

Analysis on foreign direct investment


```
# including basic libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os

# suppress warnings
from warnings import filterwarnings
filterwarnings('ignore')

pwd= os.getcwd()

raw_data=pd.read_excel(pwd+"\\FDI data modified.xlsx", sheet_name="modified sectors")
```

```
dataset=raw_data.copy()
dataset.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
0	ENERGY PRODUCTION	98.77	968.51	117.59	111.95	150.62	96.01	277.48	2508.66	1457.93	2203.30	2209.67	4480.07	2061.65	1808.47
1	ELECTRICAL MACHINES	316.49	497.62	644.83	523.83	713.40	1440.38	2732.32	2172.02	2108.20	1652.27	993.43	1557.15	720.07	1393.16
2	COMUNICACION	259.19	877.77	228.10	100.21	128.18	673.91	552.21	1549.19	3283.67	3030.09	2076.61	2673.20	707.91	1735.47
3	TRANSPORT INDUSTRIES	197.74	271.05	455.67	261.29	186.82	226.95	421.35	1814.13	1772.77	1644.69	1818.10	1125.84	1647.64	1820.96
4	MACHINARIES AND INTRUMENTS	72.68	110.13	96.24	91.16	48.10	193.91	154.16	467.59	331.72	727.51	643.08	2100.38	797.96	931.21

▼

data cleaning

```
dataset.dtypes
```

```
Sector      object
2000-01     float64
2001-02     float64
2002-03     float64
2003-04     float64
2004-05     float64
2005-06     float64
2006-07     float64
2007-08     float64
2008-09     float64
2009-10     float64
2010-11     float64
2011-12     float64
2012-13     float64
2013-14     float64
2014-15     float64
2015-16     float64
2016-17     float64
dtype: object
```

```
round(dataset.isna().mean()*100,2)
```

```
Sector      0.0
2000-01     0.0
2001-02     0.0
2002-03     0.0
2003-04     0.0
2004-05     0.0
2005-06     0.0
2006-07     0.0
2007-08     0.0
2008-09     0.0
```

```

2009-10    0.0
2010-11    0.0
2011-12    0.0
2012-13    0.0
2013-14    0.0
2014-15    0.0
2015-16    0.0
2016-17    0.0
dtype: float64

```

we have all data in correct data types and there is no null value

✓ data transformation

calculating average of last 17 years, 15 years, 12 years and so on... as required

```
dataset.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
0	ENERGY PRODUCTION	98.77	968.51	117.59	111.95	150.62	96.01	277.48	2508.66	1457.93	2203.30	2209.67	4480.07	2061.65	1808.47
1	ELECTRICAL MACHINES	316.49	497.62	644.83	523.83	713.40	1440.38	2732.32	2172.02	2108.20	1652.27	993.43	1557.15	720.07	1393.16
2	COMUNICACION	259.19	877.77	228.10	100.21	128.18	673.91	552.21	1549.19	3283.67	3030.09	2076.61	2673.20	707.91	1735.47
3	TRANSPORT INDUSTRIES	197.74	271.05	455.67	261.29	186.82	226.95	421.35	1814.13	1772.77	1644.69	1818.10	1125.84	1647.64	1820.96
4	MACHINARIES AND INTRUMENTS	72.68	110.13	96.24	91.16	48.10	193.91	154.16	467.59	331.72	727.51	643.08	2100.38	797.96	931.21

```

lenth=dataset.shape[0]
width=dataset.shape[1]

```

```
dataset_year=dataset.iloc[:,1:width]
```

```

lenth_y=dataset_year.shape[0]
width_y=dataset_year.shape[1]

```

```
dataset_year.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	2015-16
0	98.77	968.51	117.59	111.95	150.62	96.01	277.48	2508.66	1457.93	2203.30	2209.67	4480.07	2061.65	1808.47	2634.04	1985.37
1	316.49	497.62	644.83	523.83	713.40	1440.38	2732.32	2172.02	2108.20	1652.27	993.43	1557.15	720.07	1393.16	2967.71	6557.63
2	259.19	877.77	228.10	100.21	128.18	673.91	552.21	1549.19	3283.67	3030.09	2076.61	2673.20	707.91	1735.47	3149.90	2333.74
3	197.74	271.05	455.67	261.29	186.82	226.95	421.35	1814.13	1772.77	1644.69	1818.10	1125.84	1647.64	1820.96	3265.05	3391.36
4	72.68	110.13	96.24	91.16	48.10	193.91	154.16	467.59	331.72	727.51	643.08	2100.38	797.96	931.21	1024.23	1117.33

```

def cal_avg(n):
    # n define as: last n years of average

    avg=np.round(dataset_year.iloc[:,width_y-n:width_y].mean(axis=1),2)
    return(avg)

```

```

# we have 17 years of data
dataset["avg_last_17yrs"]=cal_avg(17)

```

```
dataset.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
0	ENERGY PRODUCTION	98.77	968.51	117.59	111.95	150.62	96.01	277.48	2508.66	1457.93	2203.30	2209.67	4480.07	2061.65	1808.47
1	ELECTRICAL MACHINES	316.49	497.62	644.83	523.83	713.40	1440.38	2732.32	2172.02	2108.20	1652.27	993.43	1557.15	720.07	1393.16
2	COMUNICATION	259.19	877.77	228.10	100.21	128.18	673.91	552.21	1549.19	3283.67	3030.09	2076.61	2673.20	707.91	1735.47
3	TRANSPORT INDUSTRIES	197.74	271.05	455.67	261.29	186.82	226.95	421.35	1814.13	1772.77	1644.69	1818.10	1125.84	1647.64	1820.96
4	MACHINARIES AND INTRIMENTS	72.68	110.13	96.24	91.16	48.10	193.91	154.16	467.59	331.72	727.51	643.08	2100.38	797.96	931.21

```
dataset["avg_last_15yrs"]=cal_avg(15)
dataset["avg_last_12yrs"]=cal_avg(12)
dataset["avg_last_10yrs"]=cal_avg(10)
dataset["avg_last_8yrs"]=cal_avg(8)
dataset["avg_last_5yrs"]=cal_avg(5)
dataset["avg_last_3yrs"]=cal_avg(3)
dataset["avg_last_2yrs"]=cal_avg(2)
```

```
dataset.head()
```

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007-08	2008-09	...	2015-16	2016-17	avg_last_17yrs	avg_
0	ENERGY PRODUCTION	98.77	968.51	117.59	111.95	150.62	96.01	277.48	2508.66	1457.93	...	1985.37	2417.74	1505.17	
1	ELECTRICAL MACHINES	316.49	497.62	644.83	523.83	713.40	1440.38	2732.32	2172.02	2108.20	...	6557.63	5966.37	1938.64	
2	COMUNICATION	259.19	877.77	228.10	100.21	128.18	673.91	552.21	1549.19	3283.67	...	2333.74	7080.37	1790.57	
3	TRANSPORT INDUSTRIES	197.74	271.05	455.67	261.29	186.82	226.95	421.35	1814.13	1772.77	...	3391.36	2515.35	1343.34	
4	MACHINARIES AND INTRIMENTS	72.68	110.13	96.24	91.16	48.10	193.91	154.16	467.59	331.72	...	1117.33	735.56	567.23	

5 rows × 26 columns

```
d_list=['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']
```

```
data_FDI_year_avg=dataset.drop(d_list,axis=1)
```

```
data_FDI_year_avg
```

	Sector	avg_last_17yrs	avg_last_15yrs	avg_last_12yrs	avg_last_10yrs	avg_last_8yrs	avg_last_5yrs	avg_last_3yrs	avg
0	ENERGY PRODUCTION	1505.17	1634.70	2011.70	2376.69	2475.04	2181.45	2345.72	
1	ELECTRICAL MACHINES	1938.64	2142.85	2521.73	2608.80	2725.97	3520.99	5163.90	
2	COMUNICACION	1790.57	1953.52	2403.86	2762.02	2848.41	3001.48	4188.00	
3	TRANSPORT INDUSTRIES	1343.34	1491.20	1788.68	2081.59	2153.62	2528.07	3057.25	
4	MACHINARIES AND INTRUMENTS	567.23	630.68	768.72	887.66	1009.66	921.26	959.04	
5	AGRICULTURAL SECTOR	954.63	1066.27	1304.04	1497.32	1734.60	1147.29	1399.22	
6	MEDICAL SECTOR	1213.13	1363.64	1658.66	1942.17	1783.07	1913.32	1988.62	
7	FOOD SECTOR	649.64	715.42	864.54	998.43	1160.14	1637.96	910.75	
8	TEXTILES AND GOODS SECTOR	371.64	416.88	505.81	569.28	651.05	831.46	839.83	
9	SEVICES SECTOR	4308.10	4857.50	5949.54	6570.26	6394.70	7102.21	8093.12	
10	EDUCATION	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	

```
path=os.path.abspath("FDI data modified.xlsx")
path
```

```
'C:\\Users\\neham\\git practice\\FD-investment-analysis\\FDI data modified.xlsx'
```

```
from openpyxl import load_workbook
```

```
# as I have already run this once hence i am making it a comment line to avoid saving it more than once or any other error"
```