

IMPORT LIBRARIES

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```


read the dataset

```
In [112]: df1 = pd.read_csv('FDI data.csv', index_col=[0])
df1.head()
```

Out[112]:

	METALLURGICAL INDUSTRIES	MINING	POWER	NON- CONVENTIONAL ENERGY	COAL PRODUCTION	PETROLEUM & NATURAL GAS	ANI GENI
Sector							
2000-01	22.69	1.32	89.42	0.00	0.00	9.35	
2001-02	14.14	6.52	757.44	0.00	0.00	211.07	
2002-03	36.61	10.06	59.11	1.70	0.00	56.78	
2003-04	8.11	23.48	27.09	4.14	0.04	80.64	
2004-05	200.38	9.92	43.37	1.27	0.00	102.78	

5 rows × 16383 columns



```
In [113]: df1.drop(df1.iloc[:,63:],axis=1,inplace=True)
df1.head()
```

Out[113]:

	METALLURGICAL INDUSTRIES	MINING	POWER	CONVENTIONAL ENERGY	COAL PRODUCTION	PETROLEUM & NATURAL GAS	AN GENI
Sector							
2000-01	22.69	1.32	89.42	0.00	0.00	9.35	
2001-02	14.14	6.52	757.44	0.00	0.00	211.07	
2002-03	36.61	10.06	59.11	1.70	0.00	56.78	
2003-04	8.11	23.48	27.09	4.14	0.04	80.64	
2004-05	200.38	9.92	43.37	1.27	0.00	102.78	

5 rows × 63 columns



```
In [114]: df=df1.T
df.head(3)
```

Out[114]:

Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	2009-10
METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.75	959.94	419.88
MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.36	34.16	174.40
POWER	89.42	757.44	59.11	27.09	43.37	72.69	157.15	988.68	907.66	1271.79



```
In [115]: df.shape
```

Out[115]: (63, 17)

data preprocessing

```
In [116]: nv = df1.isnull().sum()
          nv
          nv.dropna()
```

```
Out[116]: METALLURGICAL INDUSTRIES
          0
          MINING
          0
          POWER
          0
          NON-CONVENTIONAL ENERGY
          0
          COAL PRODUCTION
          0

          ..
          PRINTING OF BOOKS (INCLUDING LITHO PRINTING INDUSTRY)
          0
          COIR
          0
          CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES
          0
          CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and co
          nstruction-development projects      0
          MISCELLANEOUS INDUSTRIES
          0
          Length: 63, dtype: int64
```

```
In [117]: # handling null values
          def fetch_count_per_null_val(data):
              nv = data.isnull().sum()
              nv = nv[nv>0]
              nv_df = pd.DataFrame({'Feature':nv.index, 'Count_Null':nv.values,
                                   'Per_Null':(nv.values/data.shape[0])*100})
              nv_df = nv_df.sort_values('Per_Null', ascending=False)
              return nv_df
```

```
In [118]: nv_df = fetch_count_per_null_val(df)
          nv_df
```

```
Out[118]:
```

Feature	Count_Null	Per_Null
---------	------------	----------

```
In [119]: # handling duplicates
          df.duplicated().sum()
```

```
Out[119]: 0
```

Statistical analysis

In [120]: df1.describe()

Out[120]:

	METALLURGICAL INDUSTRIES	MINING	POWER	NON- CONVENTIONAL ENERGY	COAL PRODUCTION	PETROLEUM & NATURAL GAS
count	17.000000	17.000000	17.000000	17.000000	17.000000	17.000000
mean	607.678824	133.637059	681.713529	304.793529	1.631765	403.303529
std	590.318680	209.032041	524.452076	359.716927	3.929237	568.490900
min	8.110000	1.320000	27.090000	0.000000	0.000000	9.350000
25%	149.130000	9.920000	89.420000	1.700000	0.000000	87.710000
50%	419.880000	34.160000	757.440000	125.880000	0.000000	180.400000
75%	1098.140000	142.650000	1066.080000	615.950000	0.220000	349.290000
max	1786.140000	684.390000	1652.380000	1106.520000	14.080000	2029.980000

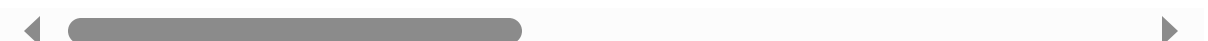
8 rows × 63 columns



In [121]: df.describe()

Out[121]:

Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000
mean	37.757302	63.931587	42.925714	34.727778	51.090317	87.932540	198.281905
std	112.227860	157.878737	86.606439	67.653735	101.934873	206.436967	686.783115
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.200000	0.215000	0.715000	1.230000	4.160000
50%	4.030000	5.070000	11.010000	6.370000	9.090000	22.620000	25.820000
75%	23.510000	44.830000	36.555000	38.660000	43.205000	63.855000	108.325000
max	832.070000	873.230000	419.960000	368.320000	527.900000	1359.970000	4713.780000



In []:

In []:

Exploratory data analysis

```
In [122]: df1['Total FDI'] = df1.iloc[:, :].sum(axis=1)
```

```
In [123]: df1_sorted = df1.sort_values(by='Total FDI', ascending=False)
```

```
In [124]: print(df1_sorted[['Total FDI']])
```

	Total FDI
Sector	
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

```
In [125]: start_year = '2000-01'  
end_year = '2010-11'
```

```
In [126]: df['Total FDI'] = df.loc[:, start_year:end_year].sum(axis=1)
```

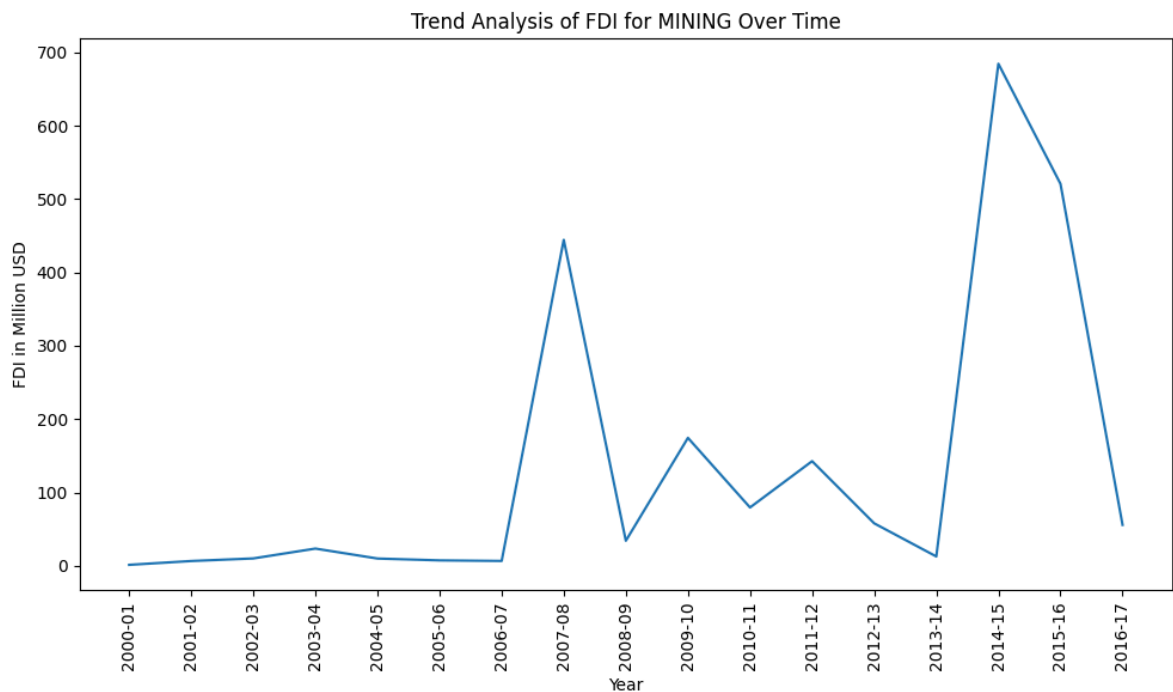
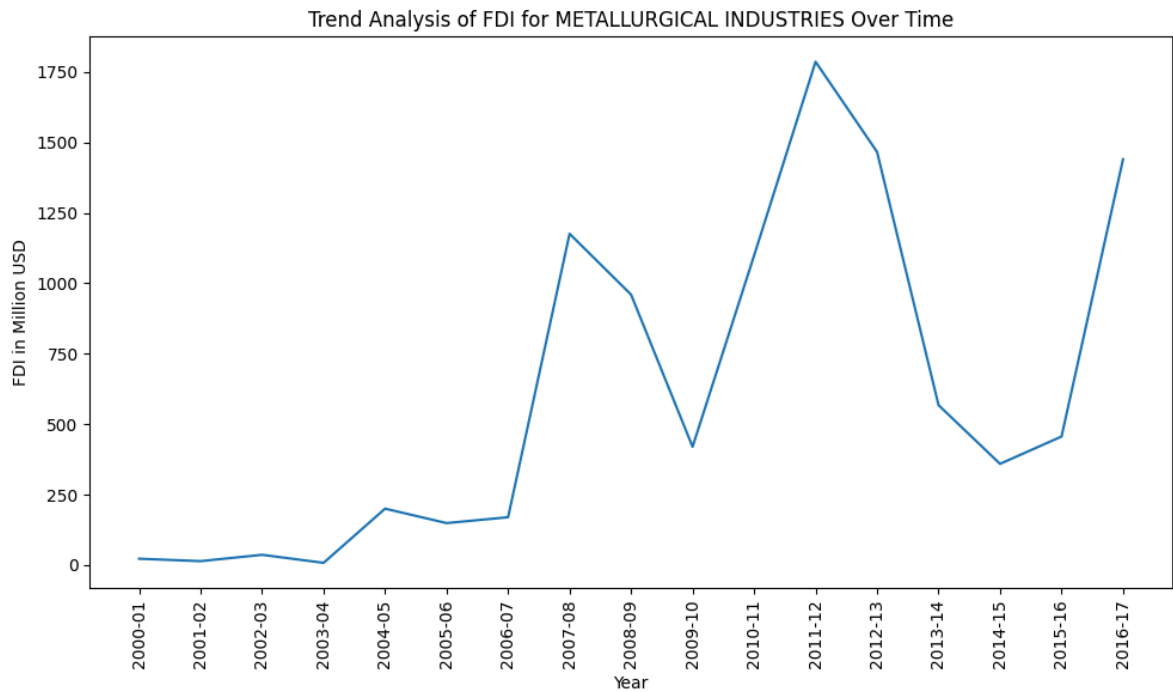
```
In [127]: df_sorted = df.sort_values(by='Total FDI', ascending=False)
```

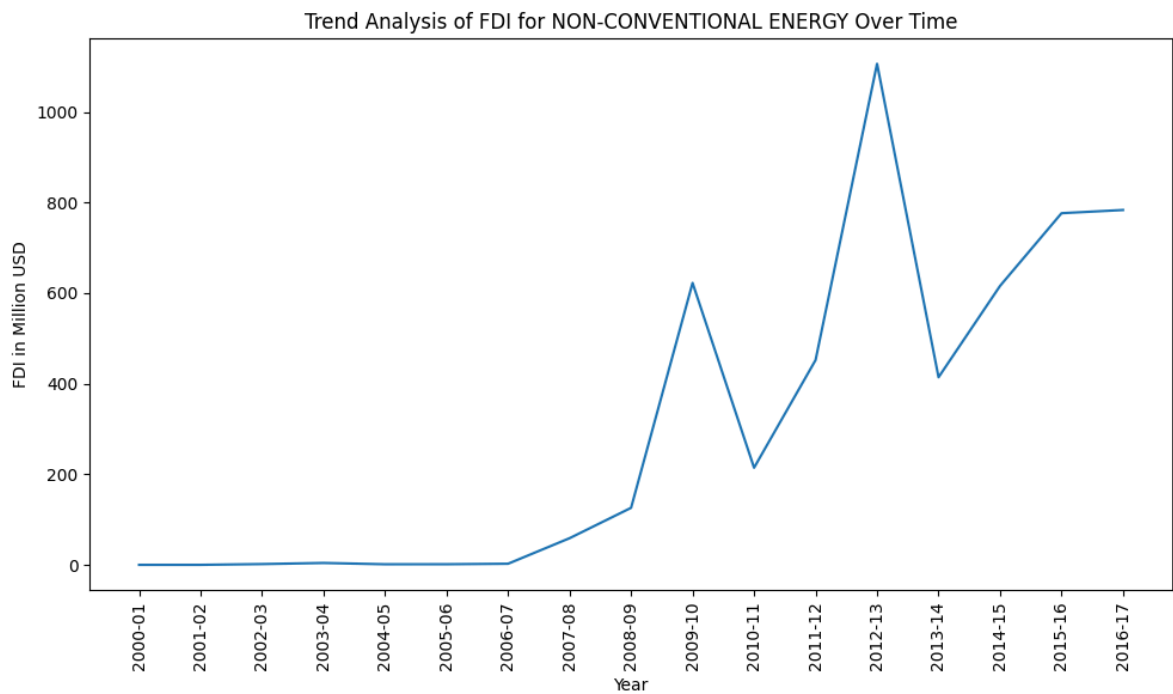
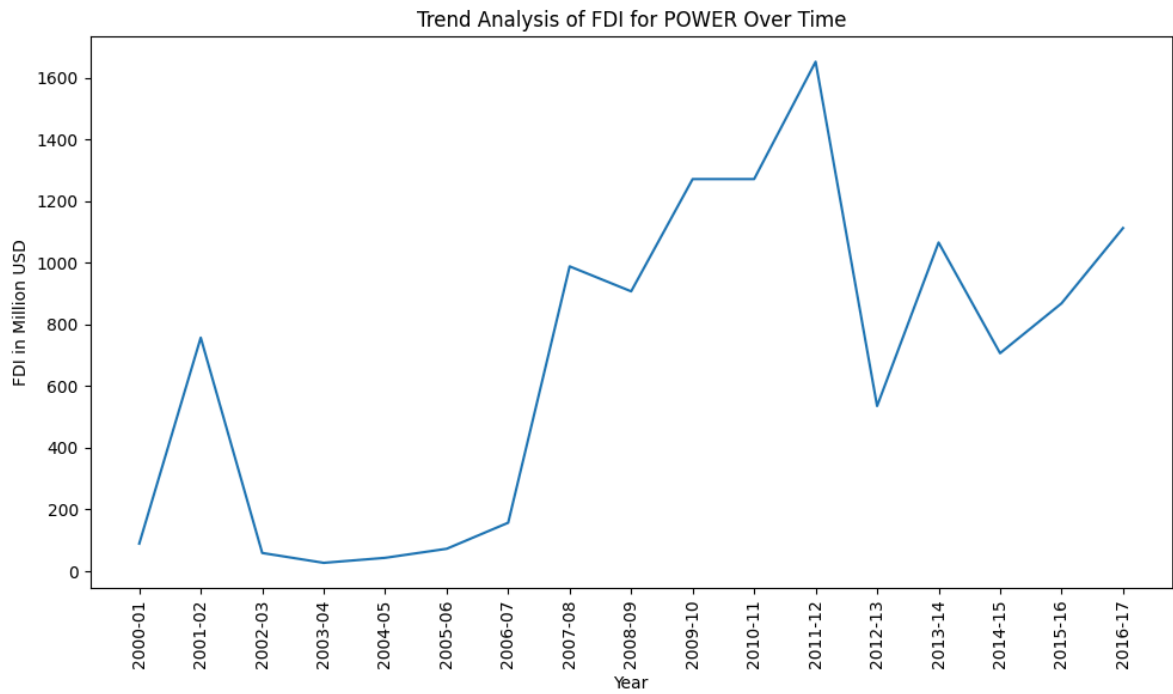
```
In [128]: print(df_sorted[['Total FDI']])
```

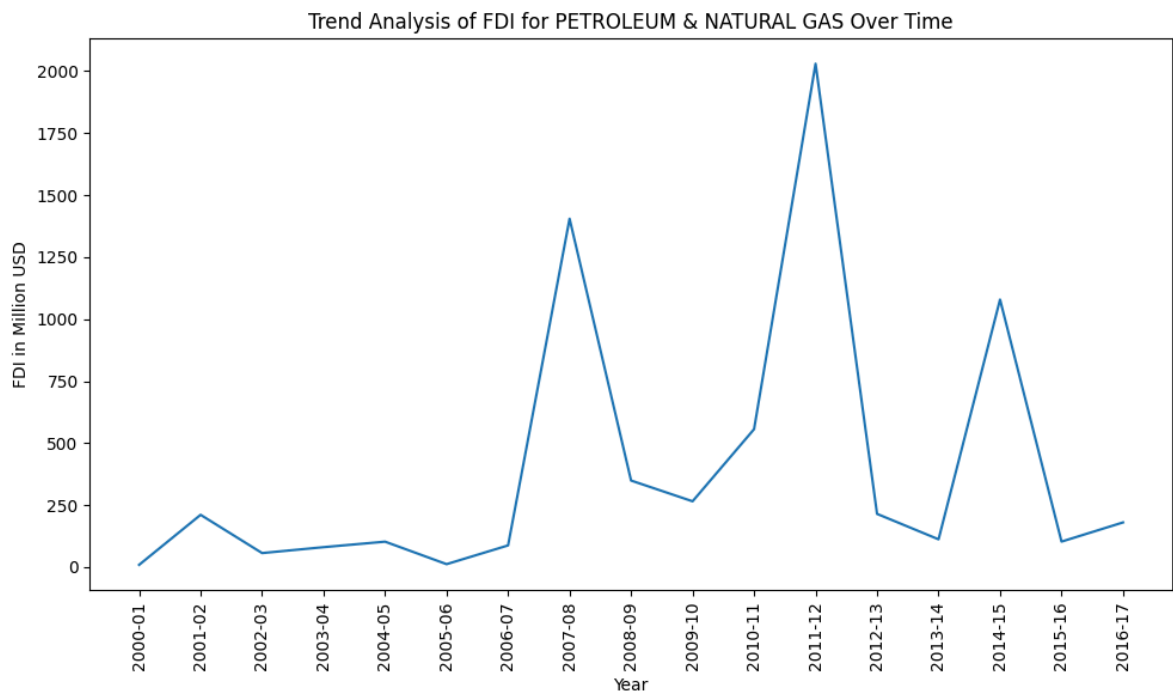
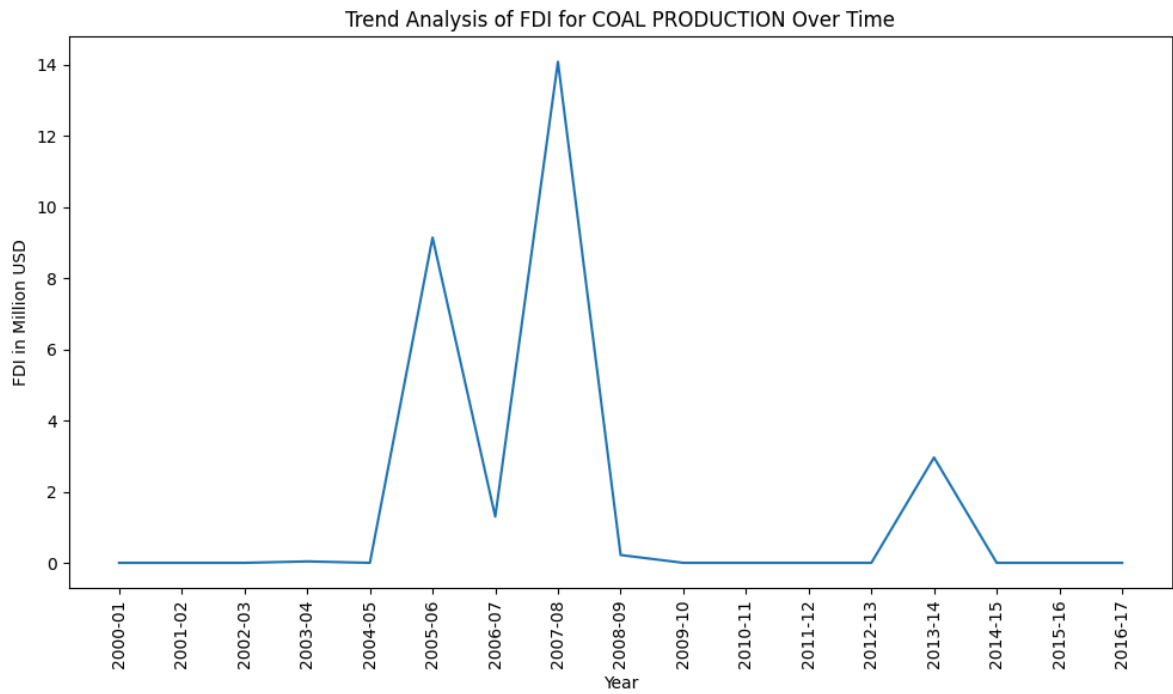
Sector	Total FDI
SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin...	27185.64
CONSTRUCTION DEVELOPMENT: Townships, housing, b...	17606.94
TELECOMMUNICATIONS	10554.92
COMPUTER SOFTWARE & HARDWARE	10408.80
MISCELLANEOUS INDUSTRIES	6800.79
...	...
BOILERS AND STEAM GENERATING PLANTS	9.99
GLUE AND GELATIN	8.72
COIR	1.46
MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS	1.27
DEFENCE INDUSTRIES	0.05

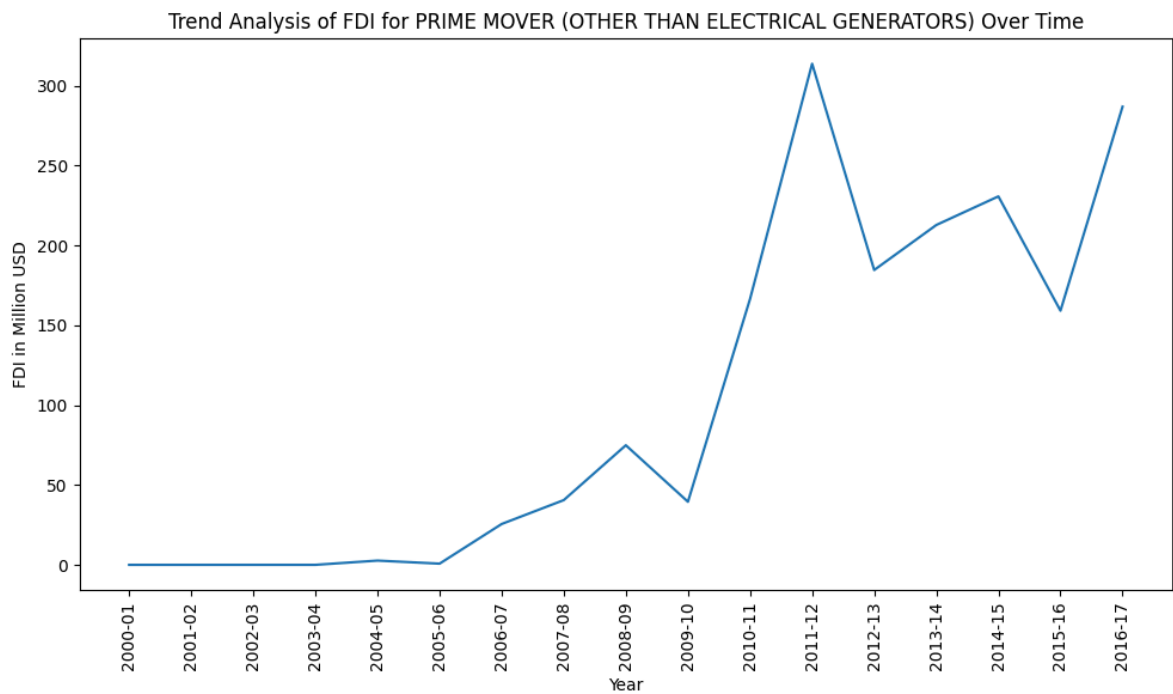
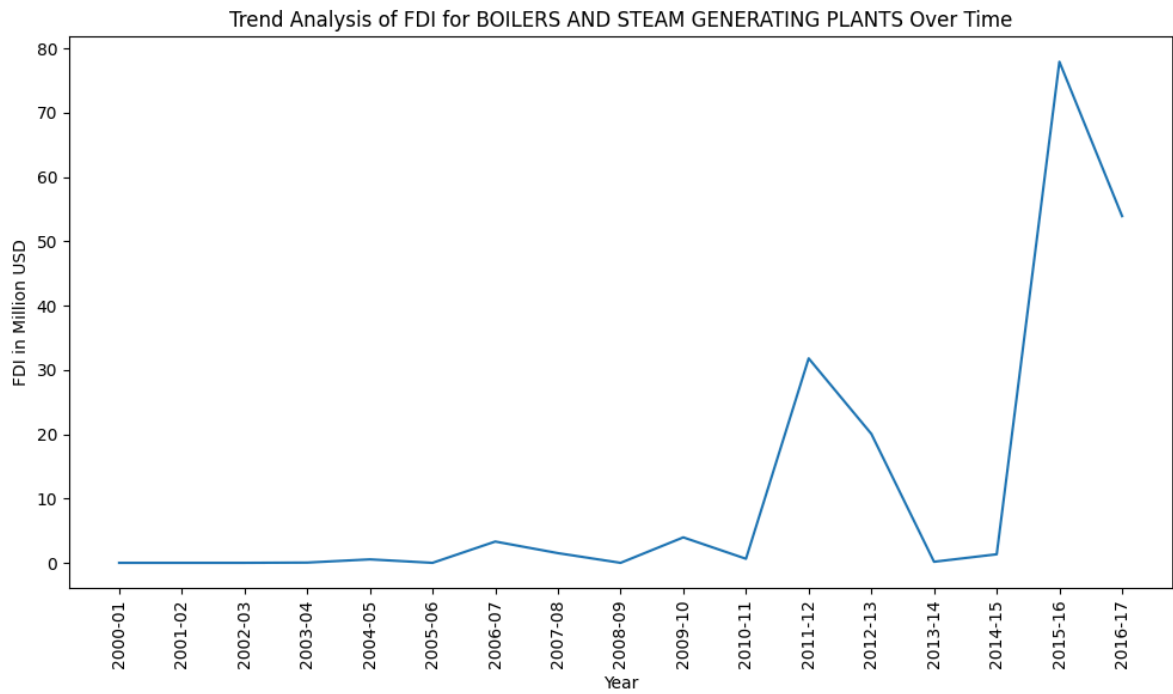
```
[63 rows x 1 columns]
```

```
In [129]: for sector in df1.columns:
plt.figure(figsize=(10, 6))
plt.plot(df1.index, df1[sector])
plt.xticks(rotation=90)
plt.xlabel('Year')
plt.ylabel('FDI in Million USD')
plt.title(f'Trend Analysis of FDI for {sector} Over Time')
plt.tight_layout()
plt.show()
```

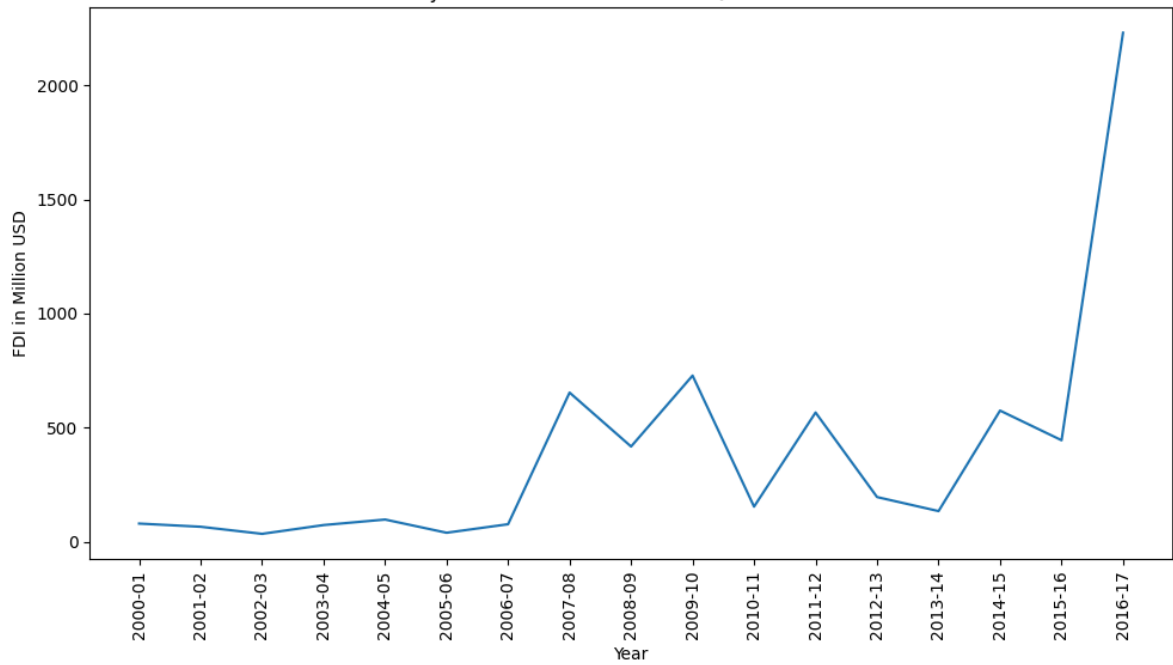




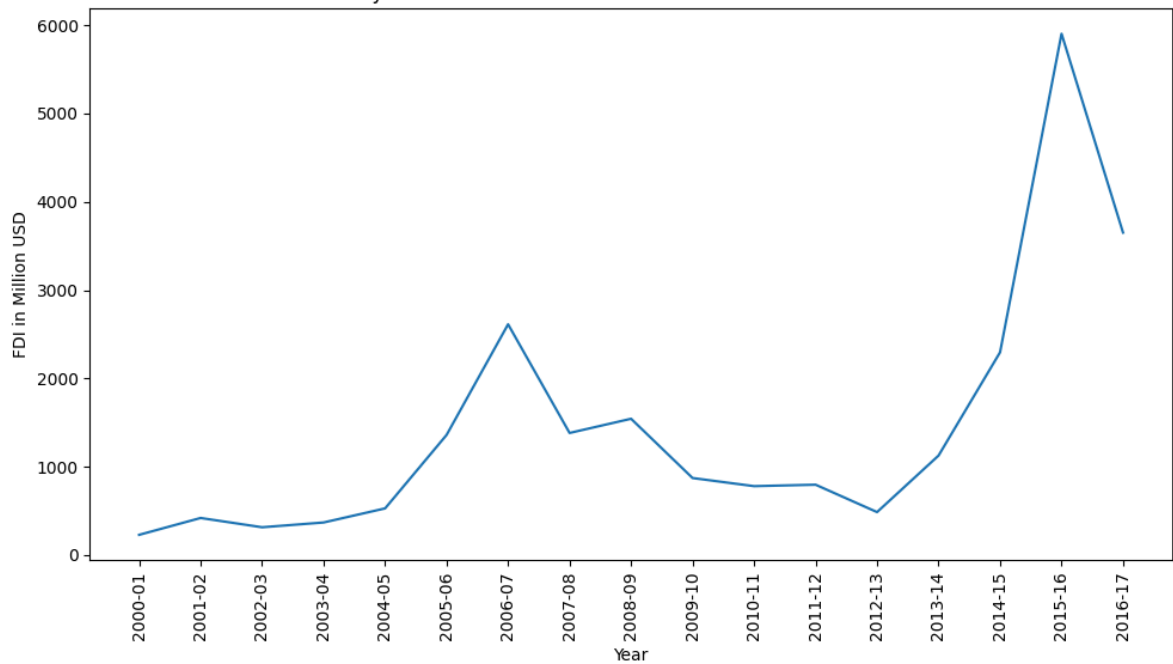


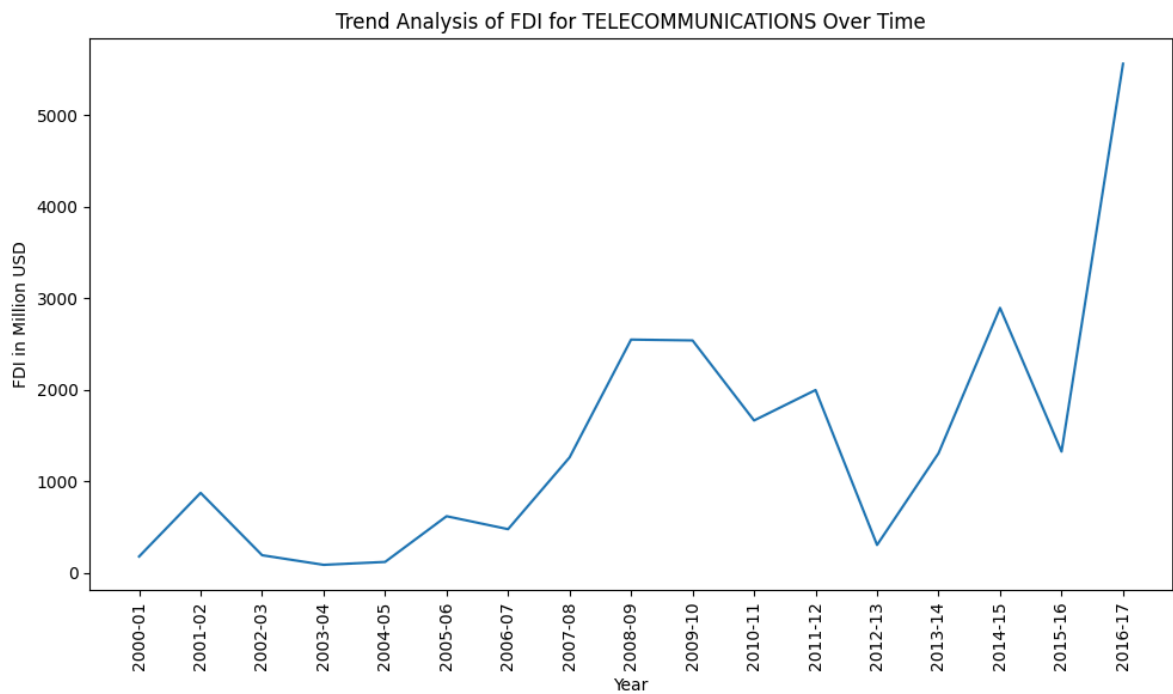
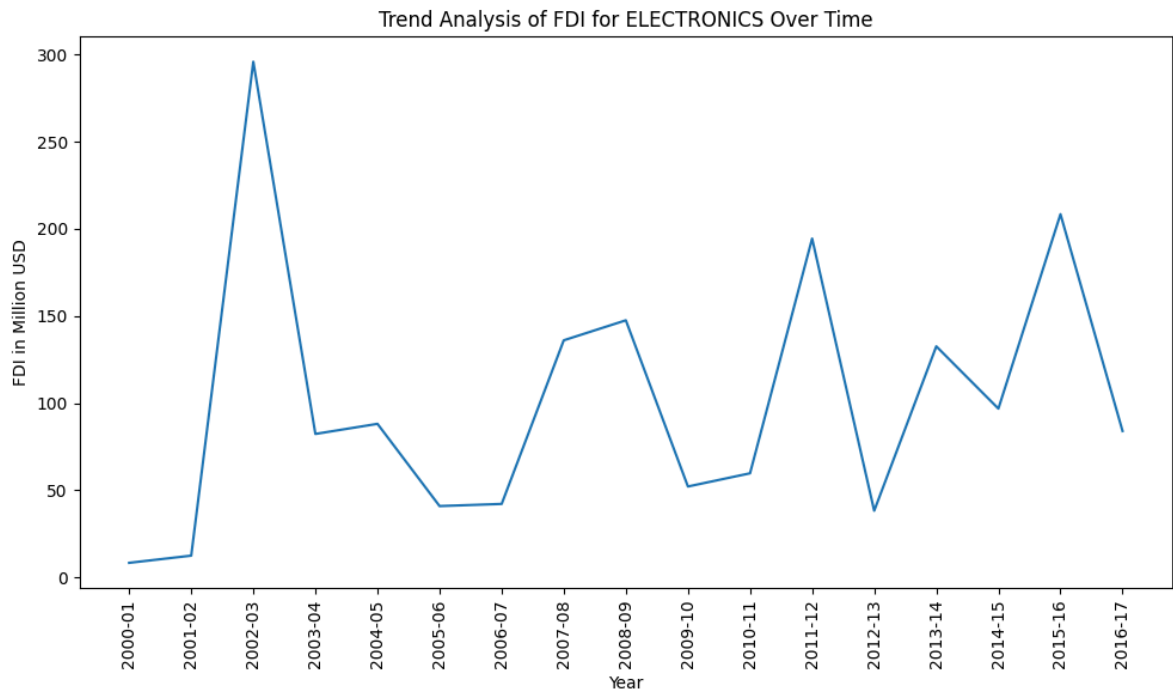


Trend Analysis of FDI for ELECTRICAL EQUIPMENTS Over Time

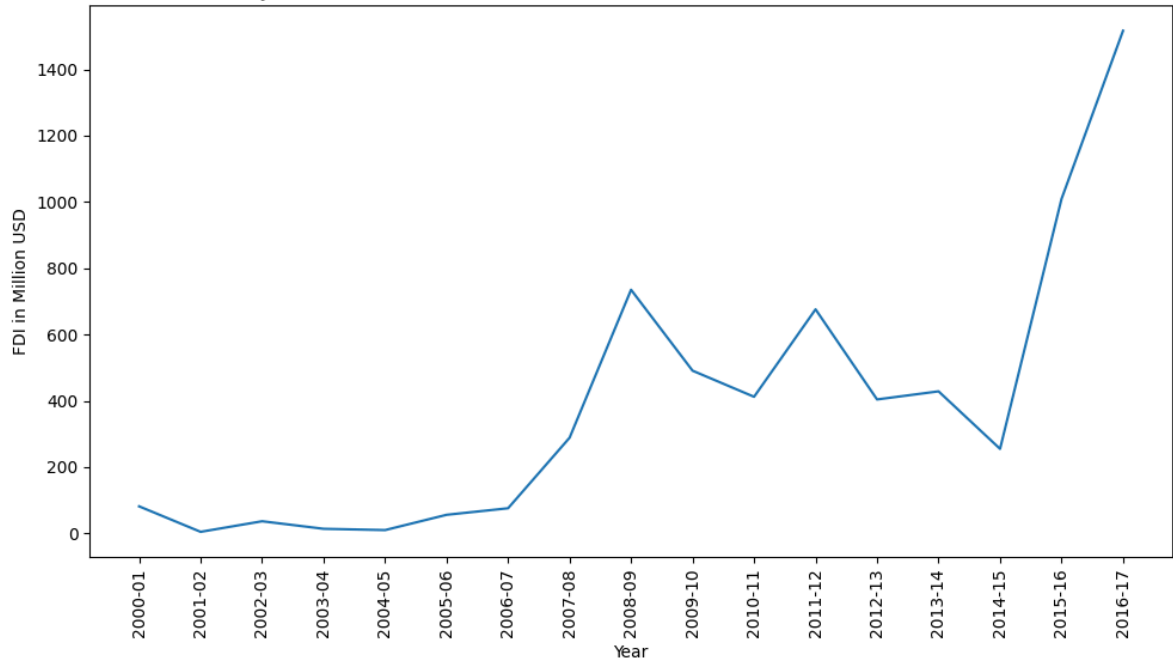


Trend Analysis of FDI for COMPUTER SOFTWARE & HARDWARE Over Time

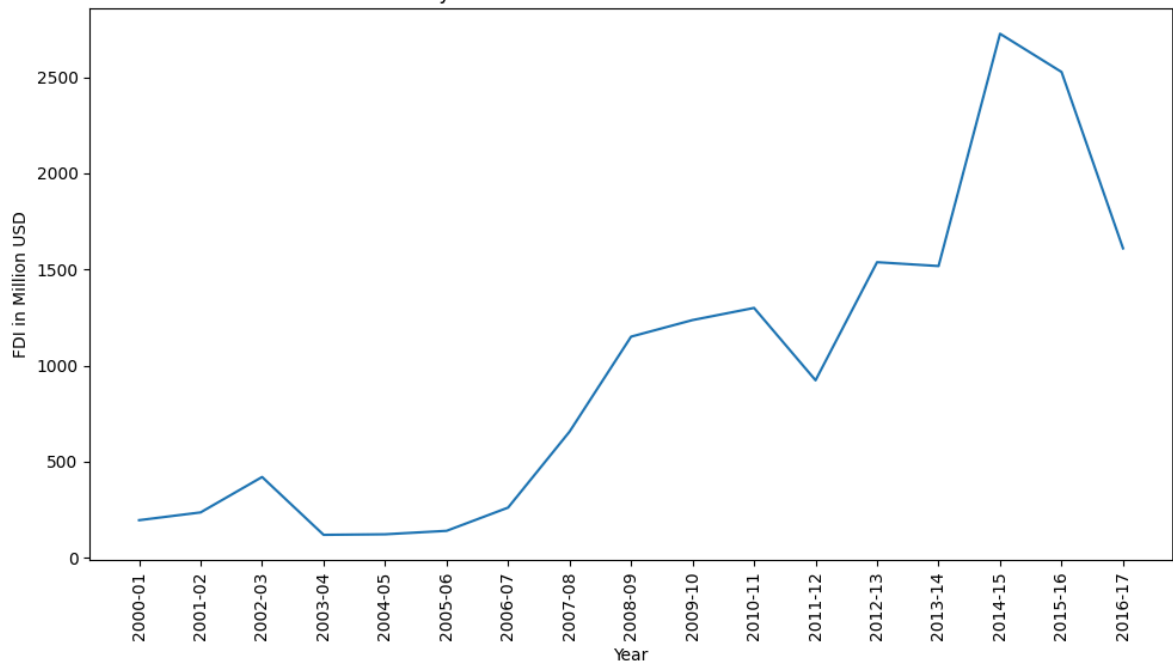




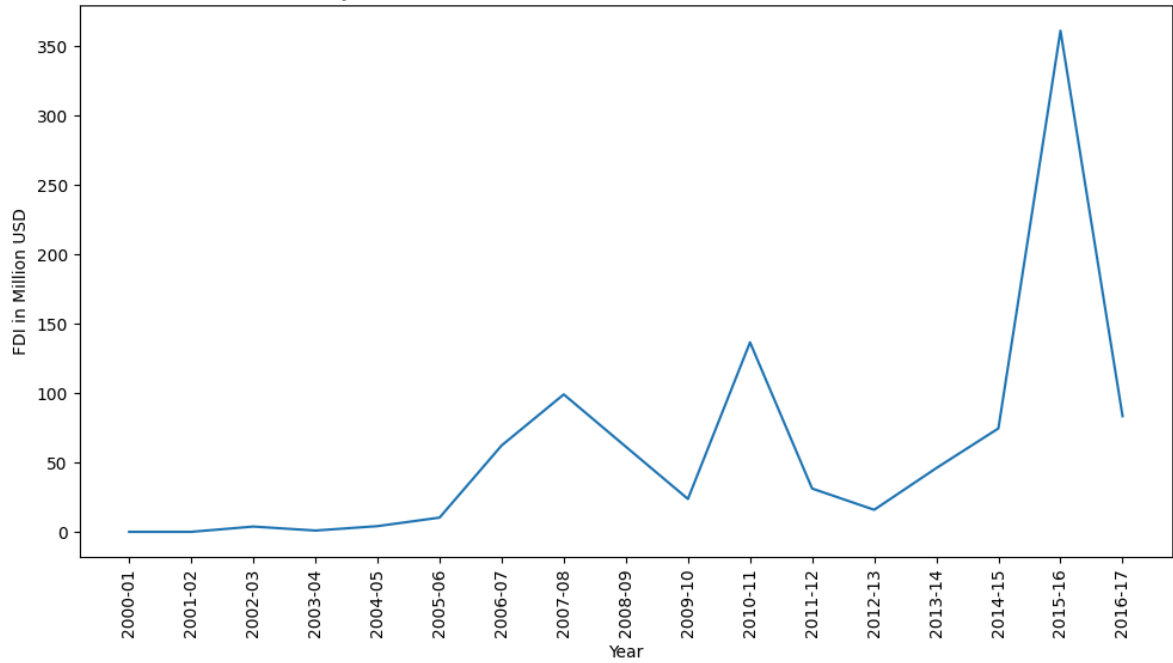
Trend Analysis of FDI for INFORMATION & BROADCASTING (INCLUDING PRINT MEDIA) Over Time



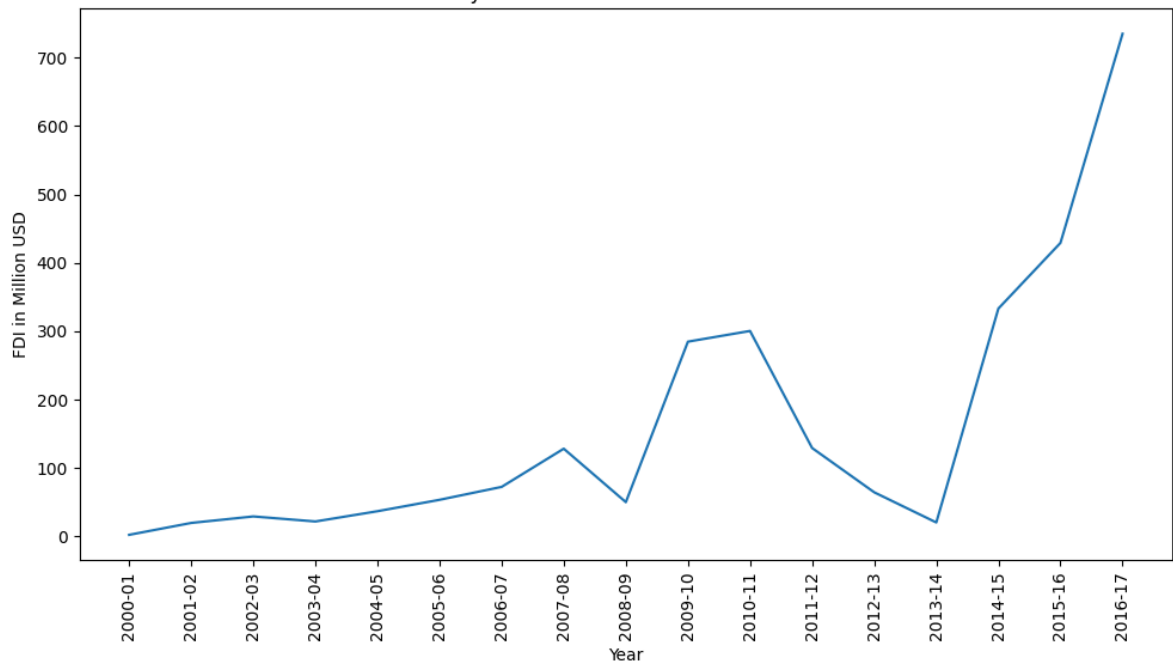
Trend Analysis of FDI for AUTOMOBILE INDUSTRY Over Time

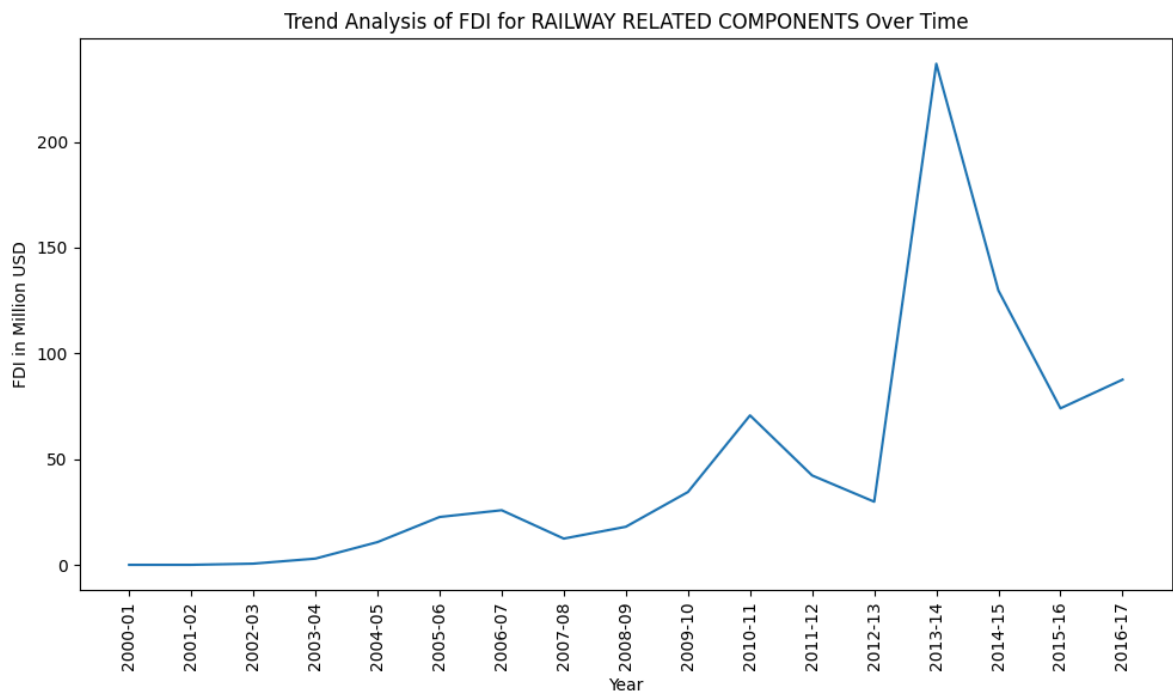
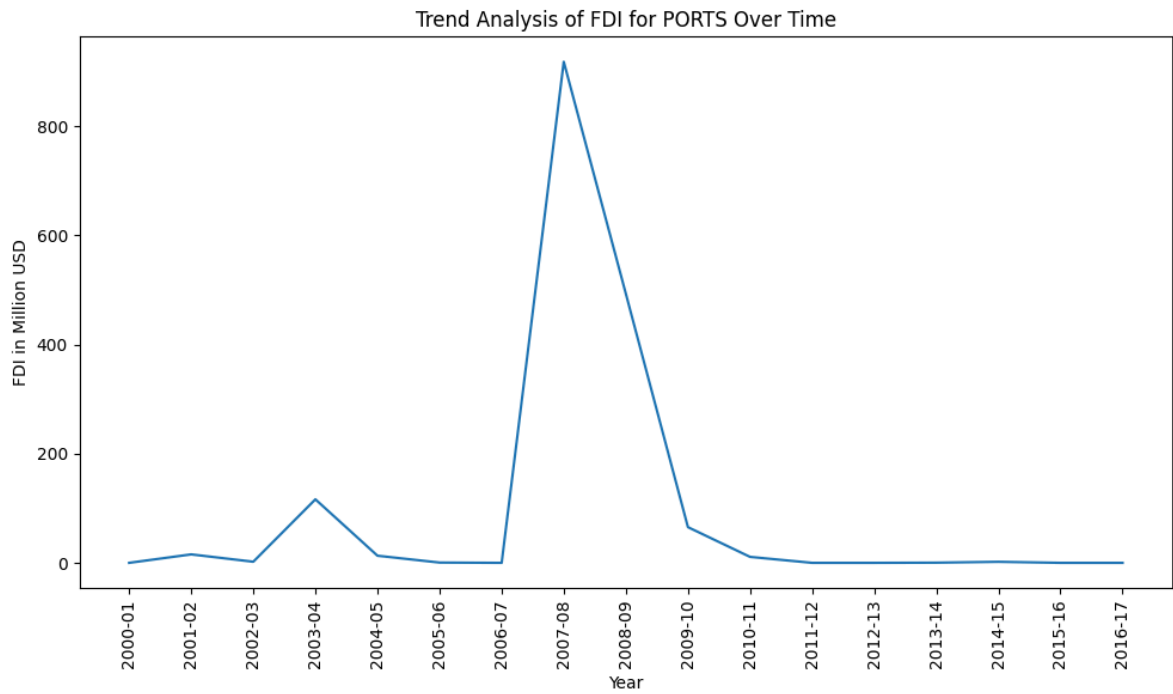


Trend Analysis of FDI for AIR TRANSPORT (INCLUDING AIR FREIGHT) Over Time

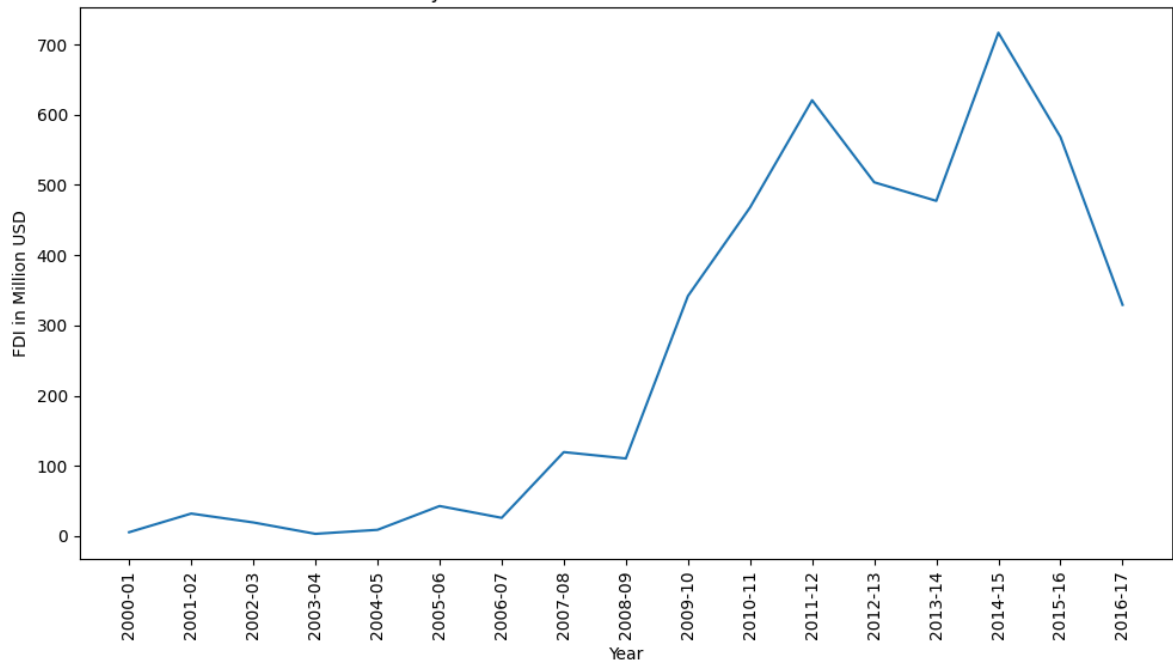


Trend Analysis of FDI for SEA TRANSPORT Over Time

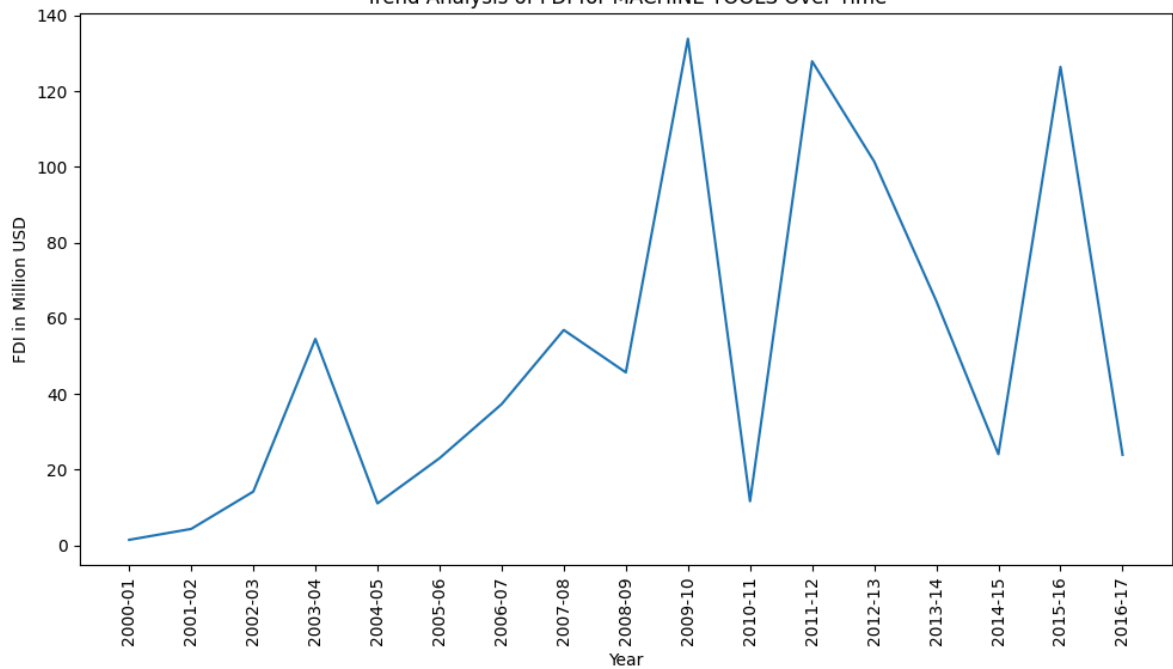


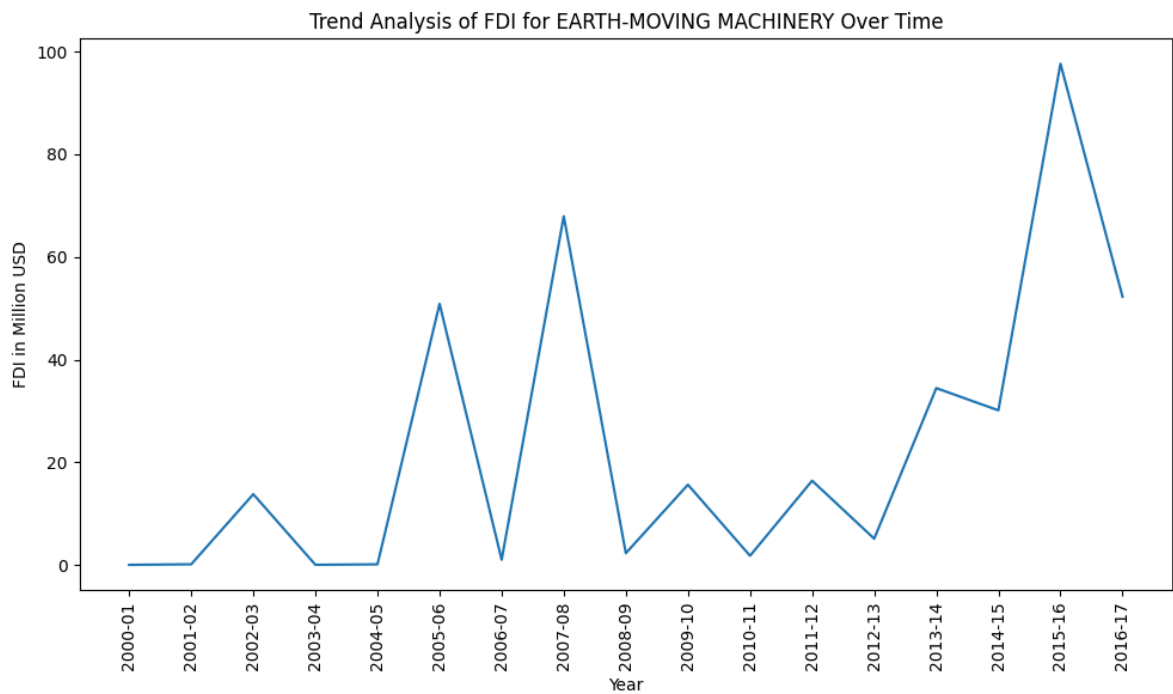
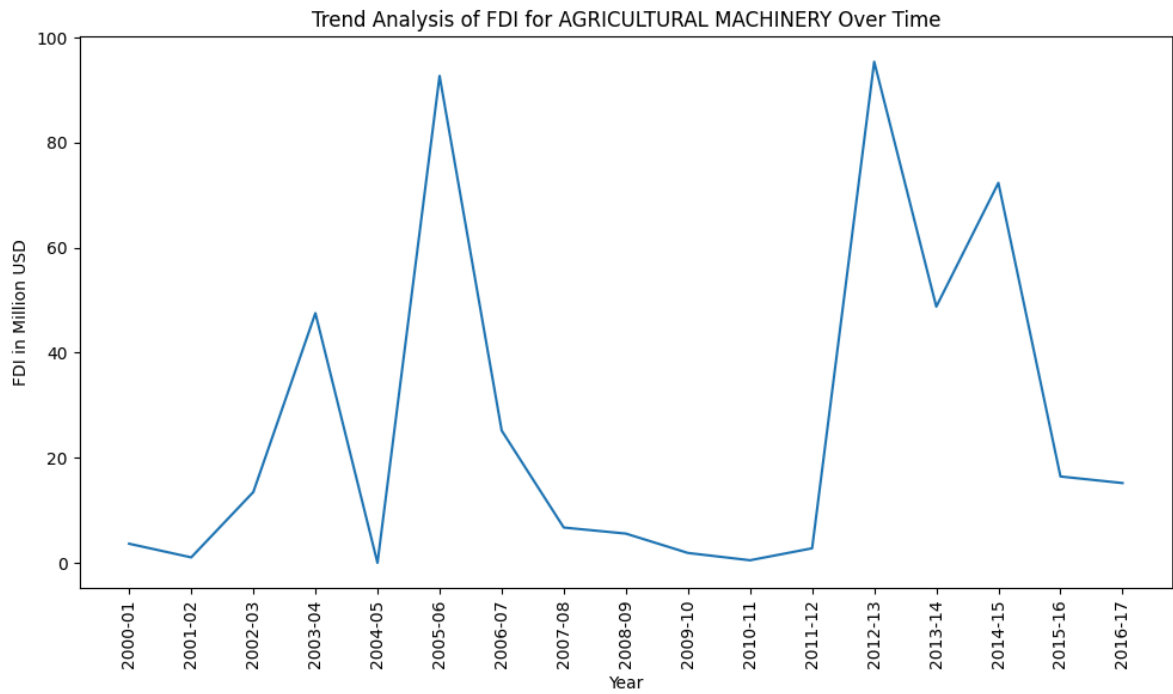


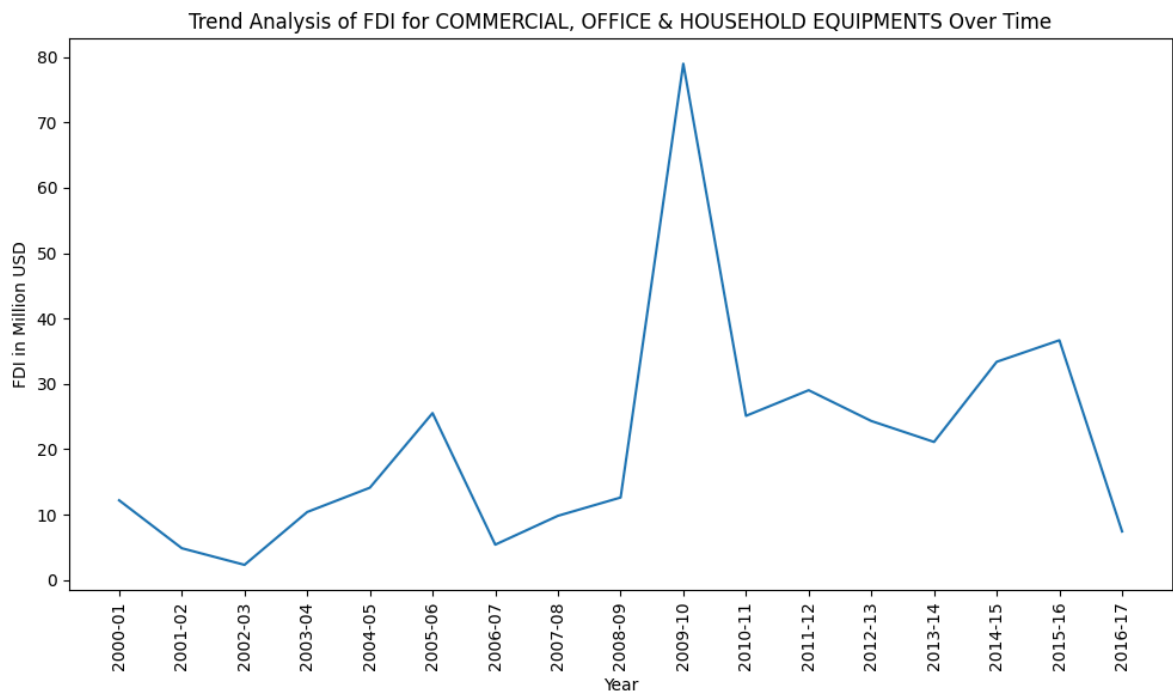
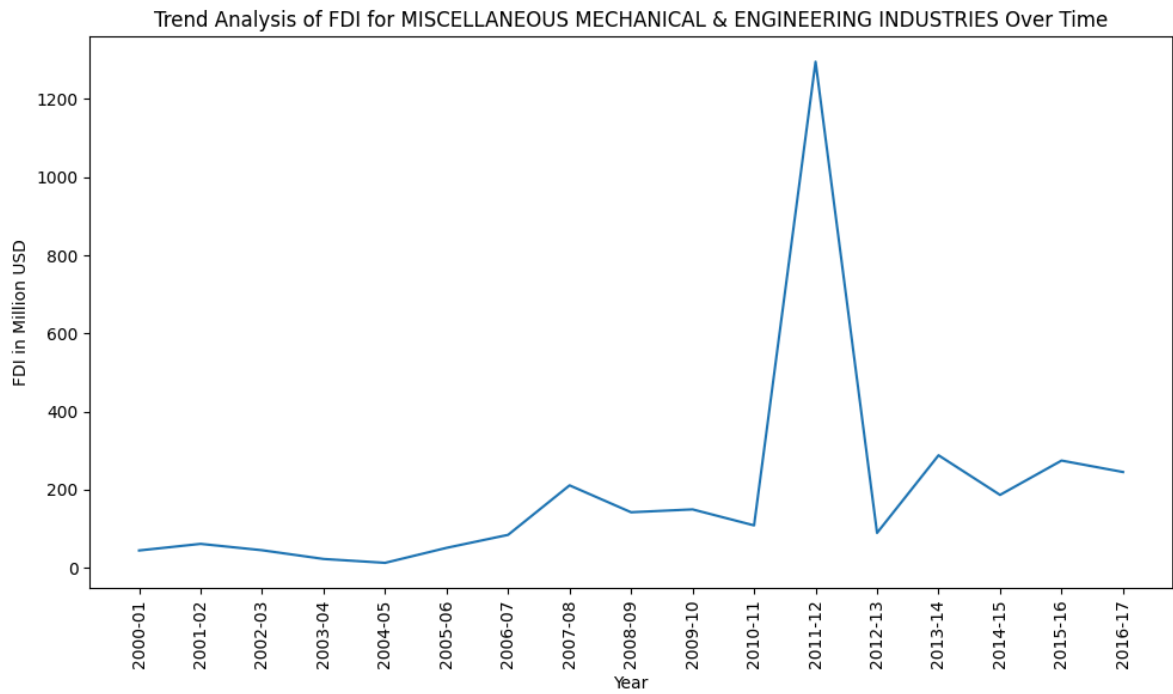
Trend Analysis of FDI for INDUSTRIAL MACHINERY Over Time

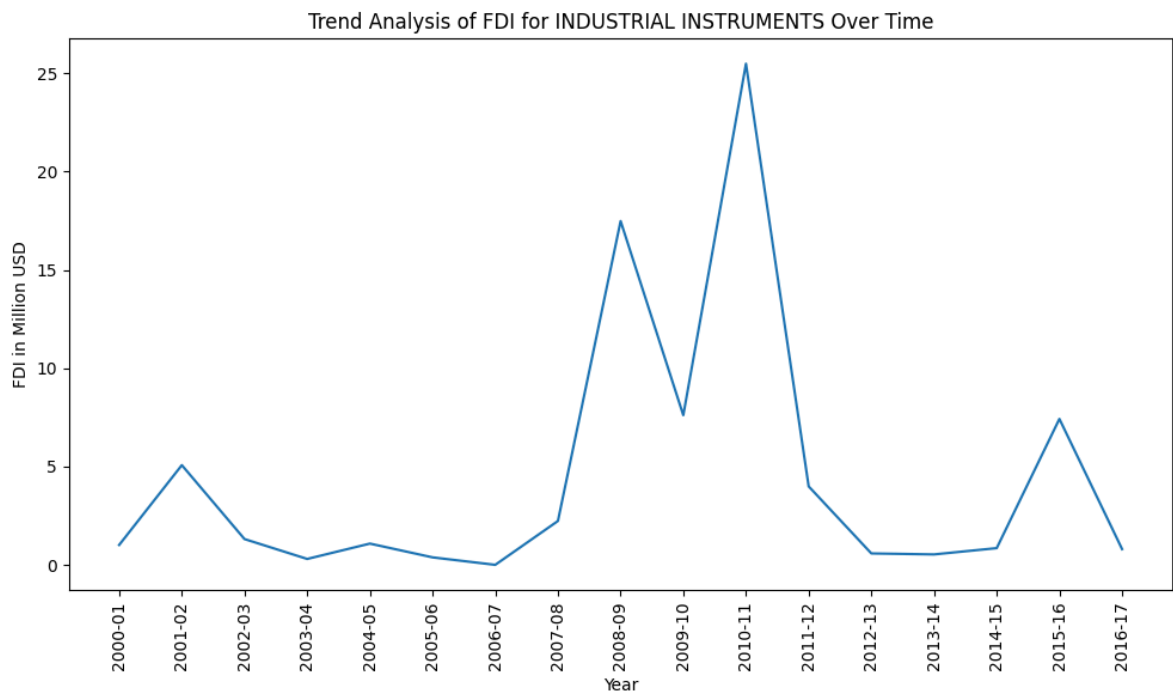
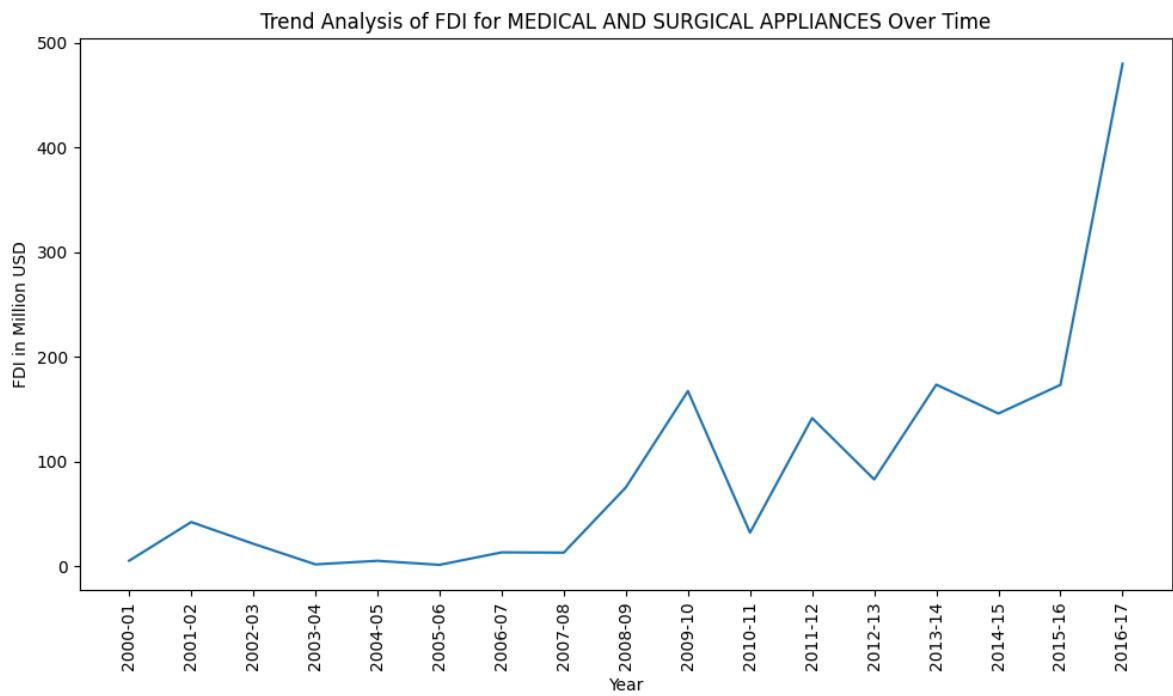


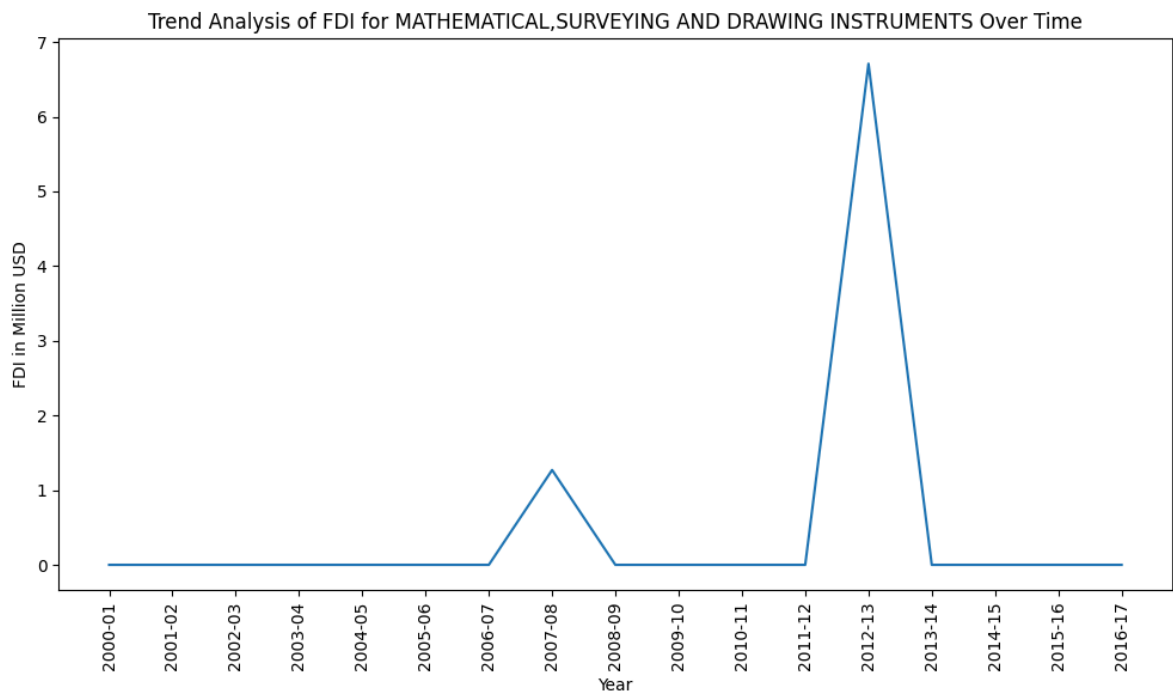
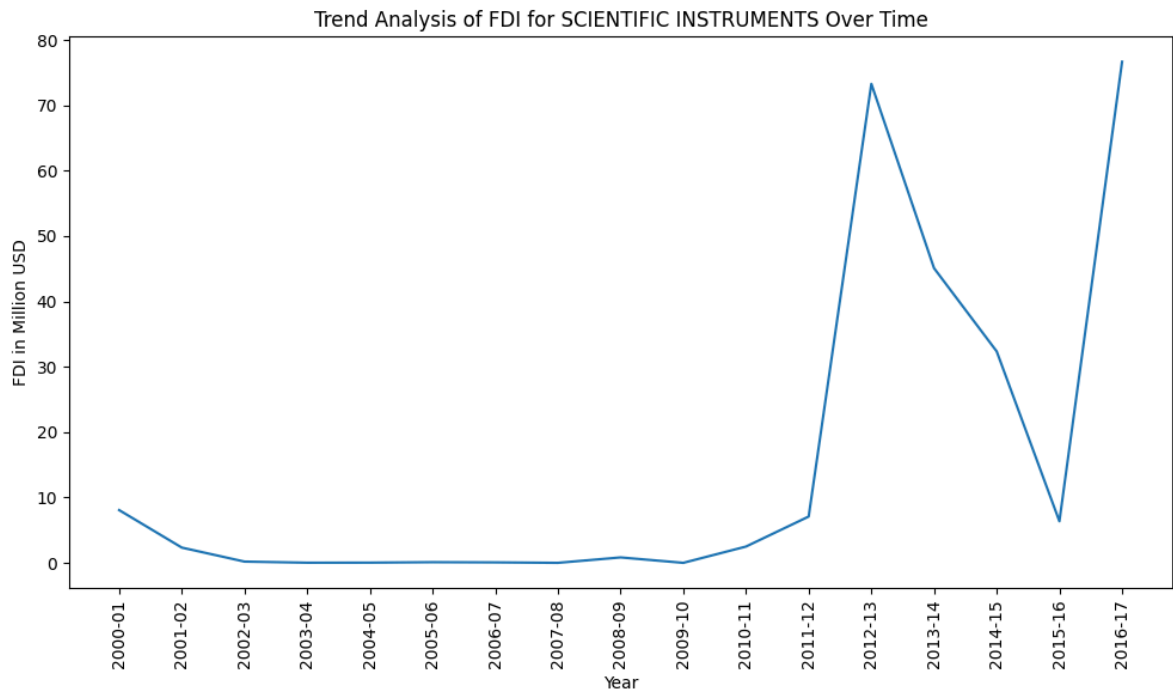
Trend Analysis of FDI for MACHINE TOOLS Over Time

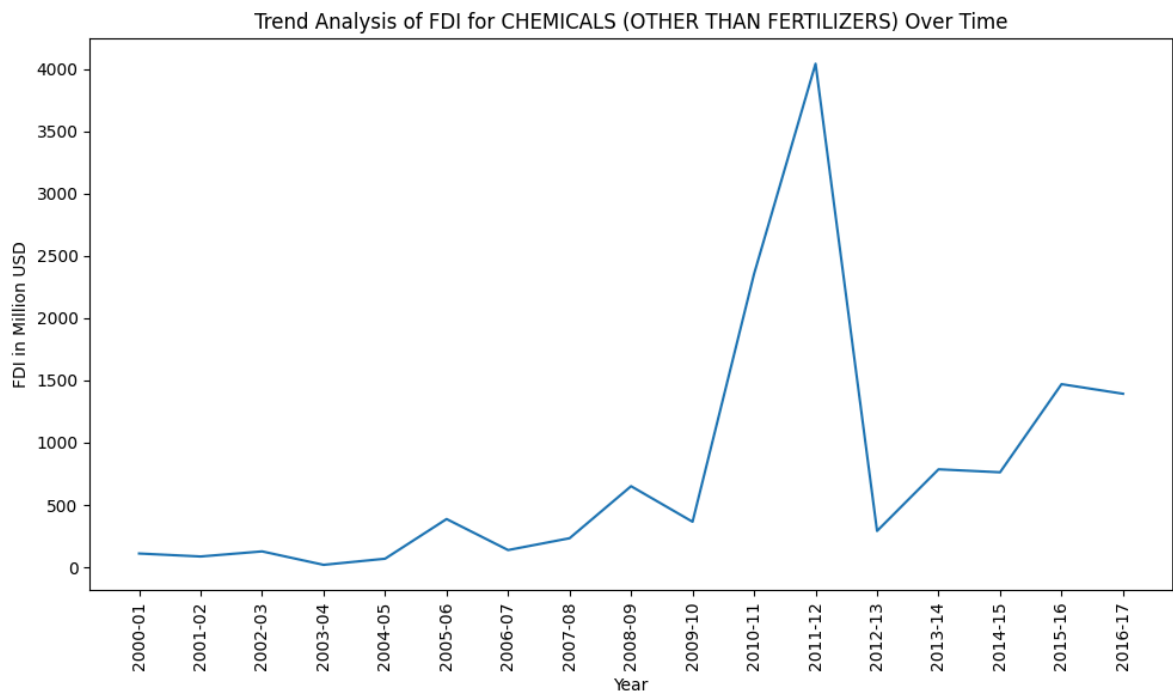
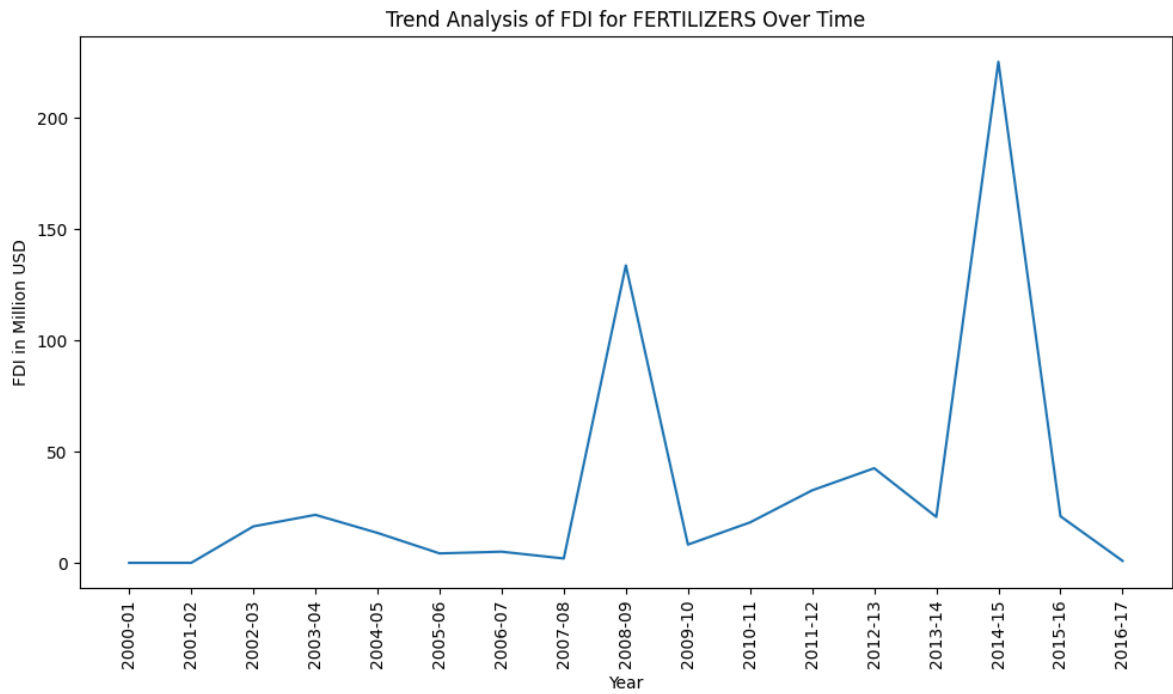




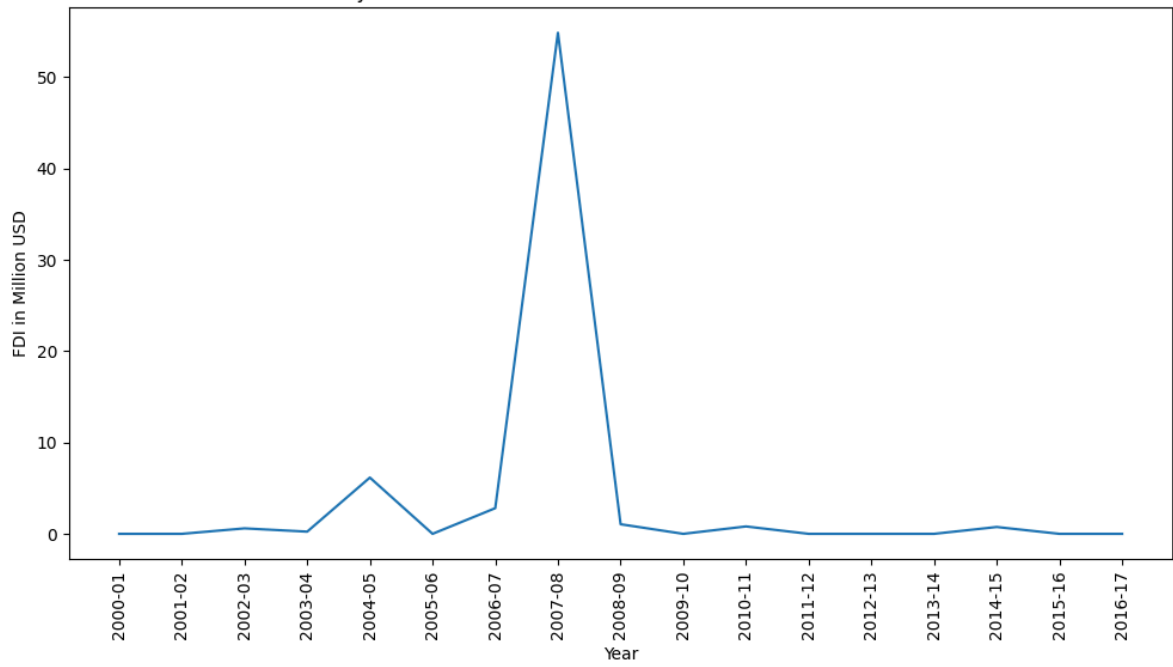




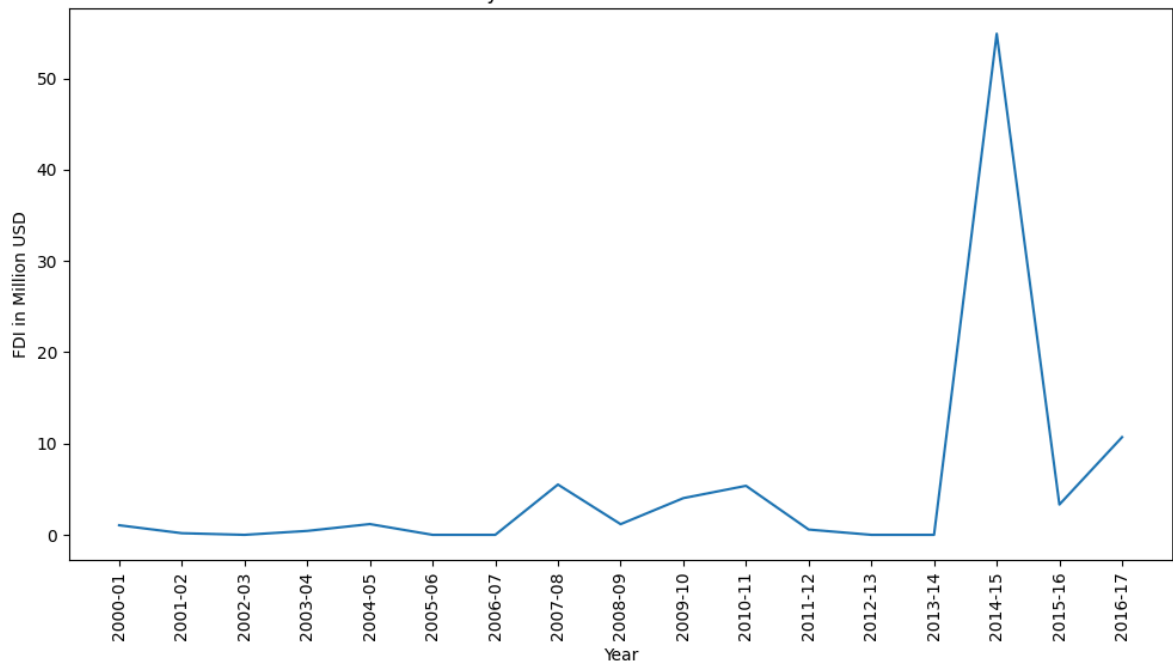


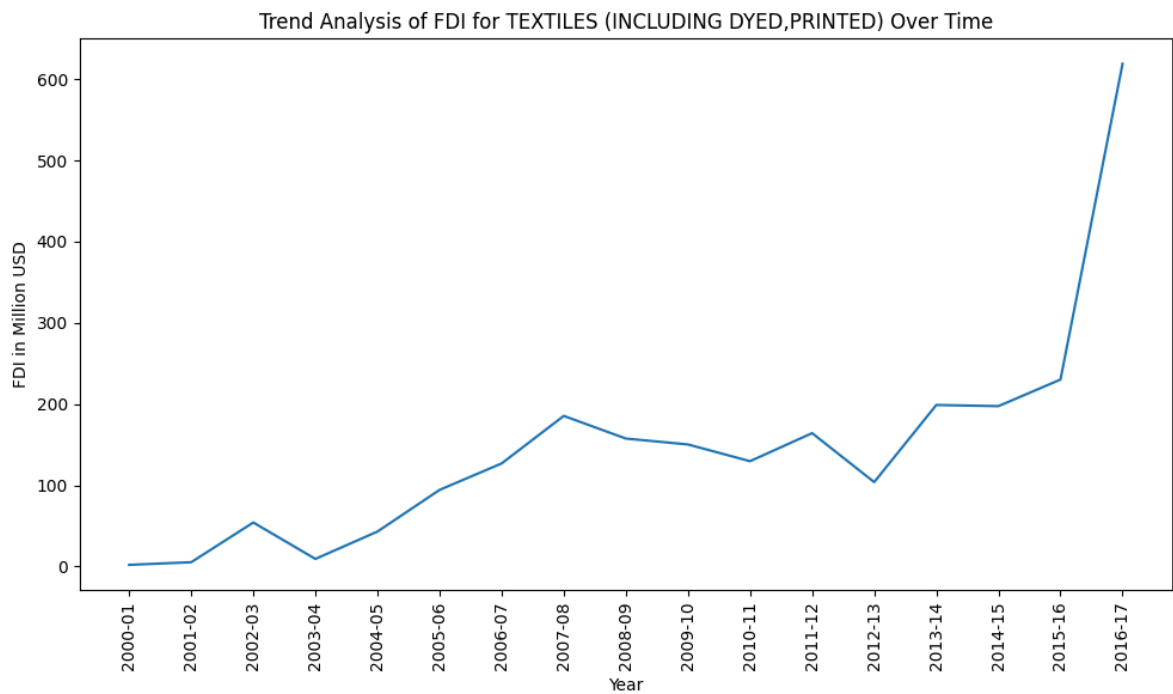
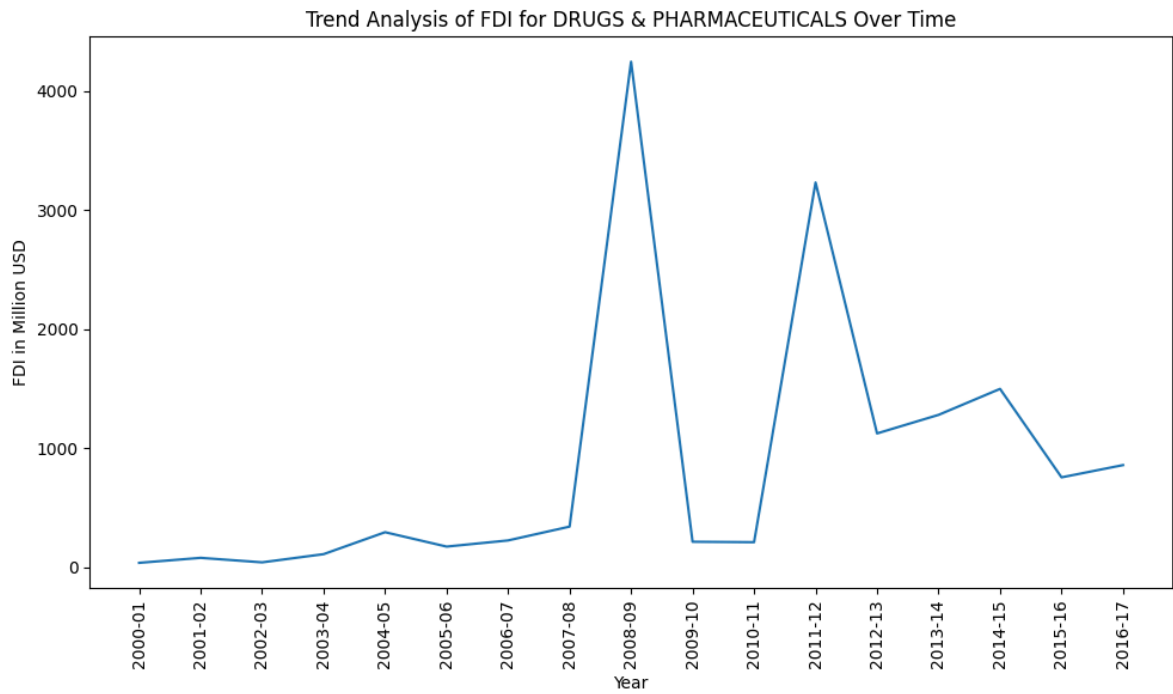


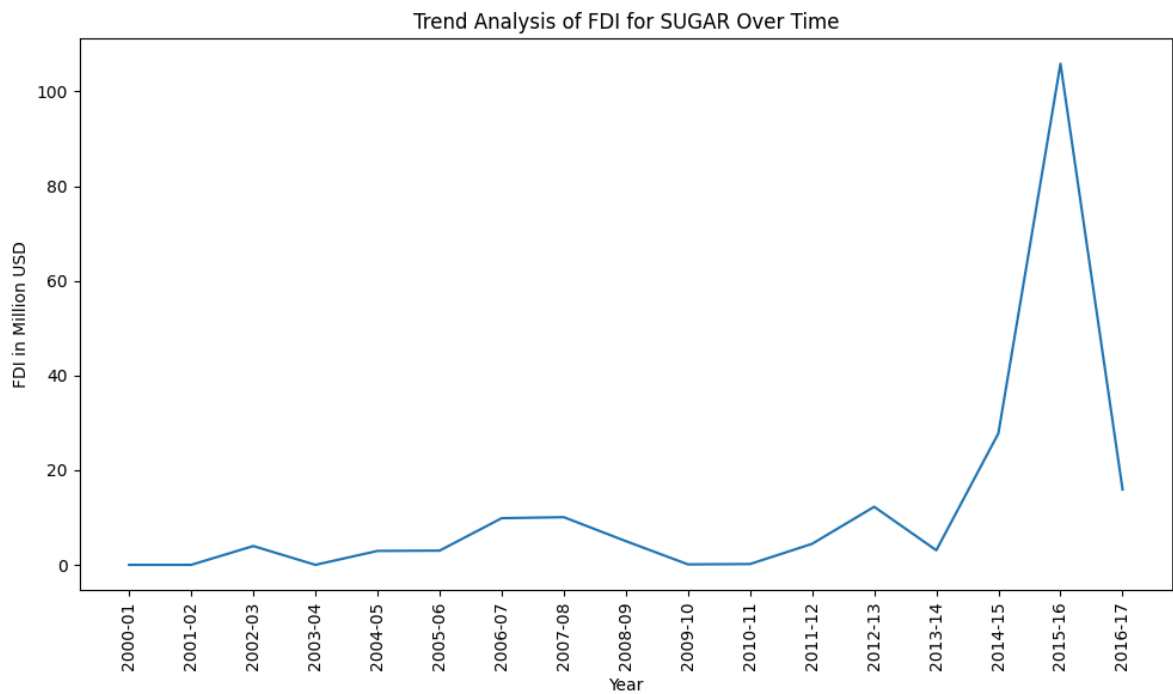
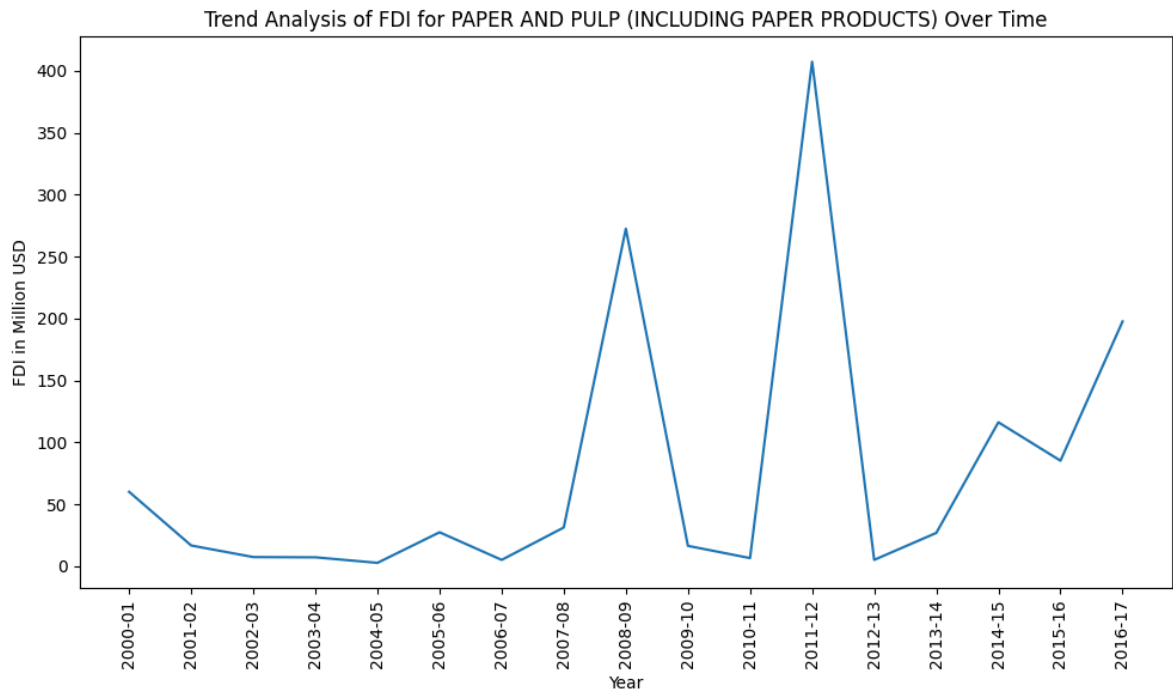
Trend Analysis of FDI for PHOTOGRAPHIC RAW FILM AND PAPER Over Time

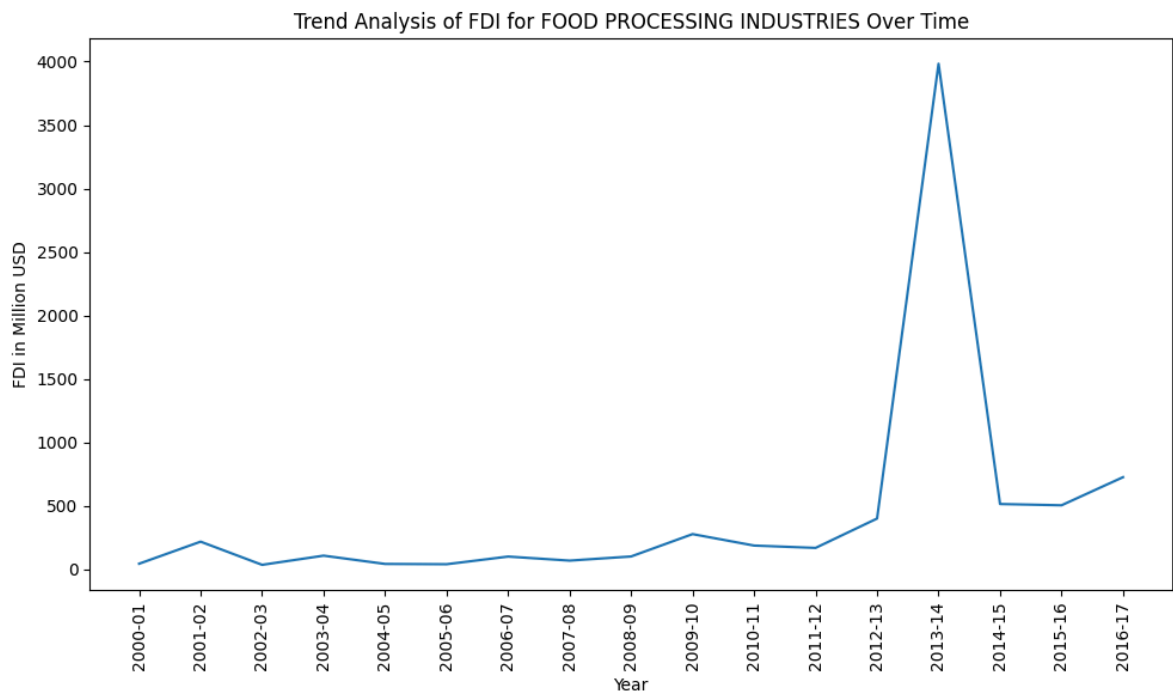
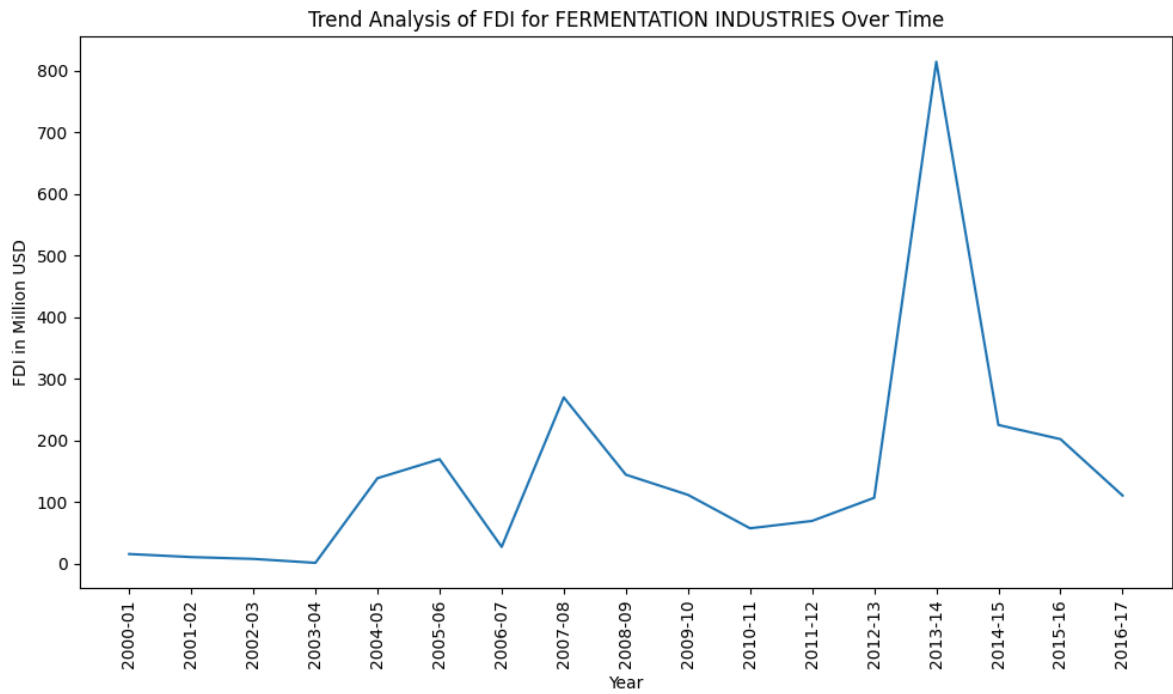


Trend Analysis of FDI for DYE-STUFFS Over Time

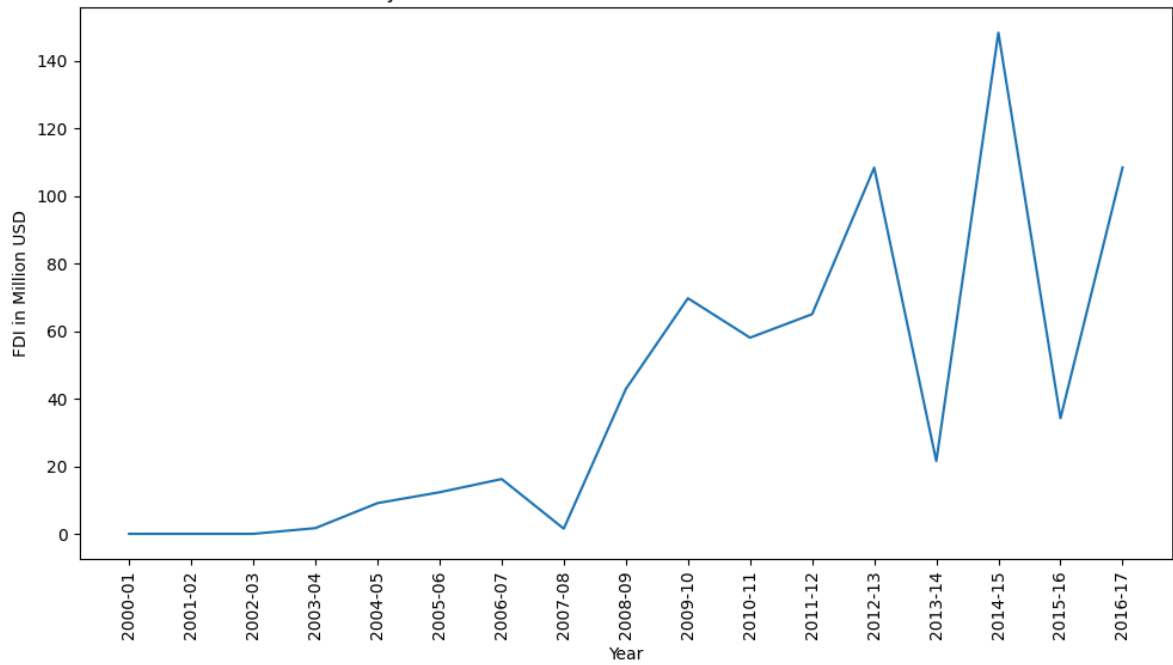




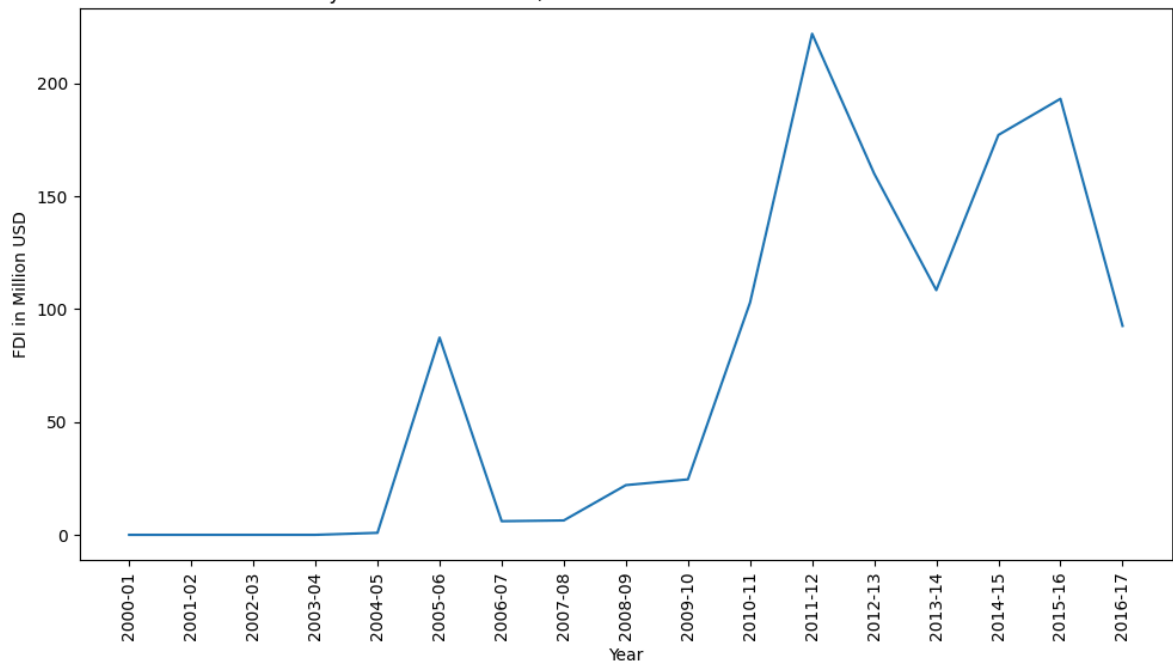


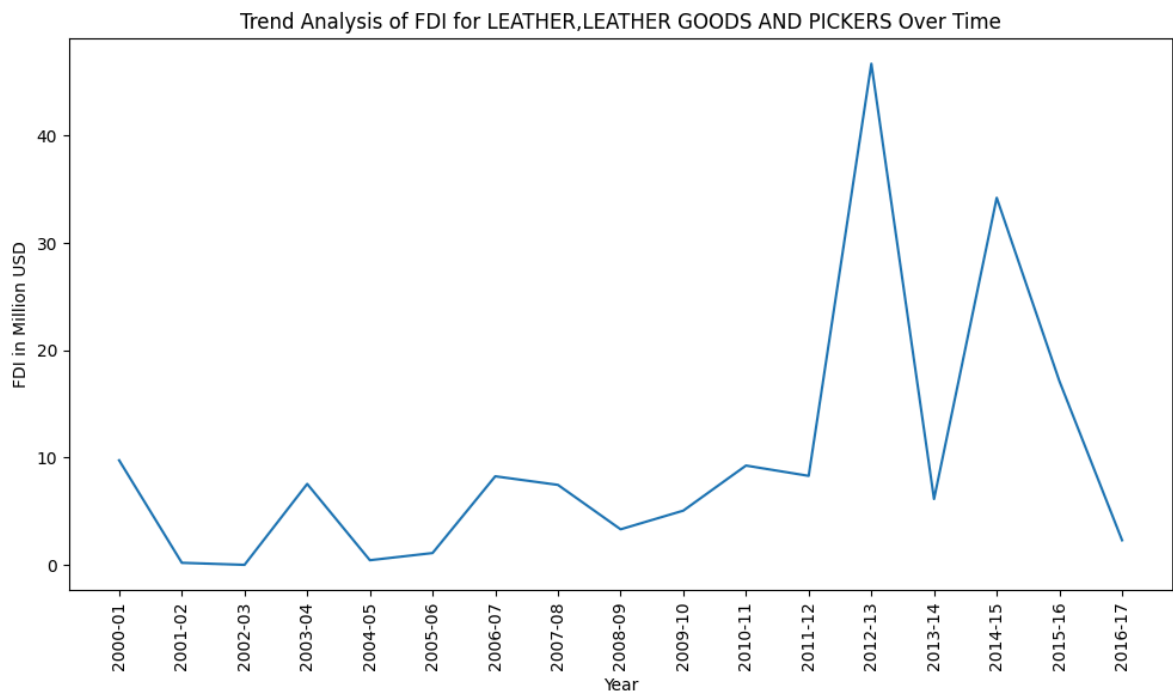
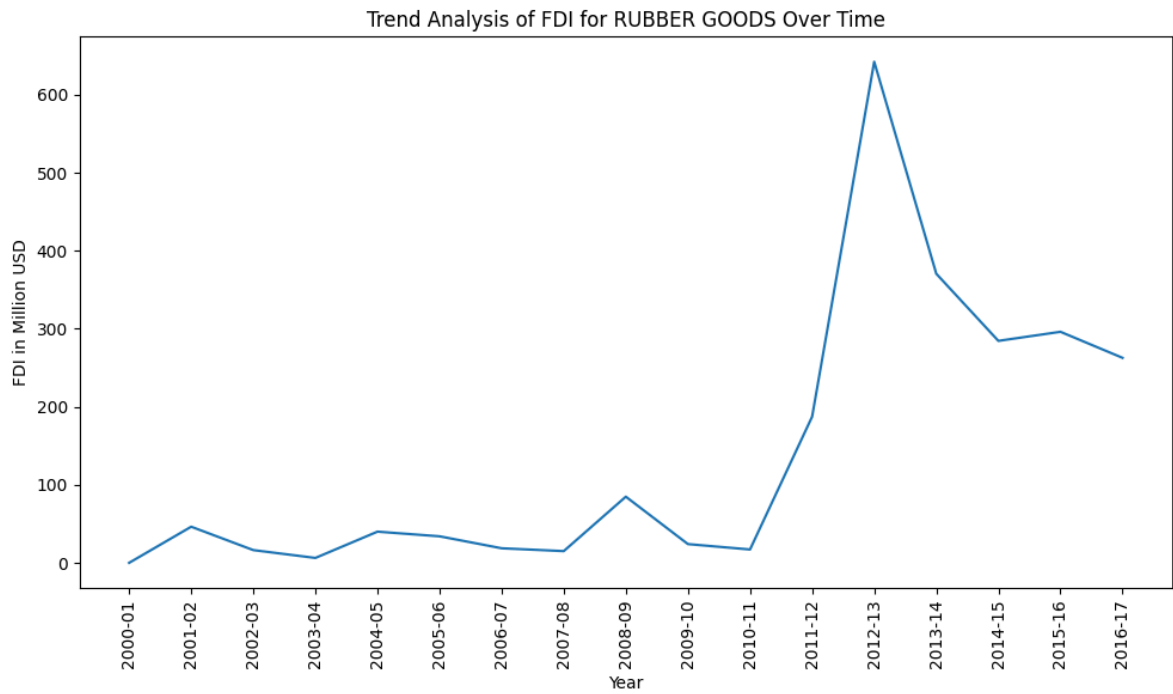


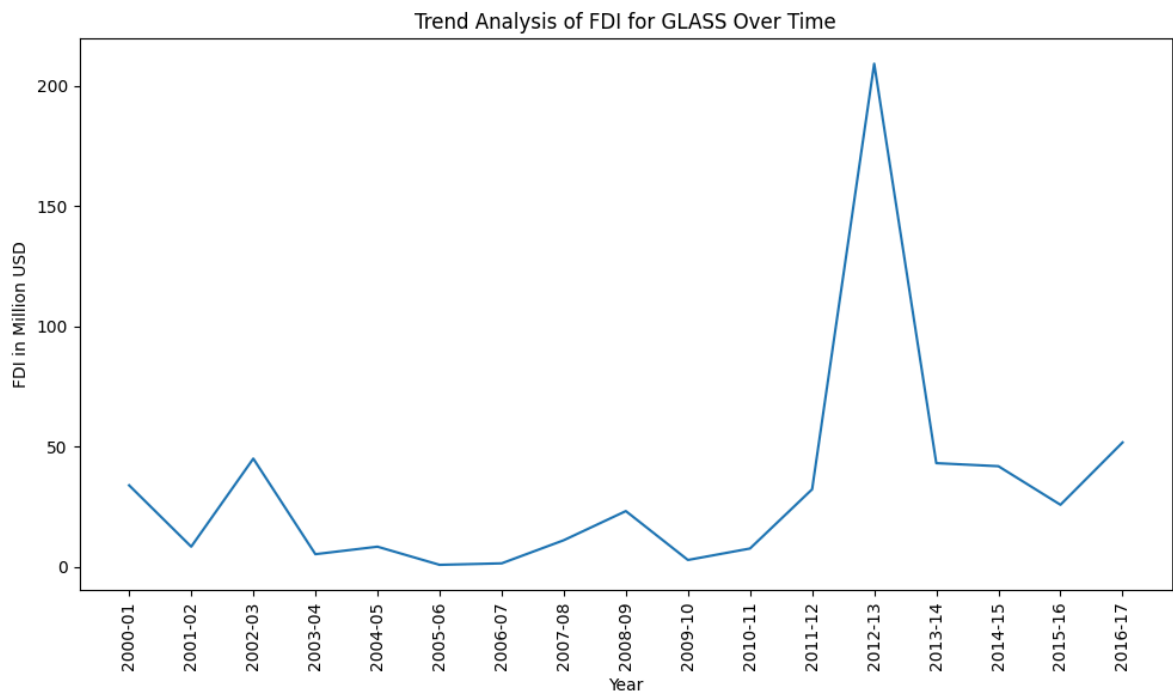
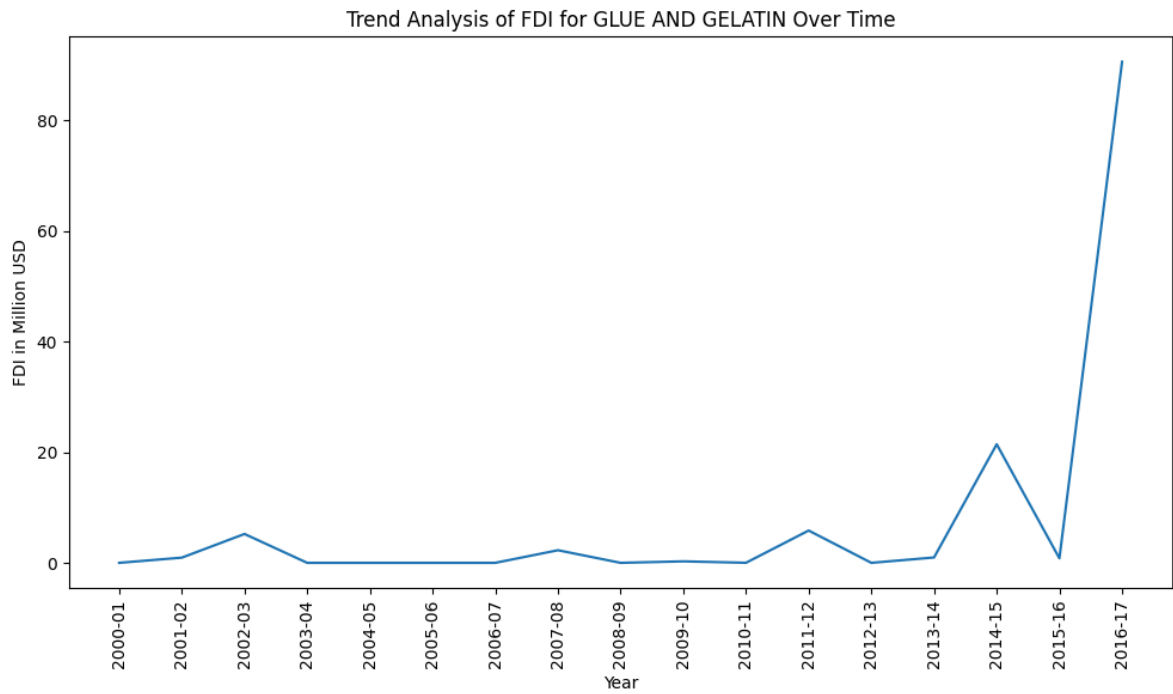
Trend Analysis of FDI for VEGETABLE OILS AND VANASPATI Over Time

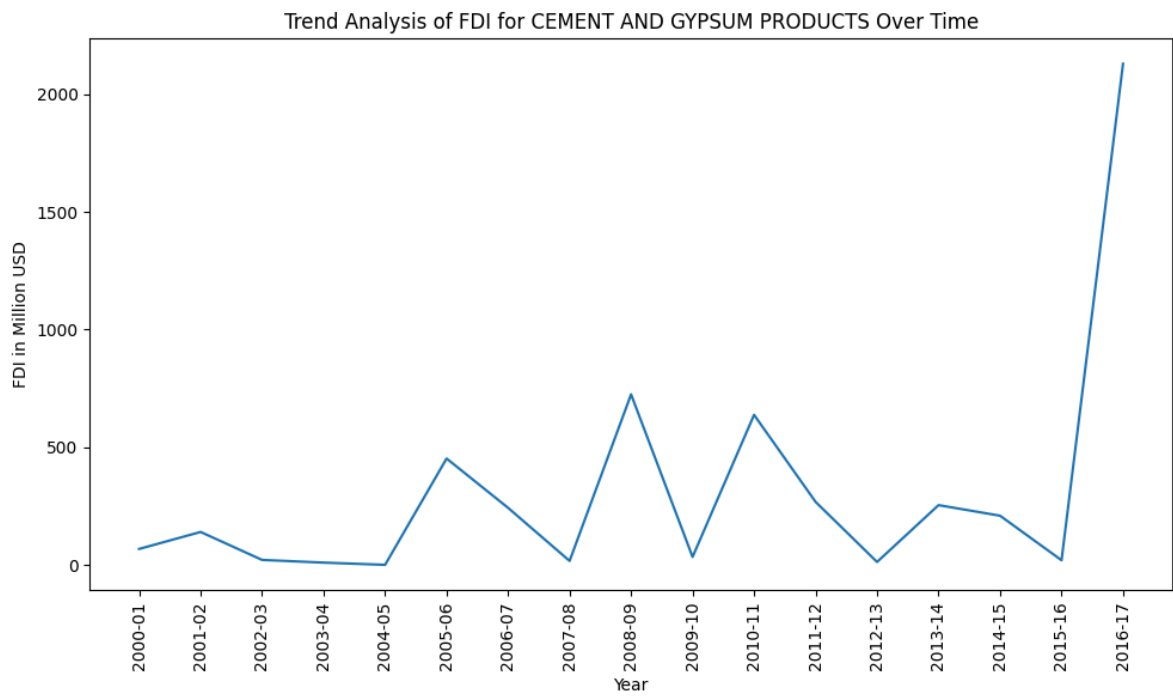
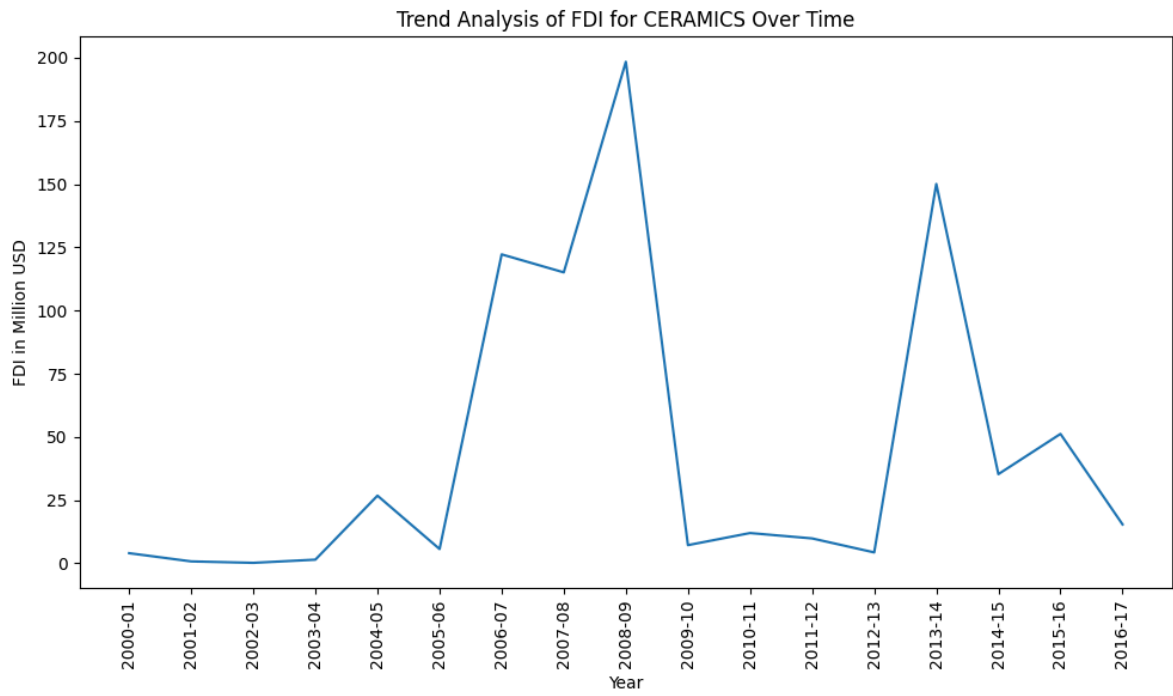


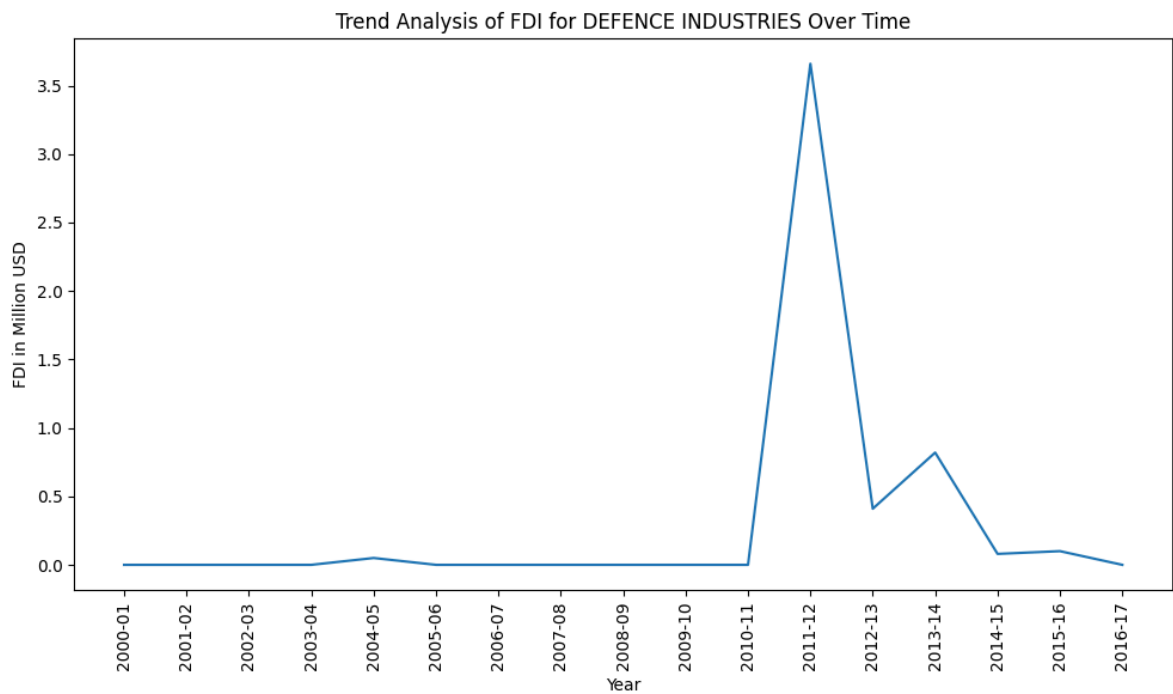
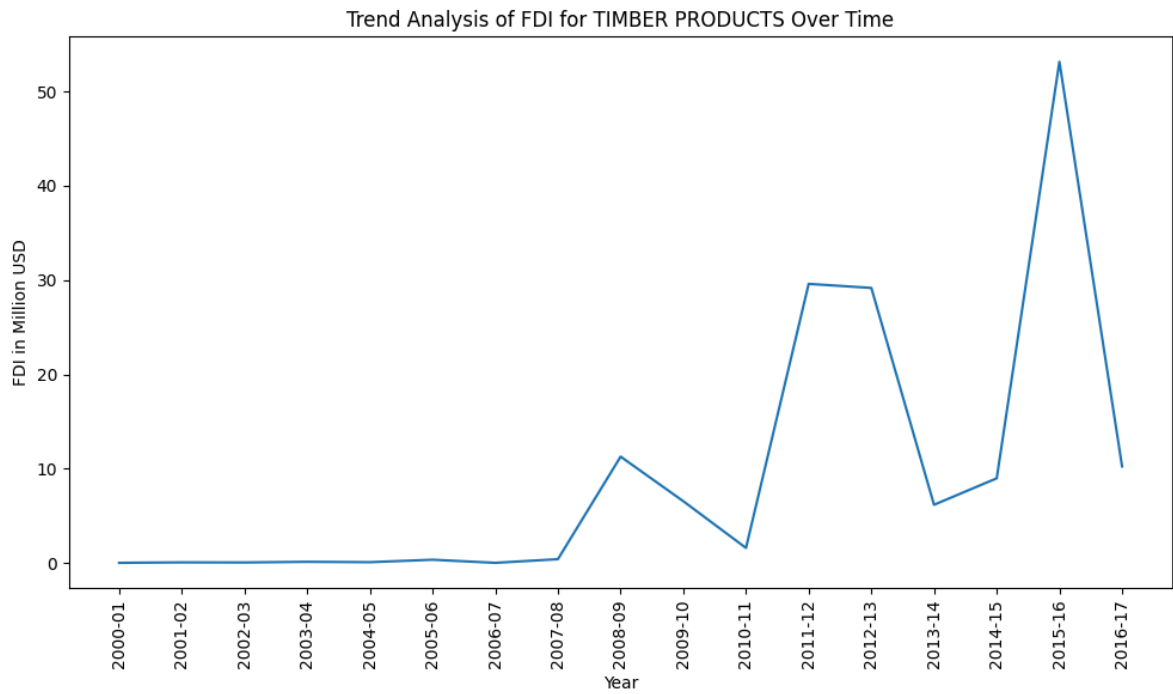
Trend Analysis of FDI for SOAPS, COSMETICS & TOILET PREPARATIONS Over Time

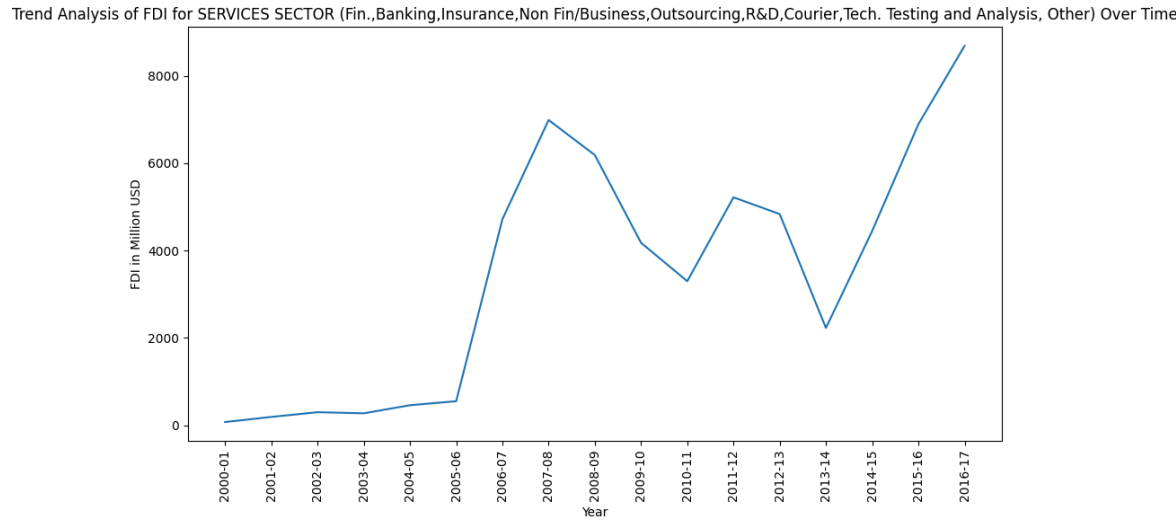
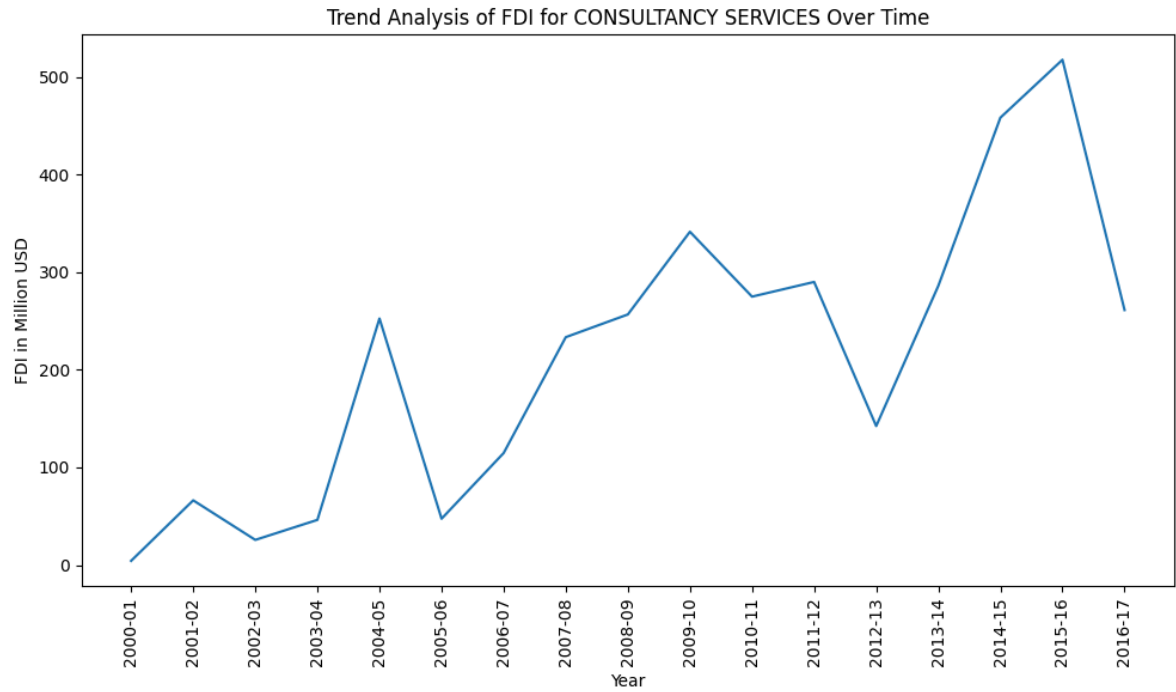




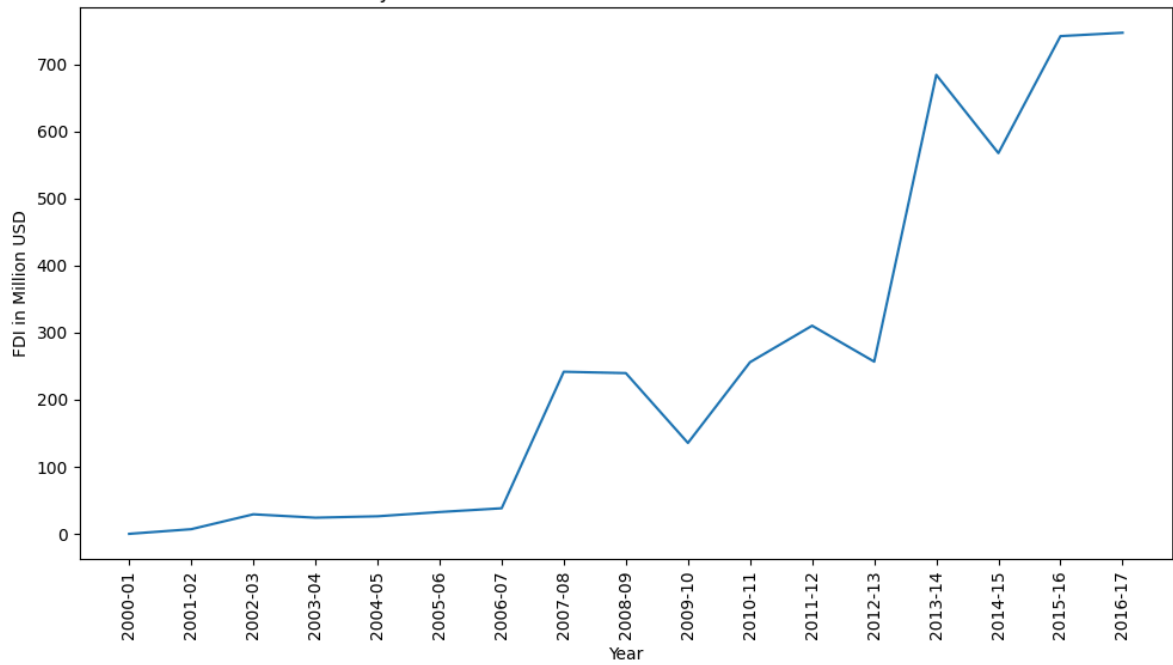




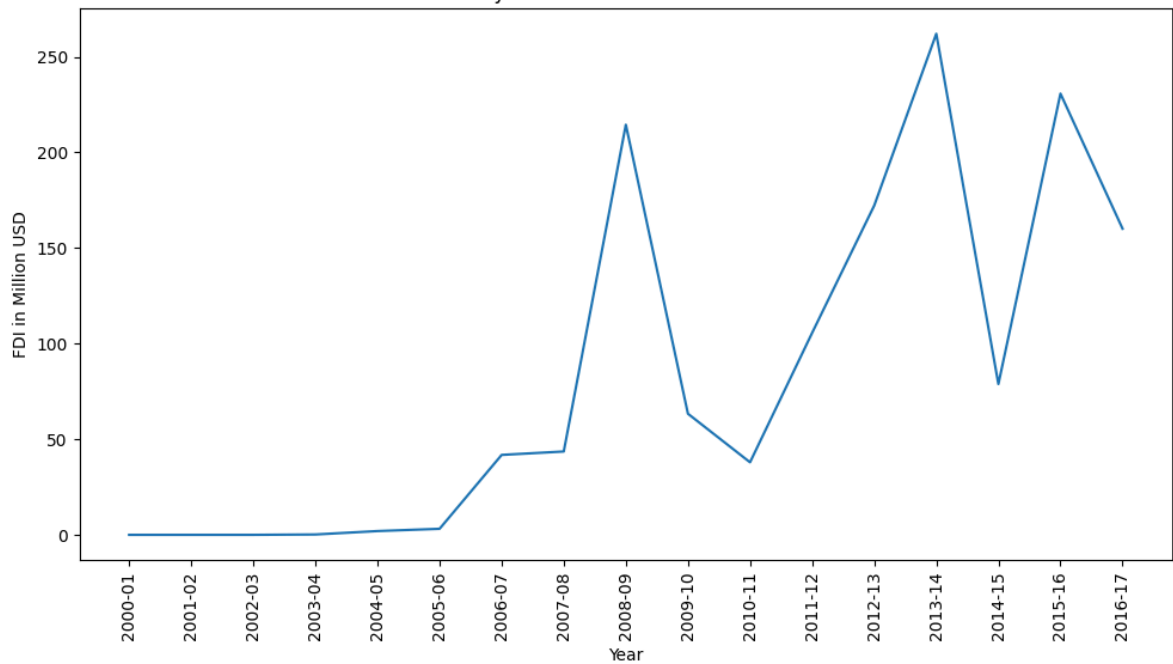


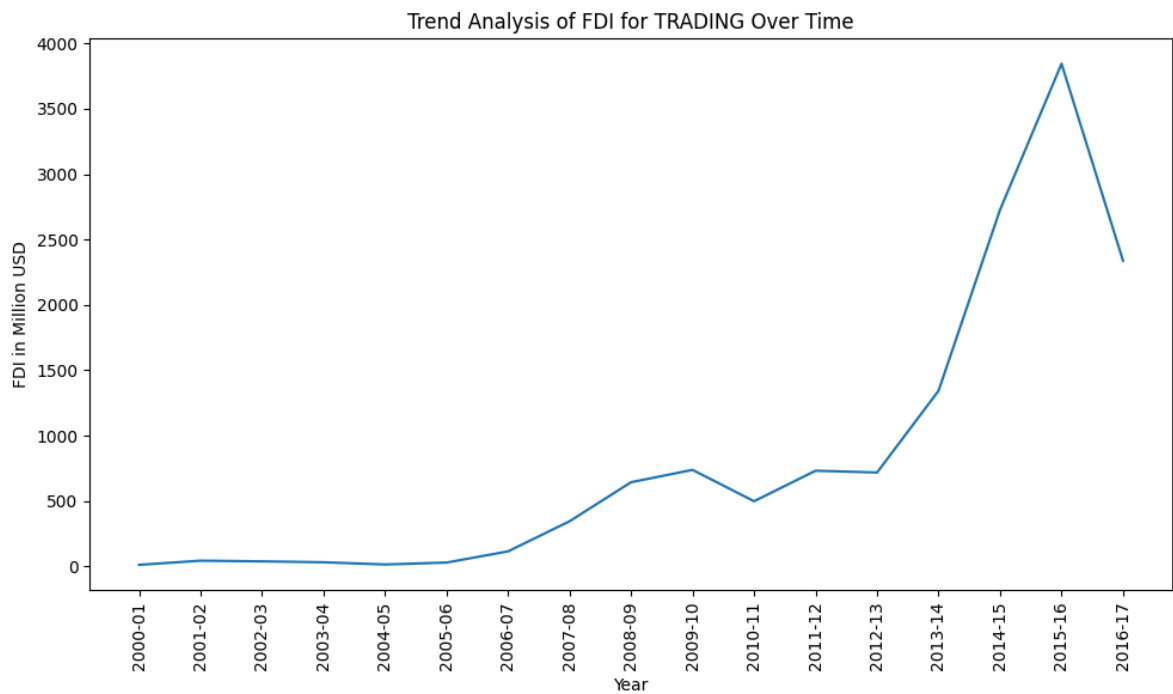
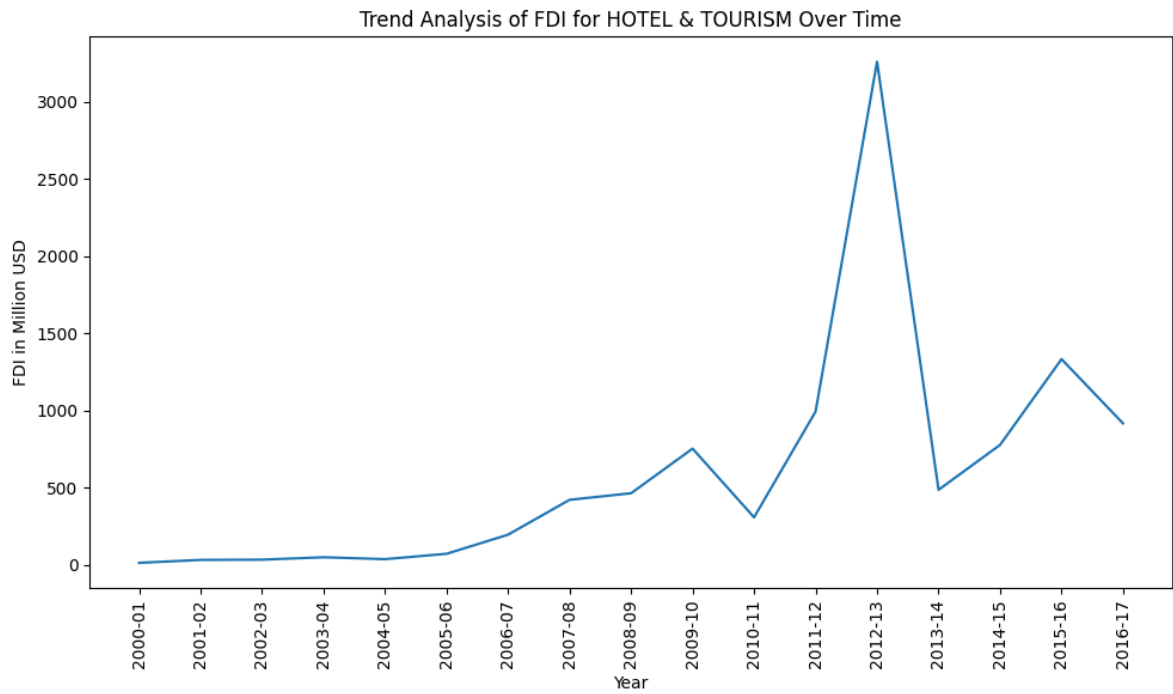


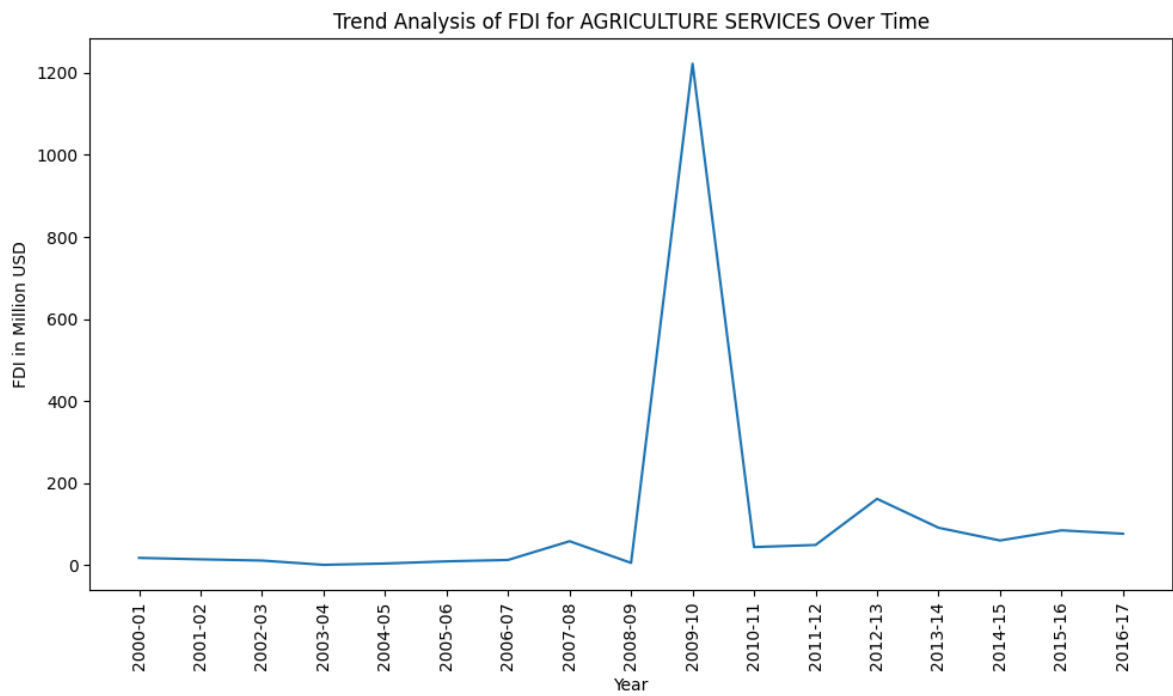
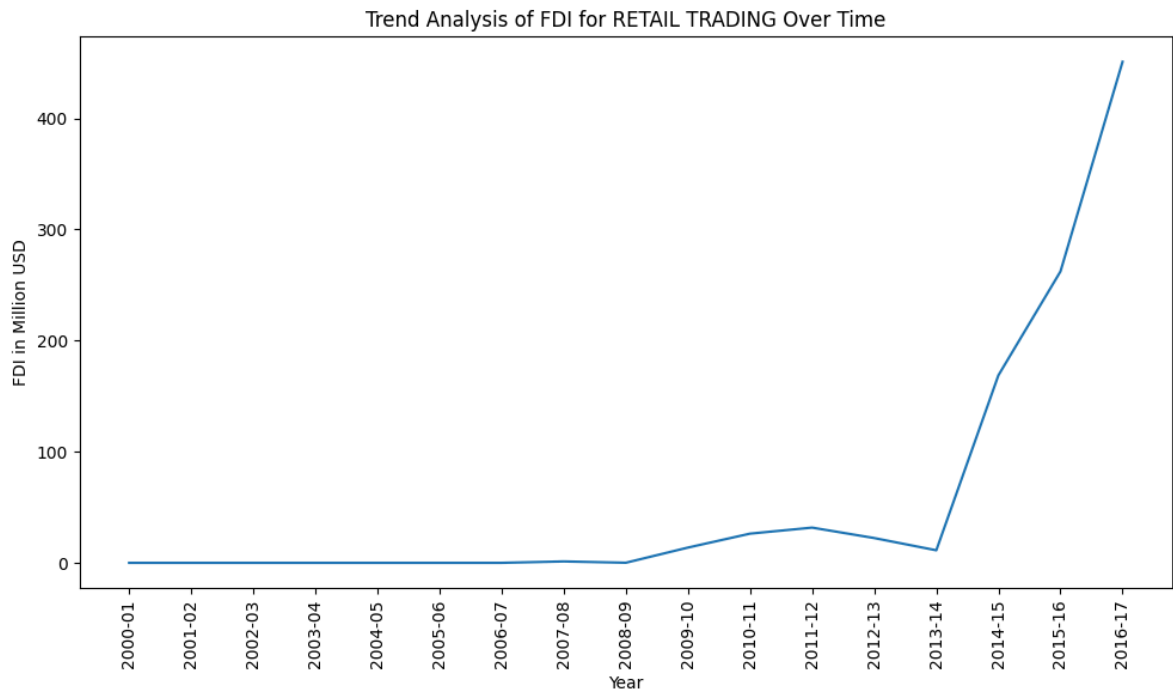
Trend Analysis of FDI for HOSPITAL & DIAGNOSTIC CENTRES Over Time



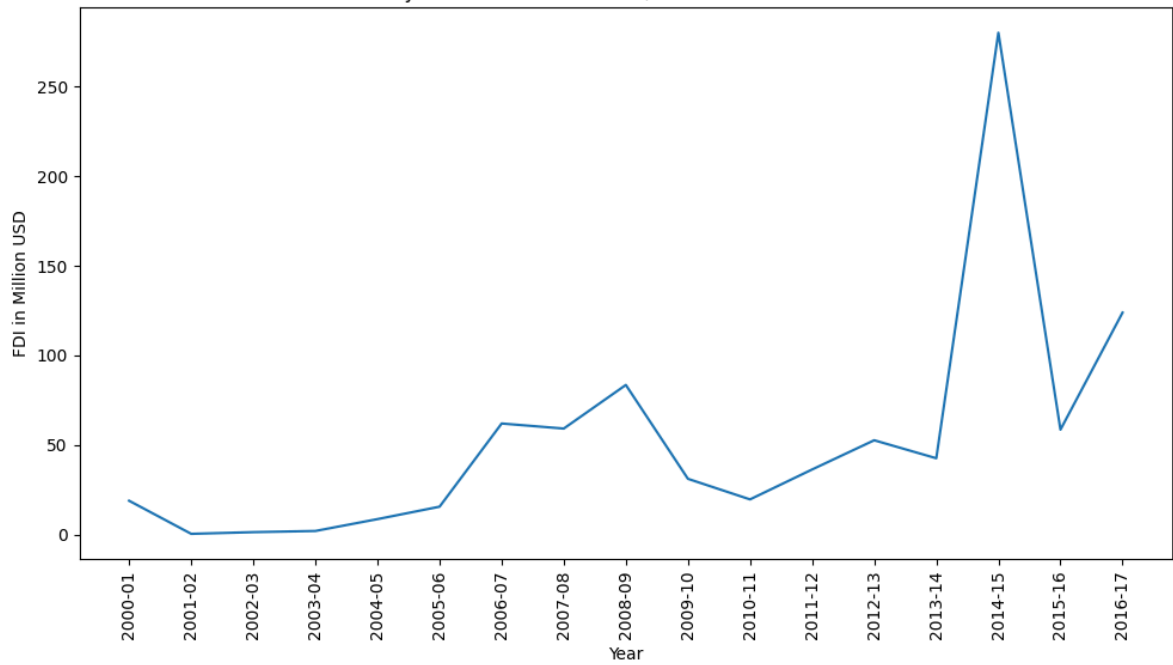
Trend Analysis of FDI for EDUCATION Over Time



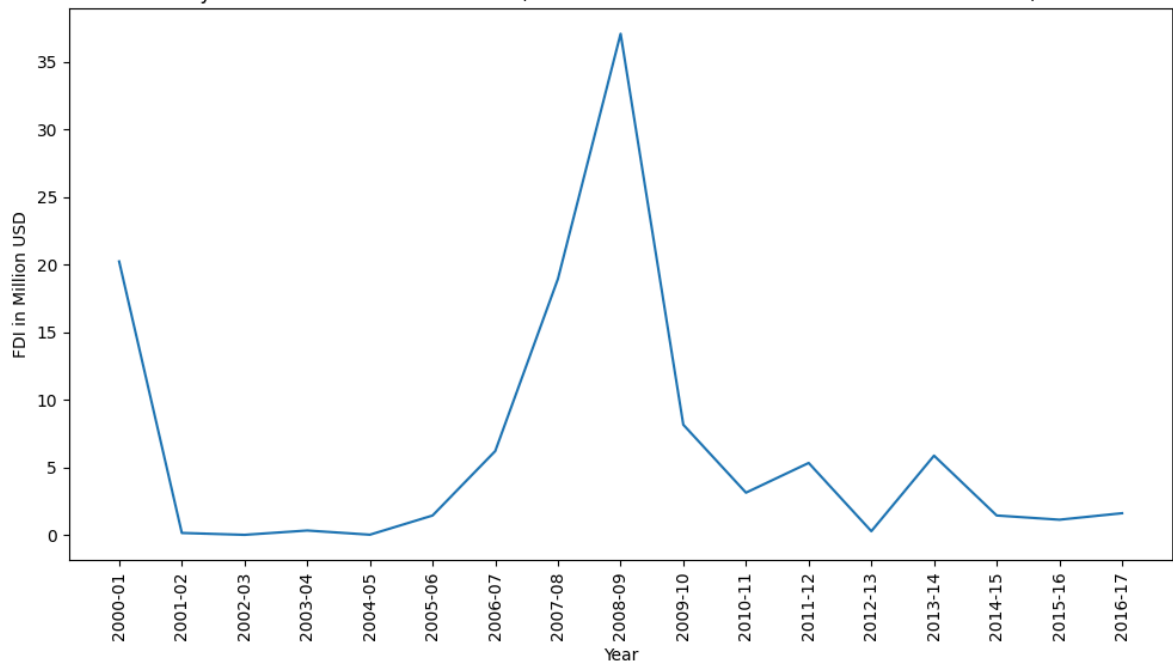


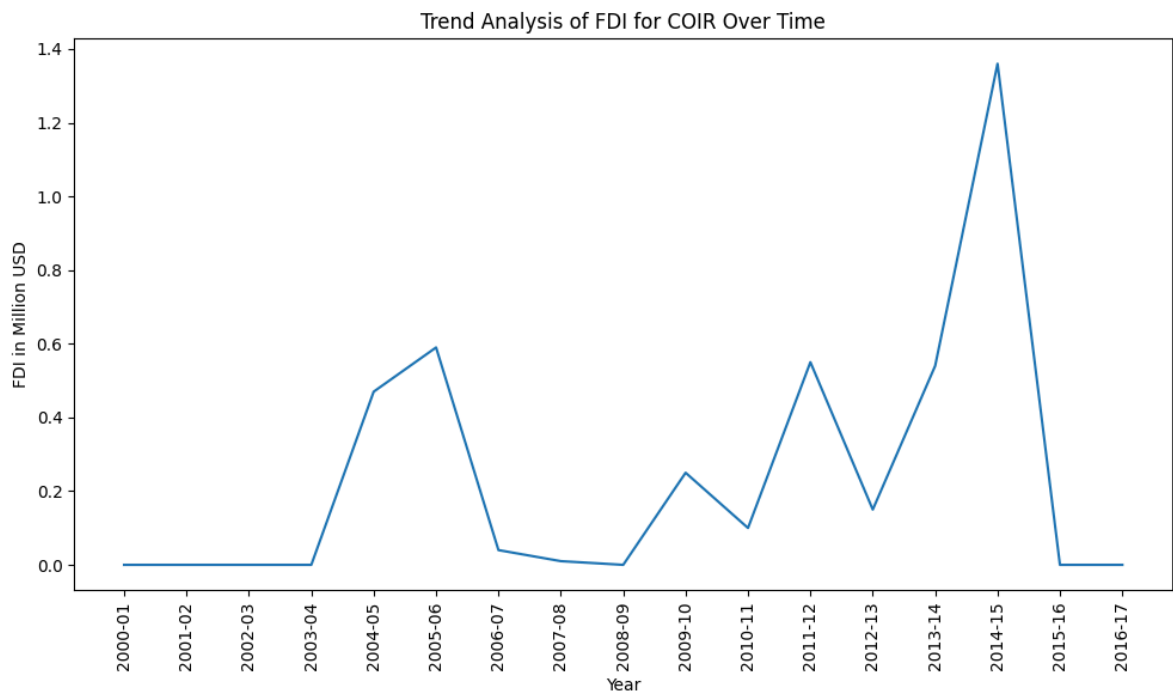
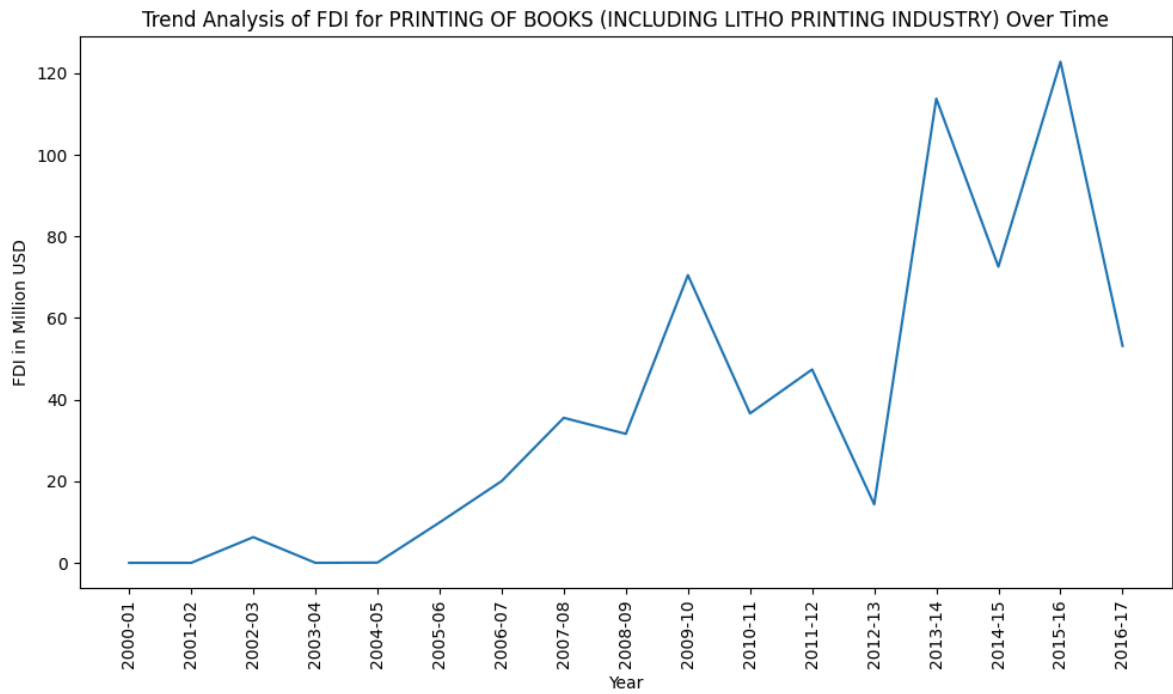


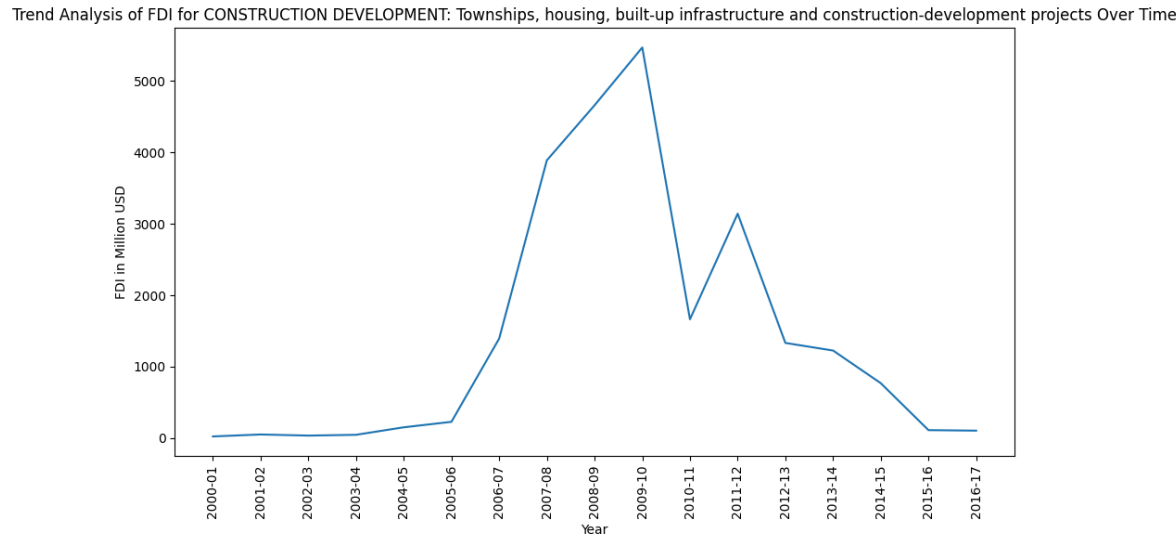
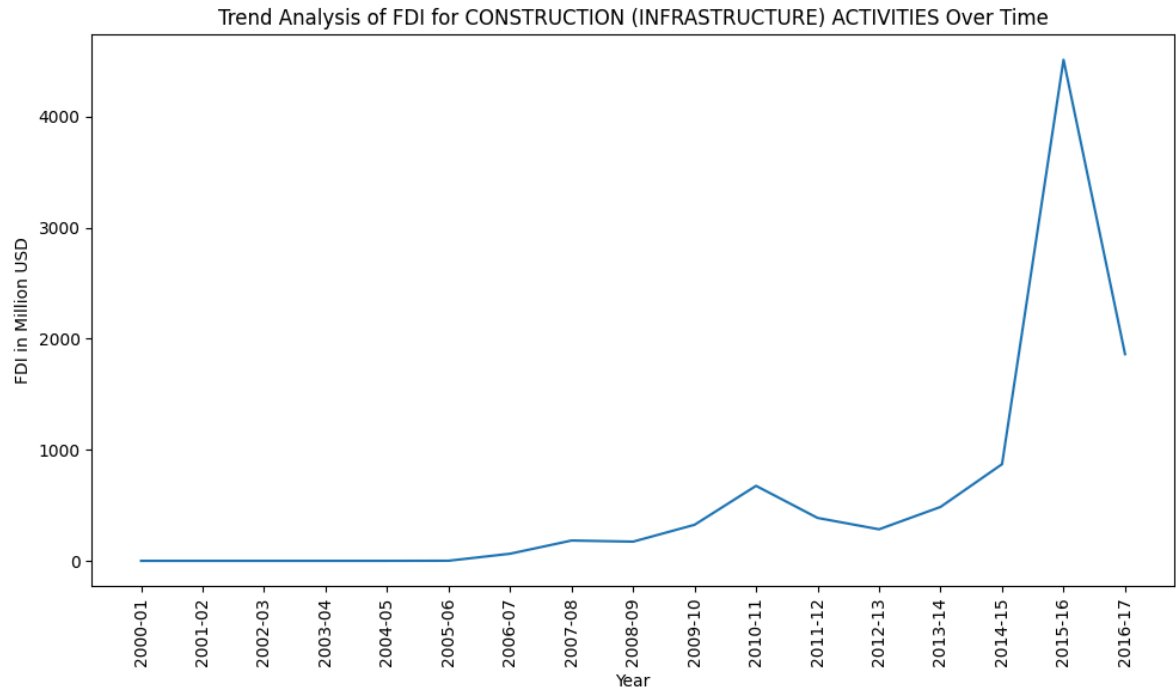
Trend Analysis of FDI for DIAMOND,GOLD ORNAMENTS Over Time

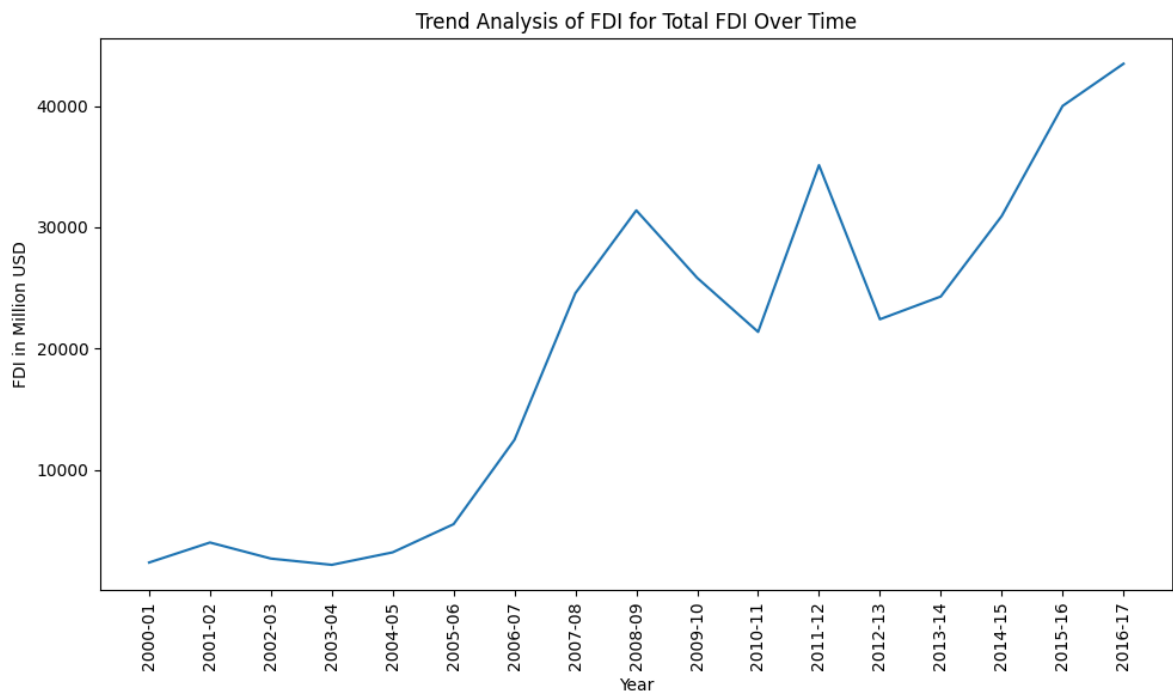
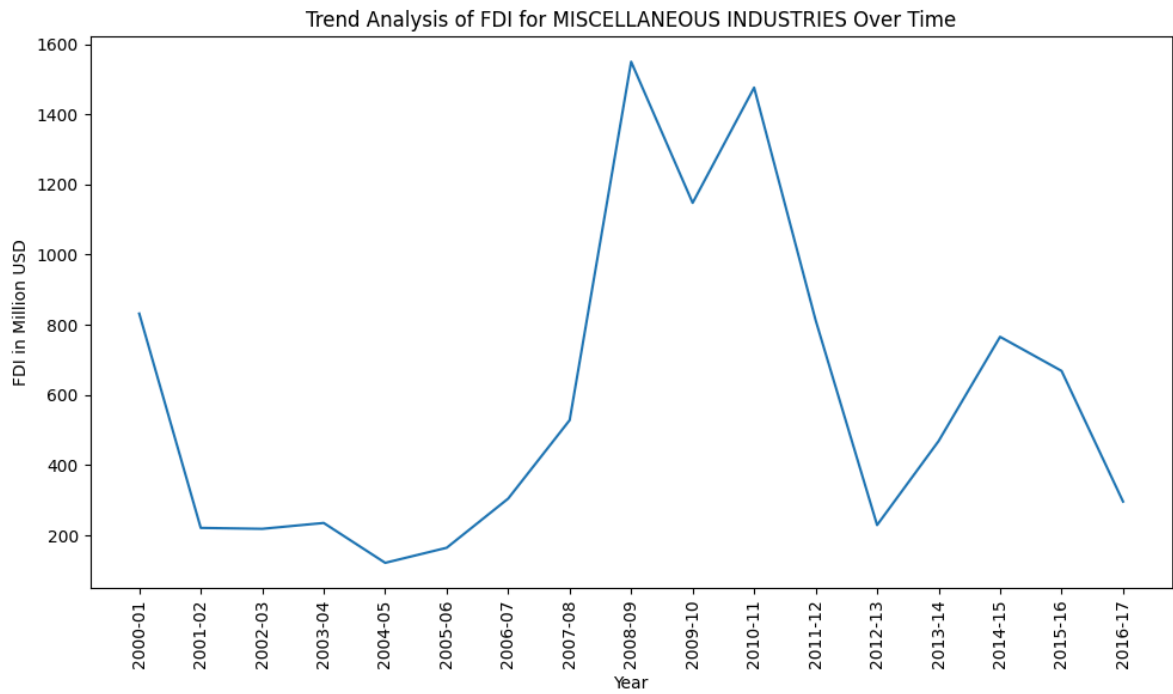


Trend Analysis of FDI for TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER) Over Time









```
In [132]: # to calculate the percentage contribution of each sector to the total FDI amount
total_fdi = df.values.sum().sum()
print(total_fdi)
# # Calculating the percentage contribution of each sector
df['Contribution (%)'] = (df.sum(axis=1) / total_fdi) * 100

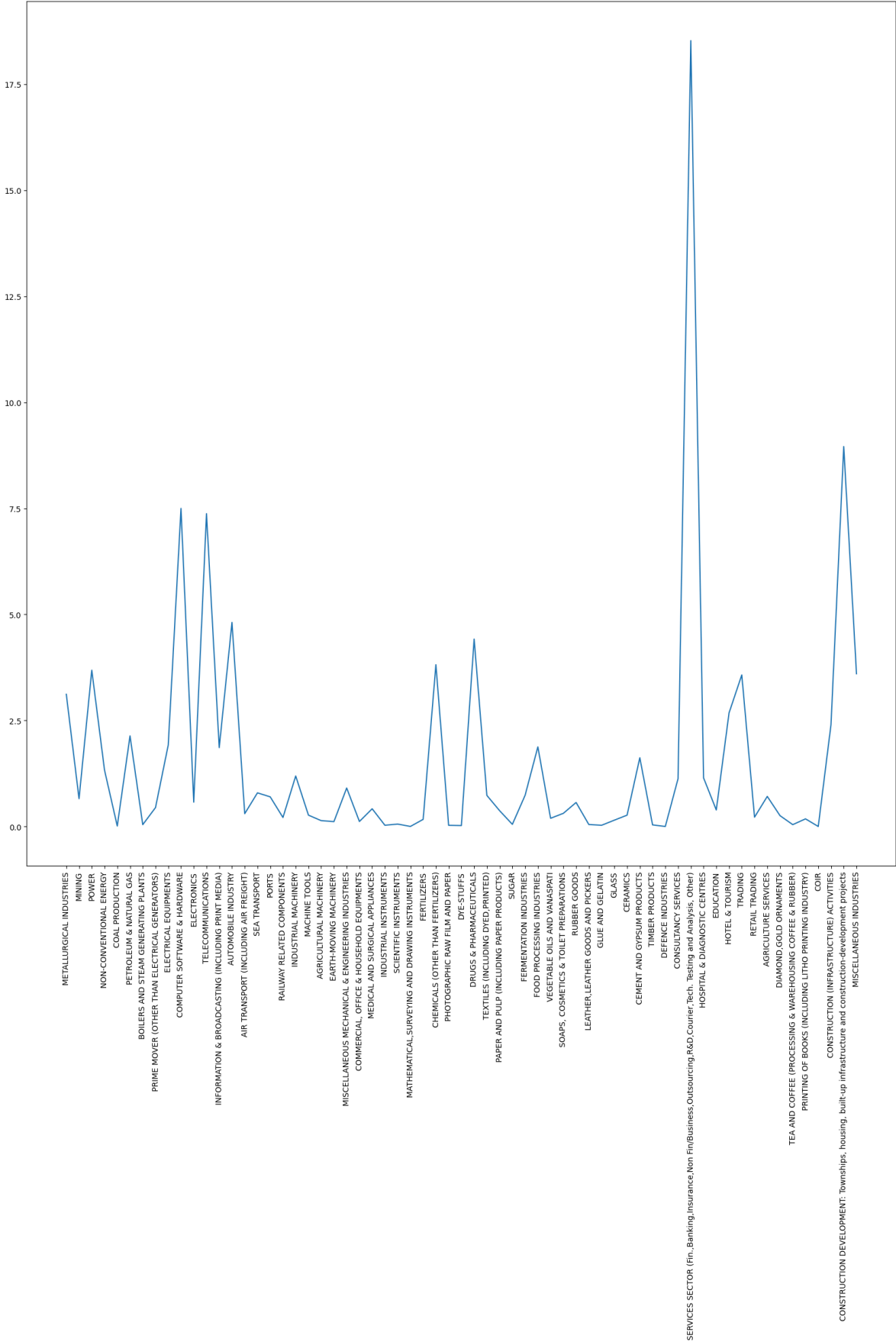
# # Displaying the resulting dataframe
print(df[['Contribution (%)']])
```

467728.57000000007

Sector	Contribution (%)
METALLURGICAL INDUSTRIES	3.118315
MINING	0.656274
POWER	3.684894
NON-CONVENTIONAL ENERGY	1.328550
COAL PRODUCTION	0.011229
...	...
PRINTING OF BOOKS (INCLUDING LITHO PRINTING IND...	0.180714
COIR	0.001180
CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES	2.402614
CONSTRUCTION DEVELOPMENT: Townships, housing, b...	8.958193
MISCELLANEOUS INDUSTRIES	3.601285

[63 rows x 1 columns]

```
In [136]: plt.figure(figsize=(20,20))
plt.plot(df[['Contribution (%)']])
plt.xticks(rotation=90)
plt.show()
```

```
In [137]: # Service sector is the highest FDI Reciever  
# Service sector is the highest FDI reciever too  
# FY 2016-17 recieved highest FDI of 43478.26M dollars
```

```
In [ ]:
```

```
In [ ]:
```