iNeuron

Low Level Report

Investment Analytics for FDI in India from 2001 to 2017

Version	1.1
Name	Pranjita Chakraborty
Project	Investment Analytics
Submission Date	24-10-21

Document Version Control

Version	Date	Author	Comments
1.0	19.10.21	Pranjita chakraborty	First document prepared
2.0	30.10.21	Pranjita Chakraborty	Added parts related to Tableau Dashboard

Contents

1. Tableau Architechture 4

2. Data Description 5

3. Colab work file

Tableau Architecture

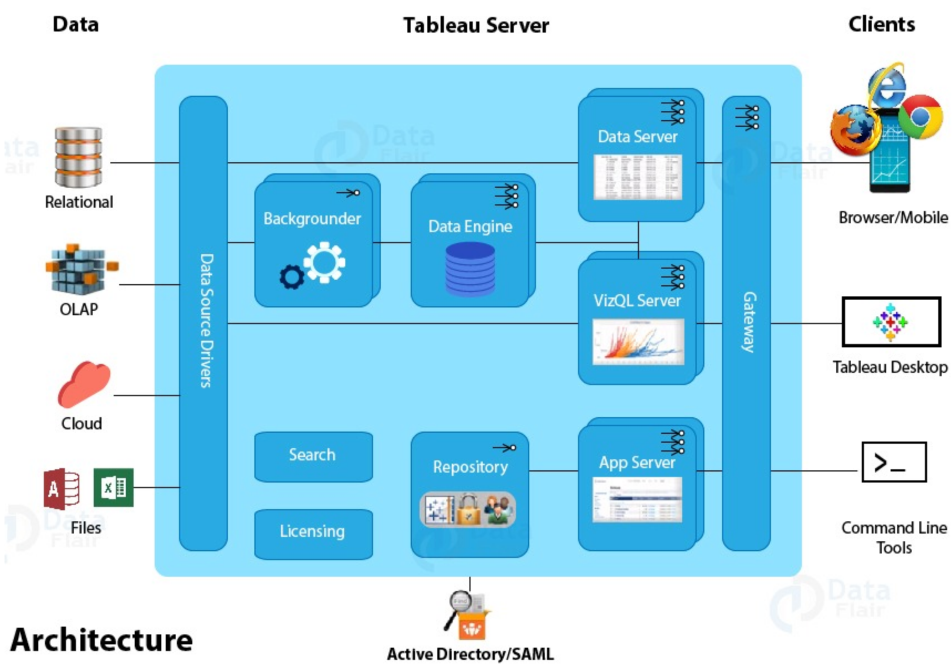


Fig: Tableau desktop structure architecture: img source: https://data-flair.training/blogs/tableau-architecture/

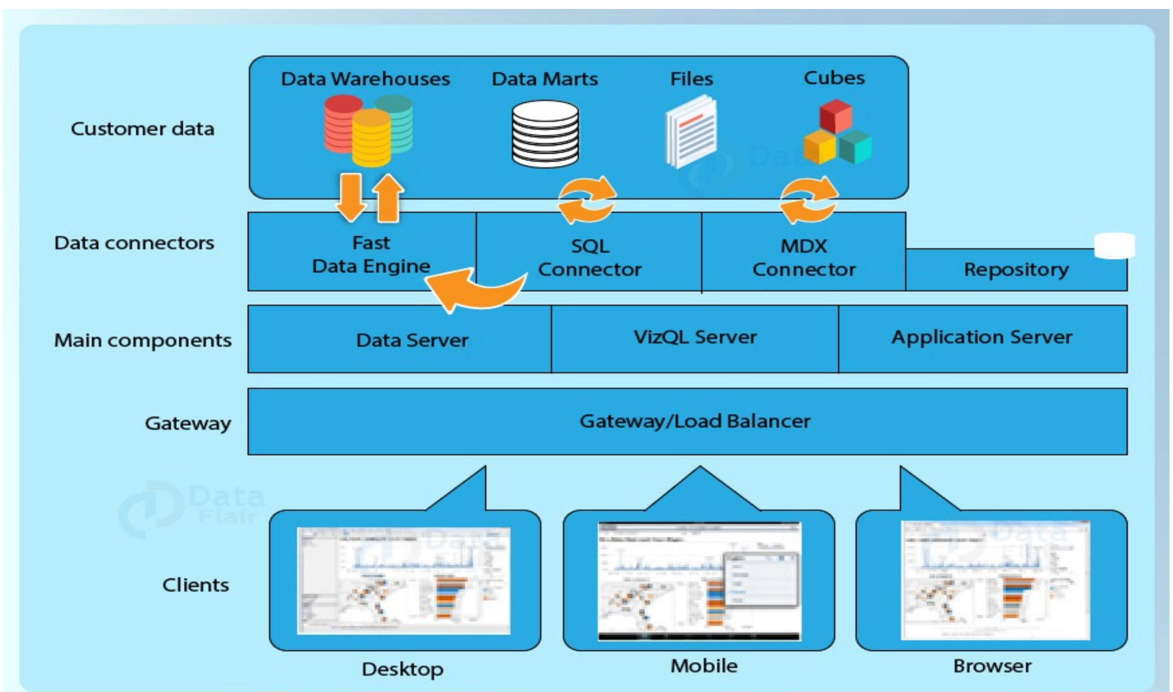


Fig: Tableau server structure architecture: img source: https://data-flair.training/blogs/tableau-architecture/

Here, first we used Tableau desktop architecture and later after posting it to Tableau Public, we user Tableau Server Architecture.

Tableau Server is essentially a communication tool which shares data connections and visualizations with the end-users or clients. So, now that we have learned about the functioning of each component in a Tableau server. Let us understand how all these components work in tandem. For this, we will club the server components into layers or tiers. So, we have five layers or sections in the Tableau Server; customer data, data connectors, main components, gateway, and clients.

The customer data layer contains all sorts of data sources available for a Tableau user like data warehouses, data marts, flat files, and multi-dimensional cubes, relational databases.

Next lies the data connectors layers which consist of a data engine, repository, SQL Connector, and MDX Connector. These components interact directly with the data sources. The Data engine processes the data requested by the user and assigns the data type, decides whether it is a measure or a dimension, and creates TDEs (data extracts). At the background of the data, engine runs an SQL Connector which creates an SQL query for all the user requests and interacts to the data sources. The SQL Connector primarily deals with data marts and flat files. Similarly, the MDX Connector deals with the multi-dimensional cubes.

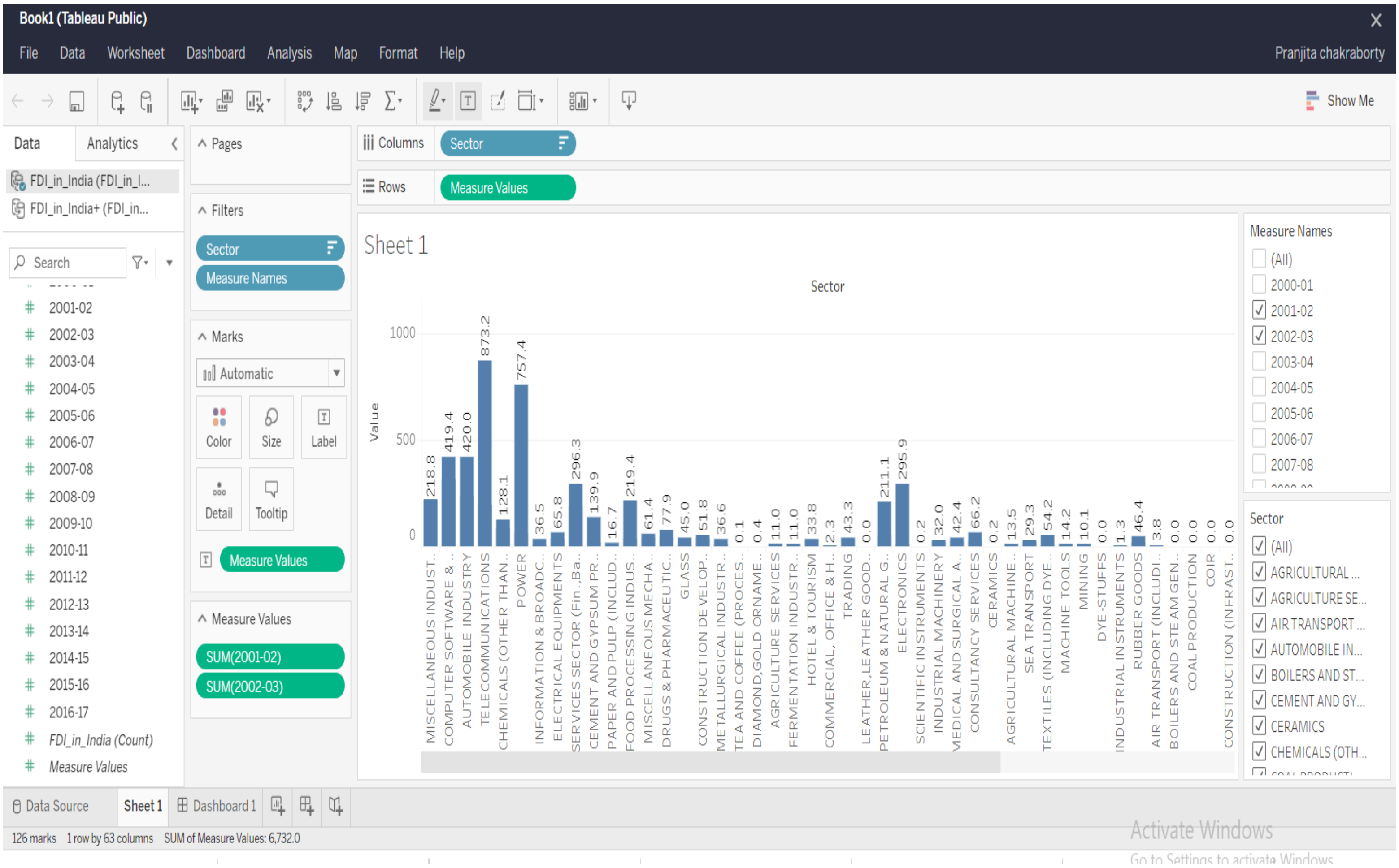
The next layer comprises of all the main components, essentially the data server which regulates and monitors the functioning of the components of the data connector layer. Along with this, it includes a VizQL Server and Application Server. The application server takes all the user requests coming from Tableau Desktop, mobile or browser for accessing the visualization. It processes the requests and detects the type of request, checks user authorization and grants access accordingly. The VizQL Server is a patented component of Tableau, where VizQL stands of Visualization Query language. It works behind the logic of Tableau visualization and creates the visualization as per your instructions on the dashboard.

The gateway, it acts as a gatekeeper of the Tableau Server and any request or query sent by the client first hits the gateway or load balancer. A gateway is nothing but a primary server which receives the queries and redirects it to an appropriate and available secondary server, known as worker server.

Data Description

The data comprises FDI data of India from period: 2000-2001 to 2016-2017. In order to contruct the dashboard, we created a pivot view of the columns under years, since having the horizontal data was not very effective for generating views on Tableau.

The chart generated without pivoting FDI years: Wanted to create two filters: one for sectors and other for the years. So, did the following as shown in the screenshot.



The rest is displayed in charts – in the video links.

Since EDA work has also been done on colab environment primarily, the colab pdf has alsobeen attached.

Here we will be looking at the trends of FDI in India 2000 to 2017.

FDI support on different sectors over the years indicates changes in interest over time.

1. Firing up colab

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv('/content/FDI_in_India.csv')
df.head(3)
```

	Sector	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07	2007 0
0	METALLURGICAL INDUSTRIES	22.69	14.14	36.61	8.11	200.38	149.13	169.94	1175.7
1	MINING	1.32	6.52	10.06	23.48	9.92	7.40	6.62	444.3

```
len(df)
```

```
63
```

```
df.info()
```

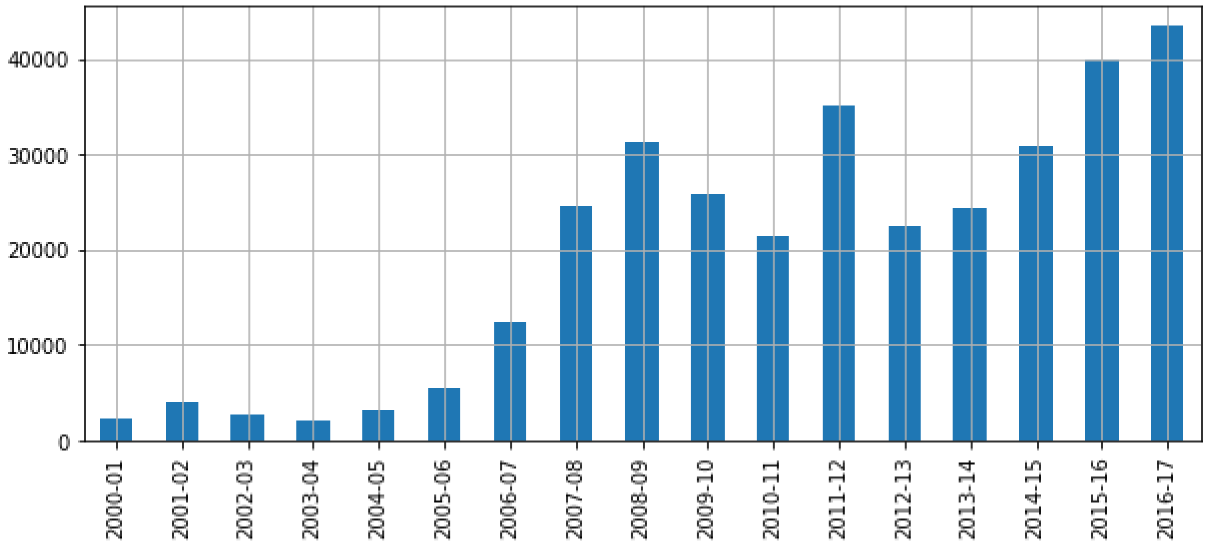
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 63 entries, 0 to 62
Data columns (total 18 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   Sector   63 non-null     object
 1   2000-01  63 non-null     float64
 2   2001-02  63 non-null     float64
 3   2002-03  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
```

2. Setting the index and subsequent

```
df.set_index('Sector', inplace = True)
```

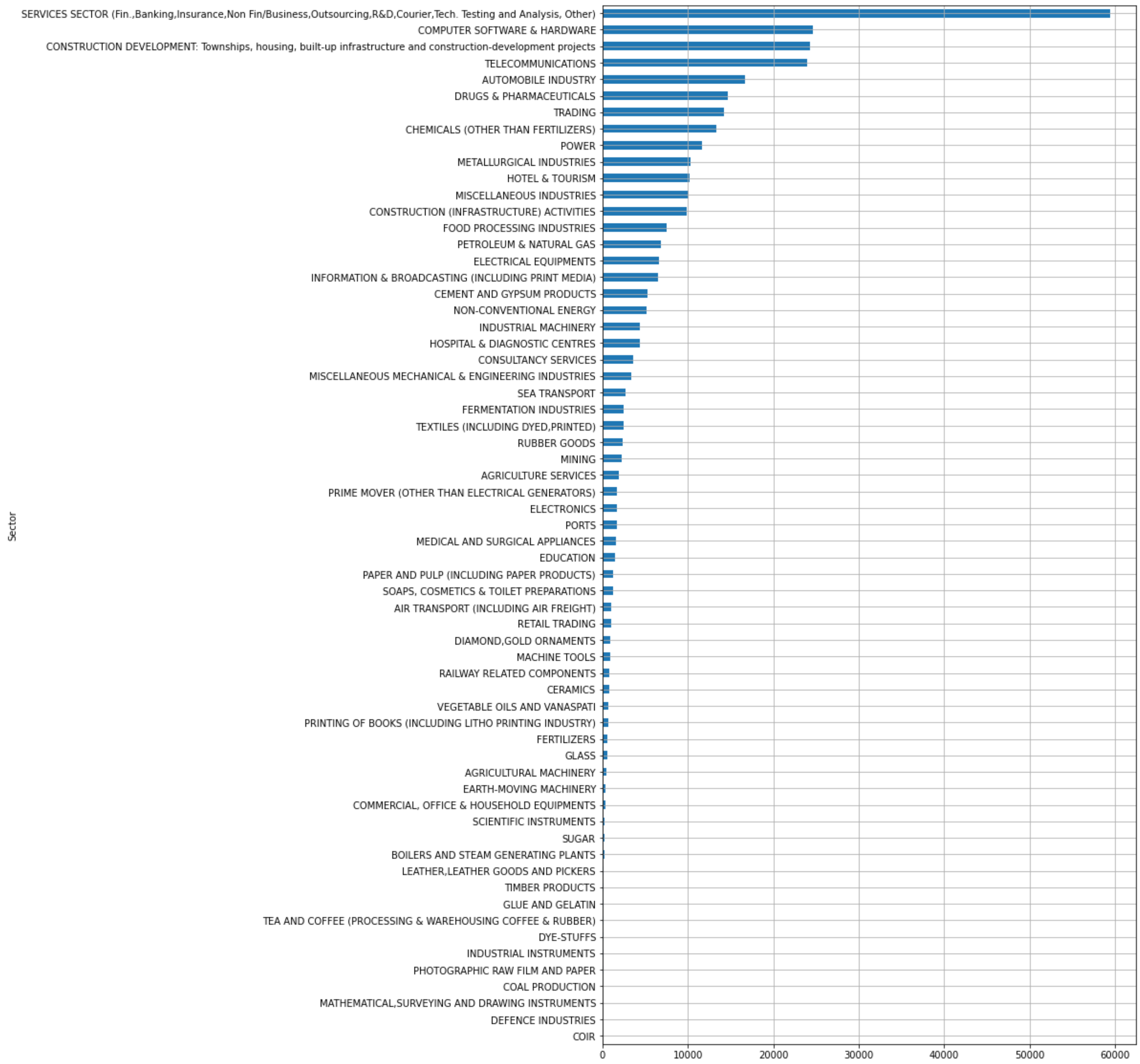
Year-wise total Investment

```
df.sum(axis=0).plot(kind='bar', figsize=(10,4))
plt.grid()
plt.show()
```



sector-wise total investment over the years

```
df.sum(axis=1).sort_values().plot(kind = 'barh', figsize=(10,20))
plt.grid()
plt.show()
```



We find that from 2007 onwards there has been an overall large increase in Investment levels.

Also, Services (Finance, Banking etc.) sector has had the max overall investment while Coir has had the least investment overall across the years.

finding trends

a. correlation

```
import numpy as np
```

```
new = df.transpose()
corrMatrix=new.corr()

corrMatrix.loc[:,:] =  np.tril(corrMatrix, k=-1)

already_in = set()
result = []
for col in corrMatrix:
    perfect_corr = corrMatrix[col][corrMatrix[col] >= 0.9].index.tolist()
    if perfect_corr and col not in already_in:
        already_in.update(set(perfect_corr))
        perfect_corr.append(col)
        result.append(perfect_corr)

result
```

```
[['TELECOMMUNICATIONS',
  'TEXTILES (INCLUDING DYED,PRINTED)',
  'GLUE AND GELATIN',
  'ELECTRICAL EQUIPMENTS'],
 ['TRADING', 'AUTOMOBILE INDUSTRY'],
 ['SUGAR',
  'CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES',
  'AIR TRANSPORT (INCLUDING AIR FREIGHT)'],
 ['RETAIL TRADING', 'SEA TRANSPORT'],
 ['SOAPS, COSMETICS & TOILET PREPARATIONS', 'INDUSTRIAL MACHINERY'],
 ['DEFENCE INDUSTRIES', 'MISCELLANEOUS MECHANICAL & ENGINEERING INDUSTRIES'],
 ['TEXTILES (INCLUDING DYED,PRINTED)', 'MEDICAL AND SURGICAL APPLIANCES'],
 ['GLASS', 'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS'],
 ['DIAMOND,GOLD ORNAMENTS', 'DYE-STUFFS'],
 ['FOOD PROCESSING INDUSTRIES', 'FERMENTATION INDUSTRIES']]
```

Above we set the correlation coefficient score to 0.9.

Now that we plot them to observe price trends considering each group as a separate segment wherein sectors have high correlation with each other.

```
X = list(df.columns)
#X.remove('Sector')
X
```

```
['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',
```

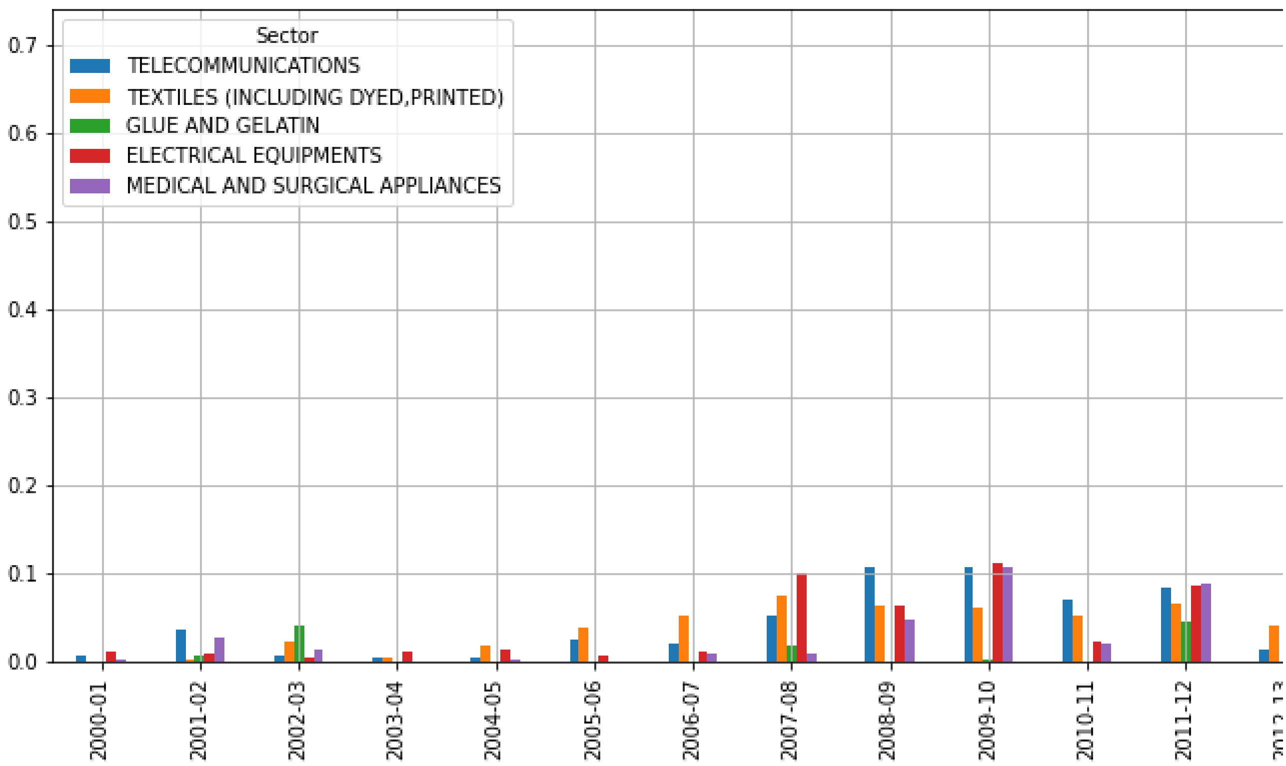
```
       '2015-16',
       '2016-17']
```

plotting graphs for each groups which are highly correlated

```
from sklearn.preprocessing import Normalizer
df.iloc[:,:] = Normalizer(norm='l1').fit_transform(df)
df.head(2)
```

	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07
Sector							
METALLURGICAL INDUSTRIES	0.002196	0.001369	0.003544	0.000785	0.019397	0.014436	0.016450

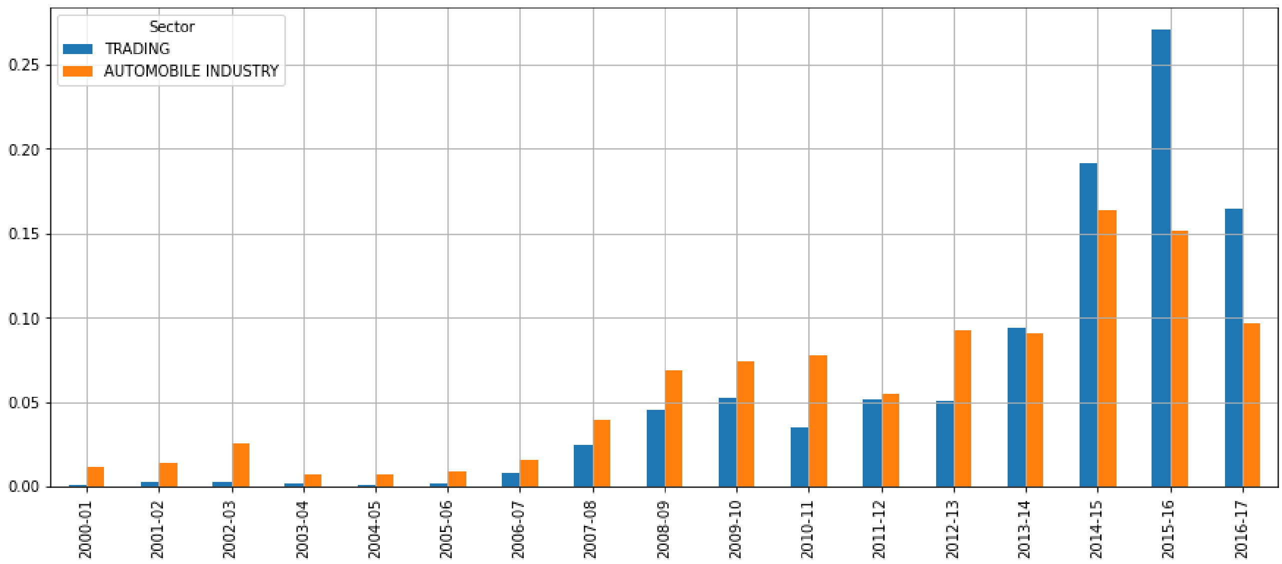
```
df_1 = df.loc[['TELECOMMUNICATIONS',
  'TEXTILES (INCLUDING DYED,PRINTED)',
  'GLUE AND GELATIN',
  'ELECTRICAL EQUIPMENTS','MEDICAL AND SURGICAL APPLIANCES'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
#df_1
```



In the figure above, we see that 6016-17 period saw the highest investment with Glue and gelatin leading the group.

```
df_1 = df.loc[['TRADING', 'AUTOMOBILE INDUSTRY'], X]
```

```
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
#df_1
```

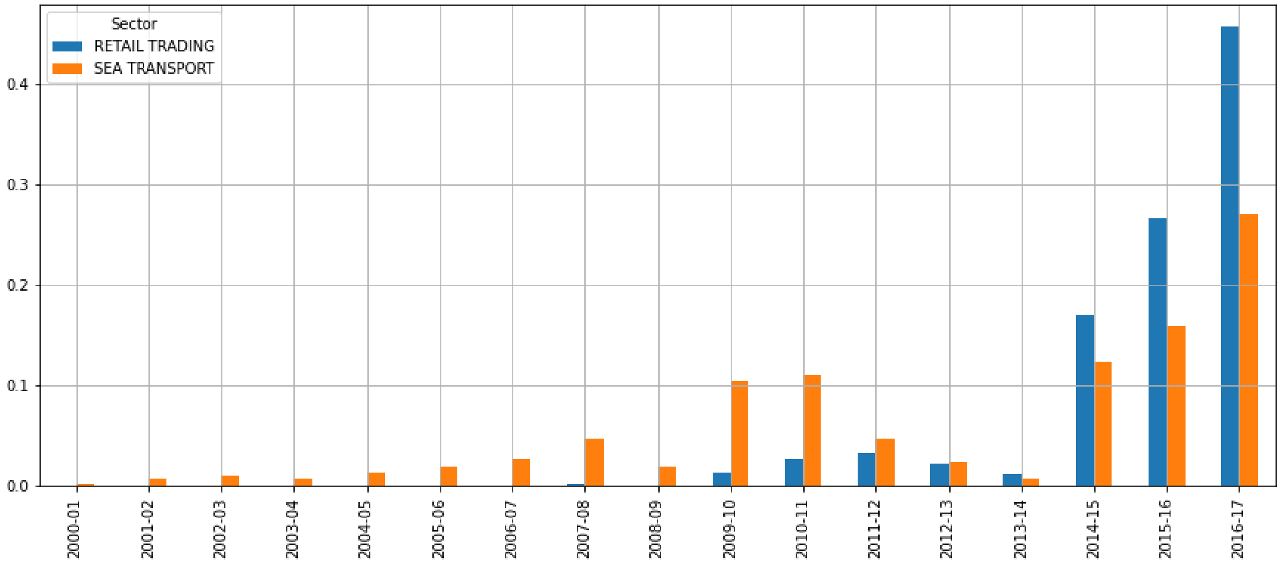


Here, while Automobile has been leading over trading, 2013-14 onwards, trading picked up more investment. Overall, a cyclic pattern with increments is observed.

```
df_1 = df.loc[['SUGAR',
  'CONSTRUCTION (INFRASTRUCTURE) ACTIVITIES',
  'AIR TRANSPORT (INCLUDING AIR FREIGHT)'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```

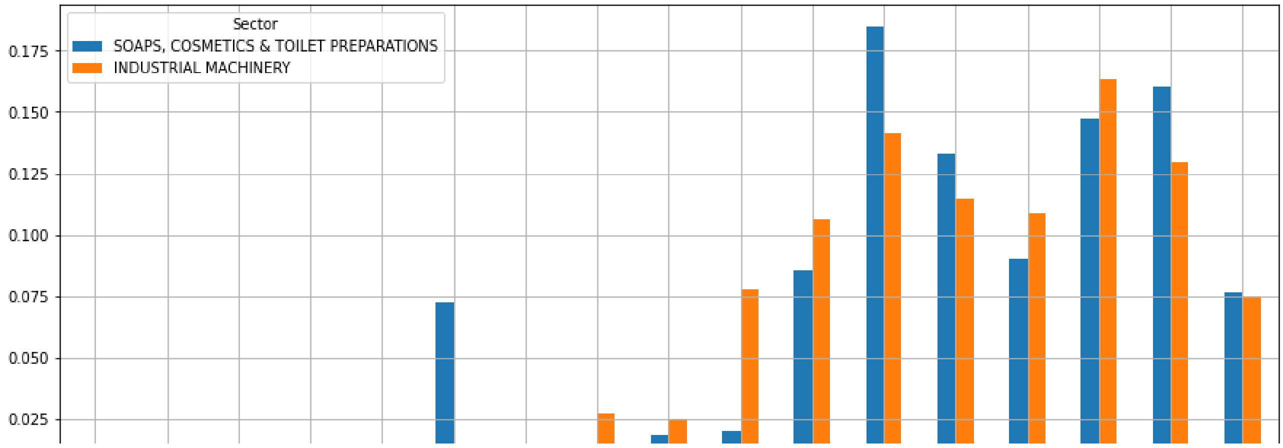
Sugar and air transport seem to be more prominent here in the beginning until 2007-08 period, with construction taking prominence post that.

```
df_1 = df.loc[['RETAIL TRADING', 'SEA TRANSPORT'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



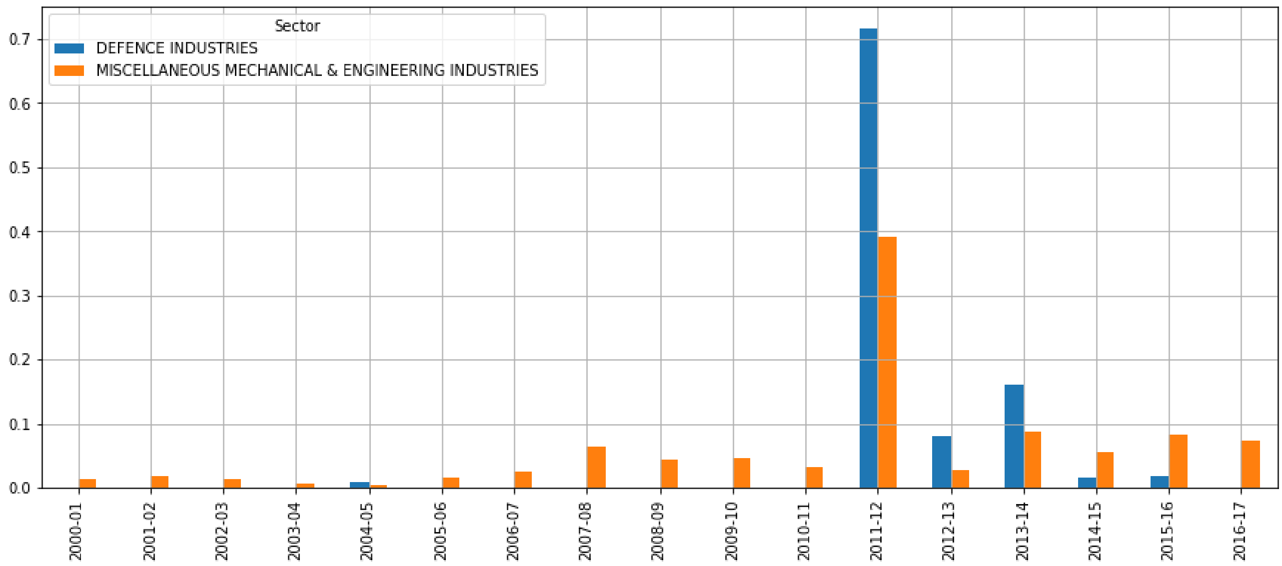
The sea transport sector oberves almost a flat normal curve (seemingly like a platykurtik curve) [upto 2013-14 from the beginining]. Post that period, retail picked up significant prominence.

```
df_1 = df.loc[['SOAPS, COSMETICS & TOILET PREPARATIONS', 'INDUSTRIAL MACHINERY'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```

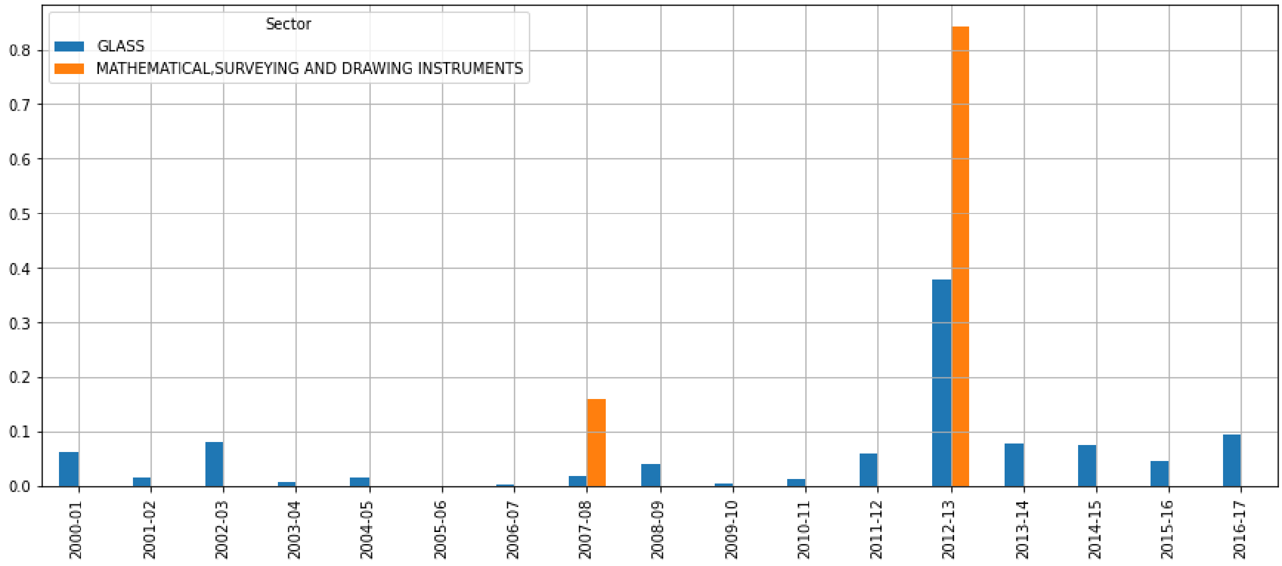


This sector has seen significant investment growth 2010-11 onwards for either sectors in this group and their seems a cyclic movement ever since in investment pattern.

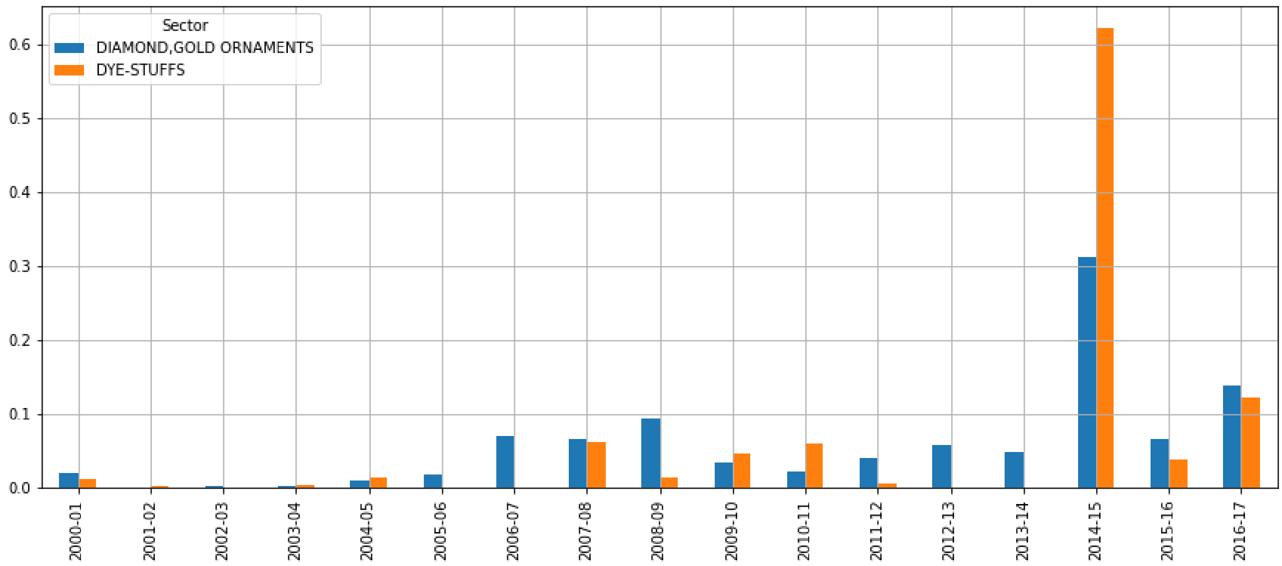
```
df_1 = df.loc[['DEFENCE INDUSTRIES', 'MISCELLANEOUS MECHANICAL & ENGINEERING INDUSTRIES'],
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



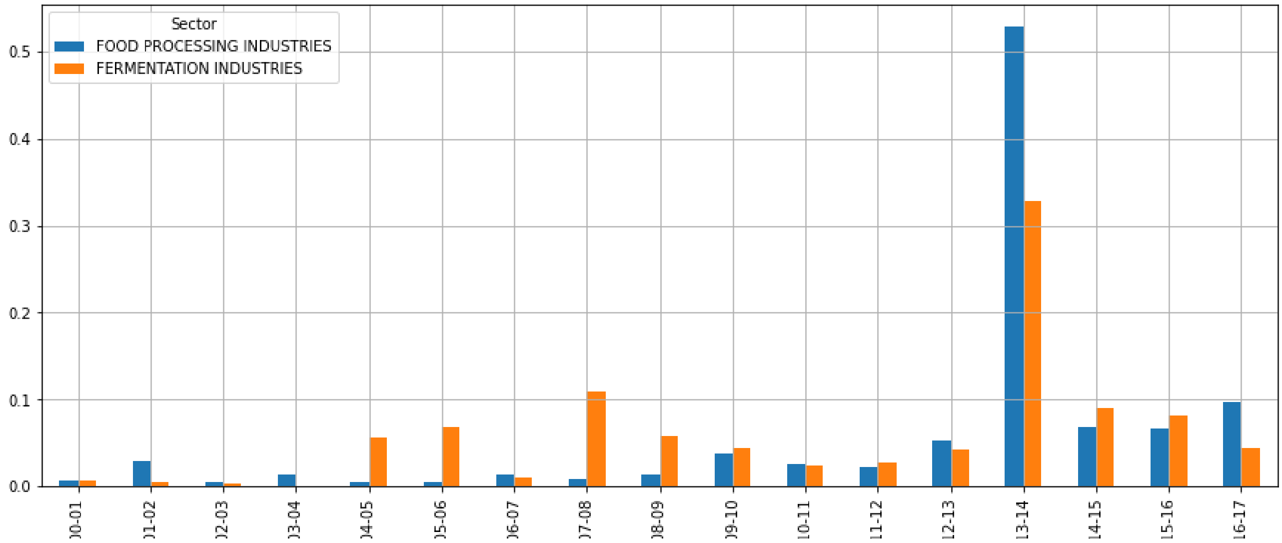

```
df_1 = df.loc[['GLASS', 'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```




```
df_1 = df.loc[['DIAMOND,GOLD ORNAMENTS', 'DYE-STUFFS'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```




```
df_1 = df.loc[['FOOD PROCESSING INDUSTRIES', 'FERMENTATION INDUSTRIES'], X]
df_1.transpose().plot(kind = 'bar', figsize=(15,6))
plt.grid()
plt.show()
```



For the above four groups, we observe spikes in investment in certain periods mostly in 2012-2014 period. Hence, we can infer that perhaps there is a connection between improvement in one sector group infleuncing growth and progress in subsequent sector groups. This area should be investigated with further data, so we can inform these aspects and sustain investments, collectively.

Top and bottom sectors invested in.

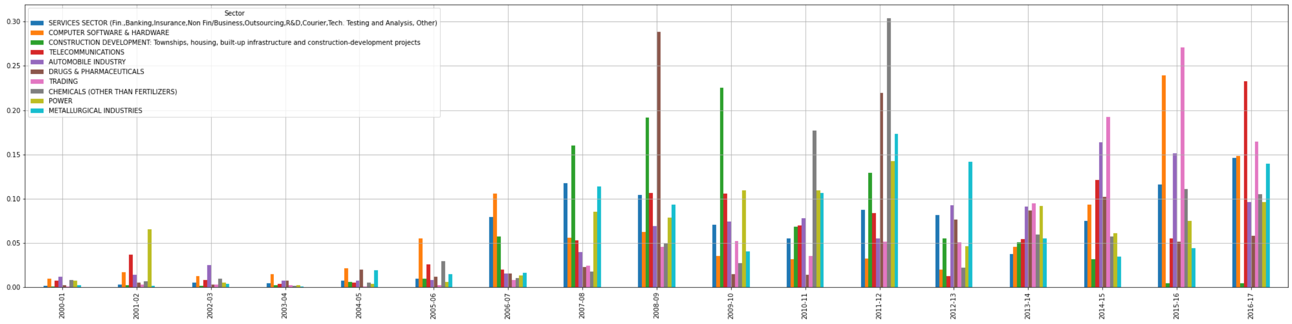
Top-most 10

```
df.sum(axis=1).nlargest(10)
```

```
Sector
SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,R&D,Courier,Tec
COMPUTER SOFTWARE & HARDWARE
CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and constructi
TELECOMMUNICATIONS
AUTOMOBILE INDUSTRY
DRUGS & PHARMACEUTICALS
TRADING
CHEMICALS (OTHER THAN FERTILIZERS)
POWER
METALLURGICAL INDUSTRIES
dtype: float64
```

```
df_1 = df.loc[['SERVICES SECTOR (Fin.,Banking,Insurance,Non Fin/Business,Outsourcing,R&D,C
'COMPUTER SOFTWARE & HARDWARE',
'CONSTRUCTION DEVELOPMENT: Townships, housing, built-up infrastructure and construction-de
'TELECOMMUNICATIONS',
'AUTOMOBILE INDUSTRY',
'DRUGS & PHARMACEUTICALS',
'TRADING',
'CHEMICALS (OTHER THAN FERTILIZERS)',
'POWER',
'METALLURGICAL INDUSTRIES'    ], X]
```

```
df_1.transpose().plot(kind = 'bar', figsize=(35,8))
plt.grid()
plt.show()
```



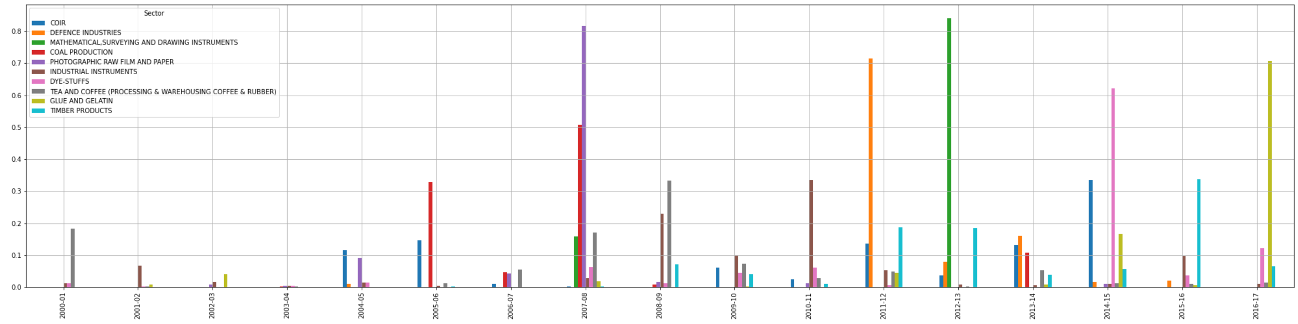
Bottom - 10

```
s = df.sum(axis=1).sort_values(ascending=True)
s.head(10)
```

```
Sector
COIR                                                          4.06
DEFENCE INDUSTRIES                                            5.12
MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS                7.98
COAL PRODUCTION                                              27.74
PHOTOGRAPHIC RAW FILM AND PAPER                              67.28
INDUSTRIAL INSTRUMENTS                                       76.12
DYE-STUFFS                                                   88.40
TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)   111.22
GLUE AND GELATIN                                            128.39
TIMBER PRODUCTS                                             157.68
dtype: float64
```

```
df_1 = df.loc[['COIR',
'DEFENCE INDUSTRIES',
'MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS',
'COAL PRODUCTION',
'PHOTOGRAPHIC RAW FILM AND PAPER',
'INDUSTRIAL INSTRUMENTS',
'DYE-STUFFS',
'TEA AND COFFEE (PROCESSING & WAREHOUSING COFFEE & RUBBER)',
'GLUE AND GELATIN',
'TIMBER PRODUCTS'], X]
```

```
df_1.transpose().plot(kind = 'bar', figsize=(35,8))
plt.grid()
plt.show()
```



Year wise Max and min Sectors invested into.

```
df.idxmax()
```

```
2000-01                              MISCELLANEOUS INDUSTRIES
2001-02                                    TELECOMMUNICATIONS
2002-03                                   AUTOMOBILE INDUSTRY
2003-04                          COMPUTER SOFTWARE & HARDWARE
2004-05                          COMPUTER SOFTWARE & HARDWARE
2005-06                          COMPUTER SOFTWARE & HARDWARE
2006-07    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2007-08    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2008-09    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2009-10    CONSTRUCTION DEVELOPMENT: Townships, housing, ...
2010-11    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2011-12    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2012-13    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2013-14                            FOOD PROCESSING INDUSTRIES
2014-15    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2015-16    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
2016-17    SERVICES SECTOR (Fin.,Banking,Insurance,Non Fi...
dtype: object
```

```
df.idxmin()
```

```
2000-01                            NON-CONVENTIONAL ENERGY
2001-02                            NON-CONVENTIONAL ENERGY
2002-03                                    COAL PRODUCTION
2003-04    PRIME MOVER (OTHER THAN ELECTRICAL GENERATORS)
```

```
2004-05                                         COAL PRODUCTION
2005-06                    BOILERS AND STEAM GENERATING PLANTS
2006-07                                                   PORTS
2007-08                                 SCIENTIFIC INSTRUMENTS
2008-09                    BOILERS AND STEAM GENERATING PLANTS
2009-10                                         COAL PRODUCTION
2010-11                                         COAL PRODUCTION
2011-12                                         COAL PRODUCTION
2012-13                                         COAL PRODUCTION
2013-14    MATHEMATICAL,SURVEYING AND DRAWING INSTRUMENTS
2014-15                                         COAL PRODUCTION
2015-16                                         COAL PRODUCTION
2016-17                                         COAL PRODUCTION
dtype: object
```

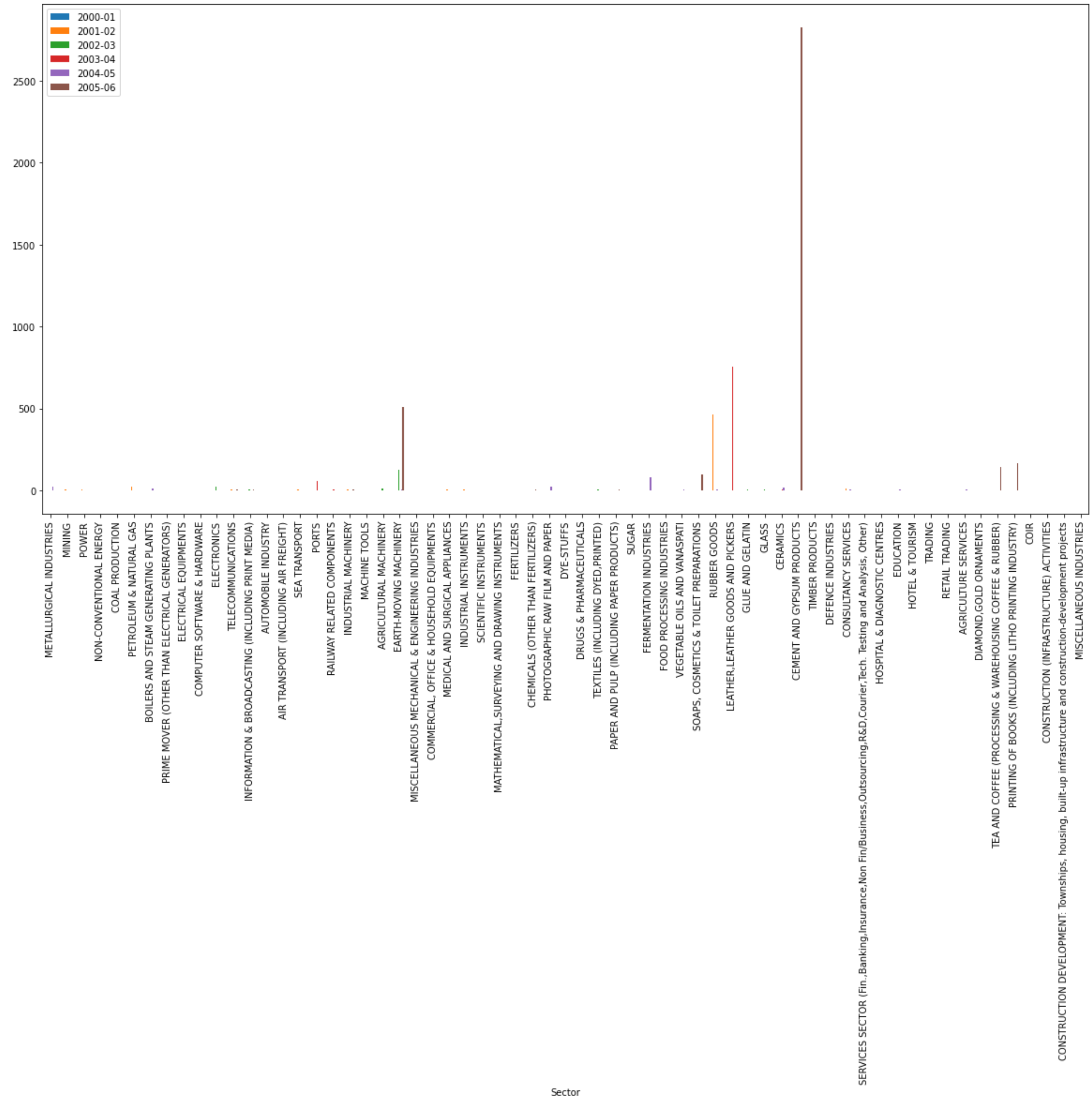
The min values shows that while COIR is sector within Bottom-10 total investment, in majority years COAL Production had least investment, while COIR didn't have the minimum investment in any year.

Percent changes in defferent time periods for the sectors.

1. 2000 - 2005

```
df.iloc[:, 0:6].pct_change(axis=1).plot(kind = 'bar', figsize=(20,10))
```

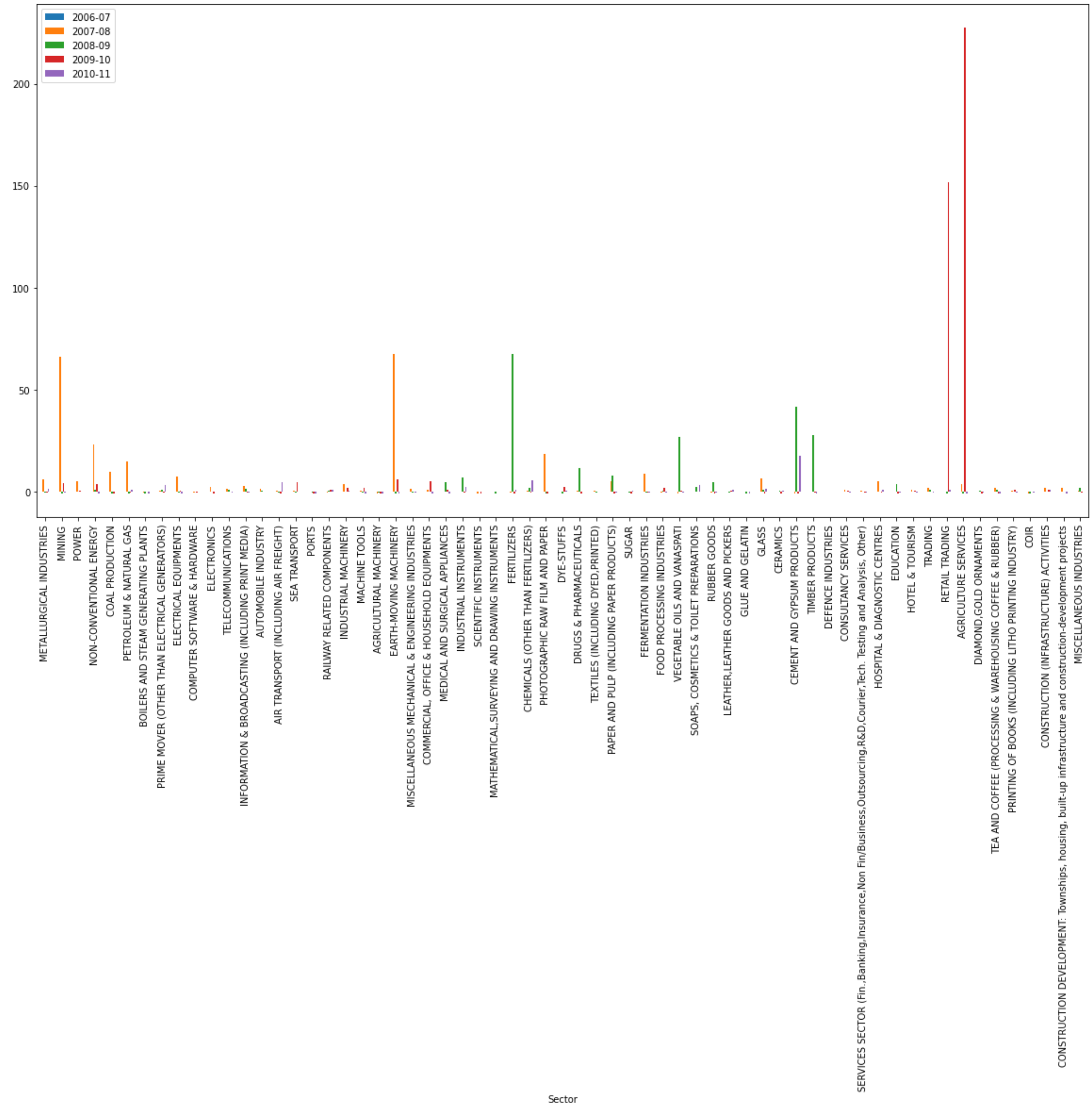
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7391dc0610>
```



2006 - 2010

```
df.iloc[:, 6:11].pct_change(axis=1).plot(kind = 'bar', figsize=(20,10))
```

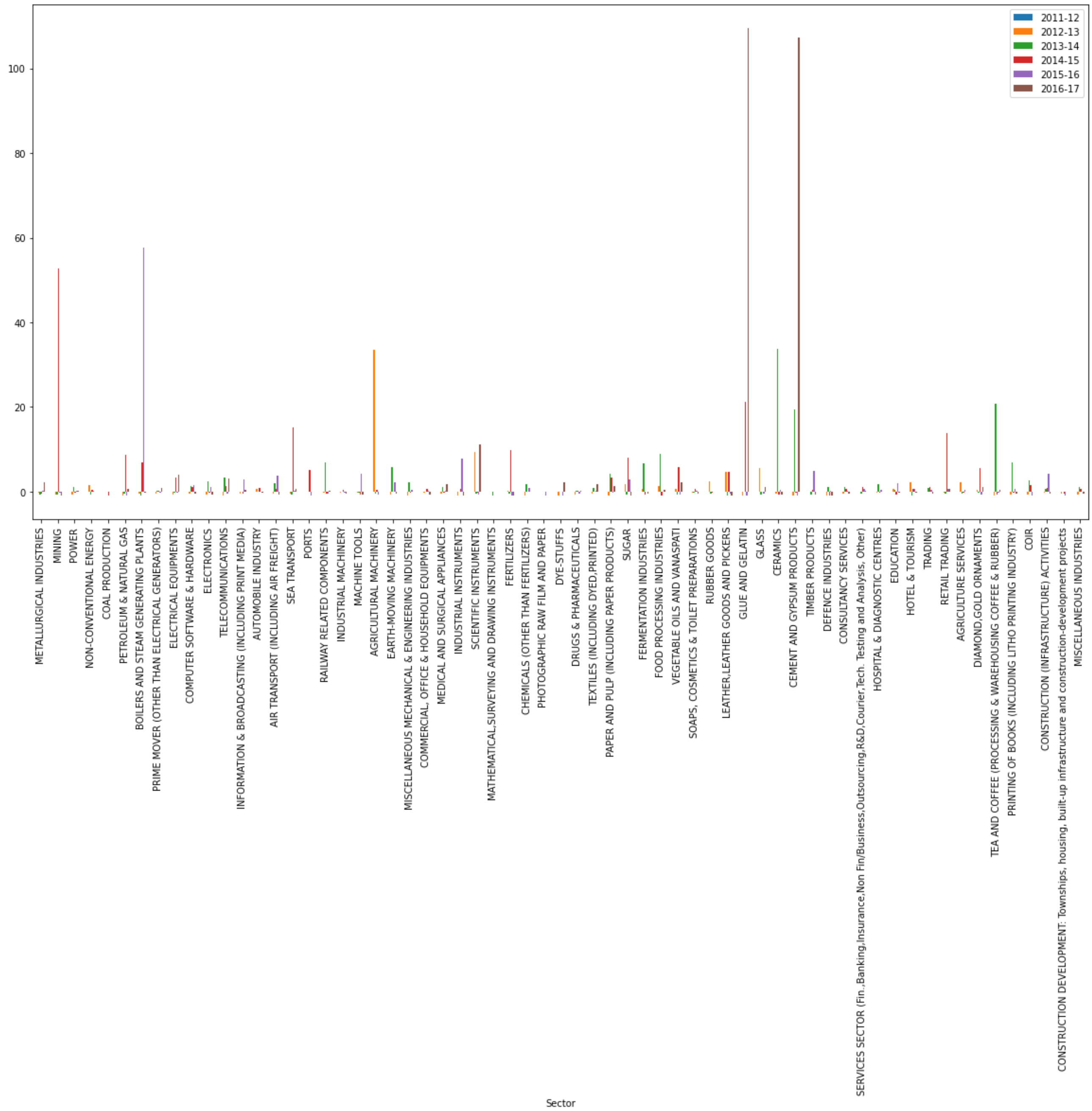
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73928530d0>
```



2011 - 2016

```
df.iloc[:, 11:].pct_change(axis=1).plot(kind = 'bar', figsize=(20,10))
```

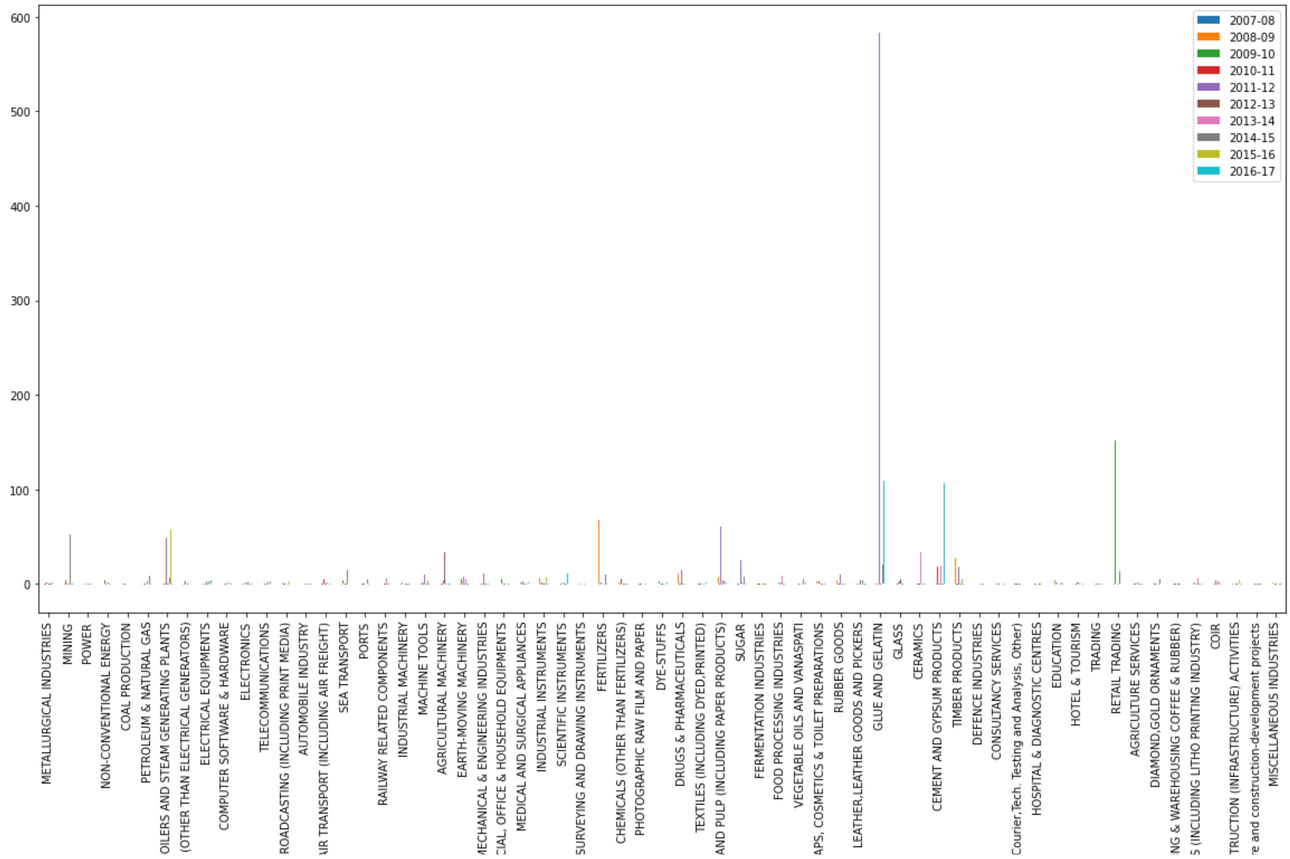
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f73931ebd50>
```



From 2007 - 2016-17, because it marked period of sharp increase

```
df.iloc[:, 7:].pct_change(axis=1).plot(kind='bar', figsize=(20,10))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7390b08d50>
```



Starting to 2006-07 period

(because there has been a sharp increase in investment post this period)

```
df.iloc[:, :7].pct_change(axis=1).plot(kind='bar', figsize=(20,10))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7390989810>
```