Low Level Design

Investment Analytics

Written By	HILAL P V
Document Version	1.0
Last Revised Date	28-07-2022

DOCUMENT CONTROL

VERSION	DATE	AUTHOR	COMMENTS
1.0	28- July - 2022	HILAL P V	Introduction and architecture defined

Reviews:

VERSION	DATE	REVIEWER	COMMENTS

Approval Status:

VERSION	REVIEW DATE	REVIEWED BY		APPROVED BY	COMMENTS

Contents

1. Introduction

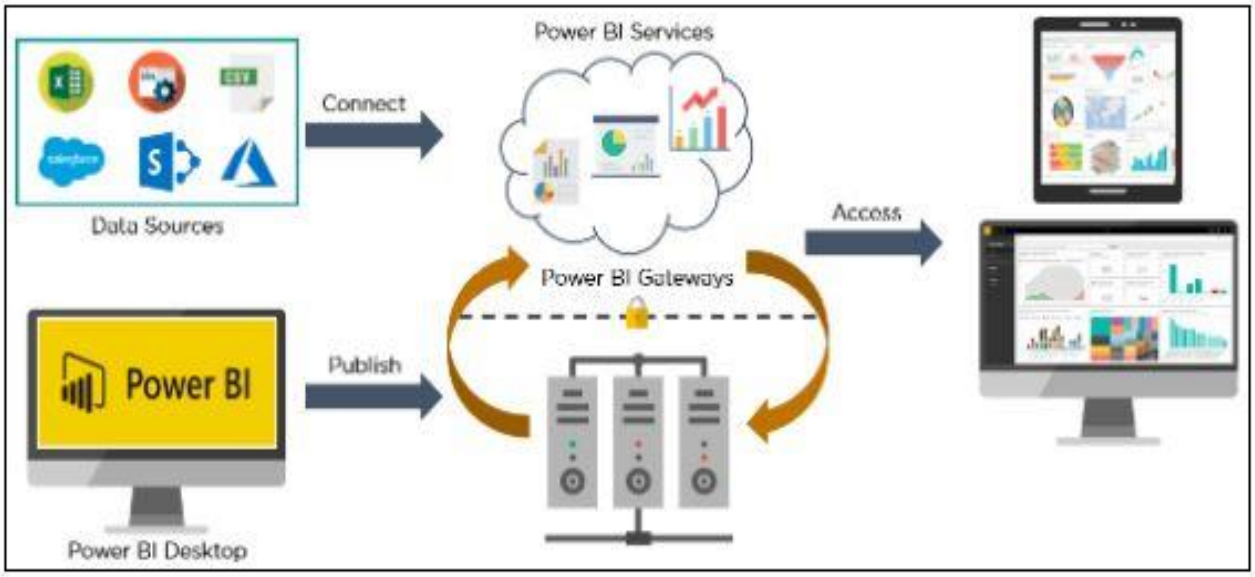
1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Investment Analytics prediction analysis. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

The LLD documentation presents the detailed structure of the Investment analytics for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design. The LLD documentation contains the complete description of the model used along with the comparisons of the proposed model/library compared with a baseline(existing) model against a set of metrics.Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture



:

3. Architecture Description

3.1. Data Description

The dataset is taken from the google drive provided by iNeuron.
Sample data sets are shown below by applying data.head() command on pandas library in Jupyter.

```
In [3]: data.head()
```

Out[3]:

	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
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	359.34	456.31	1440.18
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	684.39	520.67	55.75
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	707.04	868.80	1112.98
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	615.95	776.51	783.57
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	2.96	0.00	0.00	0.00

```
In [8]: data.tail()
```

Out[8]:

	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
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	113.78	72.58	122.81
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	0.54	1.36	0.00
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	485.37	870.25	4510.71
	CONSTRUCTION DEVELOPMENT																

Figure 1: Investment Analytics Dataset

The dataset consists of 63 individual data. There are 18 columns in the dataset which are described below.

1. Sector – There are total 63 sector's name. Some of them are

a. METALLURGICAL INDUSTRIES

b. MINING

c. POWER

d. NON-CONVENTIONAL ENERGY

e. COAL PRODUCTION etc.

2. In the 1st Column Sector name is mentioned and then the remaining 17 columns, historic data of investment are mentioned from 2000-01 to 2016-17.

3.2. Data Transformation

In the Transformation Process, we will convert our original datasets with other necessary attributes format.

1. **Data Pre-Processing**

Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data to the model to train.

This Process includes-

Handling Null/Missing Values

2. **Data Cleaning**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

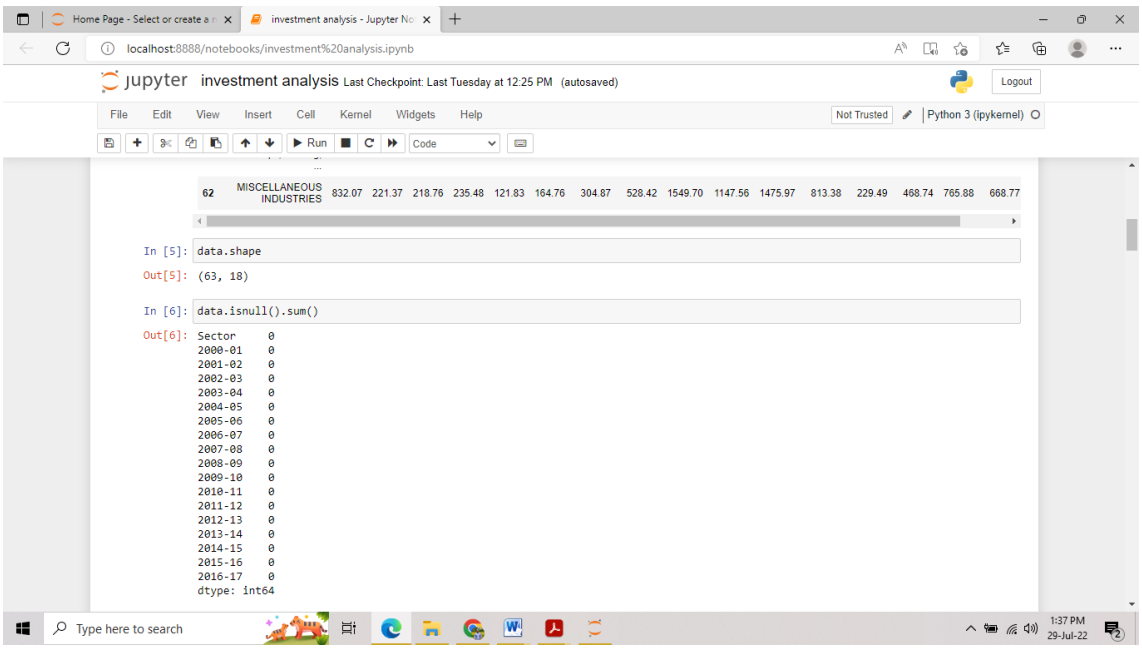
Remove duplicate or irrelevant observations

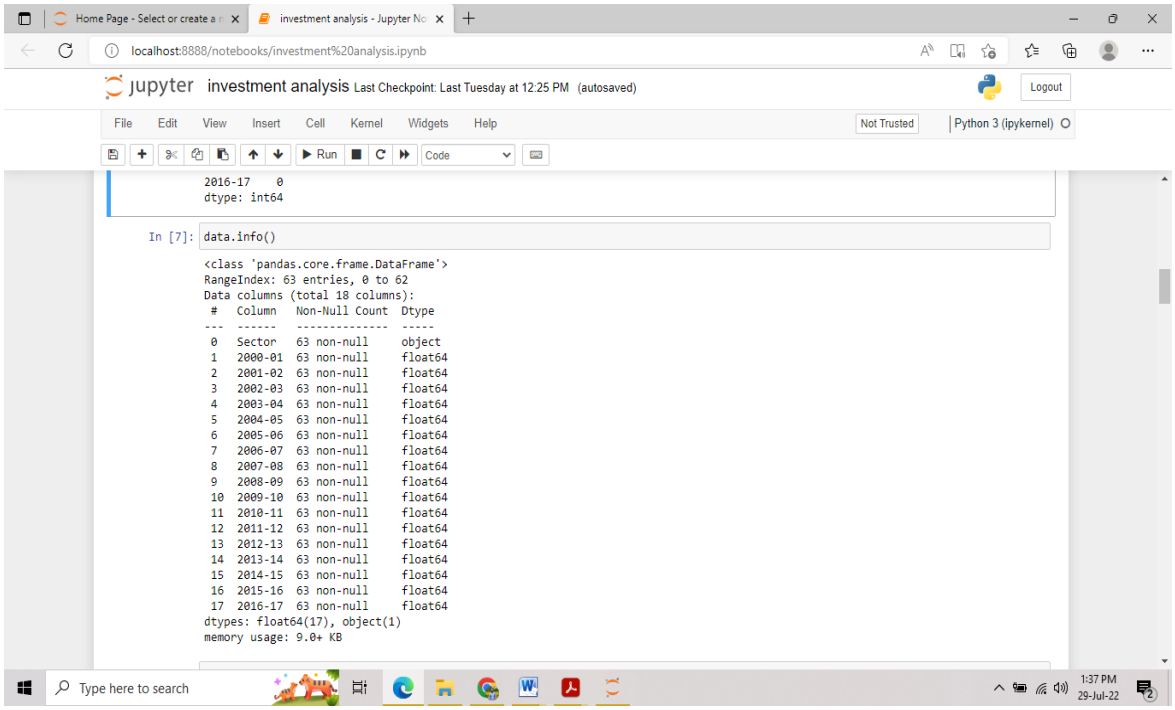
Filter unwanted outliers

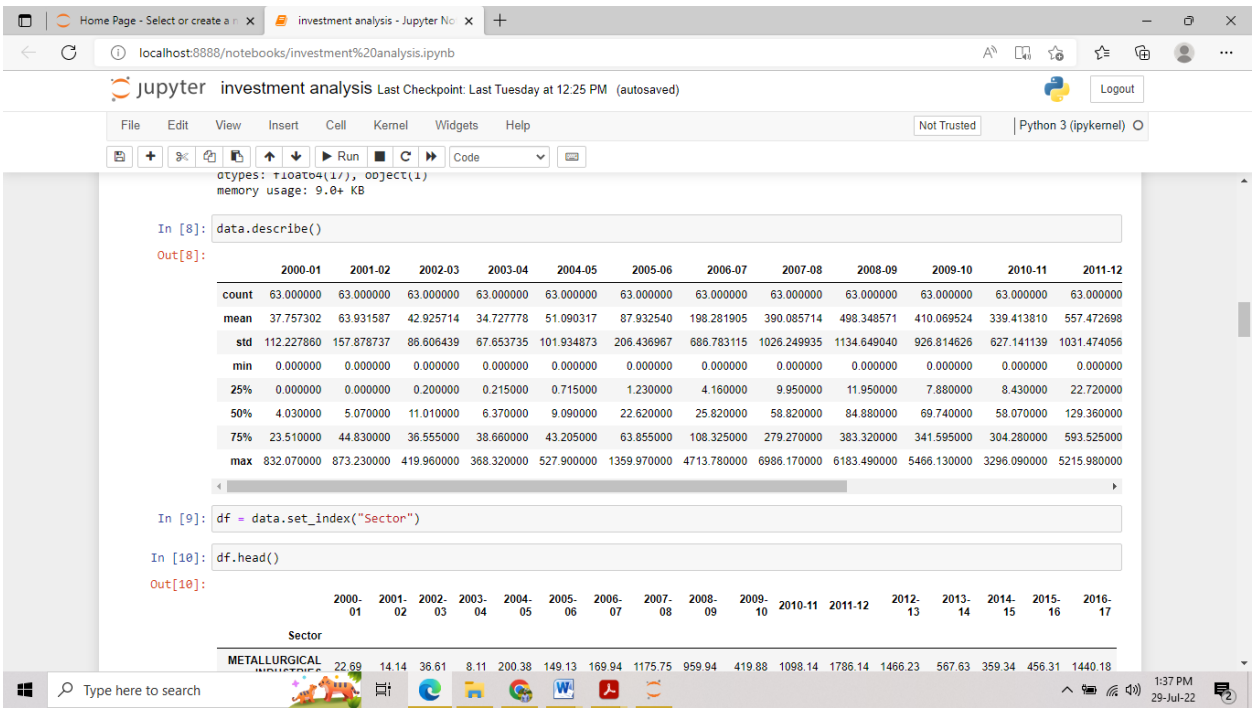
Renaming required attributes

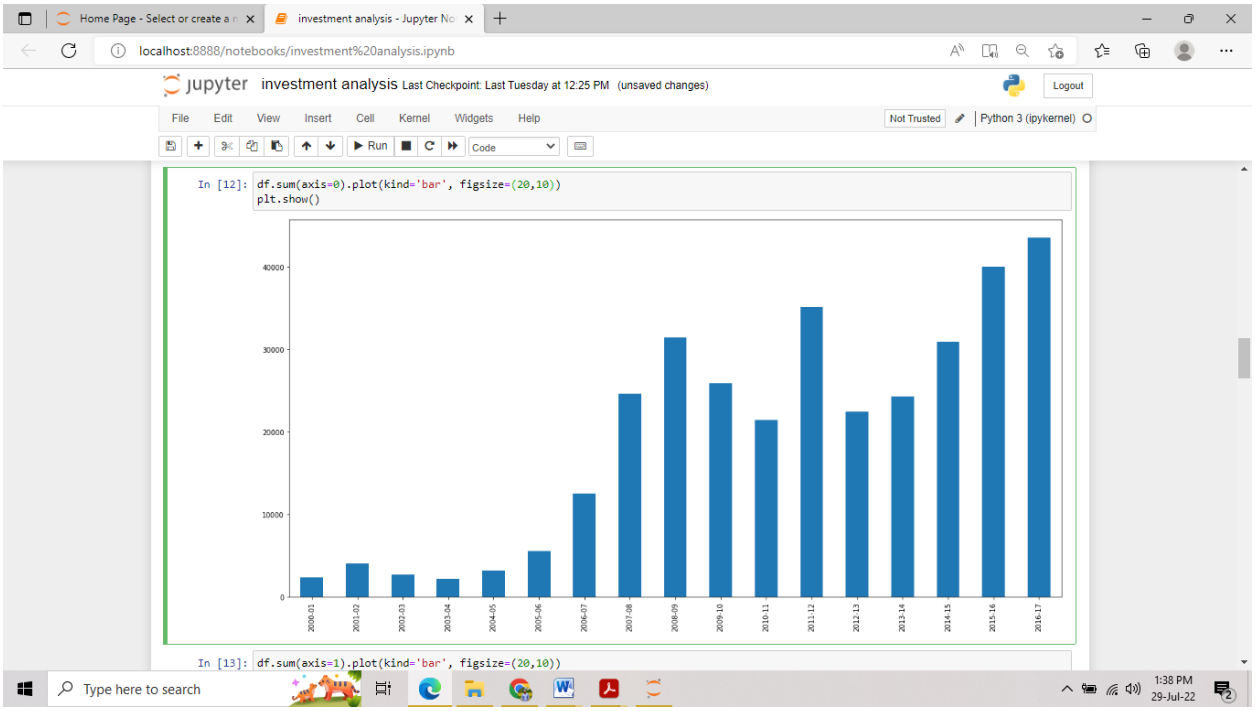
3. **Exploratory Data Analysis (EDA)**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

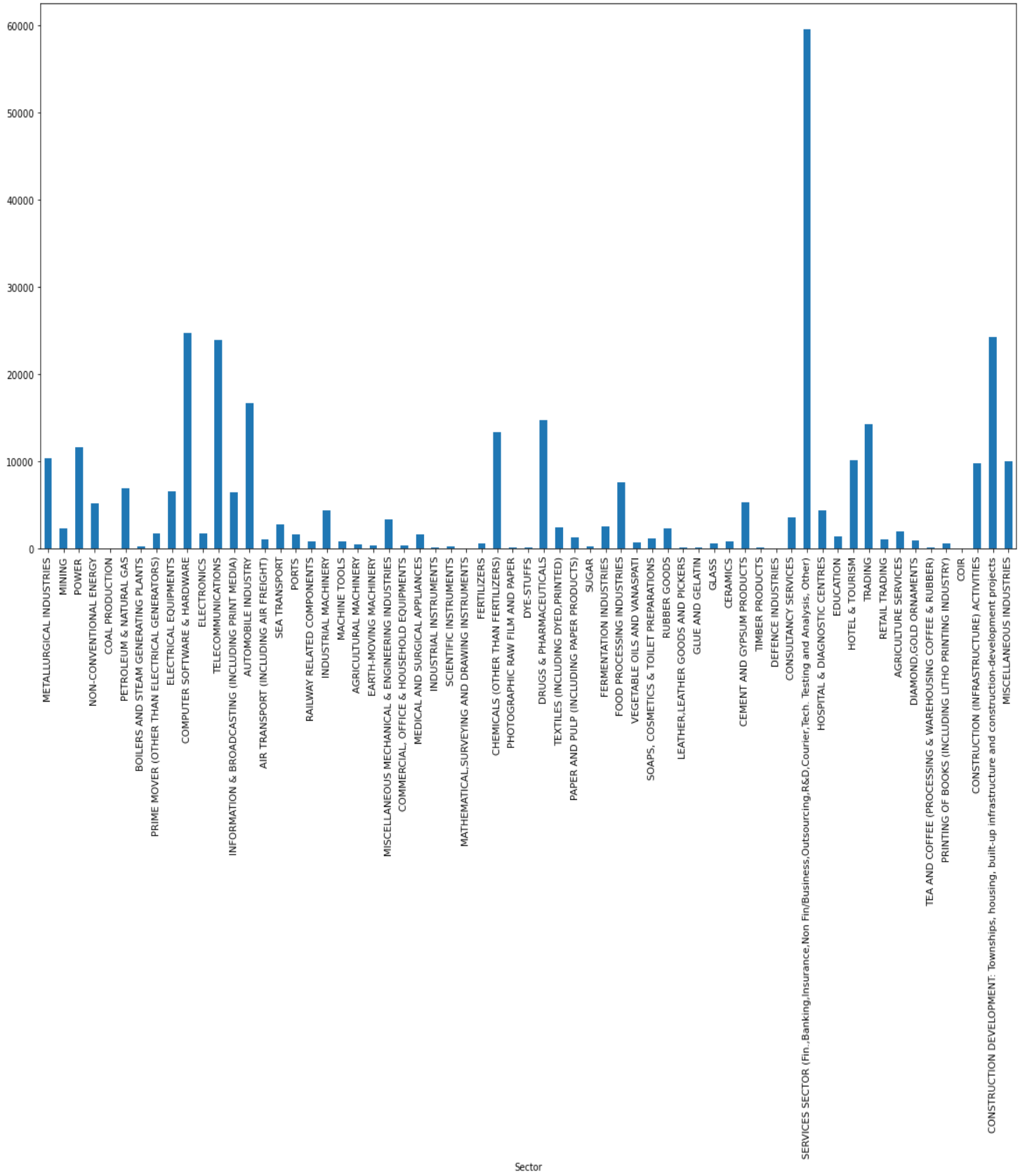

Home Page - Select or create a
investment analysis - Jupyter No
localhost:8888/notebooks/investment%20analysis.ipynb
jupyter investment analysis Last Checkpoint: Last Tuesday at 12:25 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
62 MISCELLANEOUS INDUSTRIES 832.07 221.37 218.76 235.48 121.83 164.76 304.87 528.42 1549.70 1147.56 1475.97 813.38 229.49 468.74 765.88 668.77
In [5]: data.shape
Out[5]: (63, 18)
In [6]: data.isnull().sum()
Out[6]: 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
Type here to search
1:37 PM
29-Jul-22


Home Page - Select or create a
investment analysis - Jupyter No
localhost:8888/notebooks/investment%20analysis.ipynb
jupyter investment analysis Last Checkpoint: Last Tuesday at 12:25 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
2016-17 0
dtype: int64
In [7]: data.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
Type here to search
1:37 PM
29-Jul-22

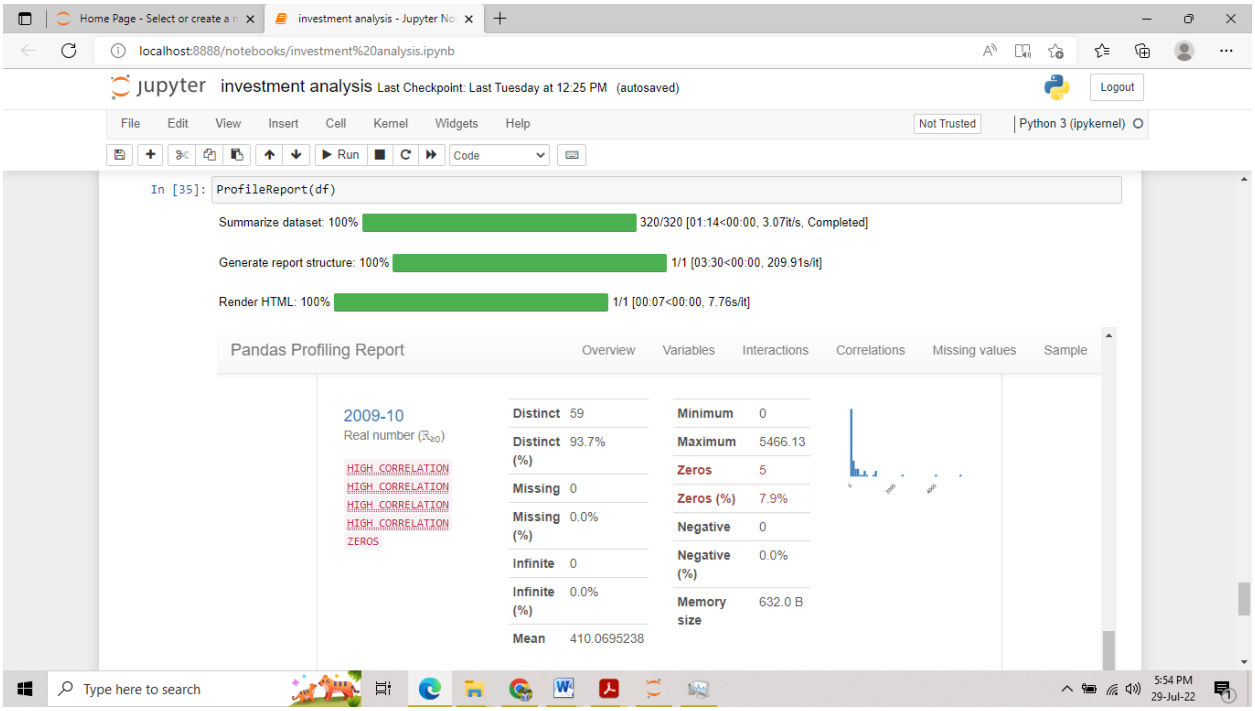

Home Page - Select or create a
investment analysis - Jupyter No
localhost:8888/notebooks/investment%20analysis.ipynb
jupyter investment analysis Last Checkpoint: Last Tuesday at 12:25 PM (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
dtypes: float64(17), object(1)
memory usage: 9.0+ KB
In [8]: data.describe()
Out[8]:
2000-01 2001-02 2002-03 2003-04 2004-05 2005-06 2006-07 2007-08 2008-09 2009-10 2010-11 2011-12
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
mean 37.757302 63.931587 42.925714 34.727778 51.090317 87.932540 198.281905 390.085714 498.348571 410.069524 339.413810 557.472698
std 112.227860 157.878737 86.606439 67.653735 101.934873 206.436967 686.783115 1026.249935 1134.649040 926.814626 627.141139 1031.474056
min 0.000000 0.000000 0.000000 0.000000 0.000000 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 9.950000 11.950000 7.880000 8.430000 22.720000
50% 4.030000 5.070000 11.010000 6.370000 9.090000 22.620000 25.820000 58.820000 84.880000 69.740000 58.070000 129.360000
75% 23.510000 44.830000 36.555000 38.660000 43.205000 63.855000 108.325000 279.270000 383.320000 341.595000 304.280000 593.525000
max 832.070000 873.230000 419.960000 368.320000 527.900000 1359.970000 4713.780000 6986.170000 6183.490000 5466.130000 3296.090000 5215.980000
In [9]: df = data.set_index("Sector")
In [10]: df.head()
Out[10]:
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
Sector
METALLURGICAL 22.69 14.14 36.61 8.11 200.38 149.13 169.94 1175.75 959.94 419.88 1098.14 1786.14 1466.23 567.63 359.34 456.31 1440.18
Type here to search
1:37 PM
29-Jul-22


Home Page - Select or create a
investment analysis - Jupyter No
localhost:8888/notebooks/investment%20analysis.ipynb
jupyter investment analysis Last Checkpoint: Last Tuesday at 12:25 PM (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [12]: df.sum(axis=0).plot(kind='bar', figsize=(20,10))
plt.show()
40000
30000
20000
10000
0
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
In [13]: df.sum(axis=1).plot(kind='bar', figsize=(20,10))
Type here to search
1:38 PM
29-Jul-22

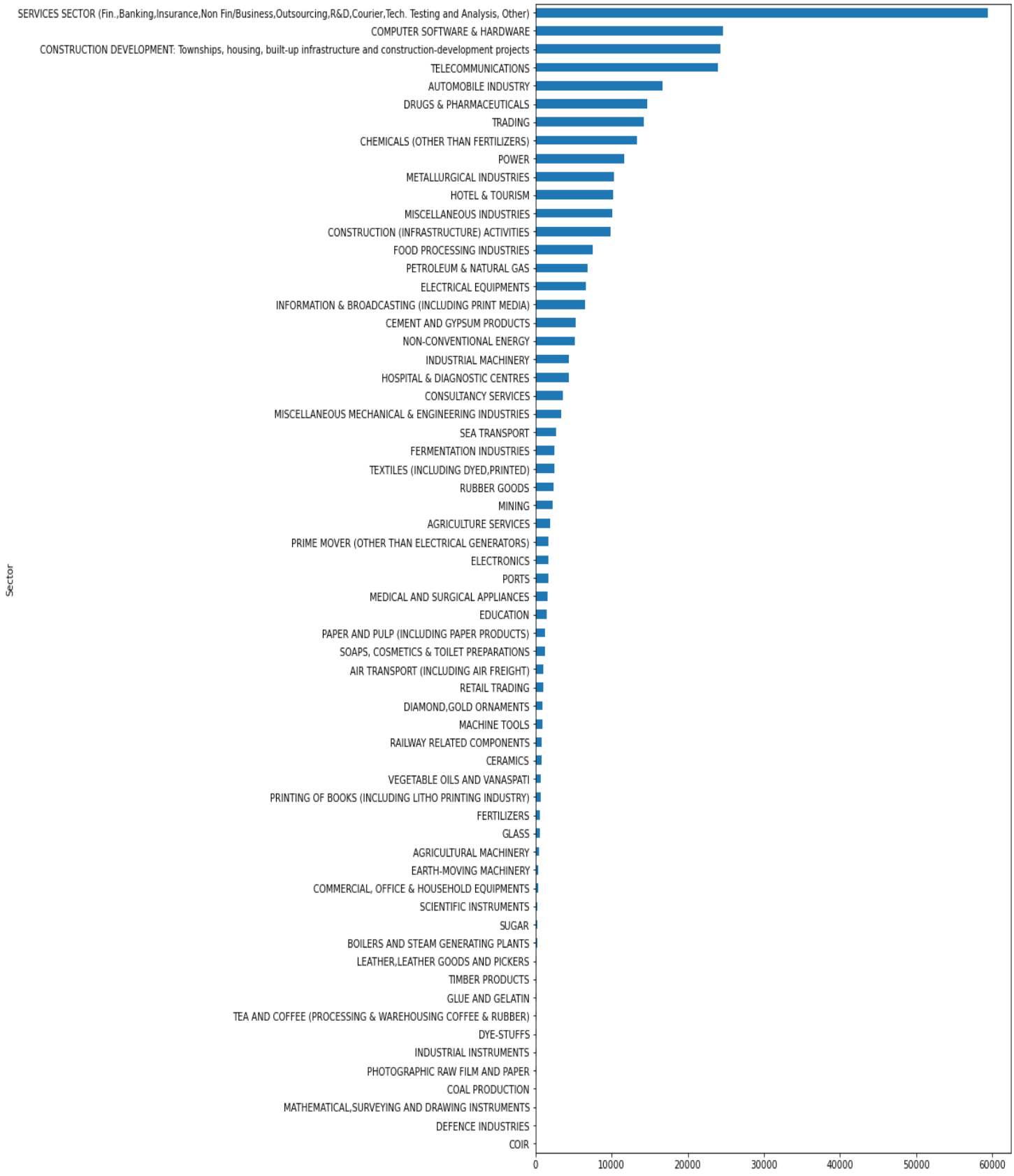
```
df.sum(axis=1).plot(kind='bar', figsize=(20,10))
plt.show()
```

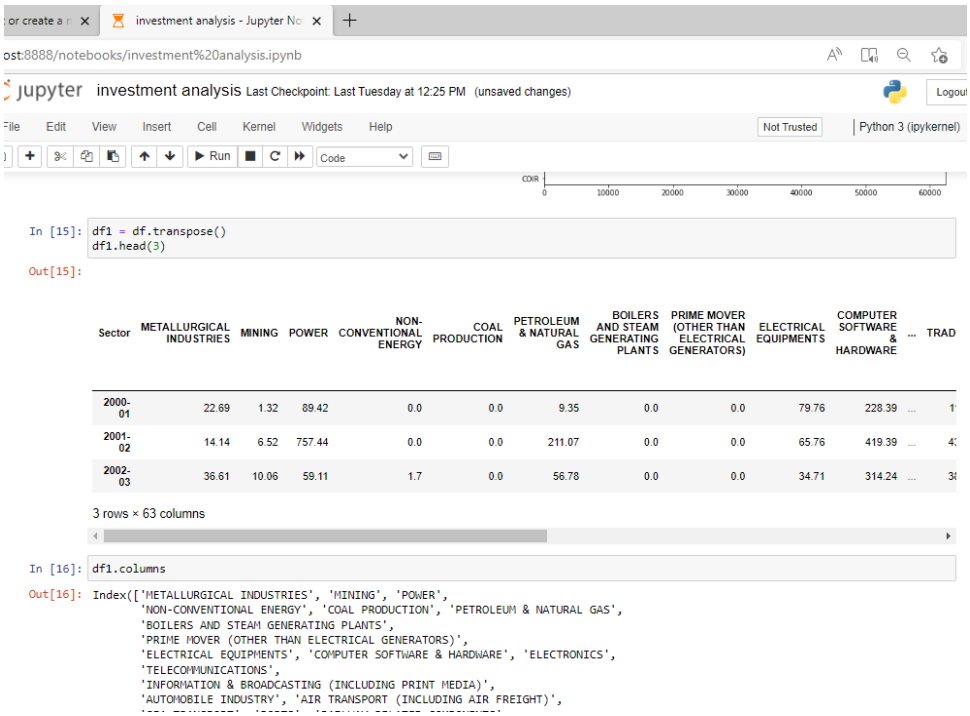


With the help of ProfileReport(df), all the details of the datas and the relationship among each datas are analysed. The corrrelation of the datas are also investigated by using ProfileReport(df) command.



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





```
In [15]: df1 = df.transpose()
         df1.head(3)
```

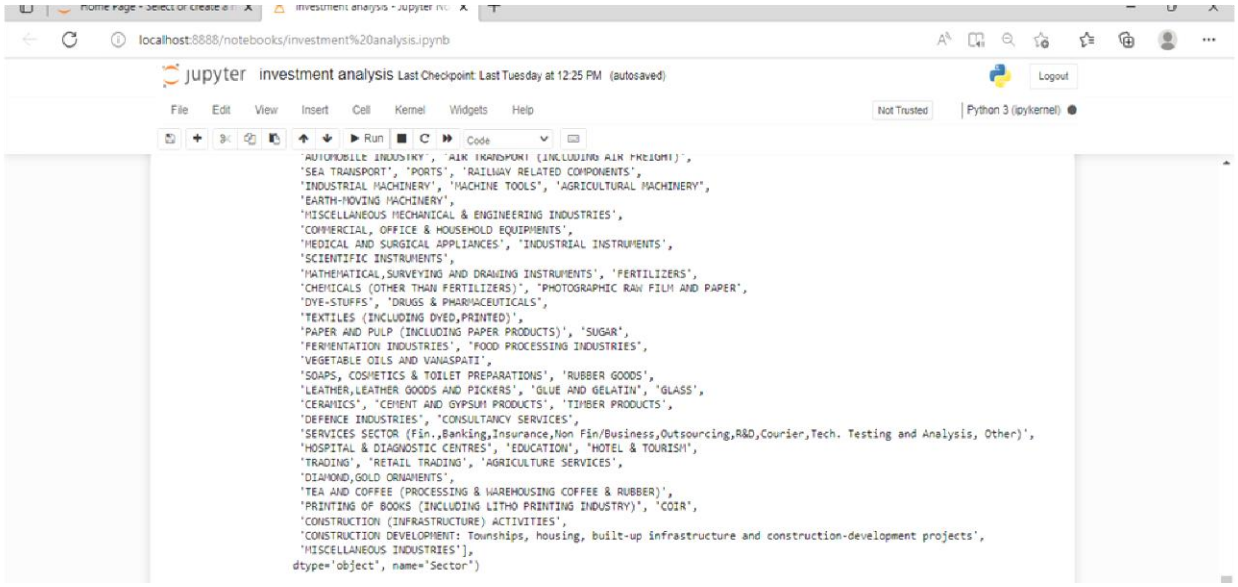
Out[15]:

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	...	TRAD
2000-01	22.69	1.32	89.42	0.0	0.0	9.35	0.0	0.0	79.76	228.39	...	1
2001-02	14.14	6.52	757.44	0.0	0.0	211.07	0.0	0.0	65.76	419.39	...	4
2002-03	36.61	10.06	59.11	1.7	0.0	56.78	0.0	0.0	34.71	314.24	...	3

3 rows × 63 columns

```
In [16]: df1.columns
```

```
Out[16]: 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')
```



In the above code, all the columns are fetched, ie, what are the different sectors in which foreign investment takes place and there are 63 different types of sectors are available.

4. Deployment

The below graph shows Year-wise total Investment from 2000-01 to 2016-17and can easily analyze which year foreign investment happened the most and the trends of investments over years.

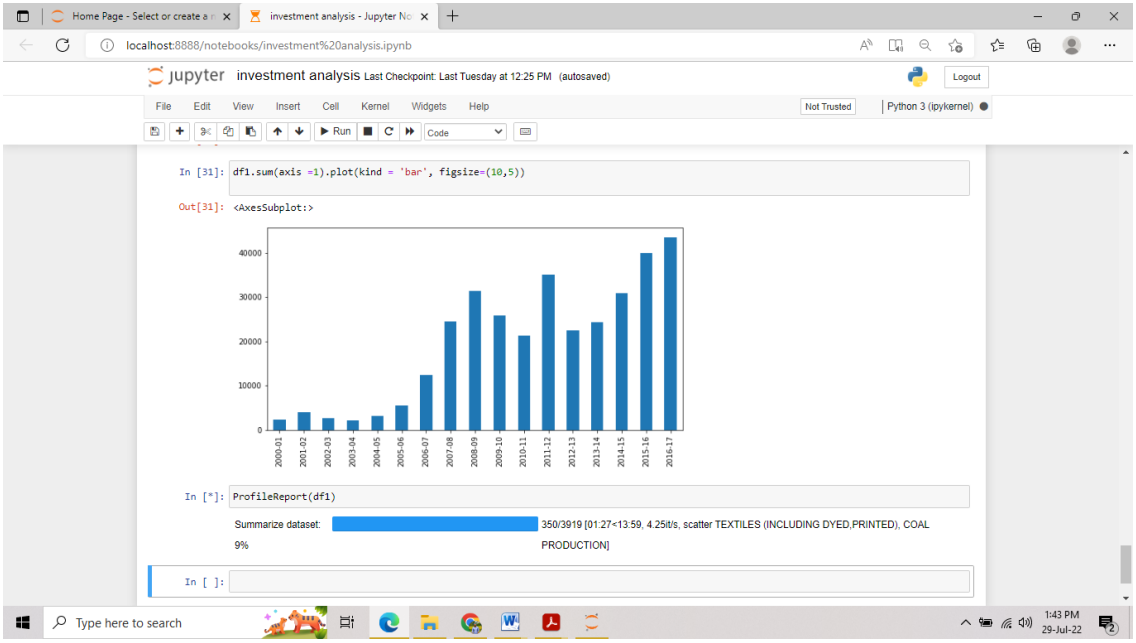


Fig.1: Year wise total investment

The top performing 5 sectors are shown below

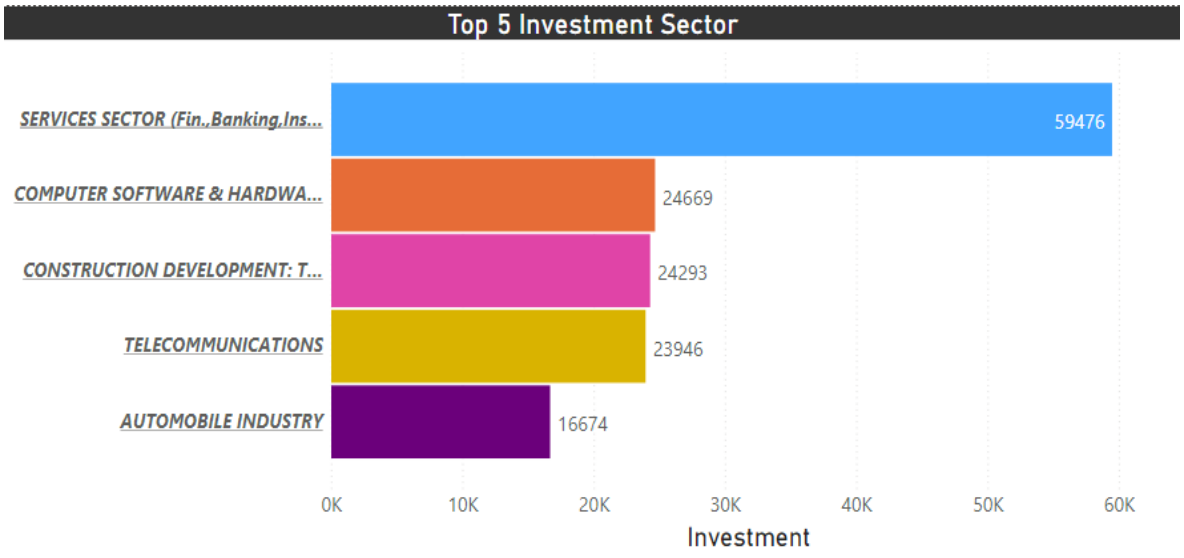


Fig 2: Top Investment Sector Wise

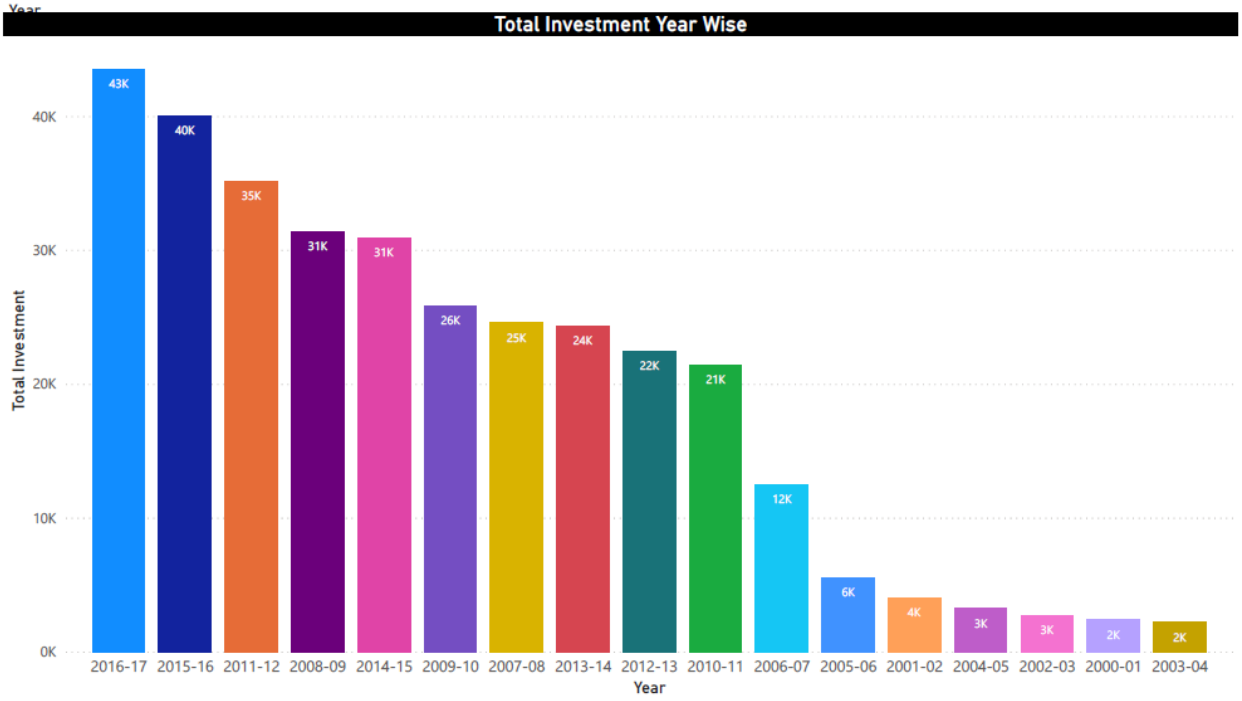


Fig 3: Total Investment Yeas wise

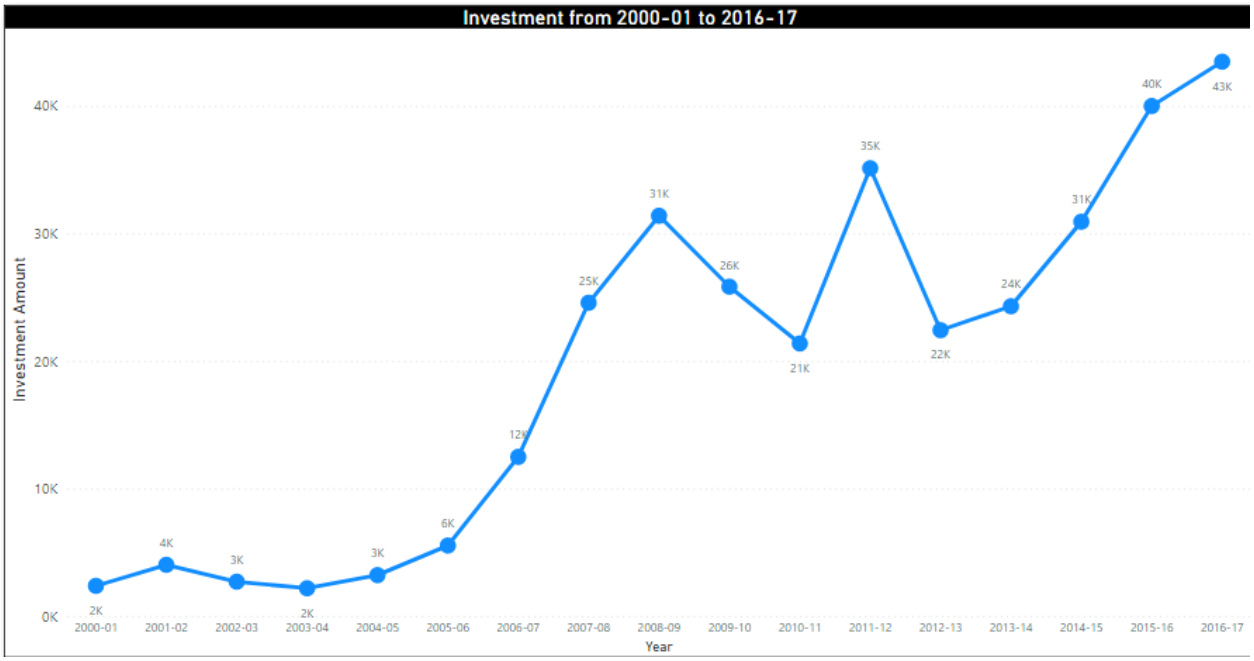


Fig 4: Year wise analytics