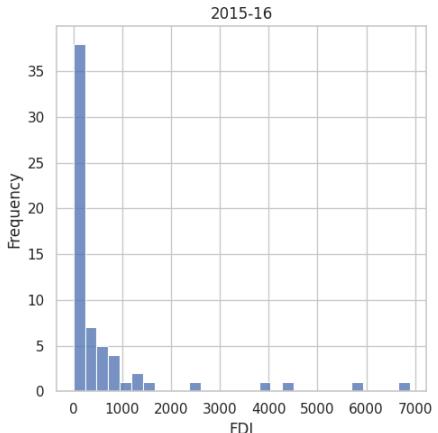
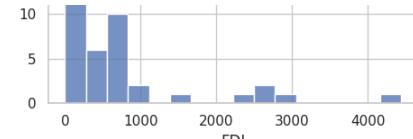
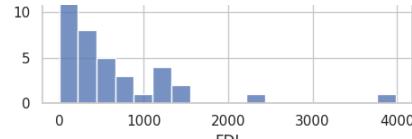
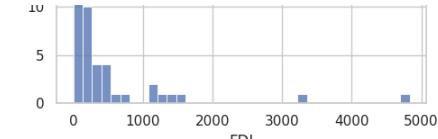
    for i, col in enumerate(columns):
        sns.histplot(df1[col], ax=axes[i])
        axes[i].set_title(col)
        axes[i].set_xlabel("FDI")
        axes[i].set_ylabel('Frequency')

    # Remove empty subplots
    for j in range(i + 1, n_rows * n_cols):
        fig.delaxes(axes[j])

    plt.tight_layout()
    plt.show()

histplots(df1)
```





```
df2 = df1.T
df2.head()
```



Sector	METALLURGICAL INDUSTRIES	MINING	POWER	CONVENTIONAL ENERGY	COAL PRODUCTION	PETROLEUM & NATURAL GAS	BOILERS AND STEAM GENERATING PLANTS	PRIME MOVER (OTHER THAN ELECTRICAL GENERATORS)	ELECTRICAL EQUIPMENTS	COMPUTER SOFTWARE & HARDWARE
2000-01	22.69	1.32	89.42	0.00	0.00	9.35	0.00	0.00	79.76	228.39
2001-02	14.14	6.52	757.44	0.00	0.00	211.07	0.00	0.00	65.76	419.39
2002-03	36.61	10.06	59.11	1.70	0.00	56.78	0.00	0.00	34.71	314.24
2003-04	8.11	23.48	27.09	4.14	0.04	80.64	0.04	0.00	73.20	368.32
2004-05	200.38	9.92	43.37	1.27	0.00	102.78	0.54	2.66	97.40	527.90

5 rows × 63 columns

df2.columns

```
Index(['METALLURGICAL INDUSTRIES', 'MINING', 'POWER',
       'NON-CONVENTIONAL ENERGY', 'COAL PRODUCTION', 'PETROLEUM & NATURAL GAS',
       'BOILERS AND STEAM GENERATING PLANTS',
       'PRIME MOVER (OTHER THAN ELECTRICAL GENERATORS)',
       'ELECTRICAL EQUIPMENTS', 'COMPUTER SOFTWARE & HARDWARE', 'ELECTRONICS',
       'TELECOMMUNICATIONS',
       'INFORMATION & BROADCASTING (INCLUDING PRINT MEDIA)',
       'AUTOMOBILE INDUSTRY', 'AIR TRANSPORT (INCLUDING AIR FREIGHT)',
       'SEA TRANSPORT', 'PORTS', 'RAILWAY RELATED COMPONENTS',
       'INDUSTRIAL MACHINERY', 'MACHINE TOOLS', 'AGRICULTURAL MACHINERY',
       'EARTH-MOVING MACHINERY',
       'MISCELLANEOUS MECHANICAL & ENGINEERING INDUSTRIES',
       'COMMERCIAL, OFFICE & HOUSEHOLD EQUIPMENTS',
       'MEDICAL AND SURGICAL APPLIANCES', 'INDUSTRIAL INSTRUMENTS',
       'SCIENTIFIC INSTRUMENTS',
       'MATHEMATICAL, SURVEYING AND DRAWING INSTRUMENTS', 'FERTILIZERS',
       'CHEMICALS (OTHER THAN FERTILIZERS)', 'PHOTOGRAPHIC RAW FILM AND PAPER',
       'DYE-STUFFS', 'DRUGS & PHARMACEUTICALS',
       'TEXTILES (INCLUDING DYED, PRINTED)',
       'PAPER AND PULP (INCLUDING PAPER PRODUCTS)', 'SUGAR',
```

```
'FERMENTATION INDUSTRIES', 'FOOD PROCESSING INDUSTRIES',
'VEGETABLE OILS AND VANASPATI',
'SOAPS, COSMETICS & TOILET PREPARATIONS', 'RUBBER GOODS',
'LEATHER, LEATHER GOODS AND PICKERS', 'GLUE AND GELATIN', 'GLASS',
'CERAMICS', 'CEMENT AND GYPSUM PRODUCTS', 'TIMBER PRODUCTS',
'DEFENCE INDUSTRIES', 'CONSULTANCY SERVICES',
'SERVICES SECTOR (Fin., Banking, Insurance, Non Fin/Business, Outsourcing, R&D, Courier, Tech. Testing and Analysis, Other)',

'HOSPITAL & DIAGNOSTIC CENTRES', 'EDUCATION', 'HOTEL & TOURISM',
'TRADING', 'RETAIL TRADING', 'AGRICULTURE SERVICES',
'DIAMOND, GOLD ORNAMENTS',
'TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)',
'PRINTING OF BOOKS (INCLUDING LITHO PRINTING INDUSTRY)', 'COIR',
'CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES',
'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-development projects',
'MISCELLANEOUS INDUSTRIES']]

dtype='object', name='Sector')
```

```
df2.rename(columns={'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-development proj': 'CONSTRUCTION DEVELOPMENT'}, inplace=True)
df2.head()
```



Sector	METALLURGICAL INDUSTRIES	MINING	POWER	NON-CONVENTIONAL ENERGY	COAL PRODUCTION	PETROLEUM & NATURAL GAS	BOILERS AND STEAM GENERATING PLANTS	PRIME MOVER (OTHER THAN ELECTRICAL GENERATORS)	ELECTRICAL EQUIPMENTS	COMPUTER SOFTWARE & HARDWARE
2000-01	22.69	1.32	89.42	0.00	0.00	9.35	0.00	0.00	79.76	228.39
2001-02	14.14	6.52	757.44	0.00	0.00	211.07	0.00	0.00	65.76	419.39
2002-03	36.61	10.06	59.11	1.70	0.00	56.78	0.00	0.00	34.71	314.24
2003-04	8.11	23.48	27.09	4.14	0.04	80.64	0.04	0.00	73.20	368.32
2004-05	200.38	9.92	43.37	1.27	0.00	102.78	0.54	2.66	97.40	527.90

5 rows x 63 columns

### FDI Over The Years For Each Sector

```
def scatterp(df2):
    columns = df2.columns
    n_cols = 3
    n_rows = -(len(columns)) // n_cols

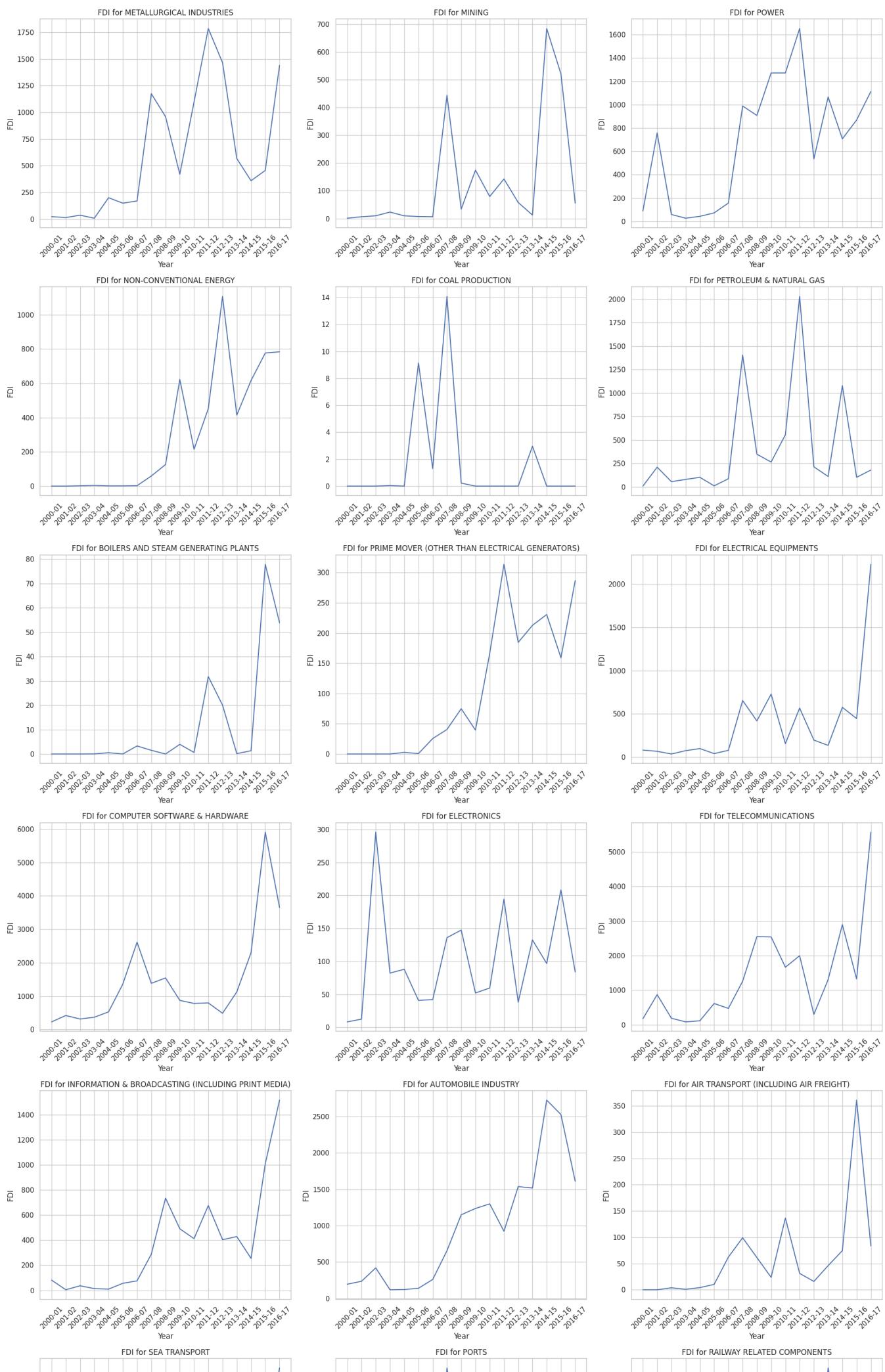
    fig, axes = plt.subplots(n_rows, n_cols, figsize=(20, 6*n_rows))
    axes = axes.flatten()

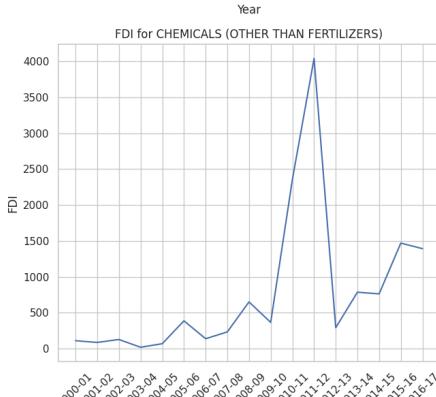
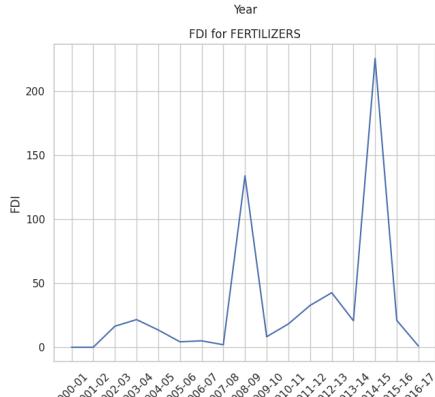
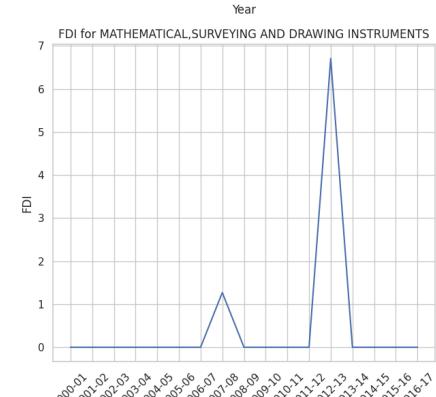
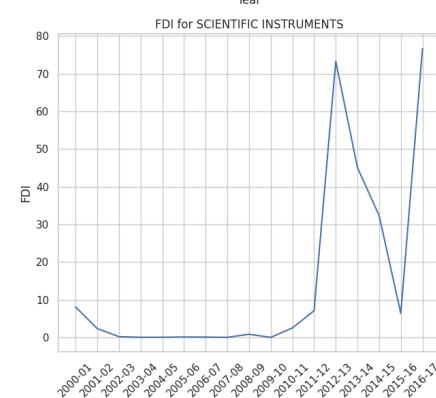
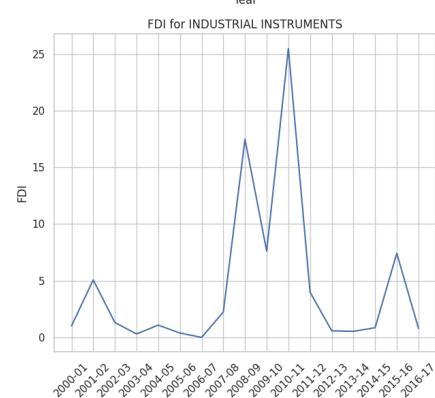
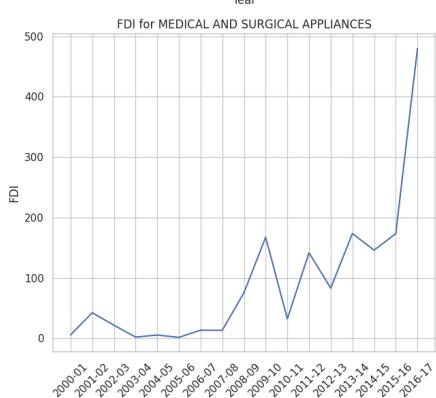
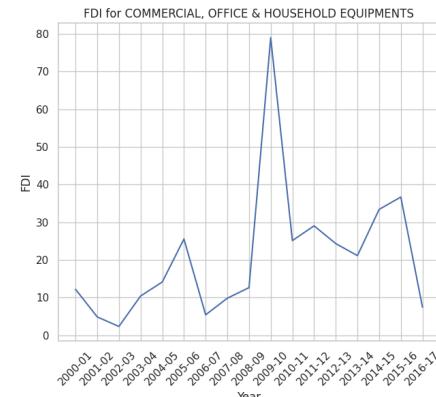
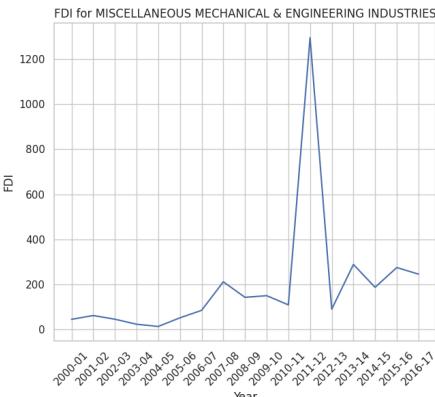
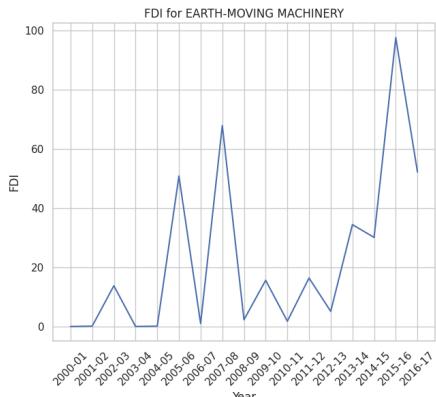
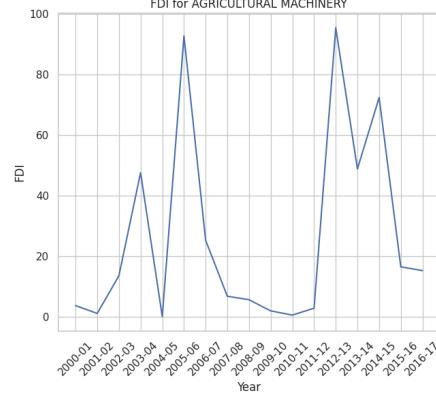
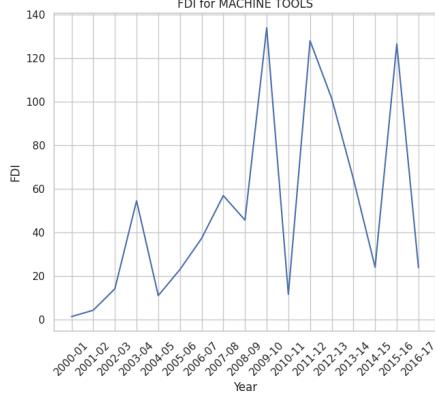
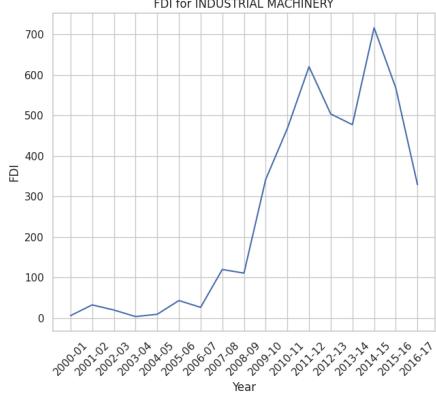
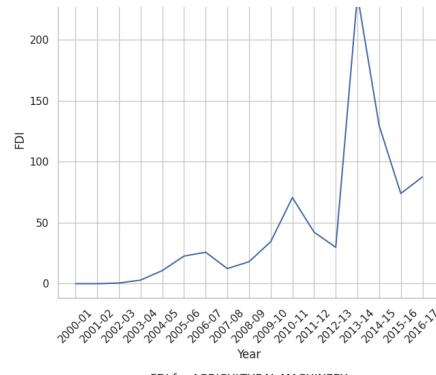
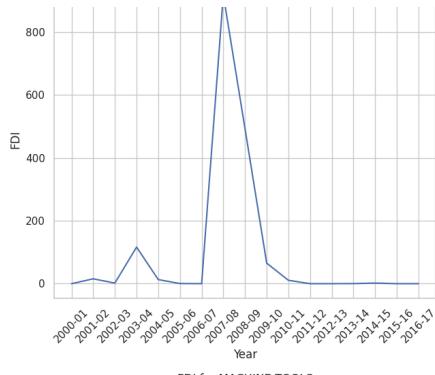
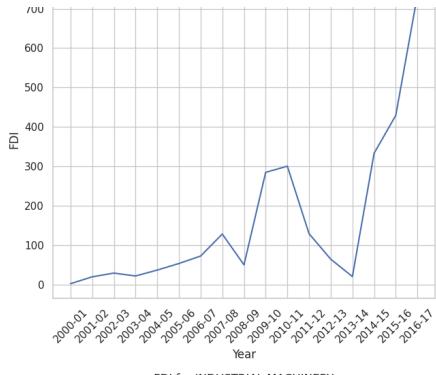
    for i, col in enumerate(columns):
        axes[i].plot(df2.index.values, df2[col])
        axes[i].set_title("FDI for " + col)
        axes[i].set_xlabel("Year")
        axes[i].set_xticklabels(df2.index.values, rotation=45)
        axes[i].set_ylabel("FDI")
        axes[i].grid(True)

    # Remove empty subplots
    for j in range(i + 1, n_rows * n_cols):
        fig.delaxes(axes[j])

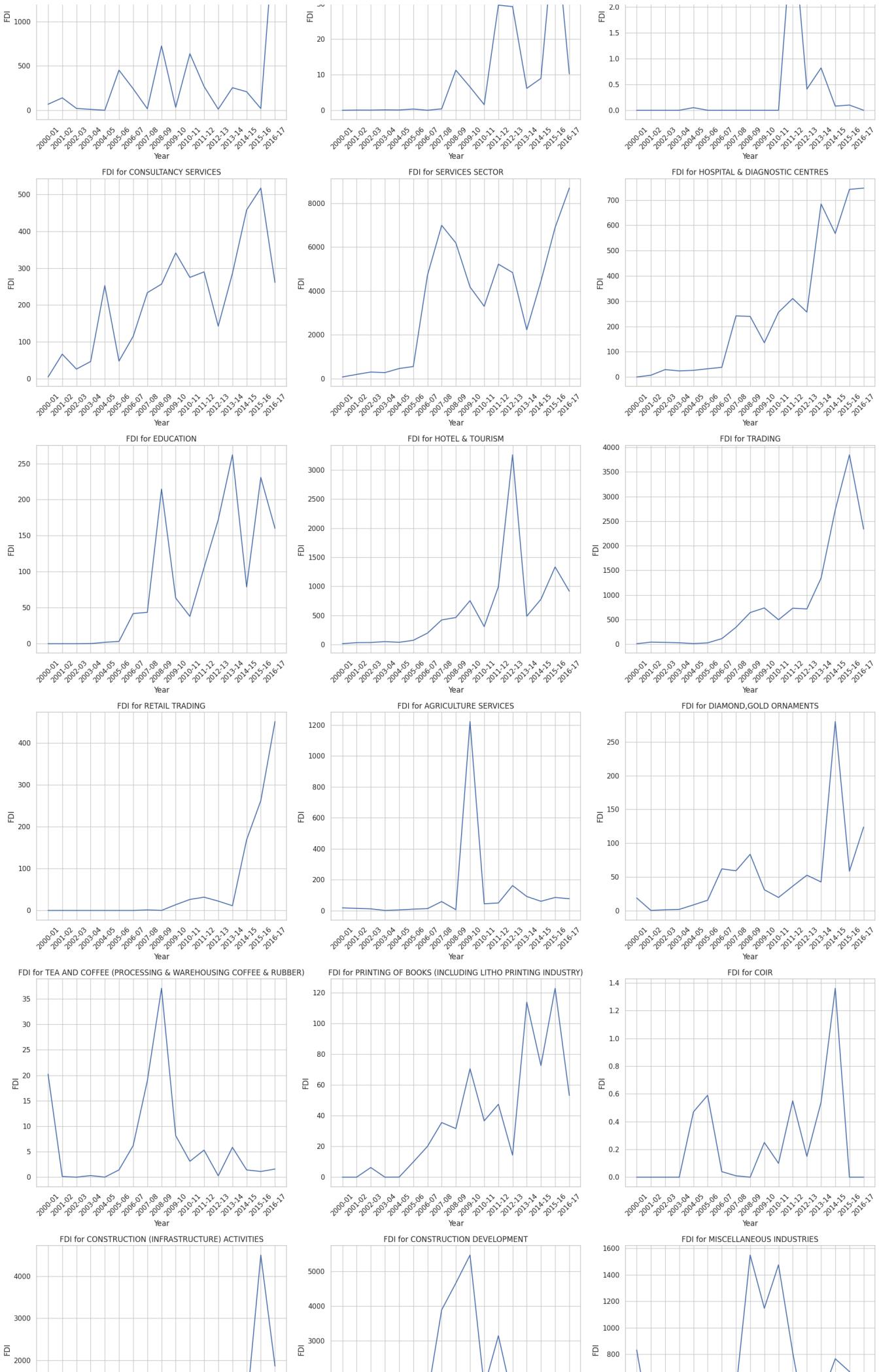
    plt.tight_layout()
    plt.show()

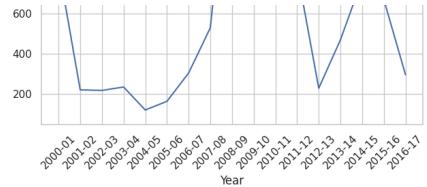
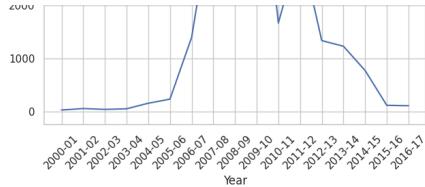
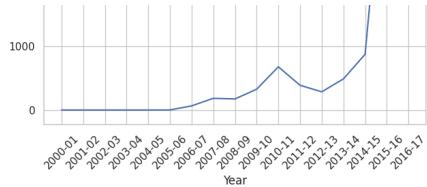
scatterp(df2)
```











The above plot shows us the how FDI changed over the years for each sector

### ✓ Year wise total FDI

```
year_total = df2.sum(axis=1)
year_total = pd.DataFrame(year_total, columns=['Total_FDI'])
year_total.sort_values(by='Total_FDI', ascending=False)
```

	Total_FDI	
2016-17	43478.26	■
2015-16	40000.99	
2011-12	35120.78	
2008-09	31395.96	
2014-15	30930.47	
2009-10	25834.38	
2007-08	24575.40	
2013-14	24299.32	
2012-13	22423.59	
2010-11	21383.07	
2006-07	12491.76	
2005-06	5539.75	
2001-02	4027.69	
2004-05	3218.69	
2002-03	2704.32	
2000-01	2378.71	
2003-04	2187.85	

### ✓ Year wise Average FDI

```
year_avg = df2.mean(axis=1)
year_avg = pd.DataFrame(year_avg, columns=['Avg_FDI'])
year_avg.sort_values(by='Avg_FDI', ascending=False)
```

	Avg_FDI	grid
2016-17	690.131111	bar
2015-16	634.936349	bar
2011-12	557.472698	bar
2008-09	498.348571	bar
2014-15	490.959841	bar
2009-10	410.069524	bar
2007-08	390.085714	bar
2013-14	385.703492	bar
2012-13	355.930000	bar
2010-11	339.413810	bar
2006-07	198.281905	bar
2005-06	87.932540	bar
2001-02	63.931587	bar
2004-05	51.090317	bar
2002-03	42.925714	bar
2000-01	37.757302	bar
2003-04	34.727778	bar

## ▼ Top 10 Sector wise total FDI

```
df3 = df2.T
sec_total = df3.sum(axis=1)
sec_total = pd.DataFrame(sec_total, columns=['Total_FDI'])
sec_total_top = sec_total.sort_values(by='Total_FDI', ascending=False).head(10)
sec_total_top
```

	Total_FDI	grid
Sector		bar
SERVICES SECTOR	59476.49	bar
COMPUTER SOFTWARE & HARDWARE	24669.49	bar
CONSTRUCTION DEVELOPMENT	24293.09	bar
TELECOMMUNICATIONS	23946.01	bar
AUTOMOBILE INDUSTRY	16673.92	bar
DRUGS & PHARMACEUTICALS	14706.90	bar
TRADING	14210.88	bar
CHEMICALS (OTHER THAN FERTILIZERS)	13293.09	bar
POWER	11589.13	bar
METALLURGICAL INDUSTRIES	10330.54	bar

Next steps: [Generate code with sec\\_total\\_top](#) [View recommended plots](#) [New interactive sheet](#)

From above data we can see that the service sector was able to draw more FDI than other sectors

## ▼ Bottom 10 Sector wise total FDI

```
sec_total_tail = sec_total.sort_values(by='Total_FDI', ascending=False).tail(10)
sec_total_tail
```



Total\_FDI



Sector



TIMBER PRODUCTS	157.68
GLUE AND GELATIN	128.39
TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)	111.22
DYE-STUFFS	88.40
INDUSTRIAL INSTRUMENTS	76.12
PHOTOGRAPHIC RAW FILM AND PAPER	67.28
COAL PRODUCTION	27.74
MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS	7.98
DEFENCE INDUSTRIES	5.12
COIR	4.06

Next steps: [Generate code with sec\\_total\\_tail](#) [View recommended plots](#) [New interactive sheet](#)

From above data we can see that the COIR sector was able to draw least FDI than other sectors

## Top 10 Sector wise Average FDI

```
sec_avg = df3.mean(axis=1)
sec_avg = pd.DataFrame(sec_avg, columns=['Avg_FDI'])
sec_avg_top = sec_avg.sort_values(by='Avg_FDI', ascending=False).head(10)
sec_avg_top
```



Avg\_FDI



Sector



SERVICES SECTOR	3498.617059
COMPUTER SOFTWARE & HARDWARE	1451.146471
CONSTRUCTION DEVELOPMENT	1429.005294
TELECOMMUNICATIONS	1408.588824
AUTOMOBILE INDUSTRY	980.818824
DRUGS & PHARMACEUTICALS	865.111765
TRADING	835.934118
CHEMICALS (OTHER THAN FERTILIZERS)	781.946471
POWER	681.713529
METALLURGICAL INDUSTRIES	607.678824

Next steps: [Generate code with sec\\_avg\\_top](#) [View recommended plots](#) [New interactive sheet](#)

## Bottom 10 Sector wise Average FDI

```
sec_avg_bottom = sec_avg.sort_values(by='Avg_FDI', ascending=False).tail(10)
sec_avg_bottom
```



Avg\_FDI



Sector



TIMBER PRODUCTS	9.275294
GLUE AND GELATIN	7.552353
TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)	6.542353
DYE-STUFFS	5.200000
INDUSTRIAL INSTRUMENTS	4.477647
PHOTOGRAPHIC RAW FILM AND PAPER	3.957647
COAL PRODUCTION	1.631765
MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS	0.469412
DEFENCE INDUSTRIES	0.301176
COIR	0.238824

Next steps: [Generate code with sec\\_avg\\_bottom](#) [View recommended plots](#) [New interactive sheet](#)

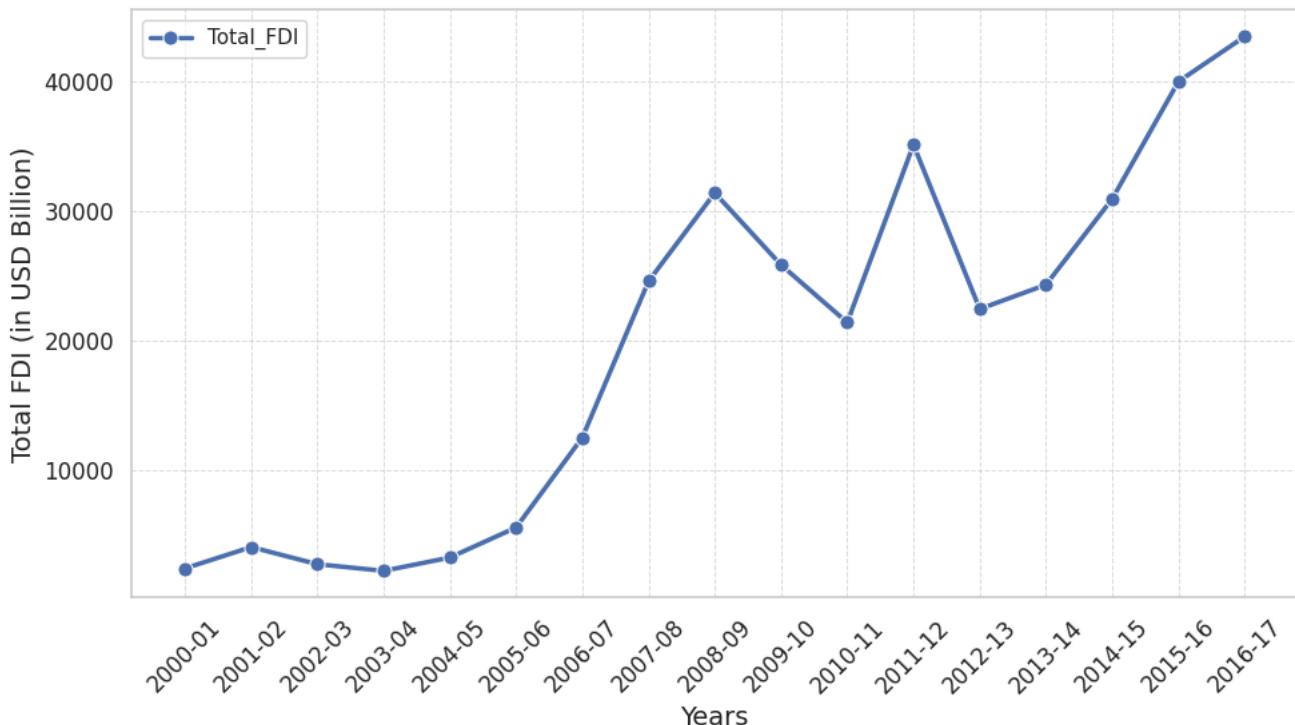
## ▼ Yearwise Total\_FDI in INDIA

```
sns.set(style="whitegrid")
plt.figure(figsize=(10, 6))

sns.lineplot(data=year_total, marker='o', color='#1f77b4', markersize=8, linewidth=2.5)

plt.title('Yearwise FDI in INDIA', fontsize=16, fontweight='bold')
plt.xlabel('Years', fontsize=14)
plt.ylabel('Total FDI (in USD Billion)', fontsize=14)
plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)

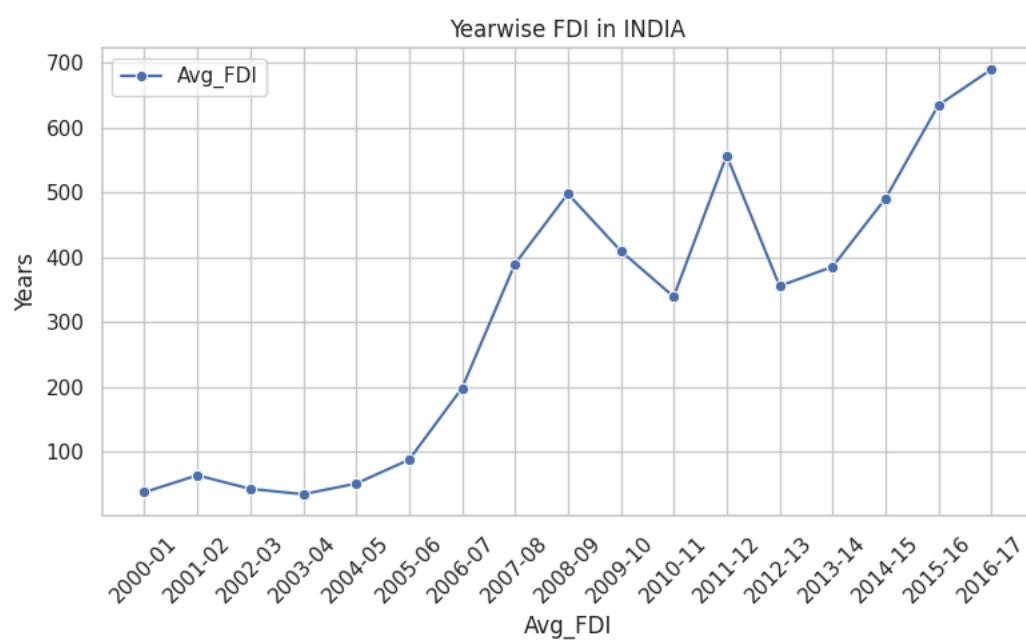
plt.grid(True, linestyle='--', linewidth=0.7, alpha=0.7)
plt.tight_layout()
plt.show()
```

**Yearwise FDI in INDIA**

## ✓ Yearwise Avg\_FDI in INDIA

```
plt.figure(figsize=(8, 5))
sns.lineplot(data=year_avg, marker='o', color='b')
plt.title('Yearwise FDI in INDIA')
plt.xlabel('Avg_FDI')
plt.ylabel('Years')
plt.grid(True)
plt.xticks(rotation=45)

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



## ✓ Top10Sectors FDI in INDIA

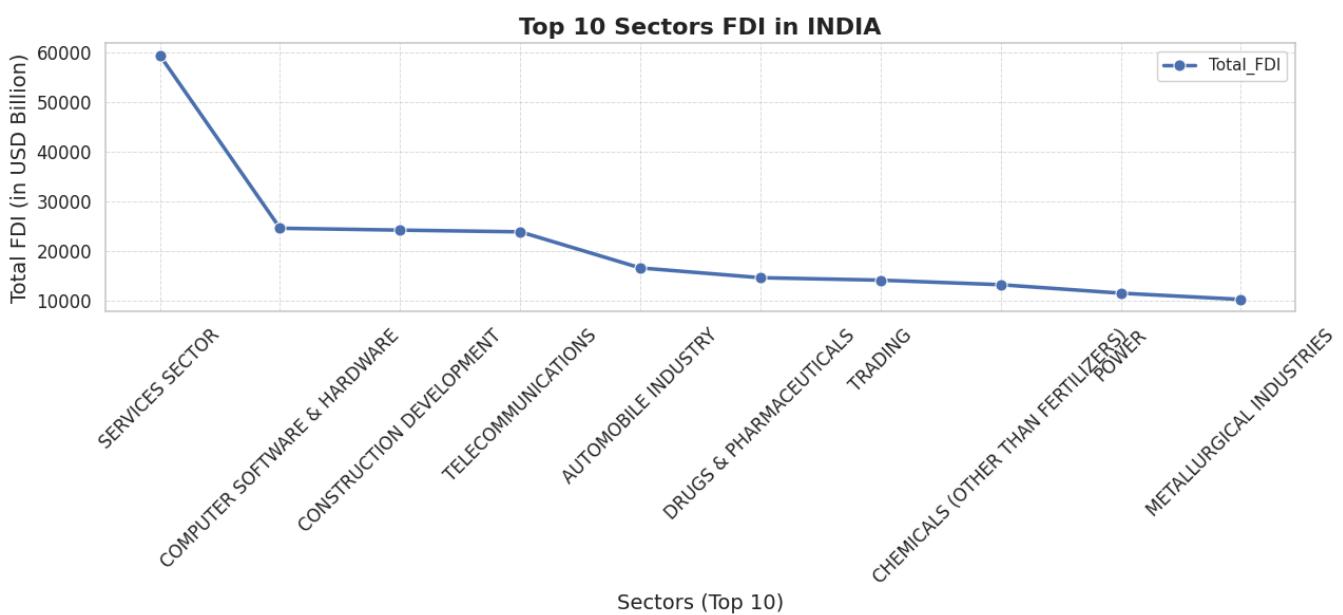
```
sns.set(style="whitegrid")

plt.figure(figsize=(13, 6))

sns.lineplot(data=sec_total_top, marker='o', color='#ff7f7f0e', markersize=8, linewidth=2.5)

plt.title('Top 10 Sectors FDI in INDIA', fontsize=16, fontweight='bold')
plt.xlabel('Sectors (Top 10)', fontsize=14)
plt.ylabel('Total FDI (in USD Billion)', fontsize=14)
plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)

plt.grid(True, linestyle='--', linewidth=0.7, alpha=0.7)
plt.tight_layout()
plt.show()
```



## ▼ Bottom10 Sectors FDI in INDIA

```

sns.set(style="whitegrid")

plt.figure(figsize=(13, 8))

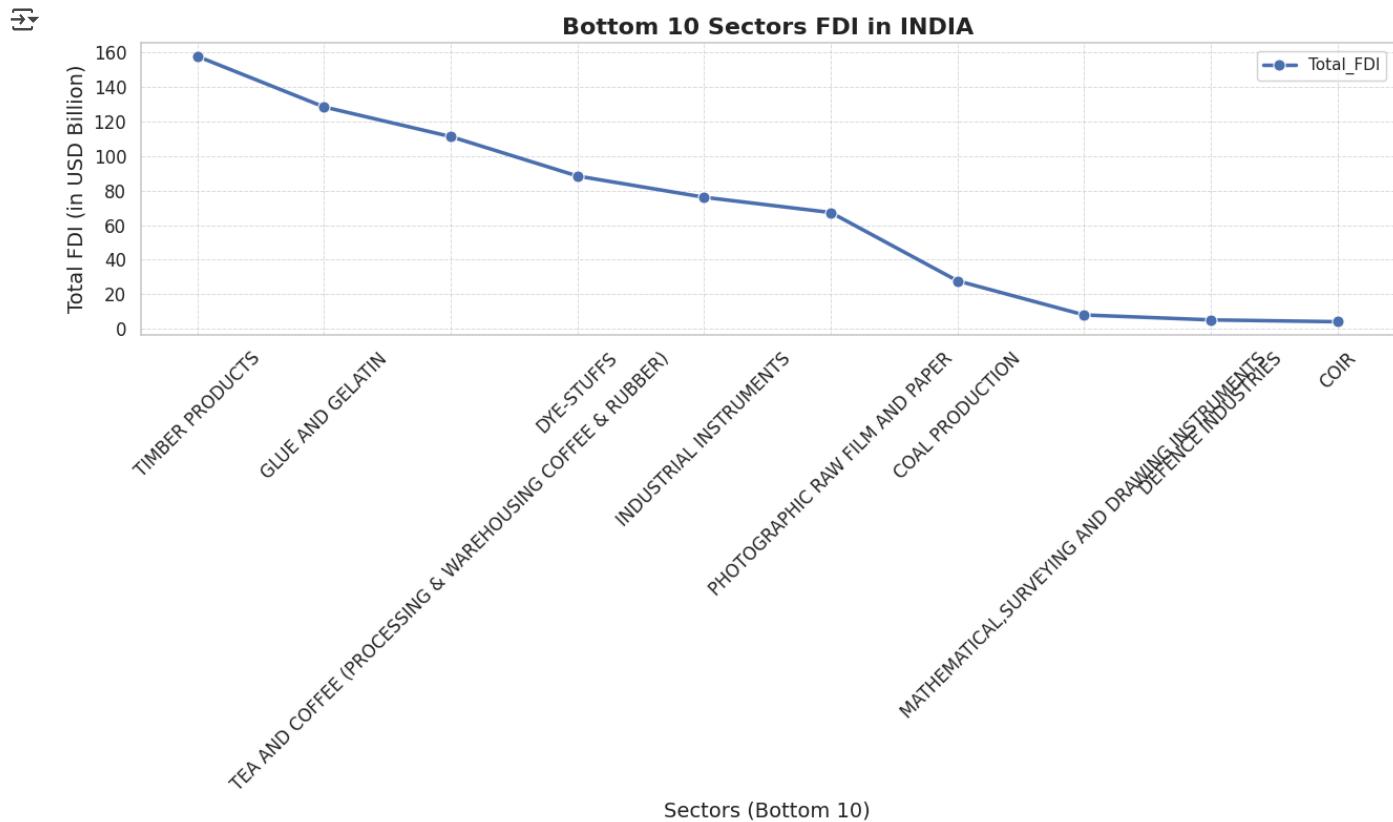
sns.lineplot(data=sec_total_tail, marker='o', color='#2ca02c', markersize=8, linewidth=2.5)

plt.title('Bottom 10 Sectors FDI in INDIA', fontsize=16, fontweight='bold')
plt.xlabel('Sectors (Bottom 10)', fontsize=14)
plt.ylabel('Total FDI (in USD Billion)', fontsize=14)

plt.xticks(rotation=45, fontsize=12)
plt.yticks(fontsize=12)

plt.grid(True, linestyle='--', linewidth=0.7, alpha=0.7)
plt.tight_layout()
plt.show()

```



```
df3.head()
```

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
METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94	419.88	1098.14	1786.14	1466.23	567.63	359.3
MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16	174.40	79.51	142.65	57.89	12.73	684.3
POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68	907.66	1271.79	1271.77	1652.38	535.68	1066.08	707.0
NON-CONVENTIONAL ENERGY	0.00	0.00	1.70	4.14	1.27	1.35	2.44	58.82	125.88	622.52	214.40	452.17	1106.52	414.25	615.6
COAL PRODUCTION	0.00	0.00	0.00	0.04	0.00	9.14	1.30	14.08	0.22	0.00	0.00	0.00	0.00	2.96	0.0

Next steps: [Generate code with df3](#) [View recommended plots](#) [New interactive sheet](#)

```
df_long = pd.melt(df, id_vars=['Sector'], var_name='Year', value_name='FDI')
df_long
```

	Sector	Year	FDI	
0	METALLURGICAL INDUSTRIES	2000-01	22.69	
1	MINING	2000-01	1.32	
2	POWER	2000-01	89.42	
3	NON-CONVENTIONAL ENERGY	2000-01	0.00	
4	COAL PRODUCTION	2000-01	0.00	
...	...	...	...	
1066	PRINTING OF BOOKS (INCLUDING LITHO PRINTING IN...	2016-17	53.17	
1067	COIR	2016-17	0.00	
1068	CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES	2016-17	1860.73	
1069	CONSTRUCTION DEVELOPMENT: Townships, housing, ...	2016-17	105.14	
1070	MISCELLANEOUS INDUSTRIES	2016-17	296.40	

1071 rows × 3 columns

Next steps: [Generate code with df\\_long](#) [View recommended plots](#) [New interactive sheet](#)

```
df_long['Growth Rate'] = df_long.groupby('Sector')['FDI'].pct_change()
df_long['Growth Rate']
```

	Growth Rate
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
...	...
1066	-0.567055
1067	NaN
1068	-0.587486
1069	-0.065837
1070	-0.556798

1071 rows × 1 columns

**dtype:** float64

```
df_long['Growth Rate'] = df_long['Growth Rate'].fillna(0)
df_long['Growth Rate']
```

	Growth Rate
0	0.000000
1	0.000000
2	0.000000
3	0.000000
4	0.000000
...	...
1066	-0.567055
1067	0.000000
1068	-0.587486
1069	-0.065837
1070	-0.556798

1071 rows × 1 columns

**dtype:** float64

```
df_long['Moving Average'] = df_long.groupby('Sector')['FDI'].transform(lambda x: x.rolling(window=3, min_periods=1).mean())
df_long['Moving Average']
```

**Moving Average**

0	22.690000
1	1.320000
2	89.420000
3	0.000000
4	0.000000
...	...
1066	82.853333
1067	0.453333
1068	2413.896667
1069	328.943333
1070	577.016667

1071 rows × 1 columns

**dtype:** float64

```
def identify_trend(series):
    return 'Uptrend' if series.mean() > 0 else 'Downtrend'
```

```
trend_df = df_long.groupby('Sector').apply(lambda x: identify_trend(x['Growth Rate'])).reset_index(name='Trend')
trend_df
```

**Sector Trend**

	Sector	Trend	
0	AGRICULTURAL MACHINERY	Uptrend	
1	AGRICULTURE SERVICES	Uptrend	
2	AIR TRANSPORT (INCLUDING AIR FREIGHT)	Uptrend	
3	AUTOMOBILE INDUSTRY	Uptrend	
4	BOILERS AND STEAM GENERATING PLANTS	Uptrend	
...	...	...	
58	TELECOMMUNICATIONS	Uptrend	
59	TEXTILES (INCLUDING DYED,PRINTED)	Uptrend	
60	TIMBER PRODUCTS	Uptrend	
61	TRADING	Uptrend	
62	VEGETABLE OILS AND VANASPATI	Uptrend	

63 rows × 2 columns

Next steps: [Generate code with trend\\_df](#) [View recommended plots](#) [New interactive sheet](#)

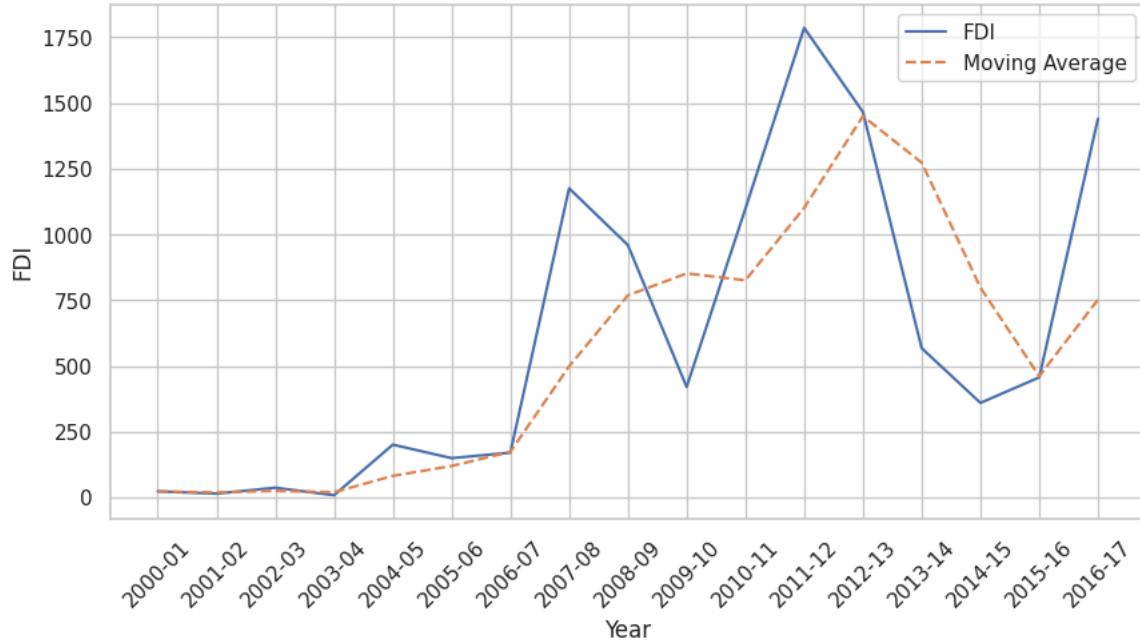
```
import matplotlib.pyplot as plt

sectors = df_long['Sector'].unique()

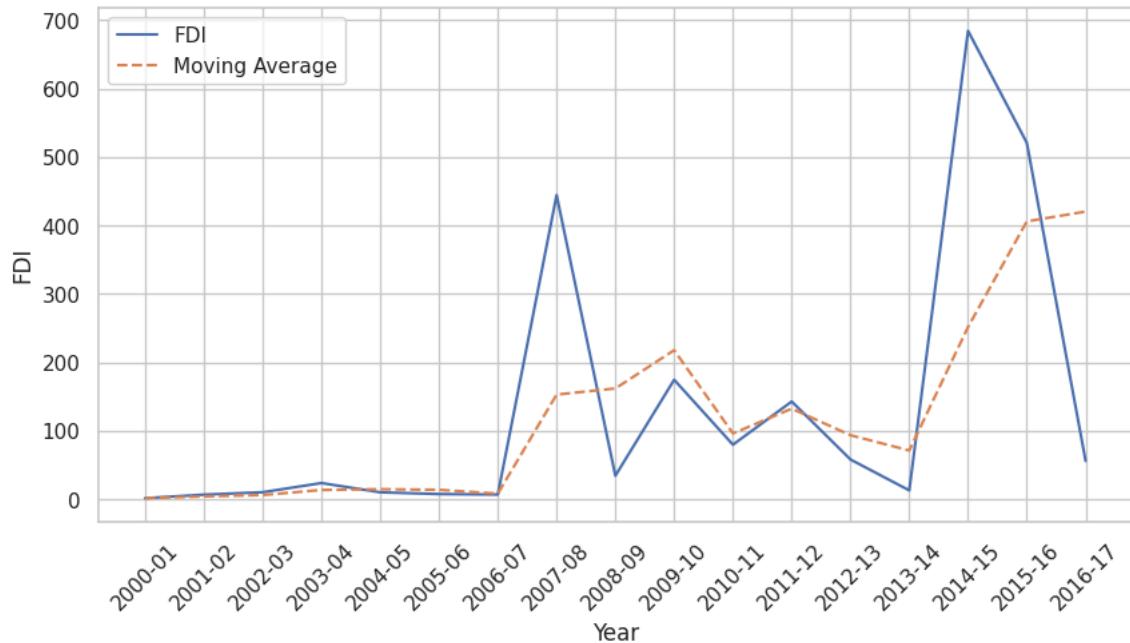
for sector in sectors:
    sector_data = df_long[df_long['Sector'] == sector]
    plt.figure(figsize=(10, 5))
    plt.plot(sector_data['Year'], sector_data['FDI'], label='FDI')
    plt.plot(sector_data['Year'], sector_data['Moving Average'], '--', label='Moving Average')
    plt.title(f'FDI Trends for {sector}')
    plt.xlabel('Year')
    plt.xticks(rotation=45)
    plt.ylabel('FDI')
    plt.legend()
    plt.grid(True)
    plt.show()
```



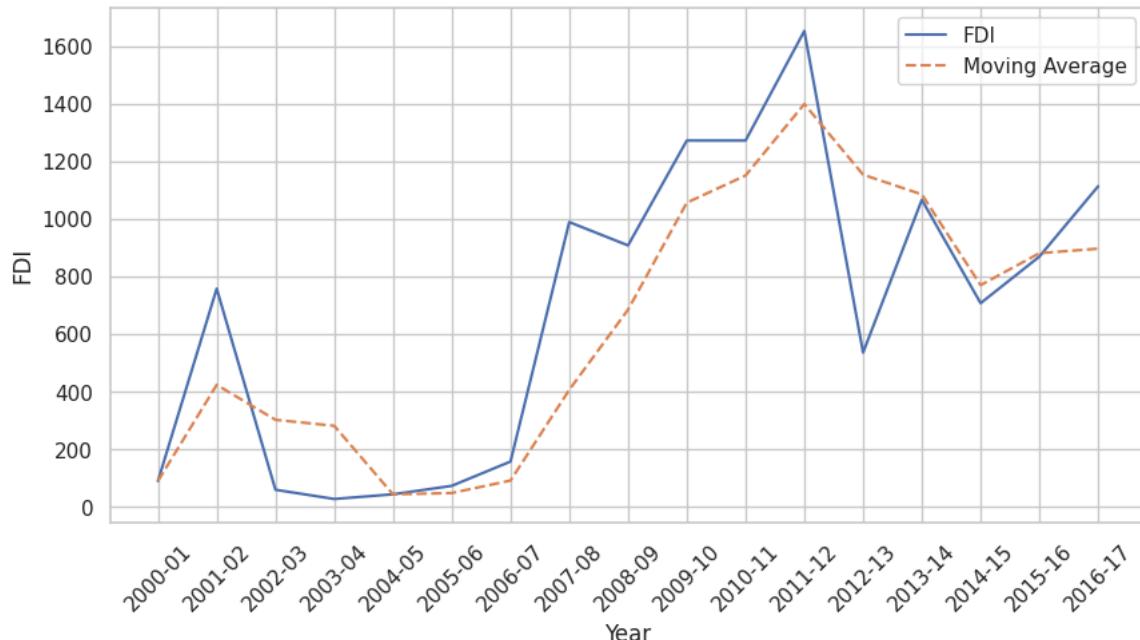
### FDI Trends for METALLURGICAL INDUSTRIES

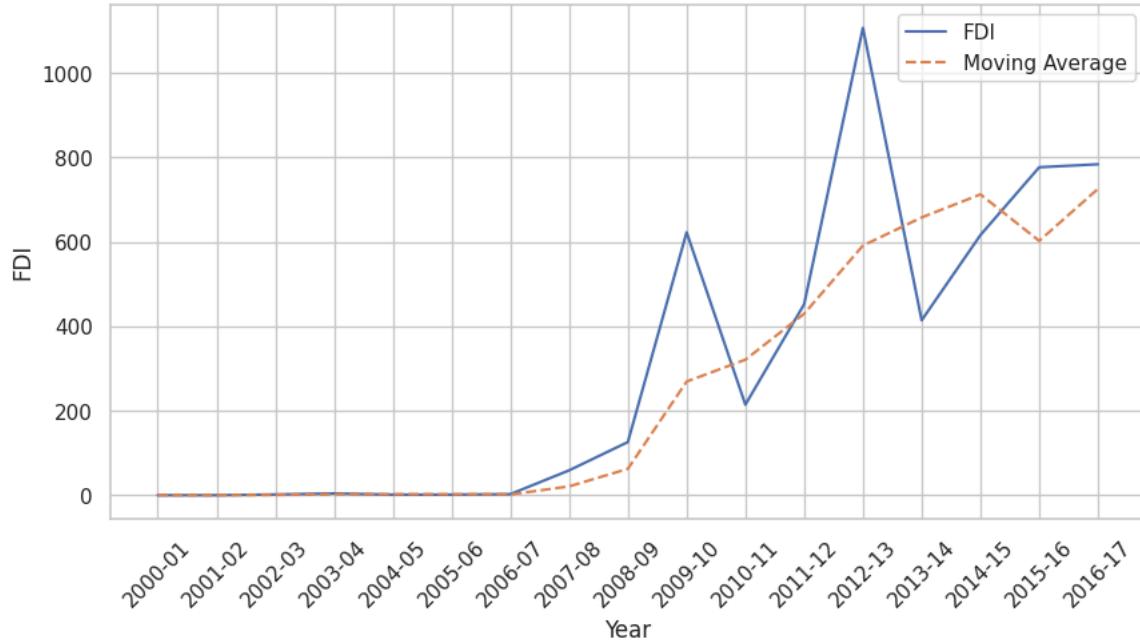


### FDI Trends for MINING

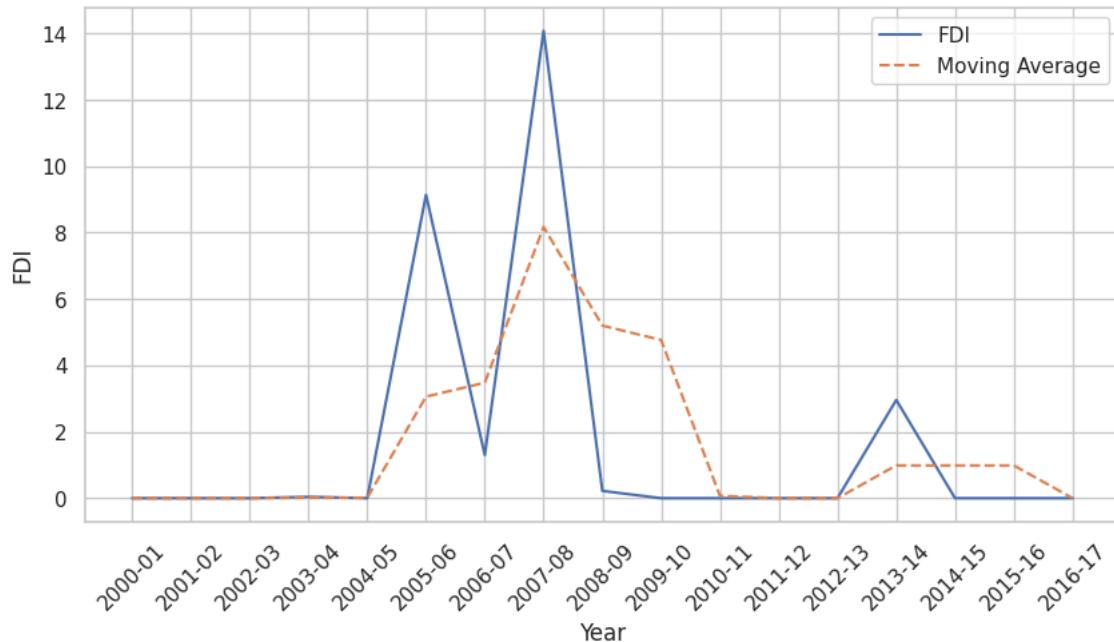


### FDI Trends for POWER

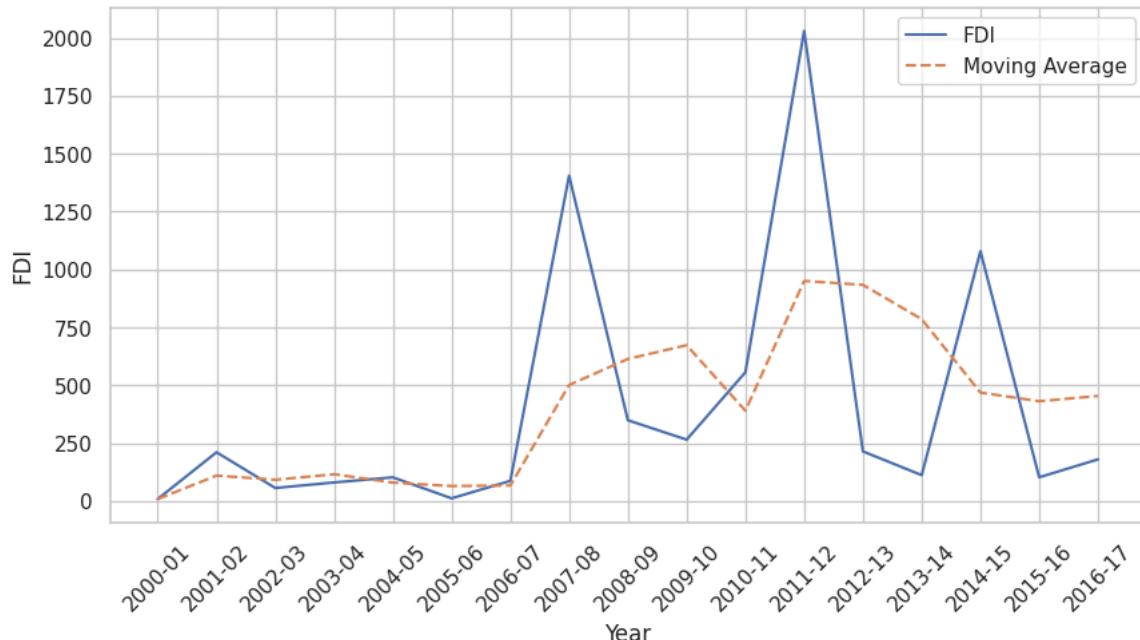




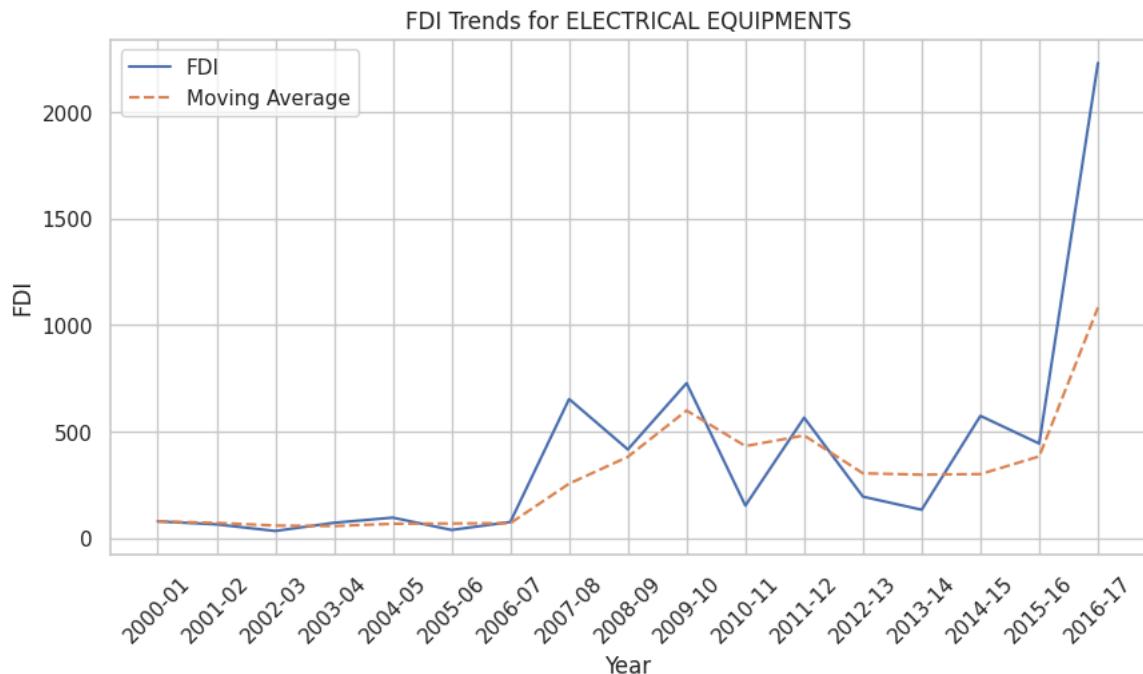
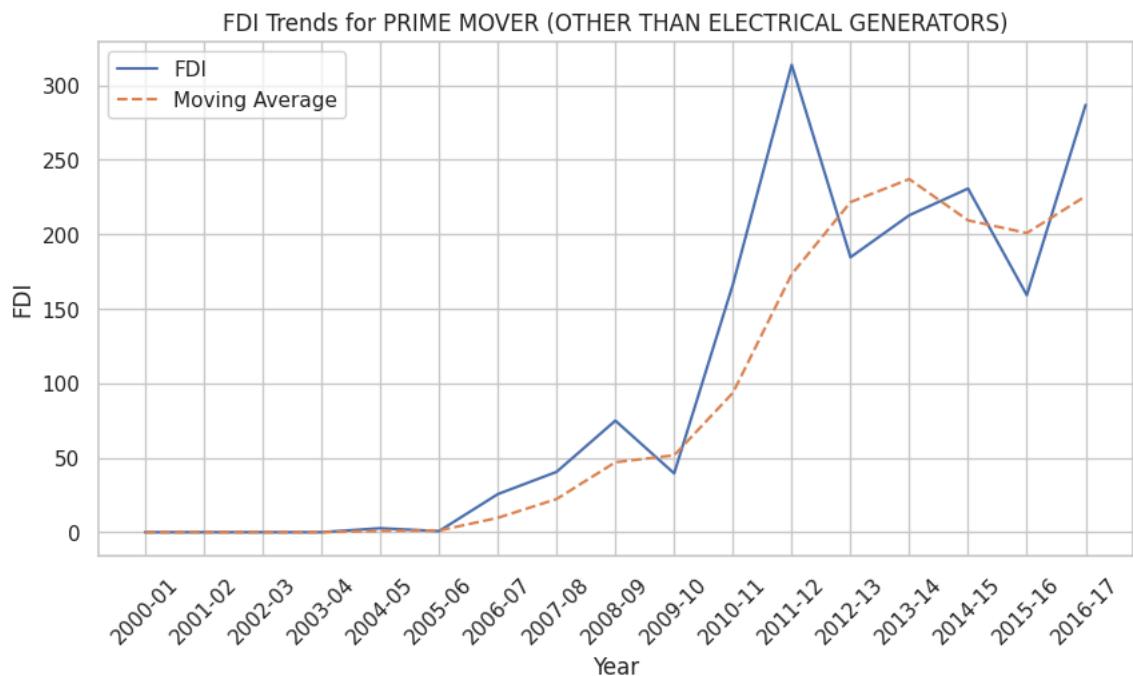
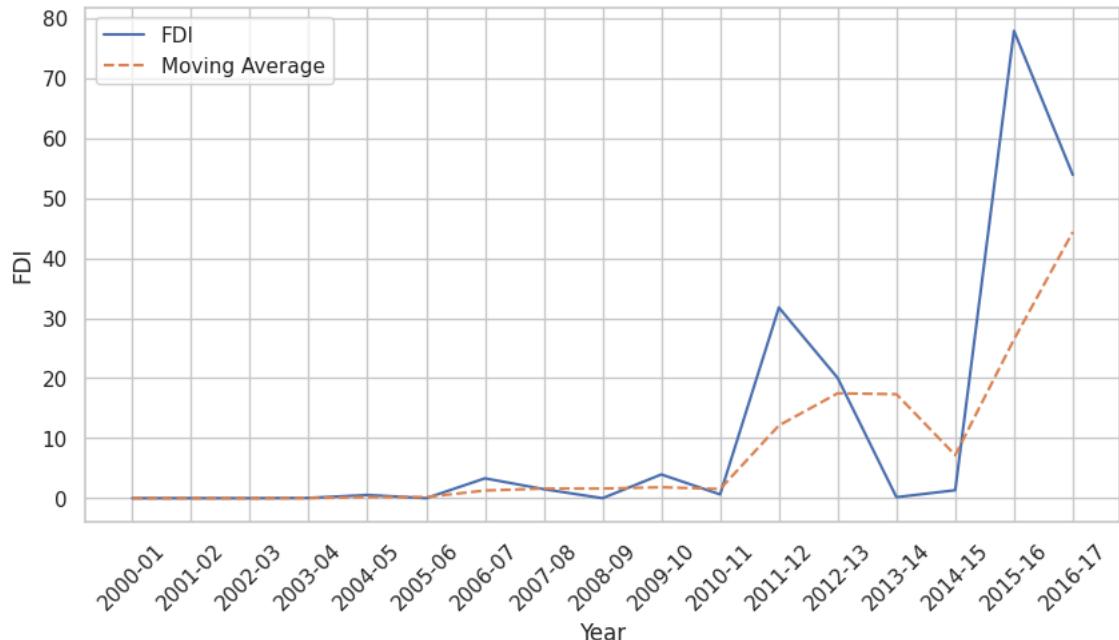
FDI Trends for COAL PRODUCTION

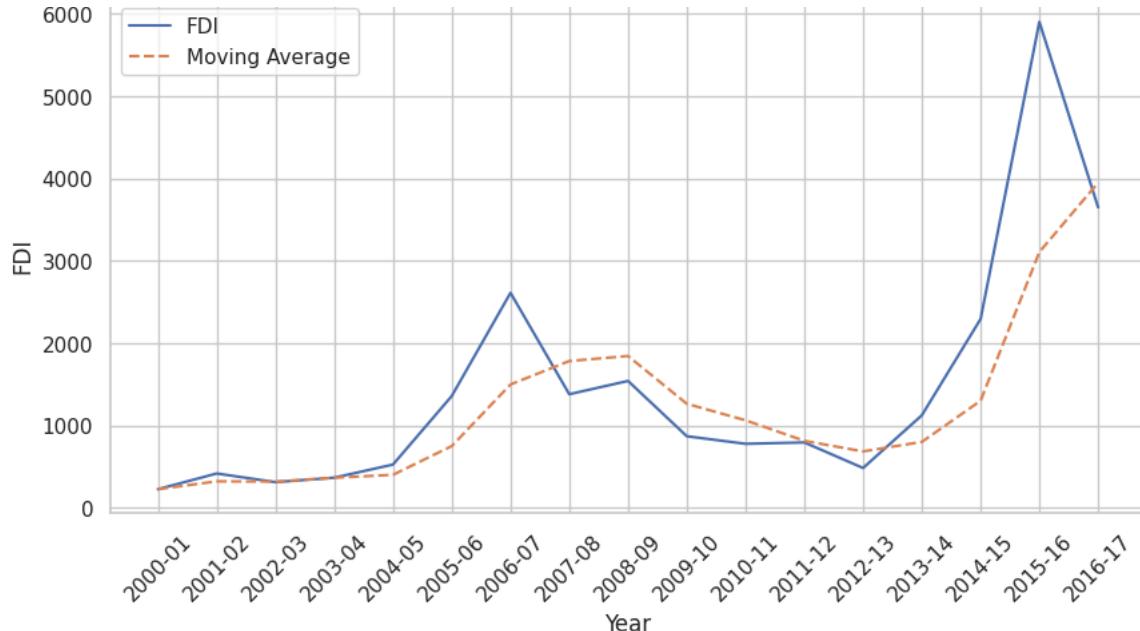


FDI Trends for PETROLEUM &amp; NATURAL GAS

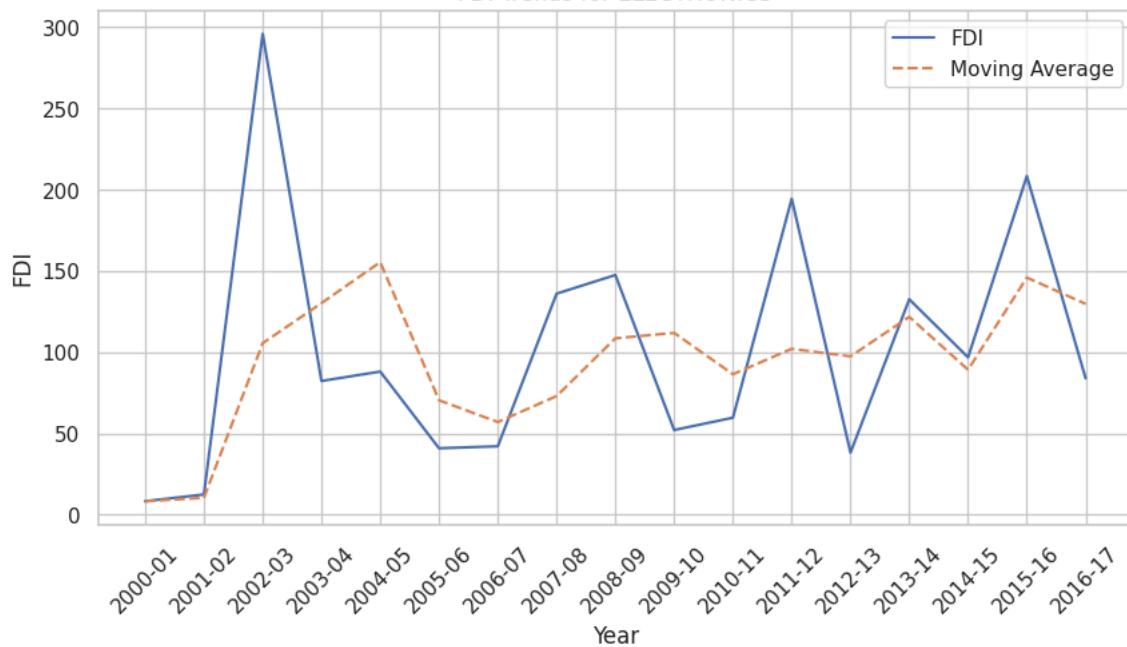


FDI Trends for BOILERS AND STEAM GENERATING PLANTS

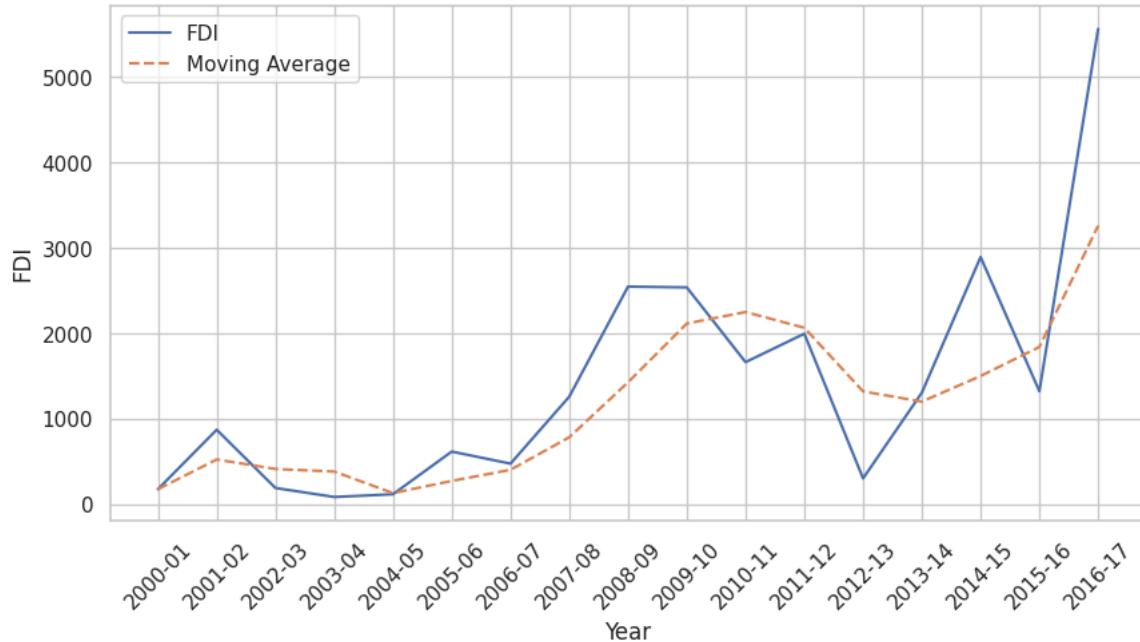




FDI Trends for ELECTRONICS

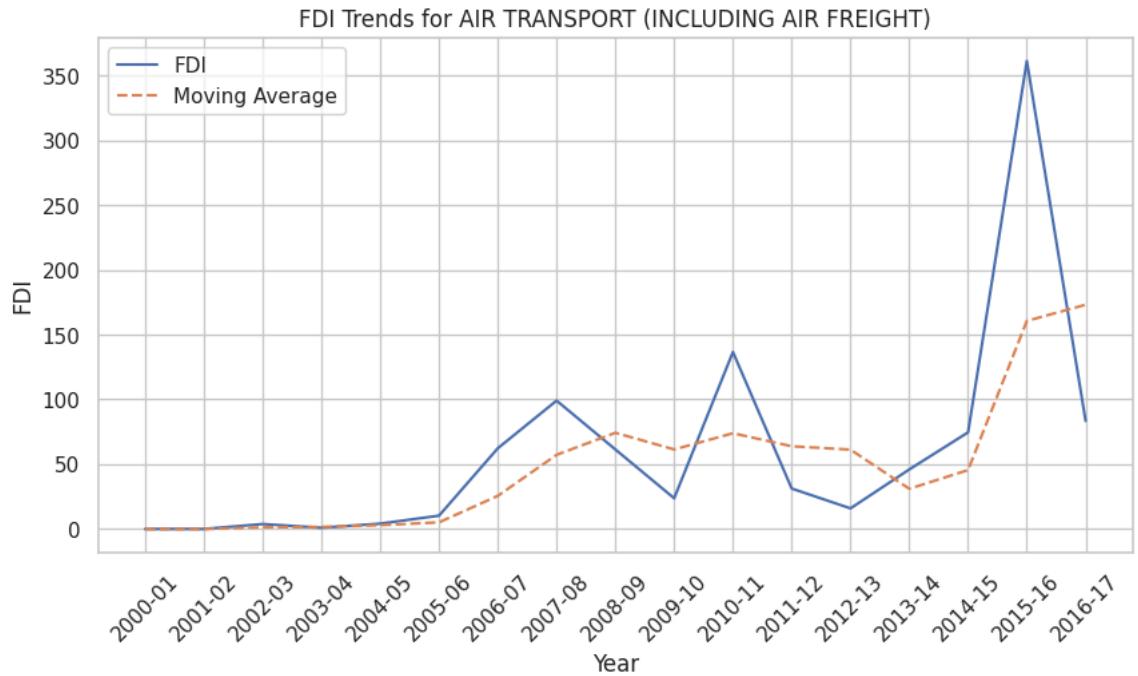
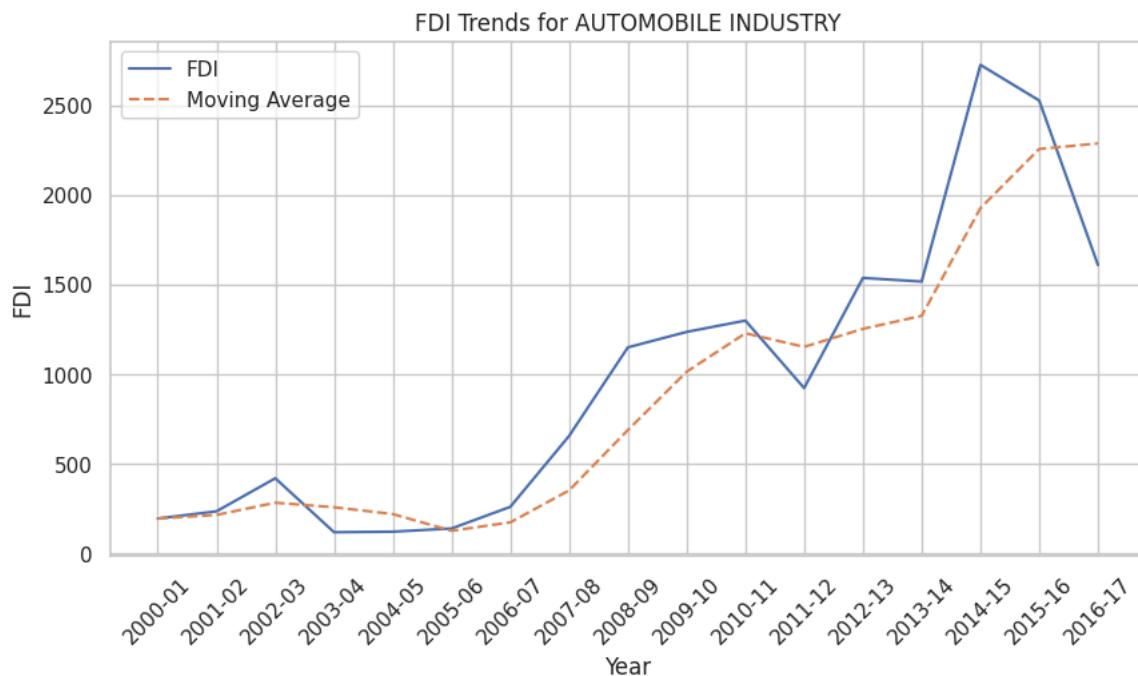
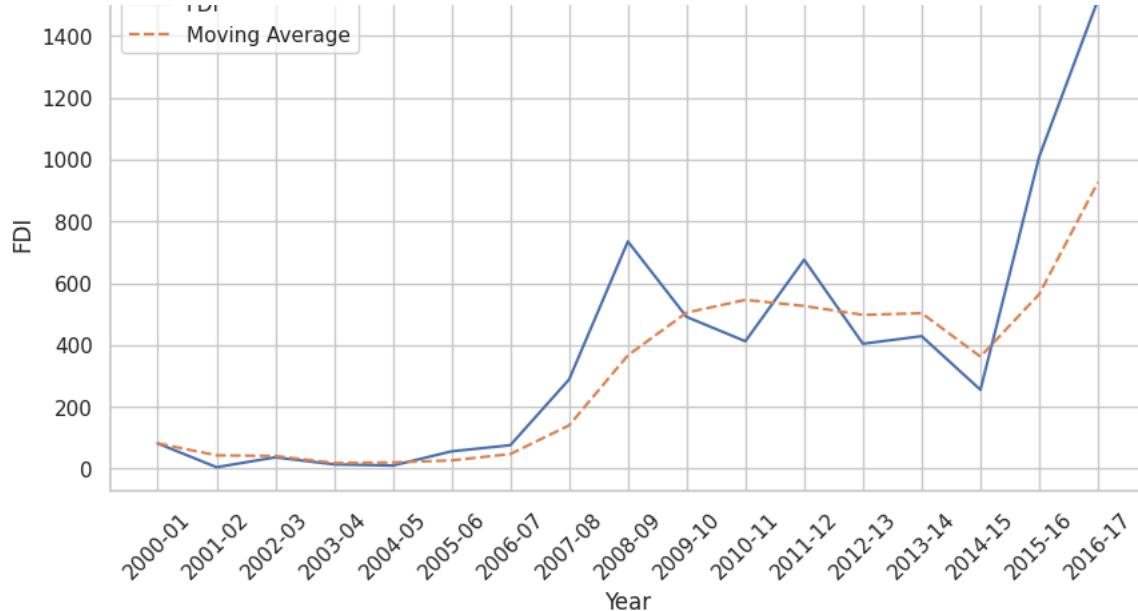


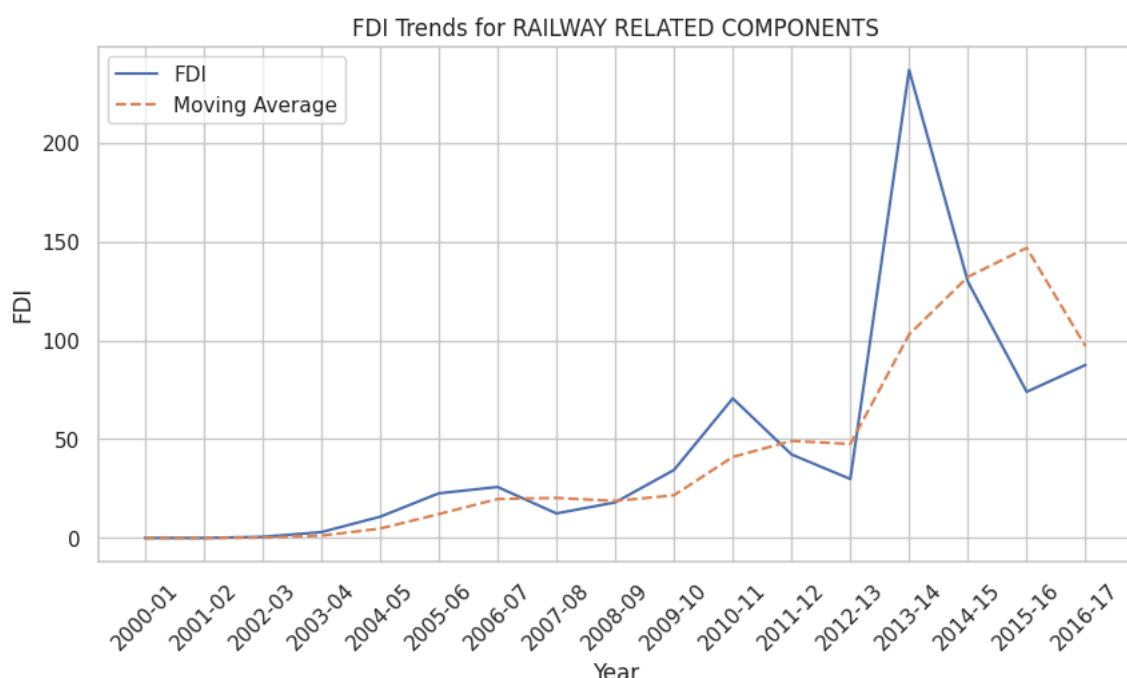
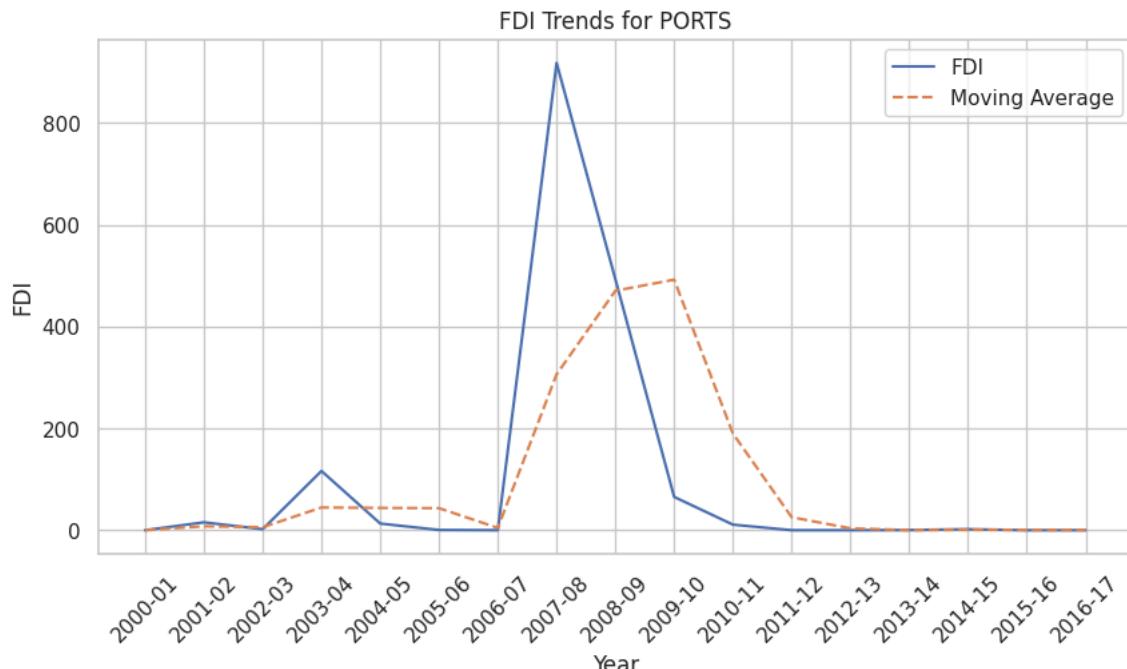
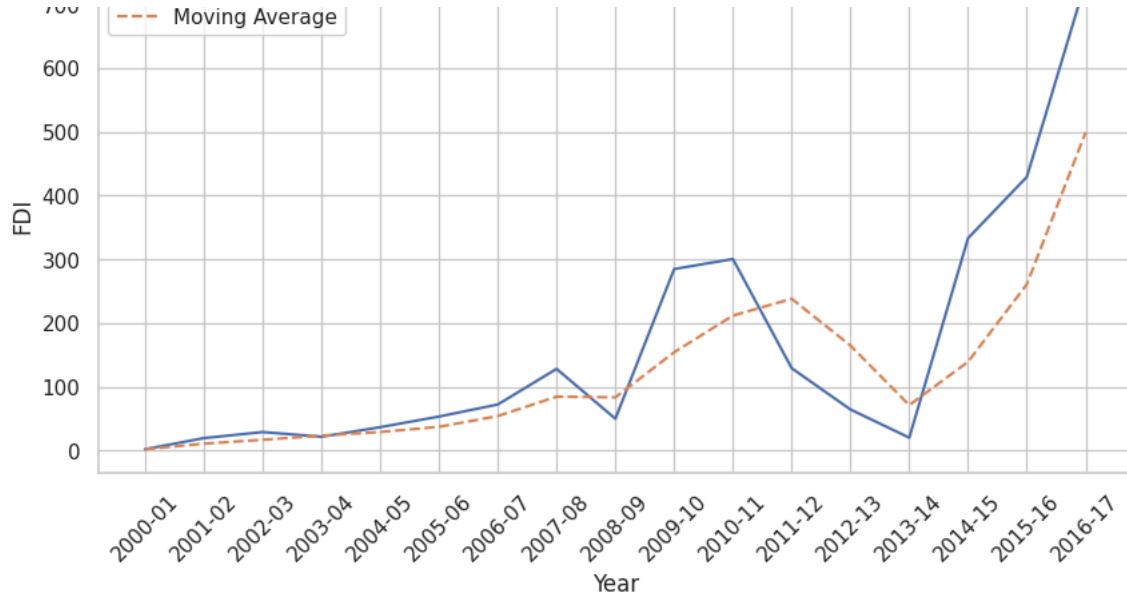
FDI Trends for TELECOMMUNICATIONS

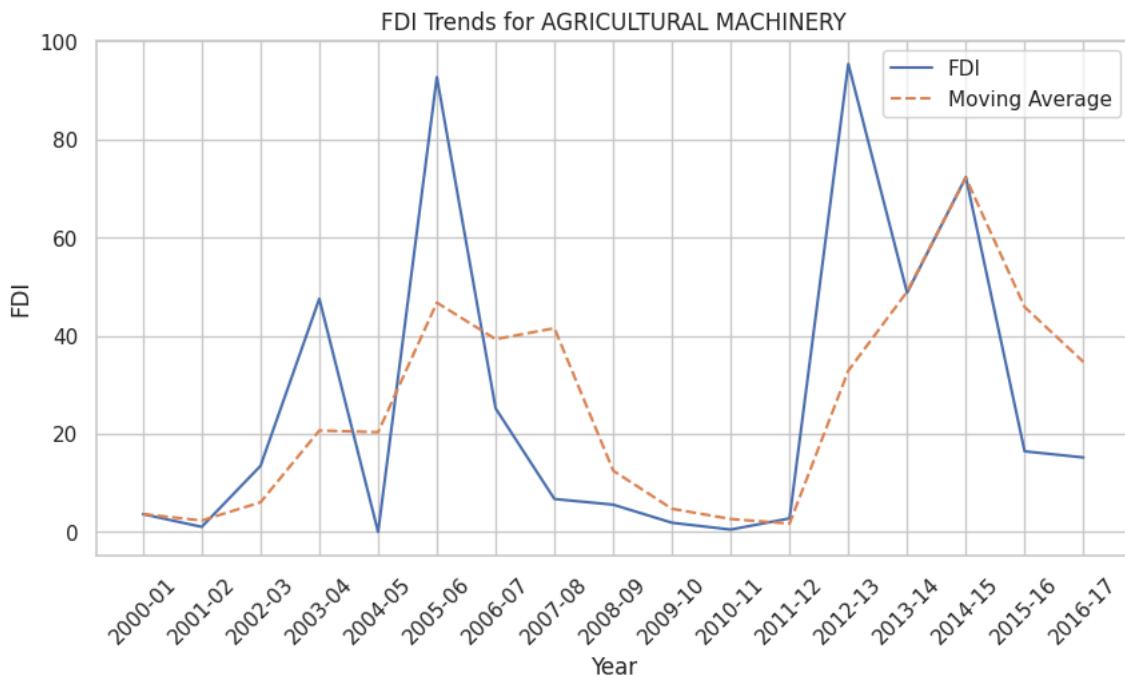
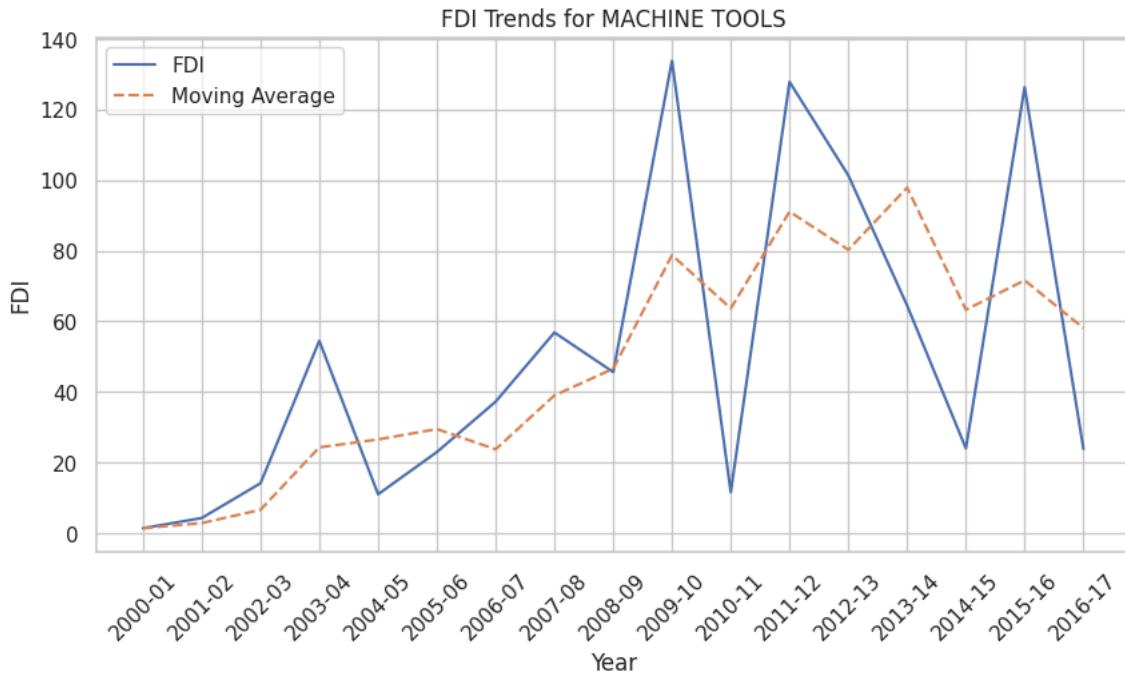
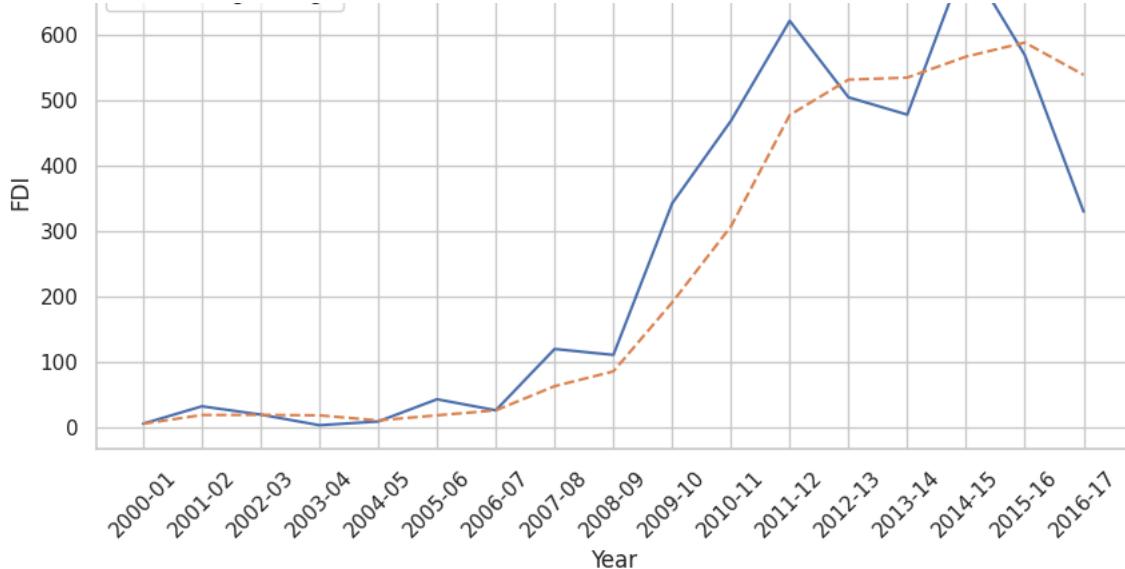


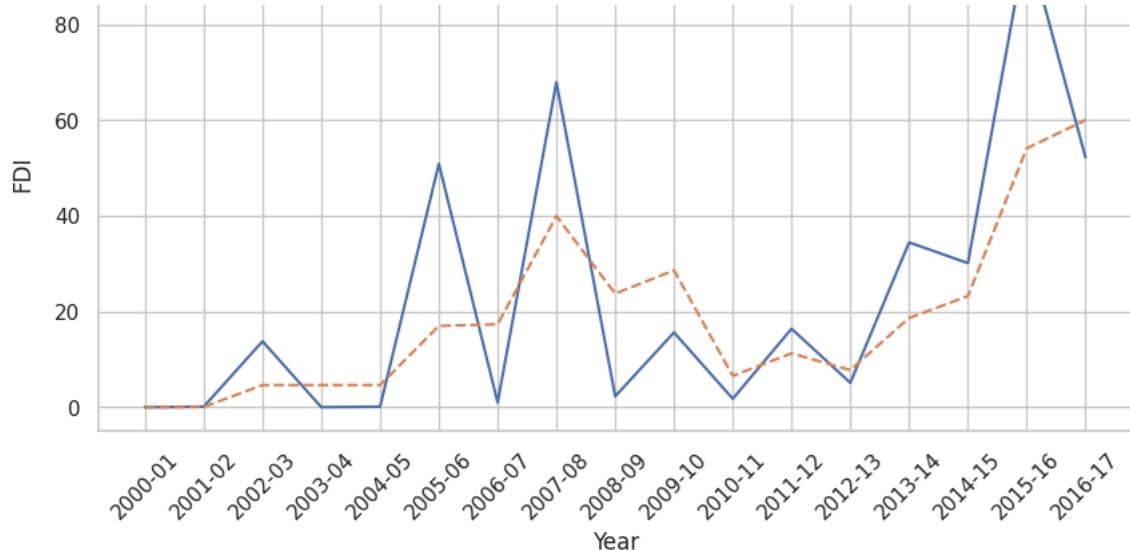
FDI Trends for INFORMATION &amp; BROADCASTING (INCLUDING PRINT MEDIA)



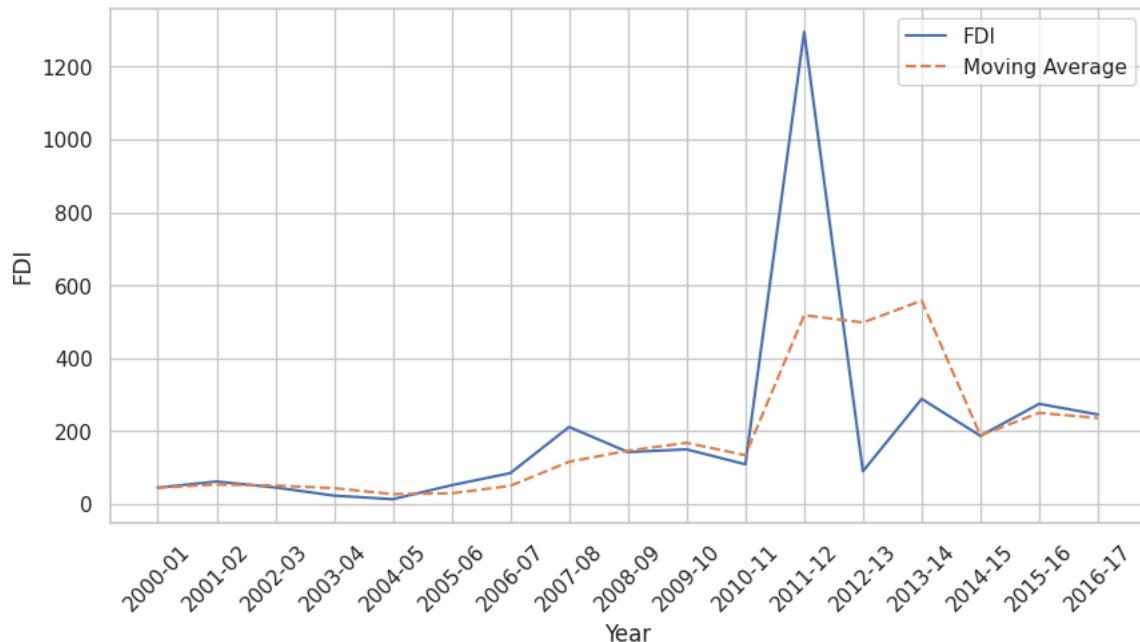




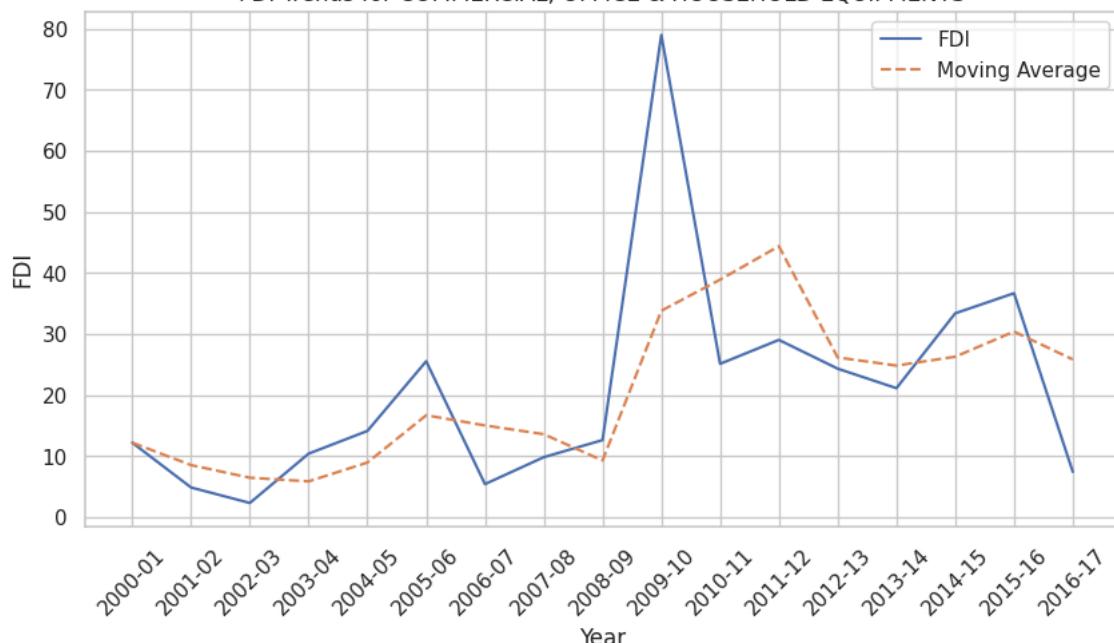




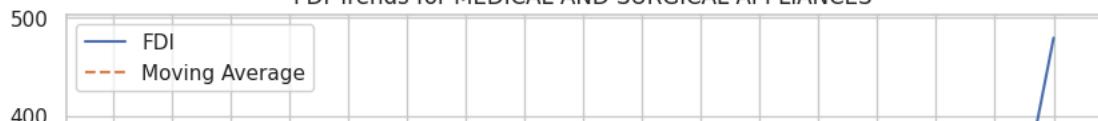
FDI Trends for MISCELLANEOUS MECHANICAL &amp; ENGINEERING INDUSTRIES

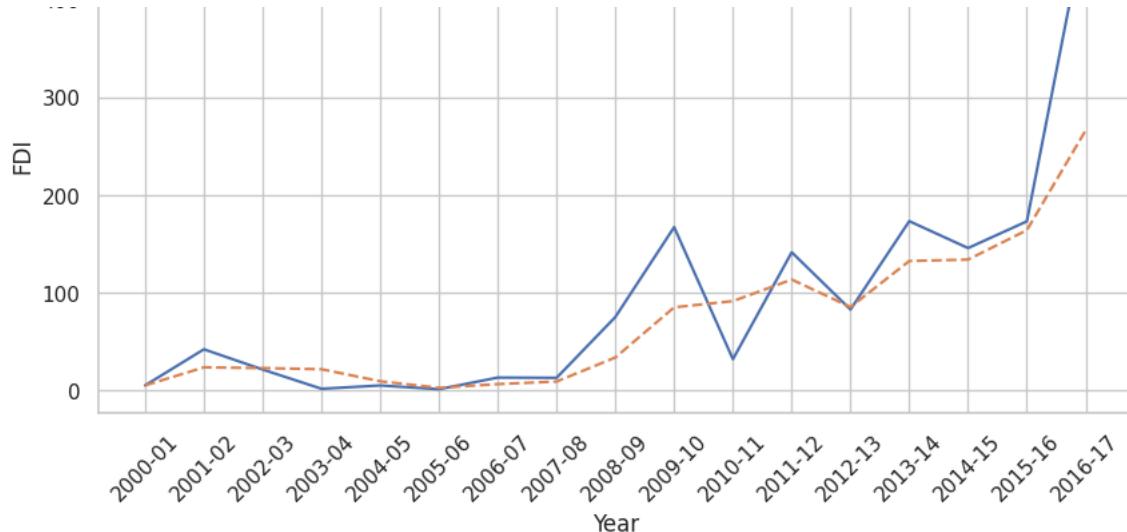


FDI Trends for COMMERCIAL, OFFICE &amp; HOUSEHOLD EQUIPMENTS

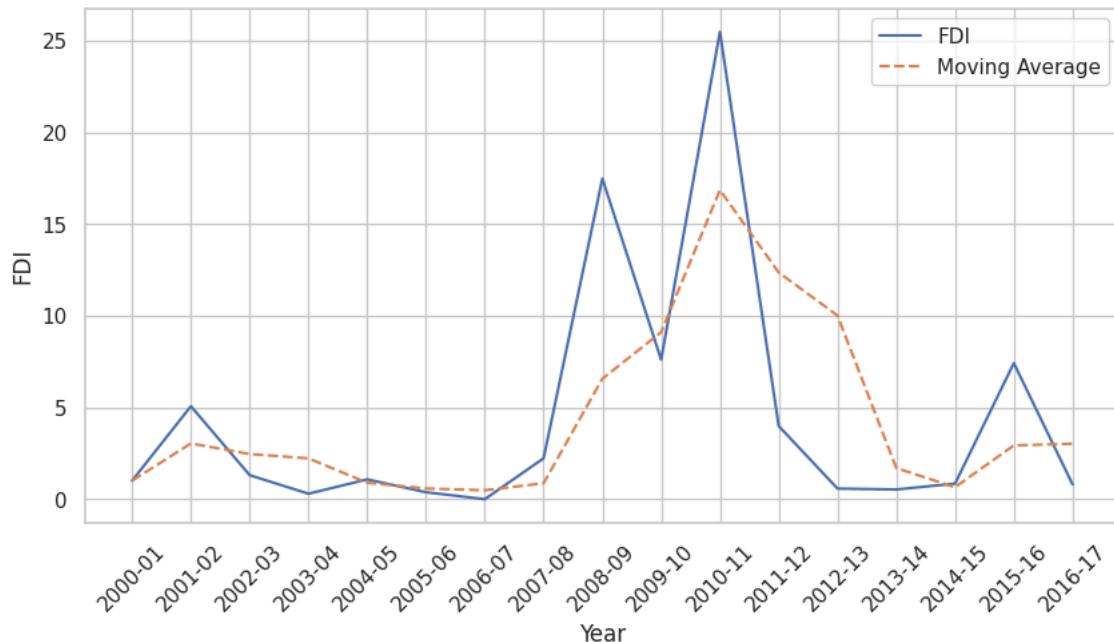


FDI Trends for MEDICAL AND SURGICAL APPLIANCES

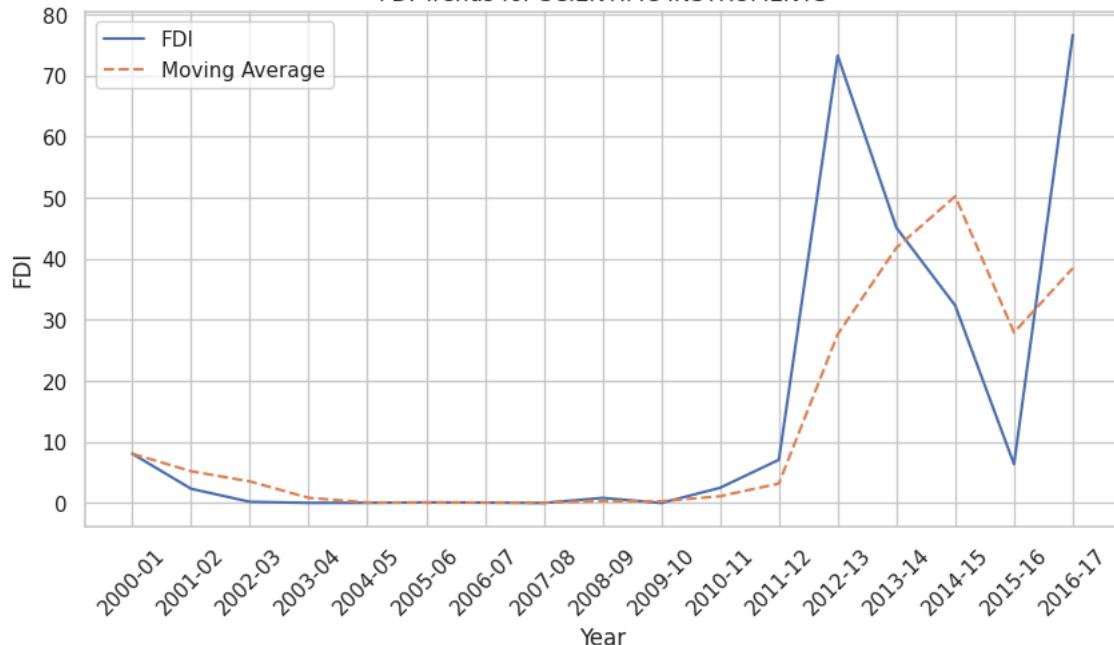




FDI Trends for INDUSTRIAL INSTRUMENTS

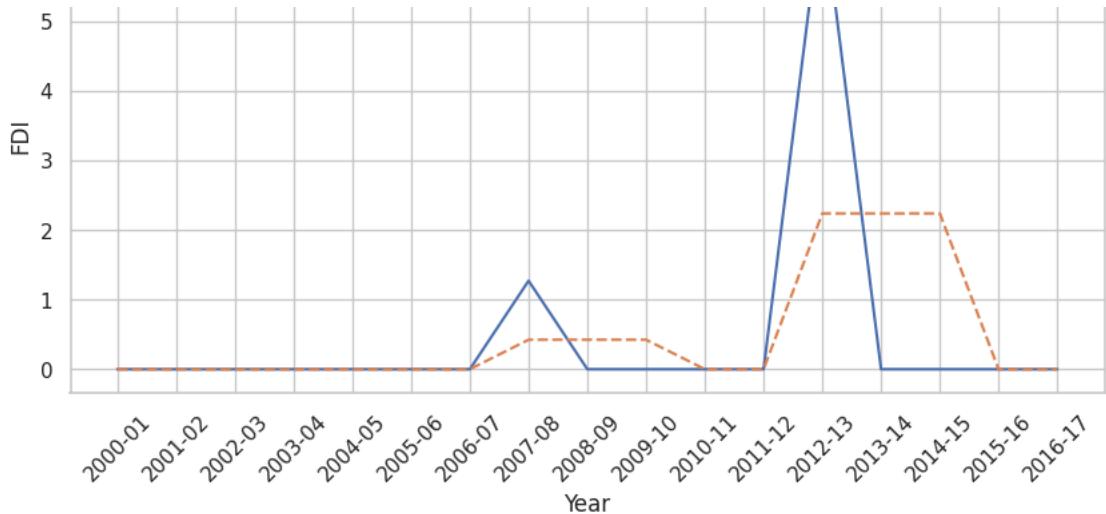


FDI Trends for SCIENTIFIC INSTRUMENTS

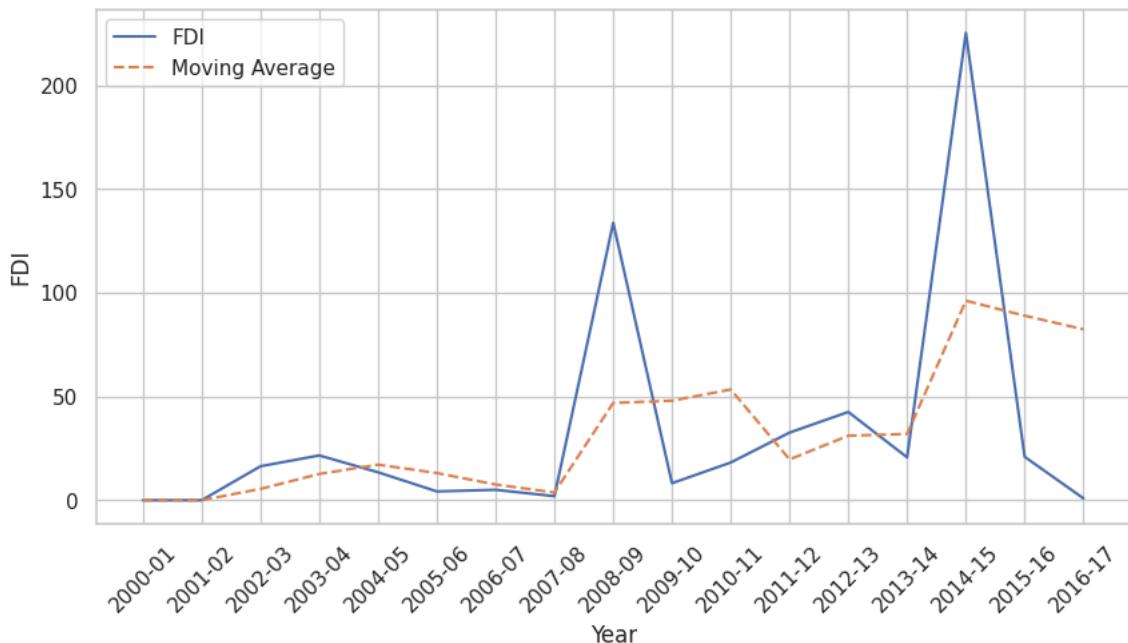


FDI Trends for MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS

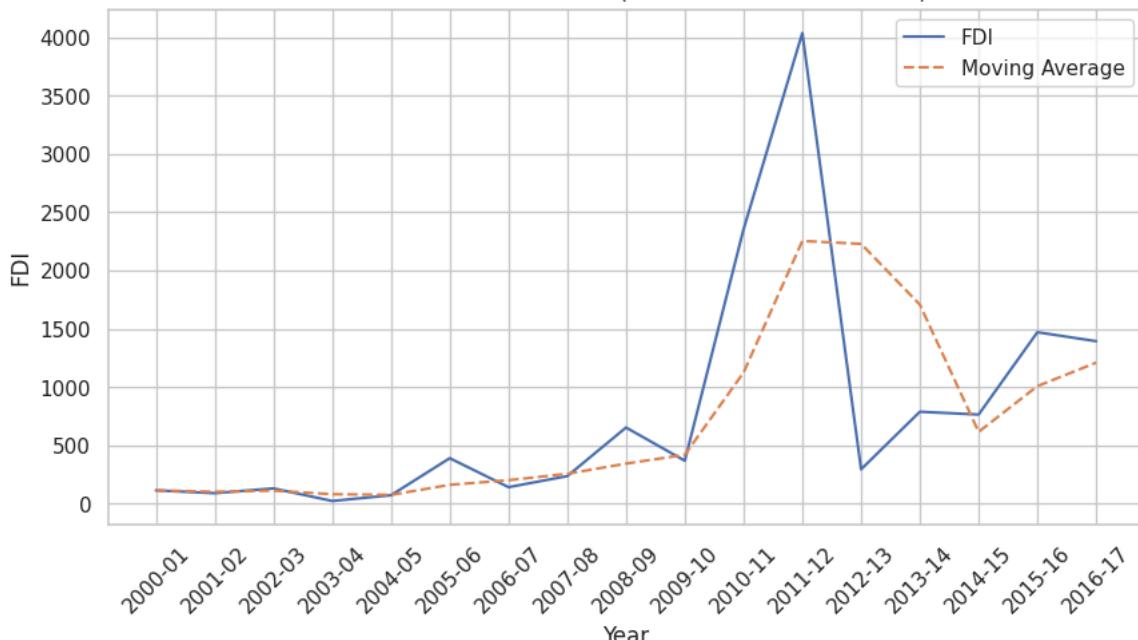




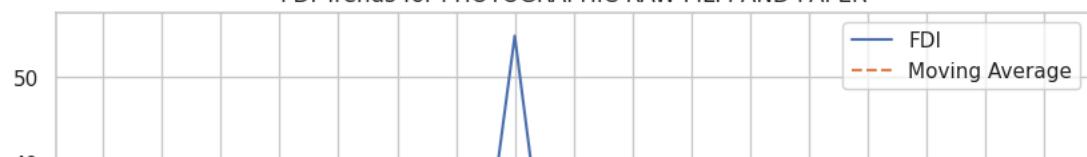
FDI Trends for FERTILIZERS

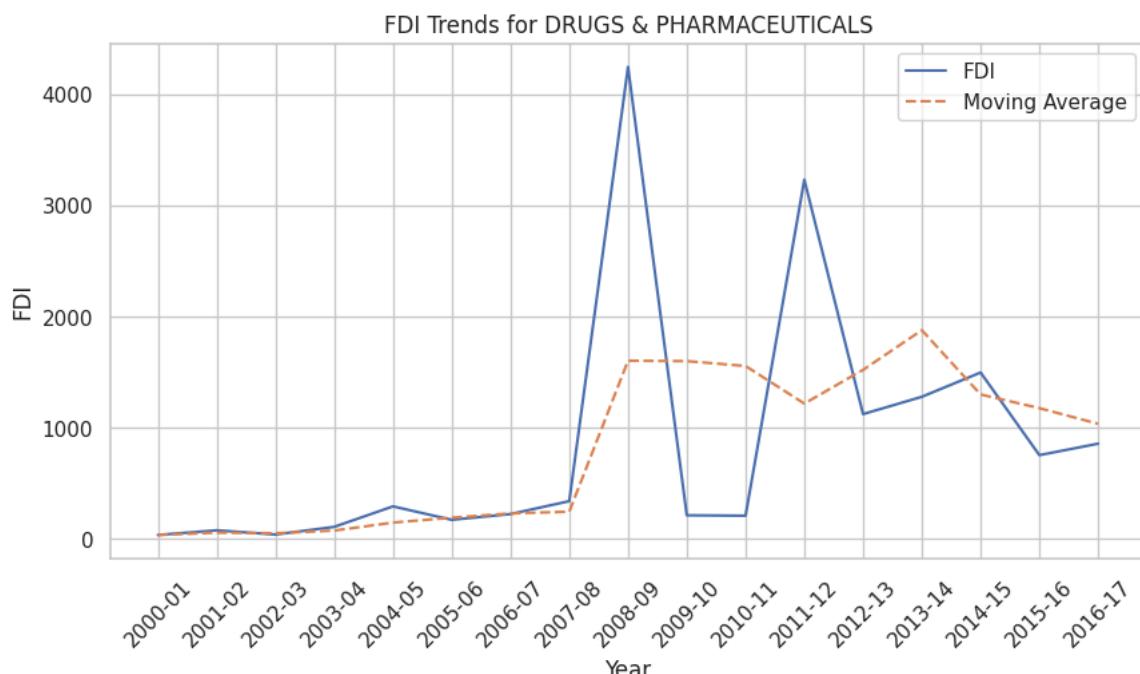
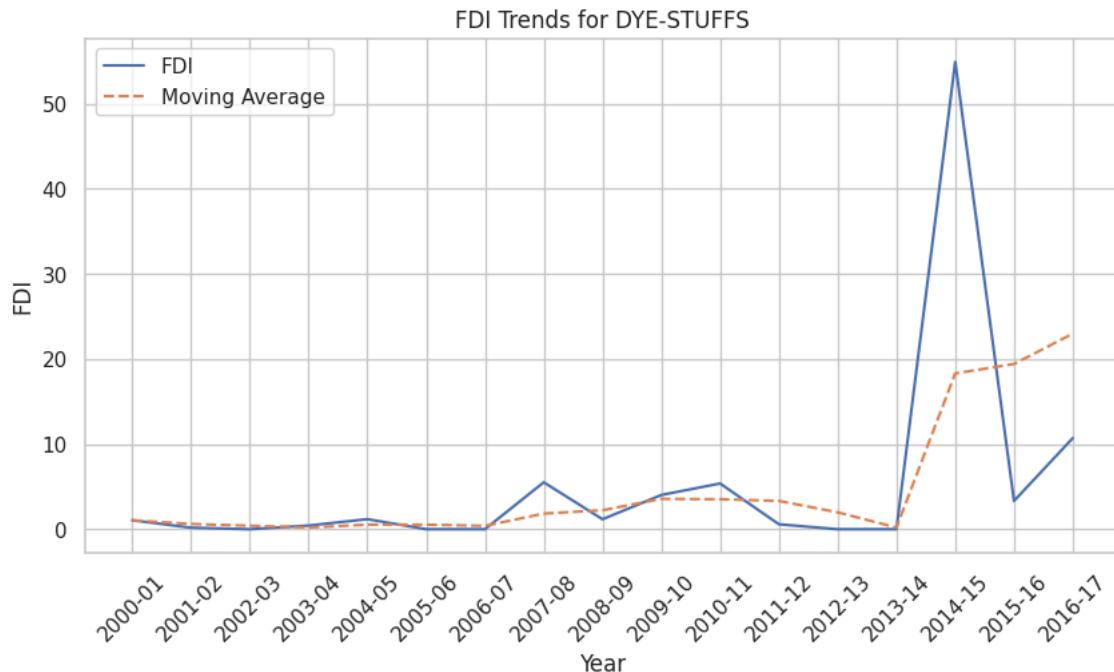
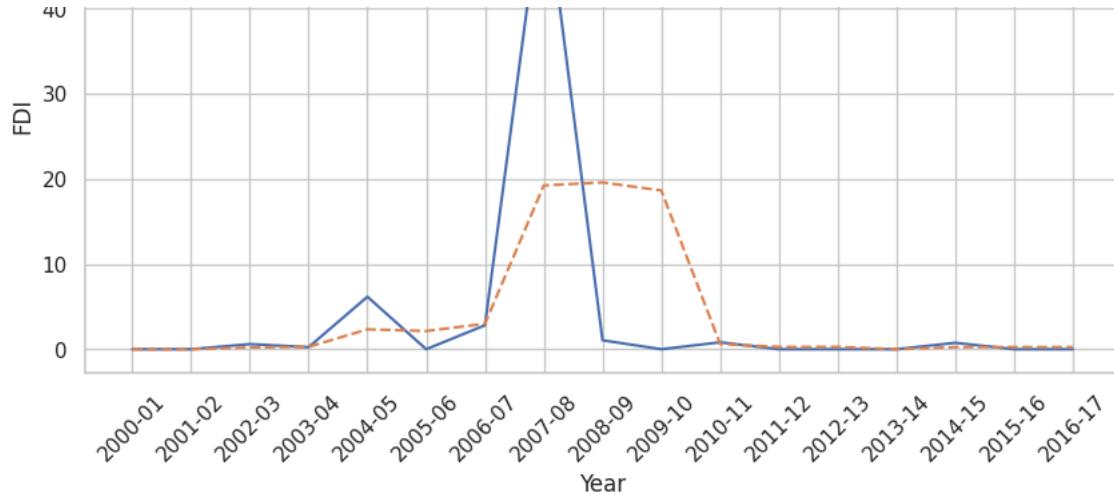


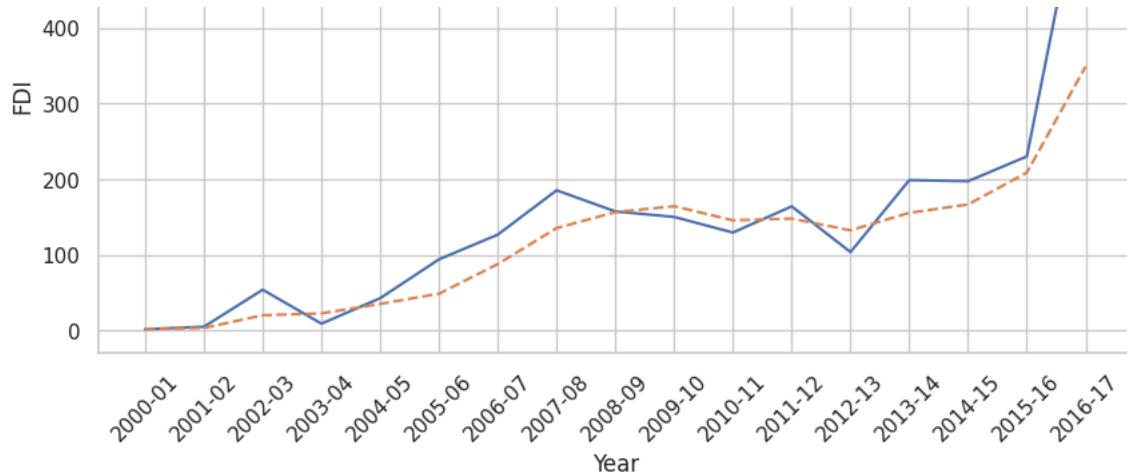
FDI Trends for CHEMICALS (OTHER THAN FERTILIZERS)



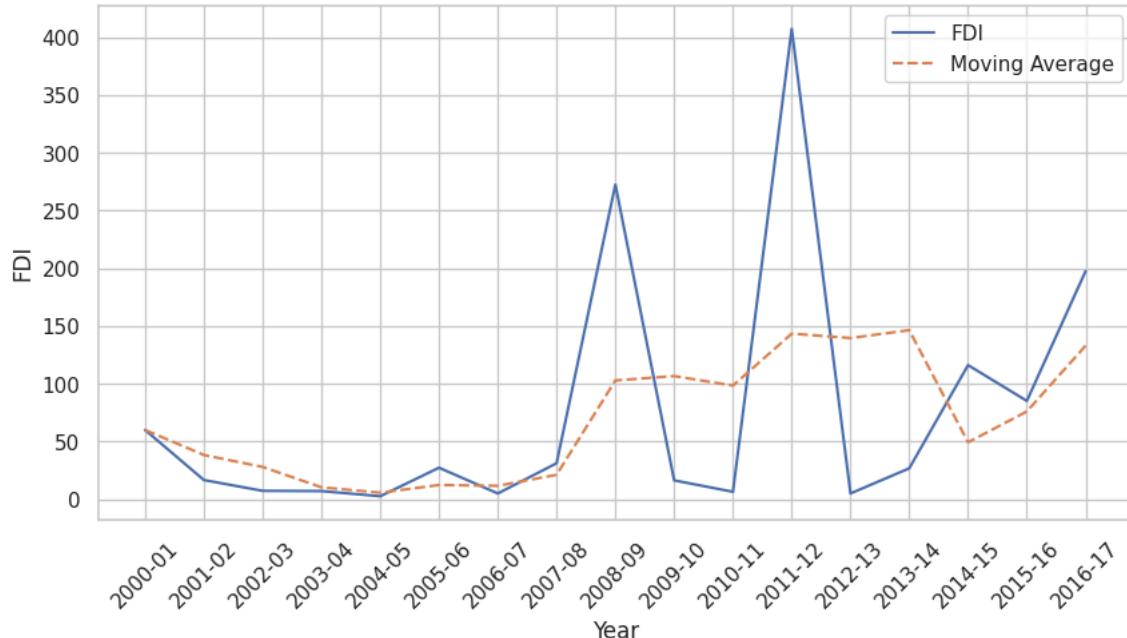
FDI Trends for PHOTOGRAPHIC RAW FILM AND PAPER



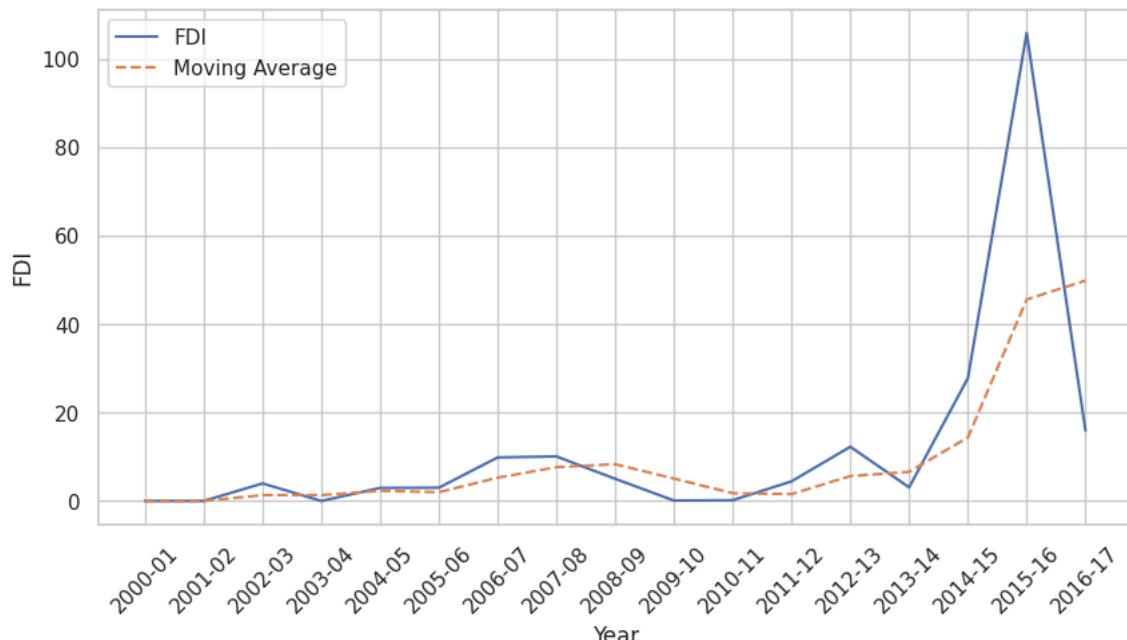




FDI Trends for PAPER AND PULP (INCLUDING PAPER PRODUCTS)

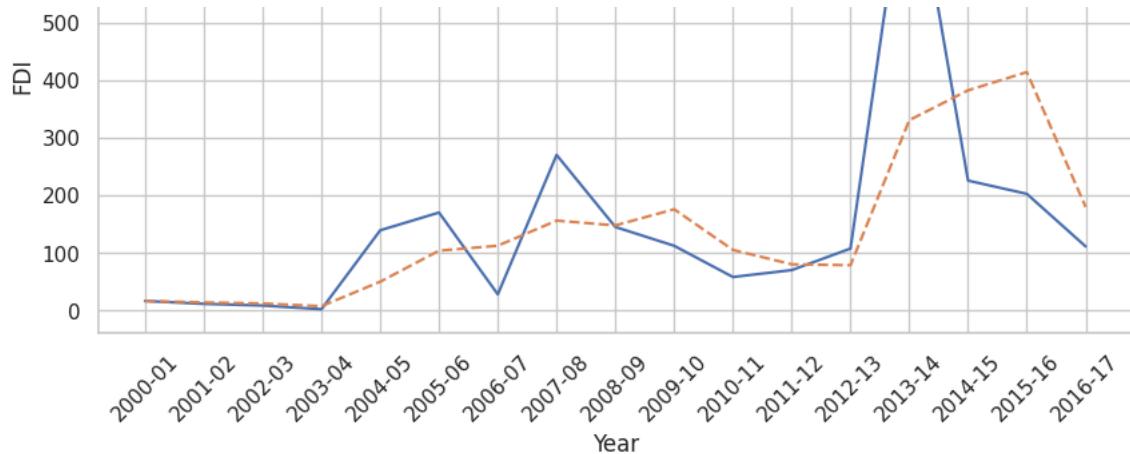


FDI Trends for SUGAR

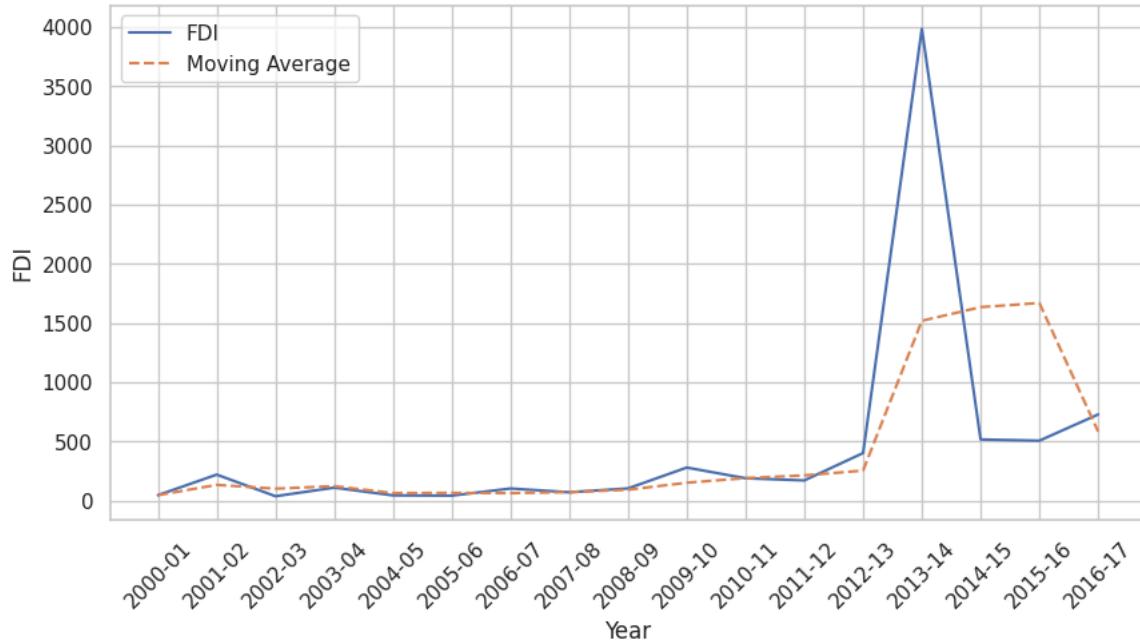


FDI Trends for FERMENTATION INDUSTRIES

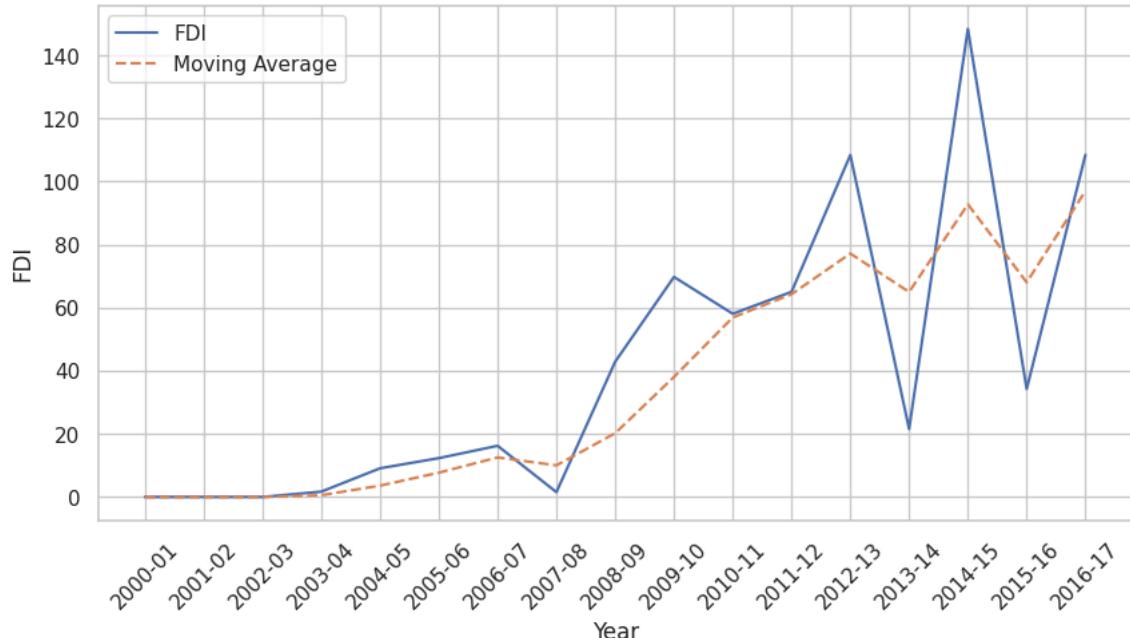




FDI Trends for FOOD PROCESSING INDUSTRIES

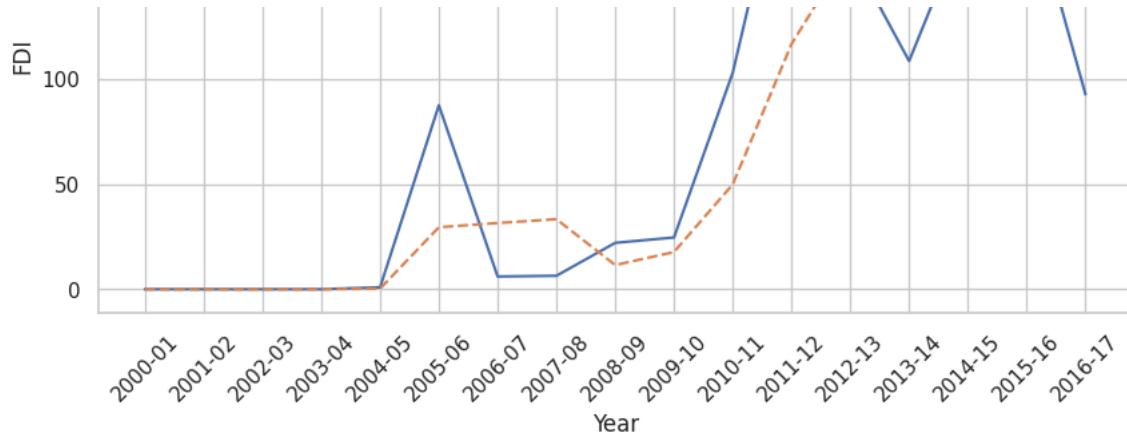


FDI Trends for VEGETABLE OILS AND VANASPATI

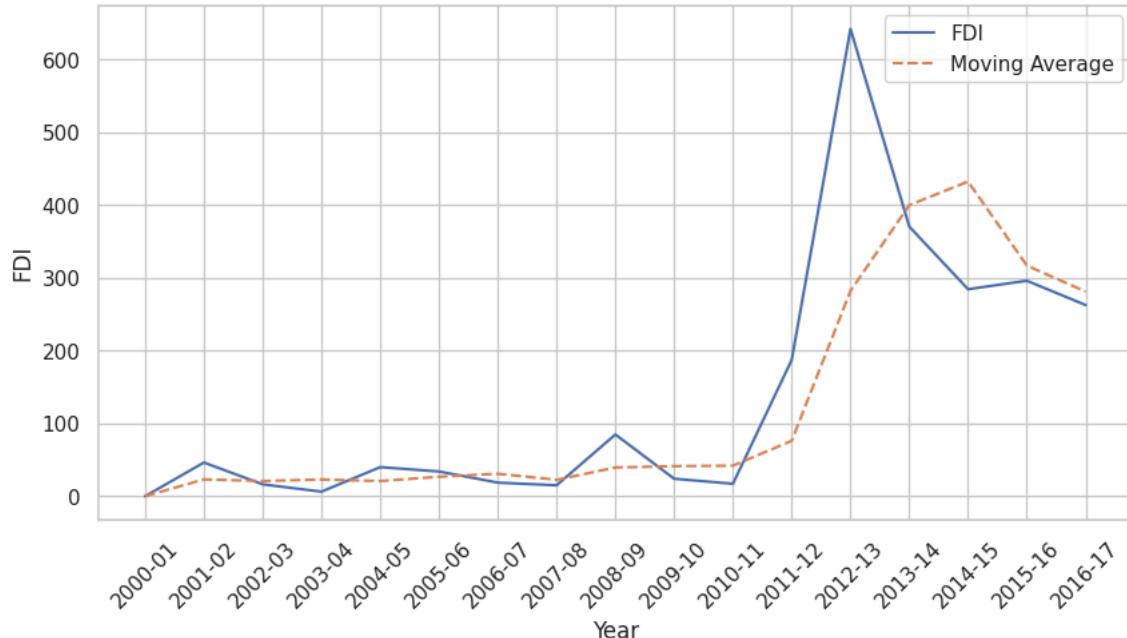


FDI Trends for SOAPS, COSMETICS &amp; TOILET PREPARATIONS

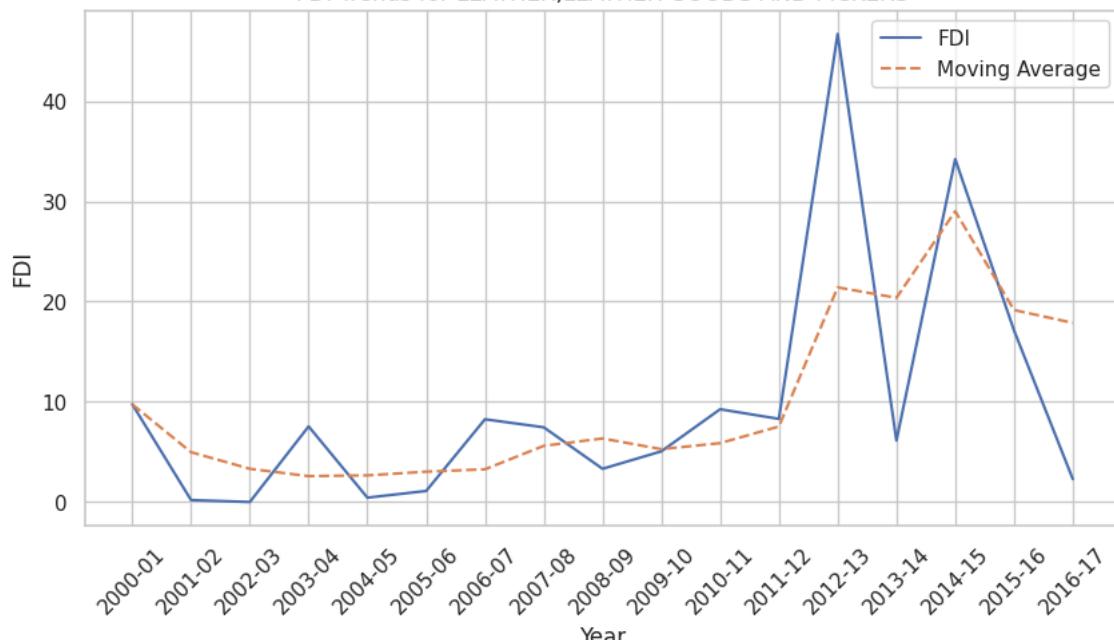




FDI Trends for RUBBER GOODS

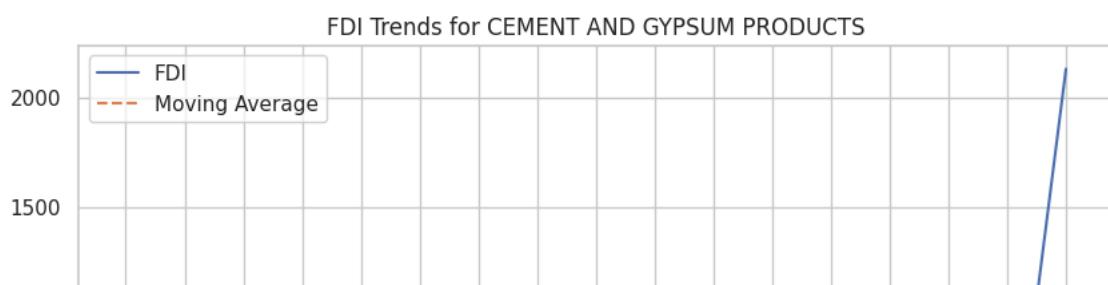
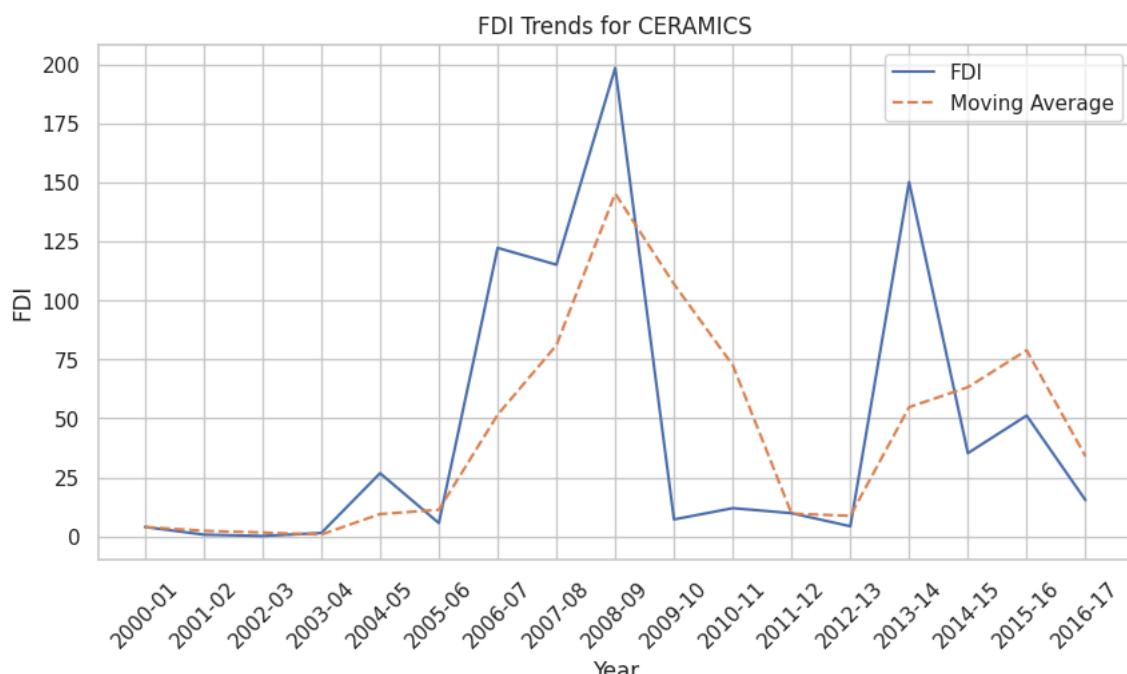
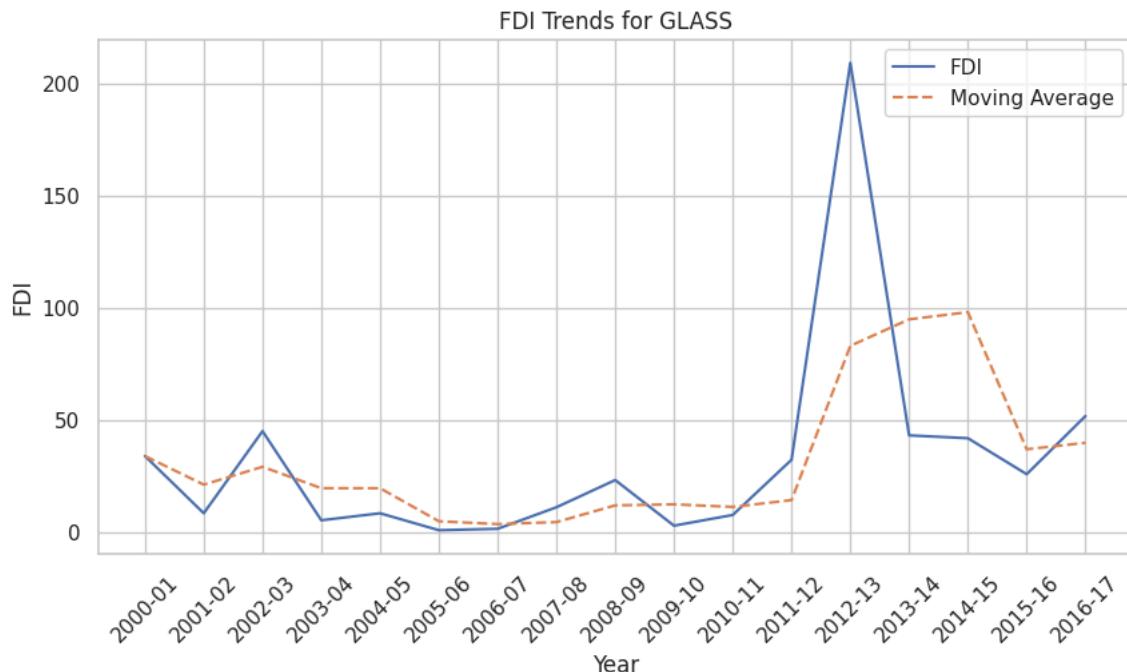
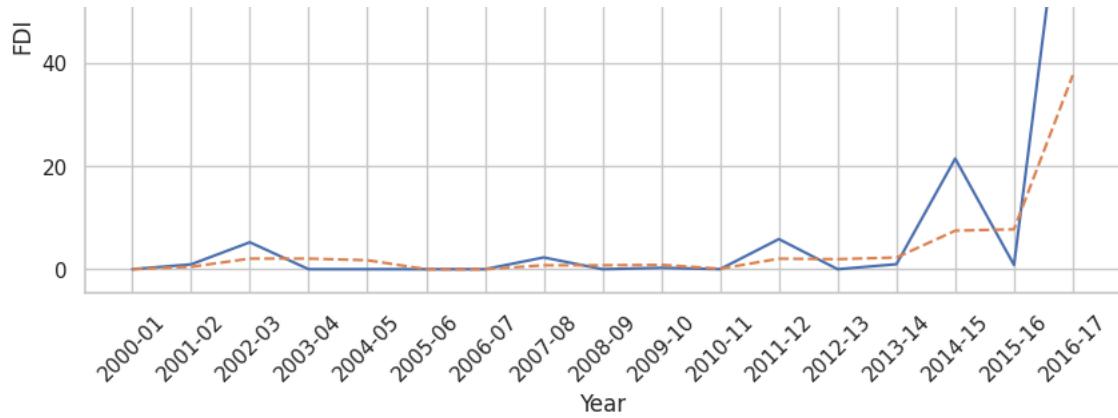


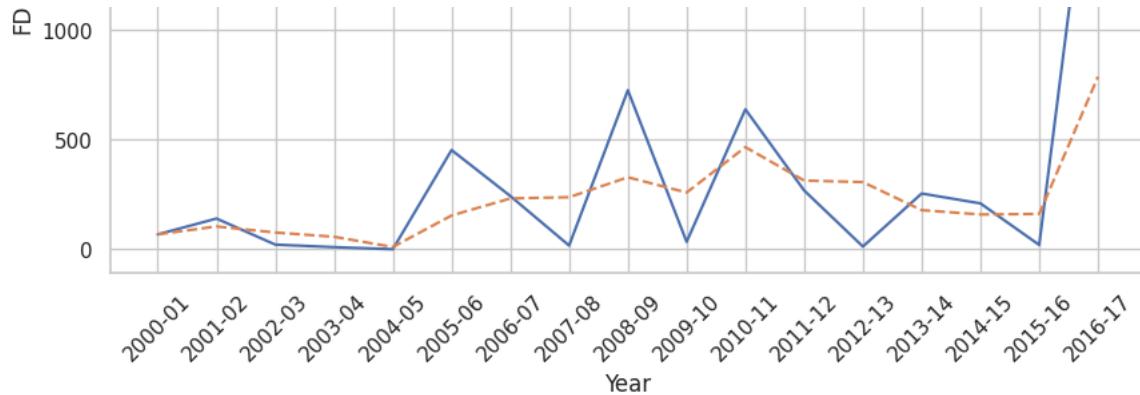
FDI Trends for LEATHER, LEATHER GOODS AND PICKERS



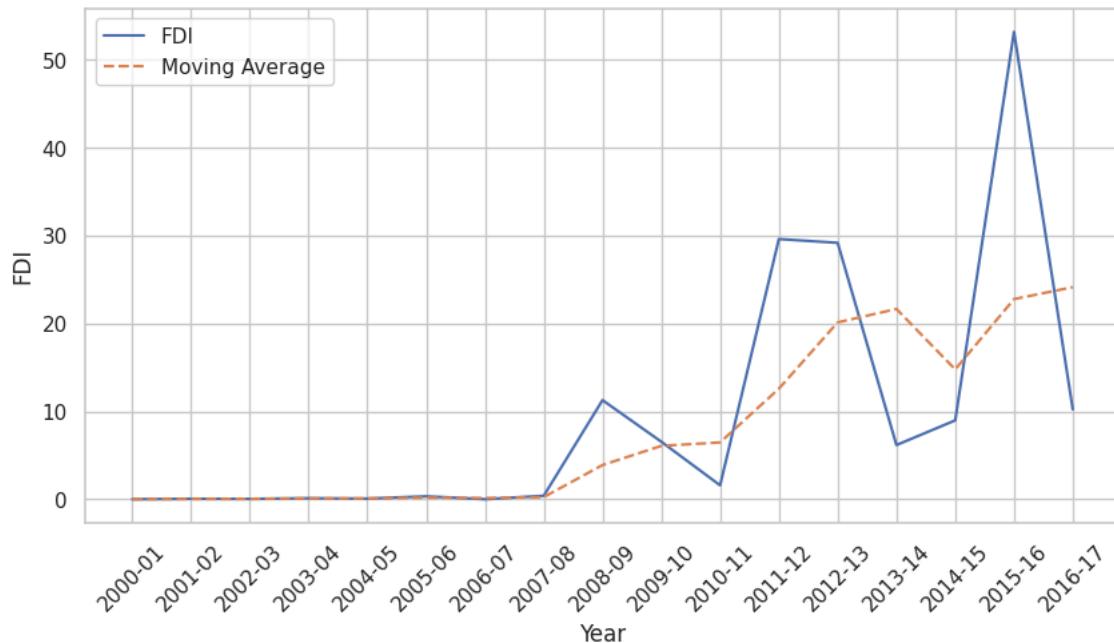
FDI Trends for GLUE AND GELATIN



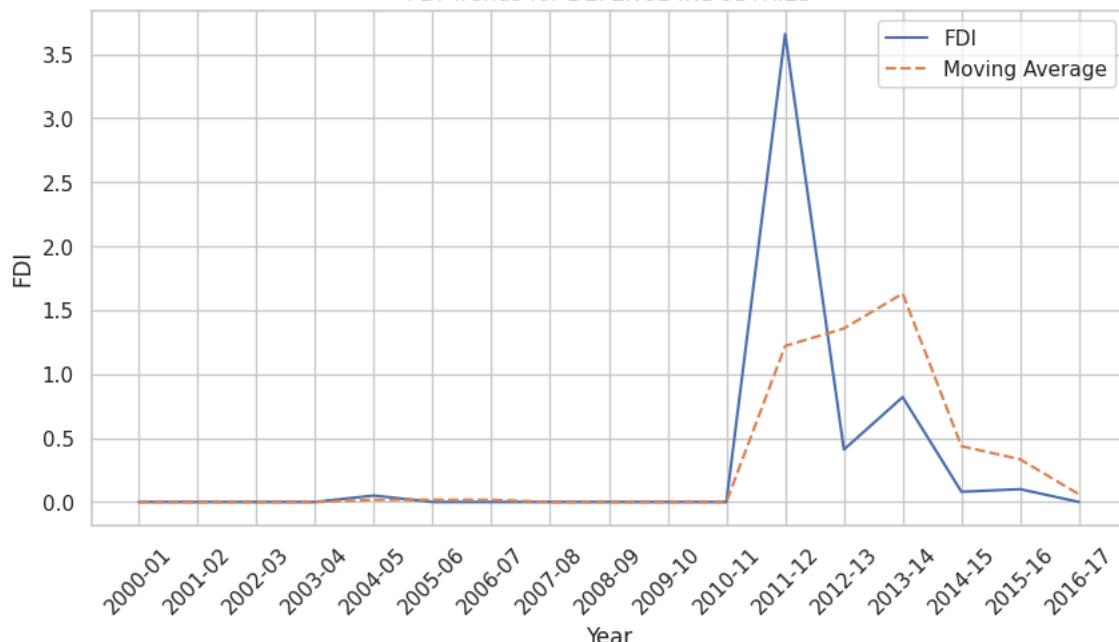




FDI Trends for TIMBER PRODUCTS

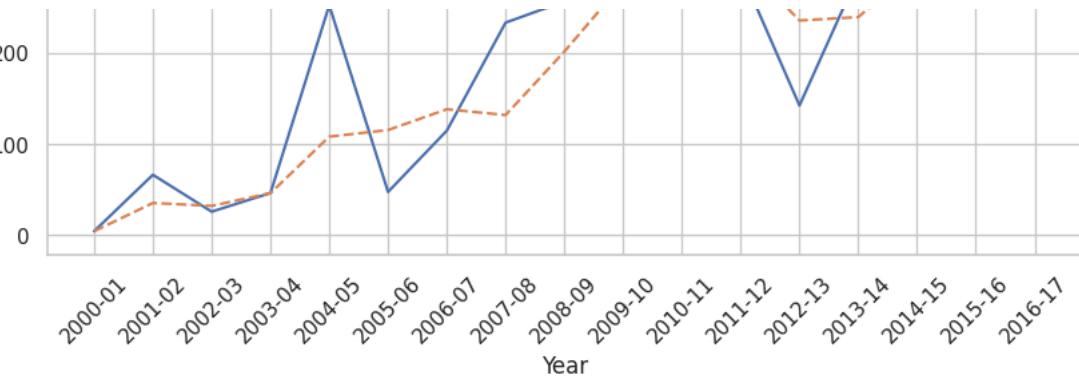


FDI Trends for DEFENCE INDUSTRIES

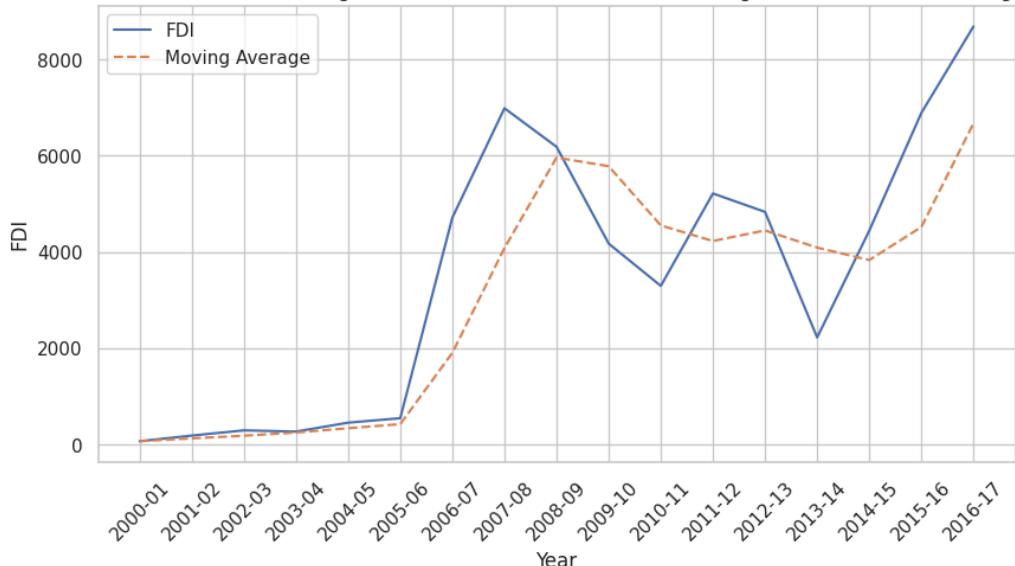


FDI Trends for CONSULTANCY SERVICES

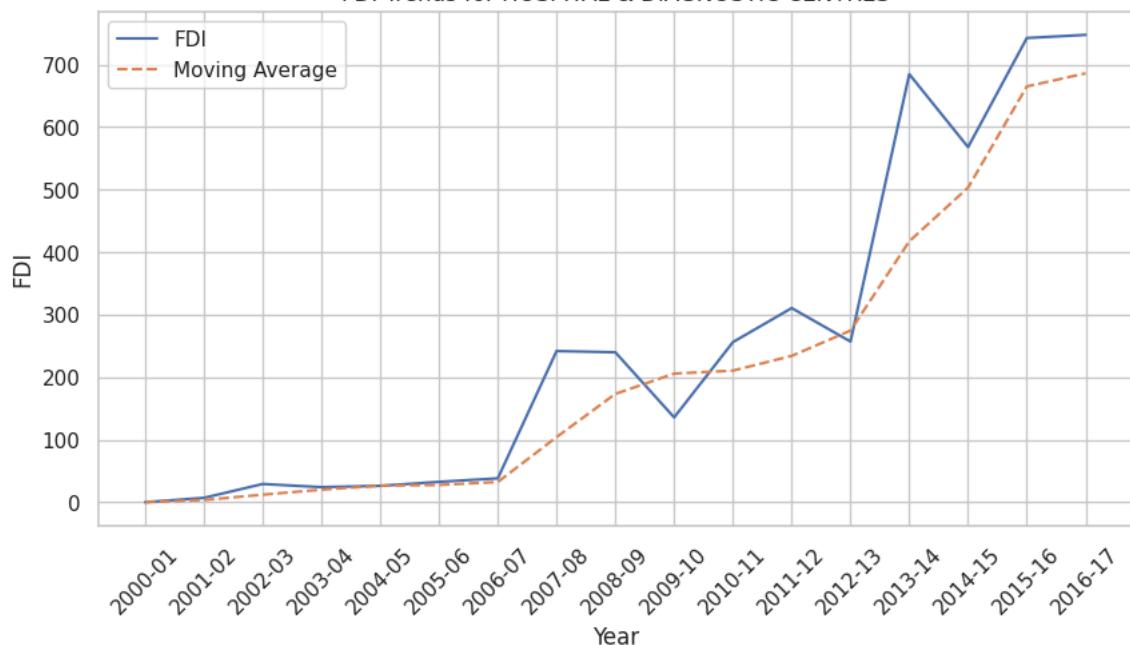




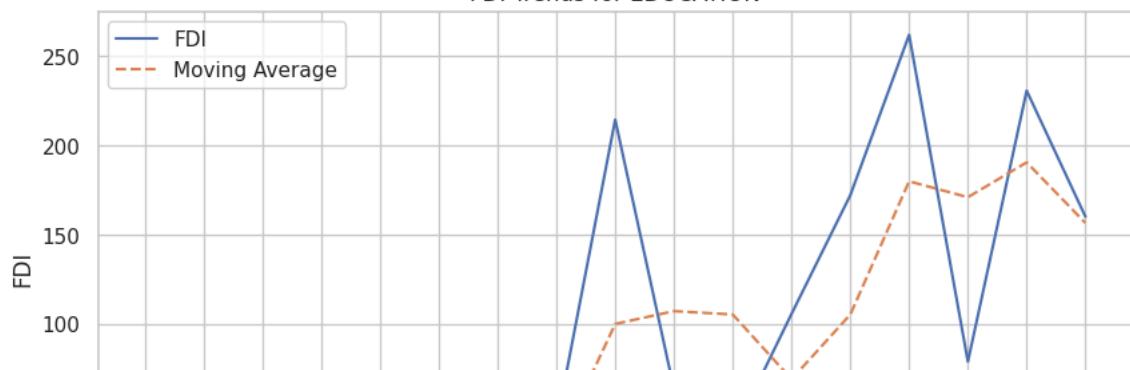
FDI Trends for SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,R&amp;D,Courier,Tech. Testing and Analysis, Other)

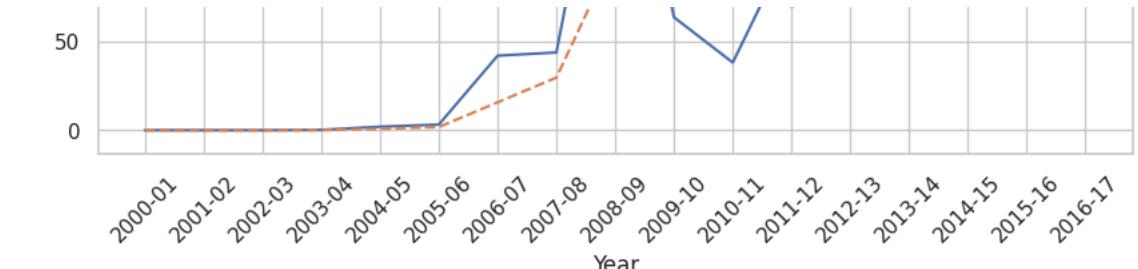


FDI Trends for HOSPITAL &amp; DIAGNOSTIC CENTRES

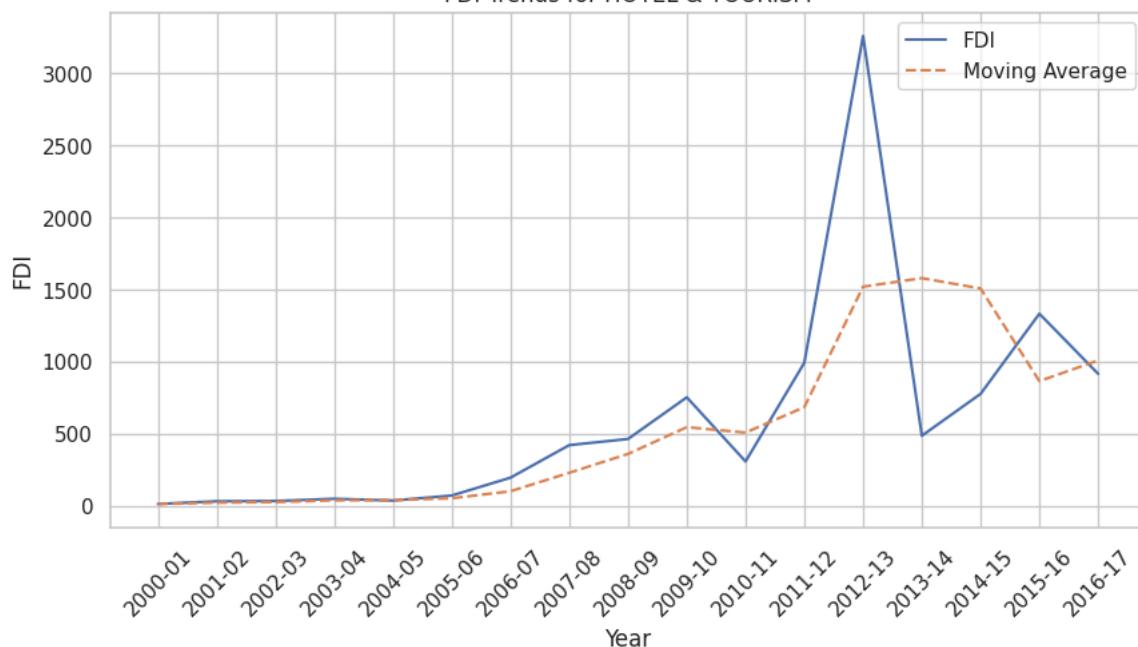


FDI Trends for EDUCATION

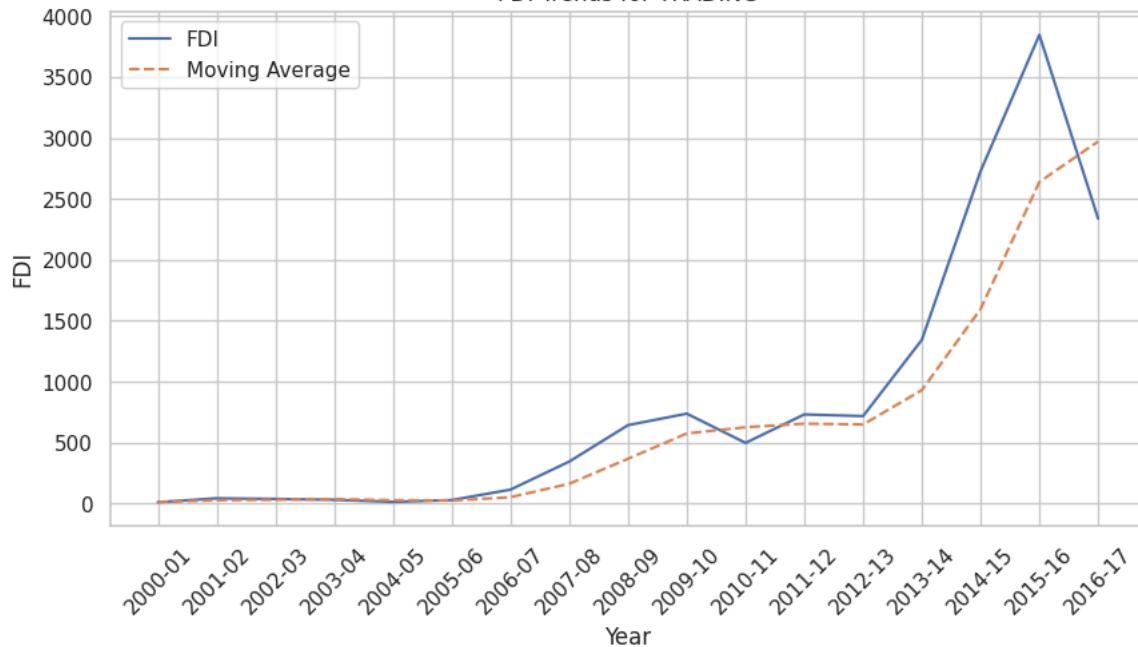




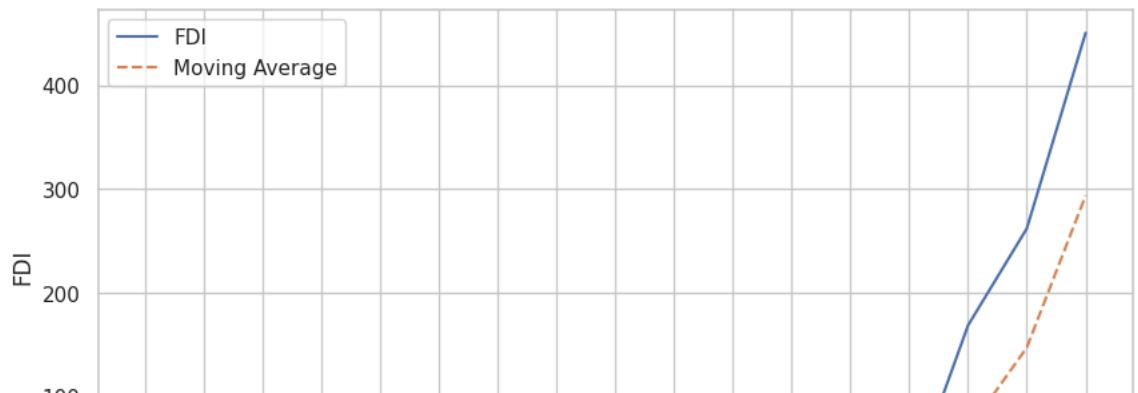
FDI Trends for HOTEL &amp; TOURISM

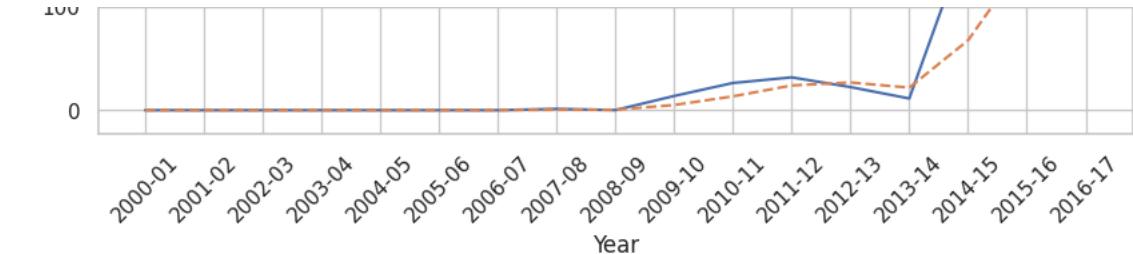


FDI Trends for TRADING

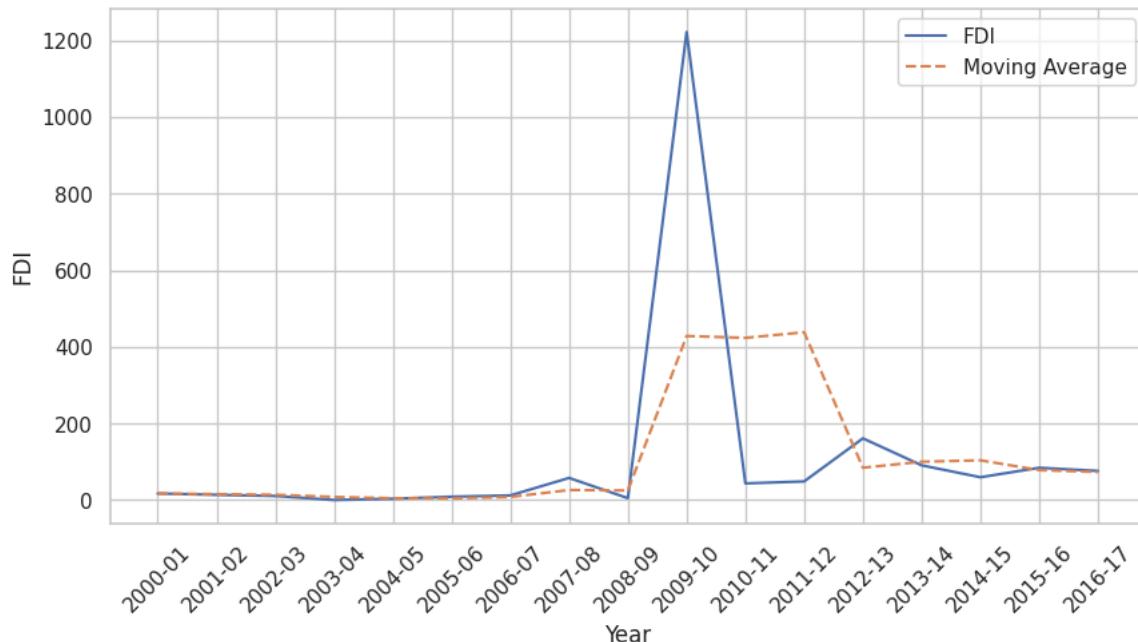


FDI Trends for RETAIL TRADING

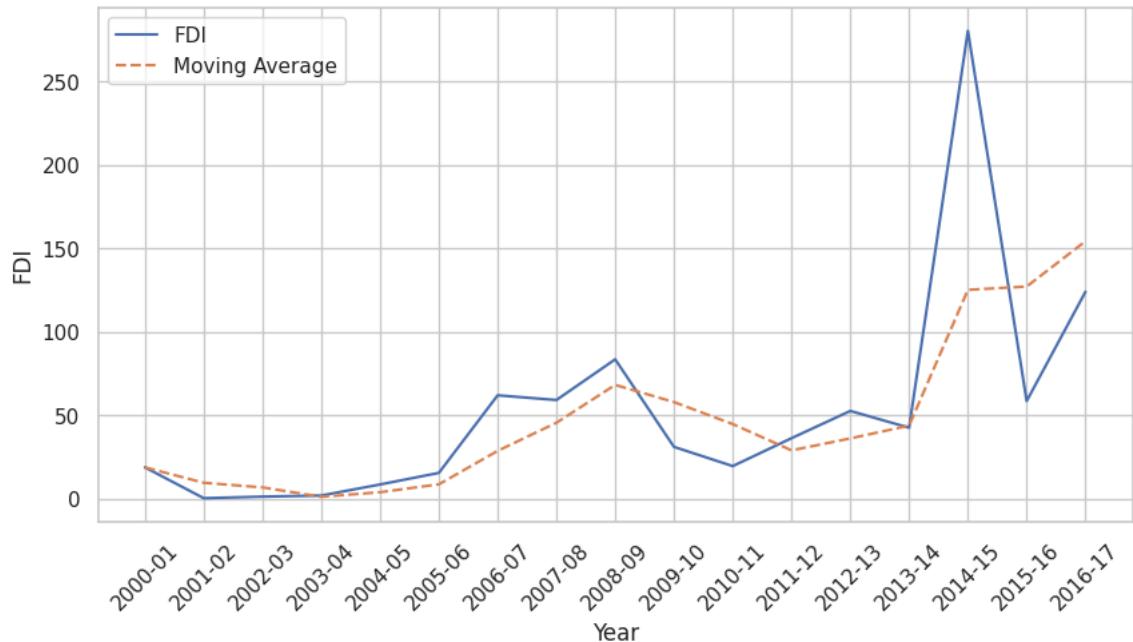




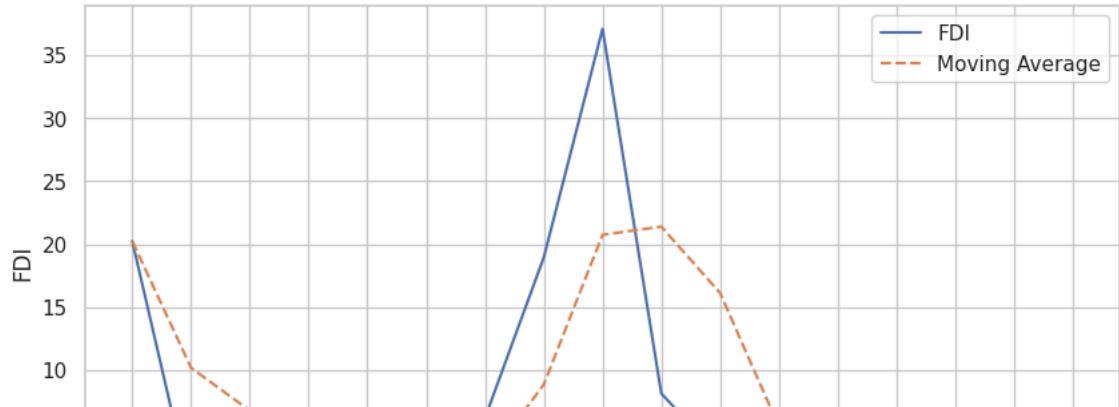
FDI Trends for AGRICULTURE SERVICES

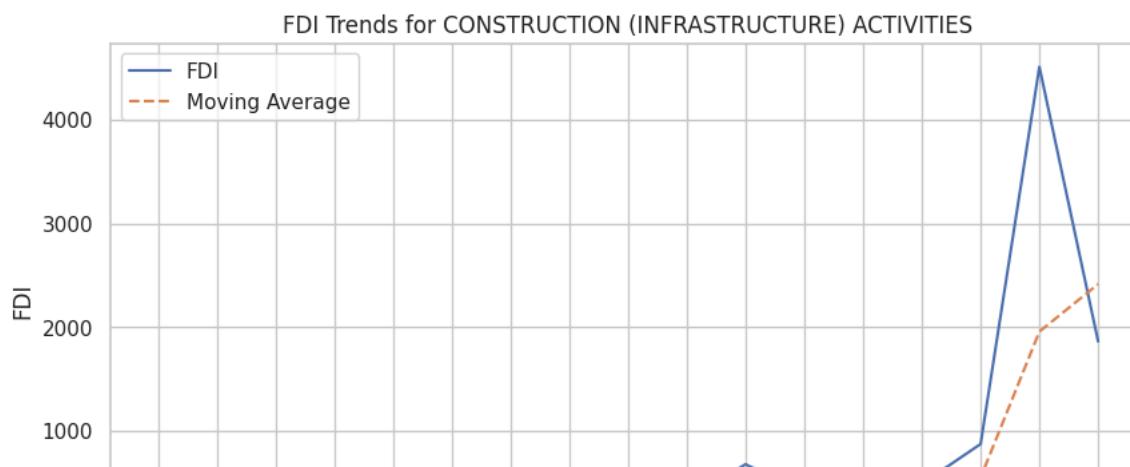
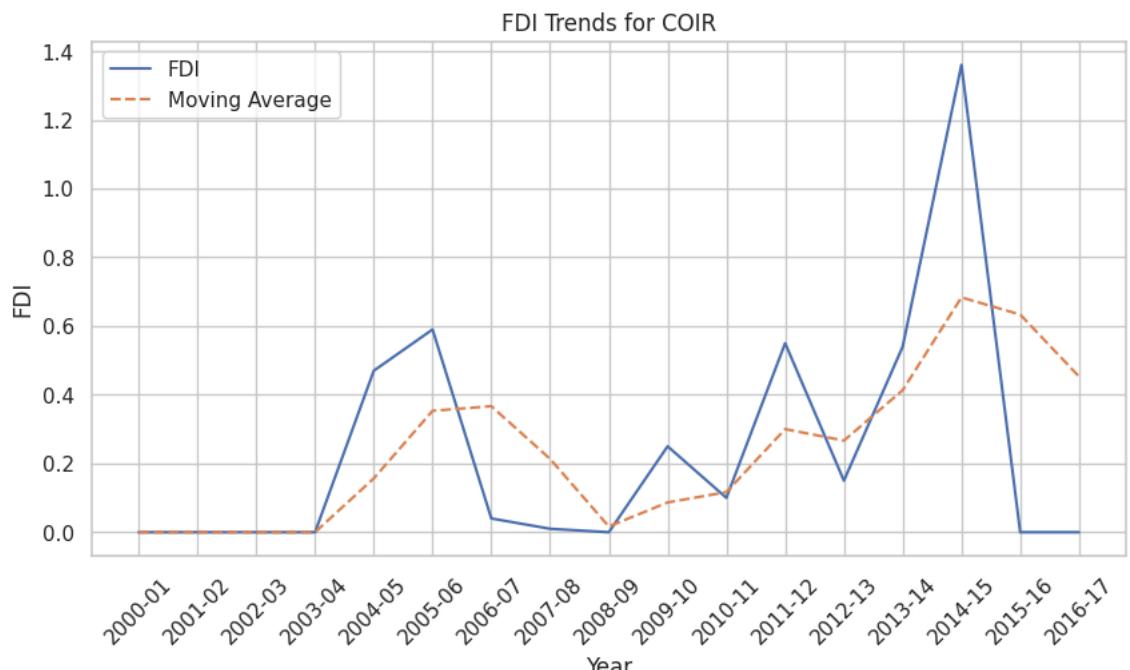
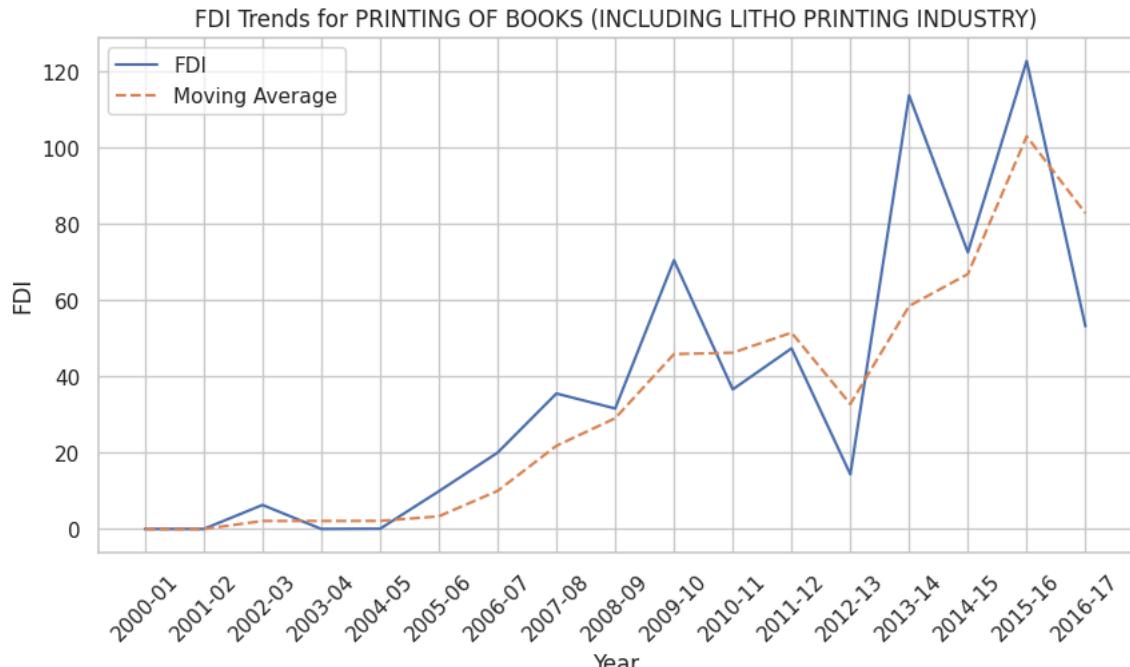
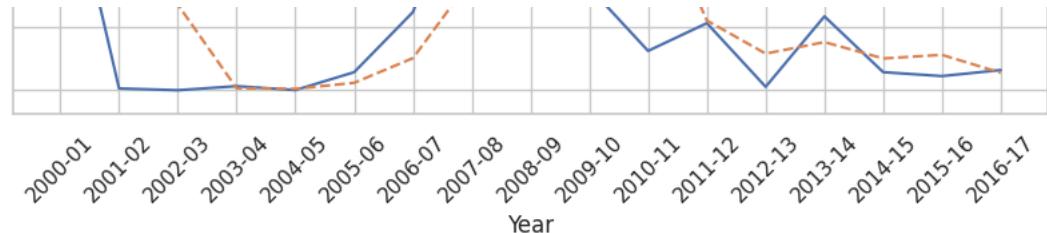


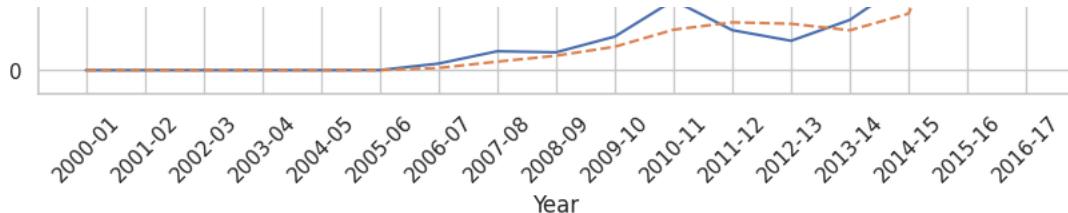
FDI Trends for DIAMOND,GOLD ORNAMENTS



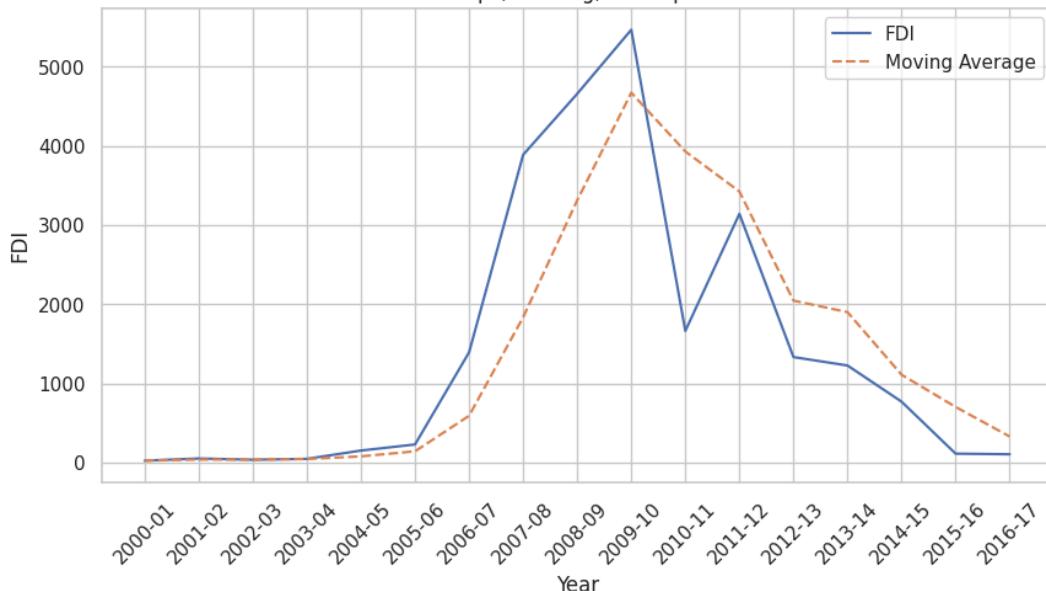
FDI Trends for TEA AND COFFEE (PROCESSING &amp; WAREHOUSING COFFEE &amp; RUBBER)



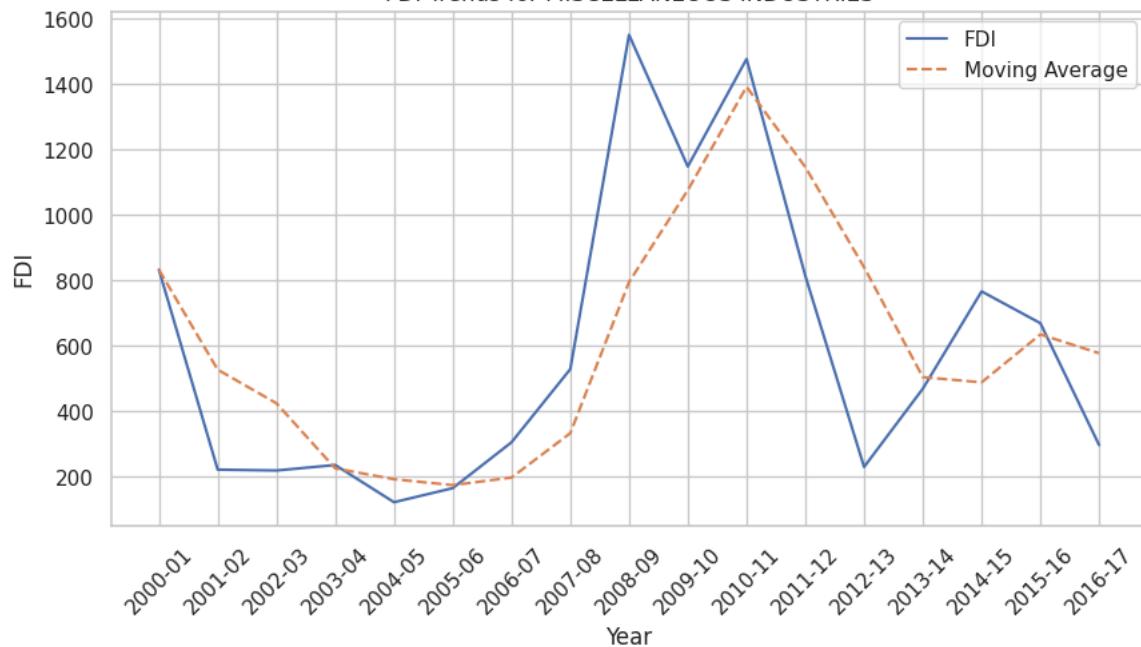




FDI Trends for CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-development projects



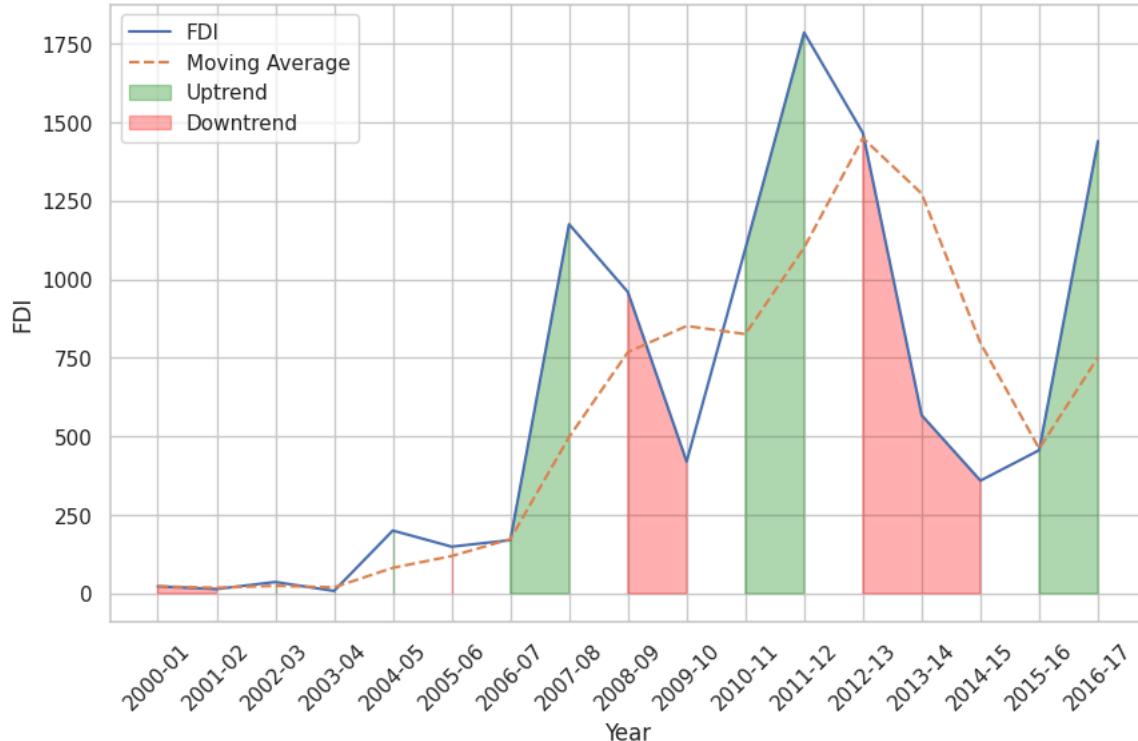
FDI Trends for MISCELLANEOUS INDUSTRIES



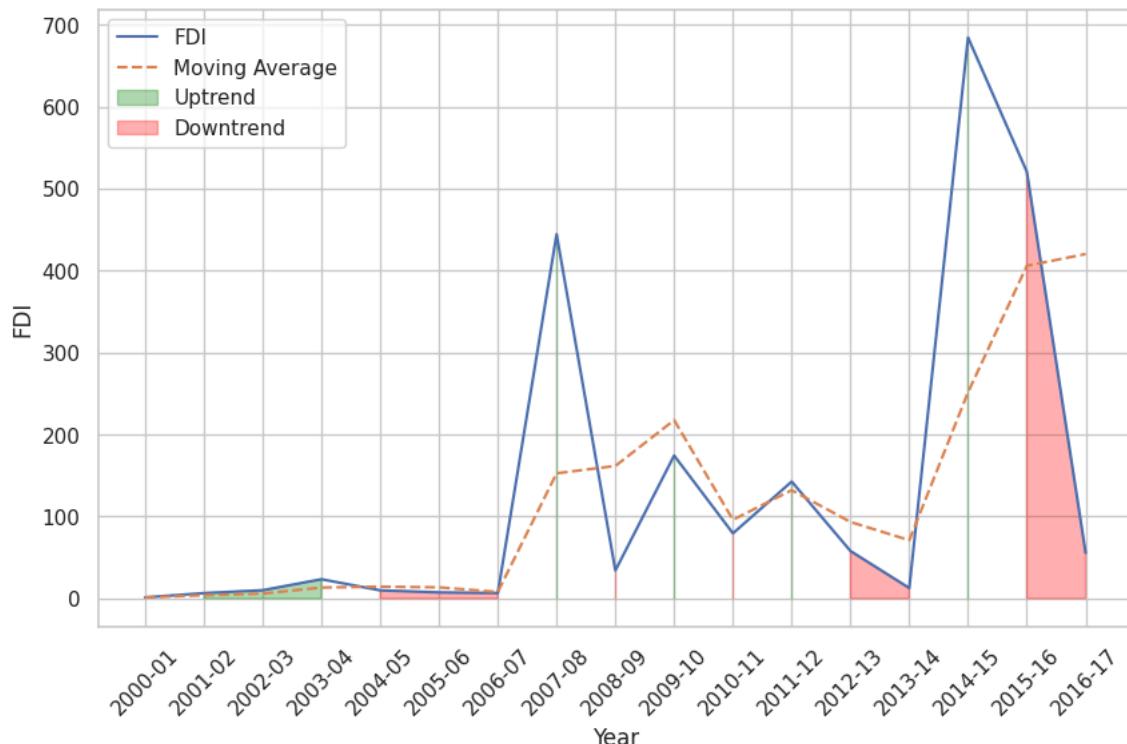
```
for sector in sectors:
    sector_data = df_long[df_long['Sector'] == sector]
    plt.figure(figsize=(10, 6))
    plt.plot(sector_data['Year'], sector_data['FDI'], label='FDI')
    plt.plot(sector_data['Year'], sector_data['Moving Average'], '--', label='Moving Average')
    plt.fill_between(sector_data['Year'], sector_data['FDI'], where=sector_data['Growth Rate'] > 0, color='green', alpha=0.3)
    plt.fill_between(sector_data['Year'], sector_data['FDI'], where=sector_data['Growth Rate'] <= 0, color='red', alpha=0.3,
    plt.title(f'FDI Trends for {sector}'))
    plt.xlabel('Year')
    plt.xticks(rotation=45)
    plt.ylabel('FDI')
    plt.legend()
    plt.grid(True)
    plt.show()
```



## FDI Trends for METALLURGICAL INDUSTRIES

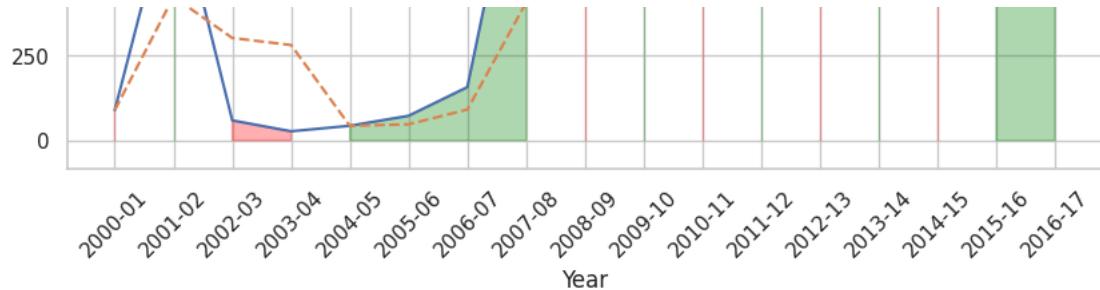


## FDI Trends for MINING

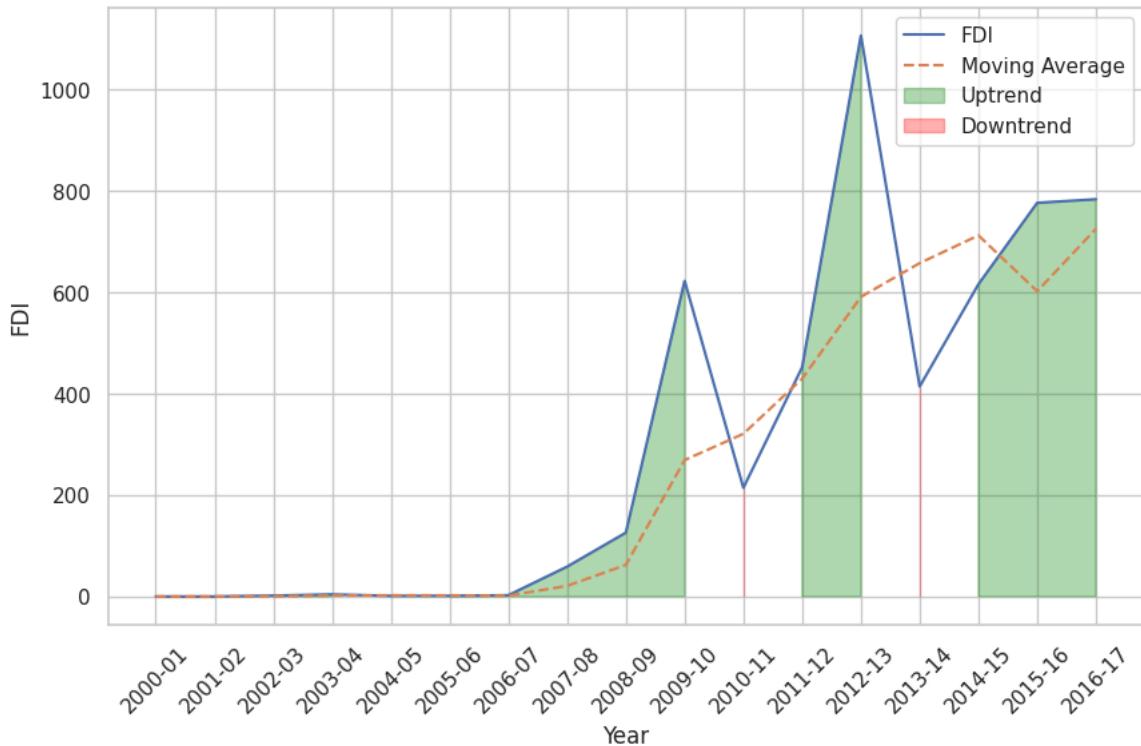


## FDI Trends for POWER

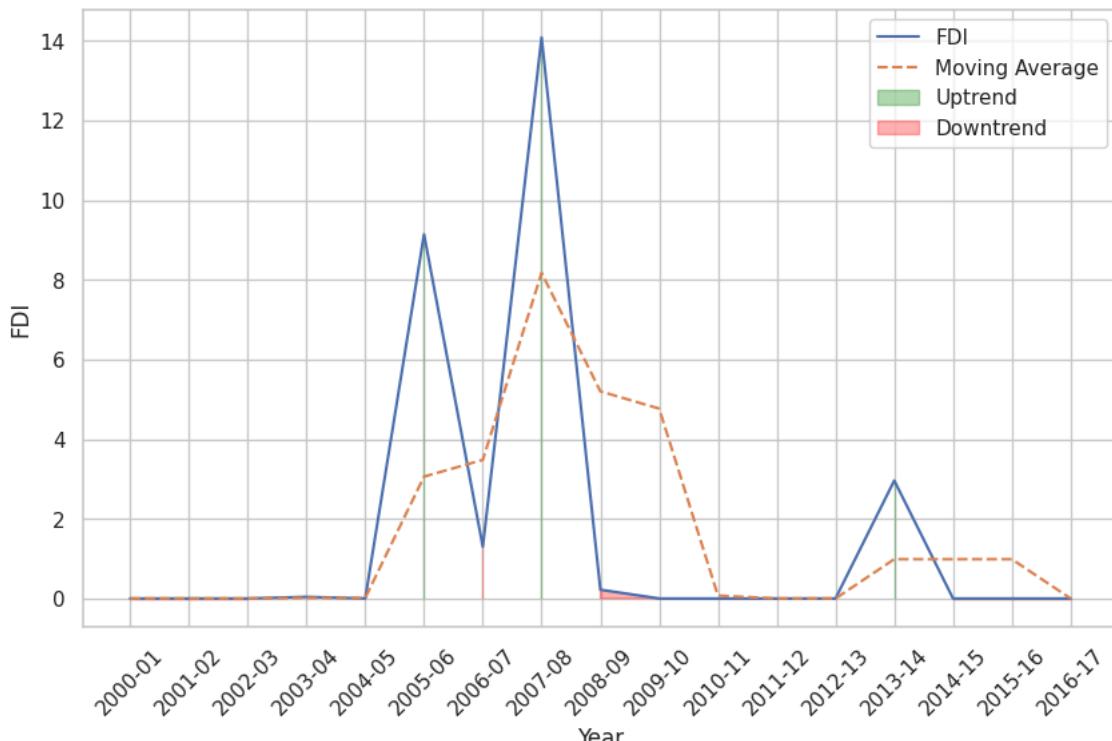




FDI Trends for NON-CONVENTIONAL ENERGY



FDI Trends for COAL PRODUCTION



FDI Trends for PETROLEUM &amp; NATURAL GAS

