Indian Institute of Technology, Kanpur

CS685A: Data Mining
Project Report

Title: Extraction and Analysis of National Institutional Ranking Framework Data

---

Supervised By: Prof. Arnab Bhattacharya

20 November 2021

Submitted By: Group 34

Chabil Kansal (2111122)	chabilk21@iitk.ac.in
Kajal Sethi (21111033)	kajals21@iitk.ac.in
Pranshu Sahijwani (21111048)	pranshus21@iitk.ac.in
Sharanya Saha (21111056)	sharanya21@iitk.ac.in
Utkarsh Srivastava (21111063)	utkarshs21@iitk.ac.in

Contents

List of Abbreviations

NIRF	National Institutional Ranking Framework
MHRD	Ministry of Human Resource Development
PDF	Portable Document File
CSV	Comma Separated values
PCS	Physically Challenged Students
EDP	Executive Development Program
PHD	Doctor of Philosophy
UGP	Undergraduate Program
PGP	Postgraduate Program

1 Abstract

The National Institutional Ranking Framework (NIRF) was approved by the MHRD and launched by Honourable Minister of Human Resource Development on 29^{th} September 2015. This framework outlines a methodology to rank institutions across the country. A lot of information is published by the board but it is not given any analytical form which could help in doing an analytical comparison between multiple institutes. Therefore we aim to extract the data present on the website in PDF format and do various sorts of analysis on the data.

2 Broad Aims of the project

The primary aim is to successfully extract raw data from PDF into a csv data so that it becomes easy to process. Thereafter, analyze the data of different colleges listed on National Institutional Ranking Framework. The analysis is performed for recent years. Finally use some data mining techniques in python to observe the patterns in the data and show it graphically.

3 Datasets

In this section, we explain the data set in terms of its source and its contents.

3.1 Data Source

The data used in the project was released by **National Institutional Ranking Framework, Ministry of Education Government of India** under ranking institutions of higher education program which seeks to provide important information about the college, which is authentic and unbiased. This framework outlines a methodology to rank institutions across the country.The ranking framework uses several parameters for ranking purposes like resources, research, and stakeholder perception. The survey is performed across the country and around 6000 institutions voluntarily participated in the rankings for the year 2021(NIRF Ranking 2021 full list: Check top 20 institutes in India, top 3 from every category). Every year, NIRF publishes list of top institutes in the country and subsequently releases the data submitted by each institute on the website for general public.

3.2 Dataset Contents

We have analysed NIRF 2021 data, NIRF 2020 data and NIRF 2019 data. Each of them consists of PDF files i.e. one PDF file for each institute. Therefore, in total analysis 300 PDF files have been extracted and used subsequently. The data contained in each PDF is listed below:

- **Sanctioned (Approved) Intake** - This table lists all the programs provided by a particular institute and the sanctioned number (strength) of students for each program over a span of 6 years.

- **Total Actual Student Strength** - This table shows the actual number of students enrolled in each program and number of students in various categories such as number of males, females, students from within the state, socially challenged students, students getting their tuition fees reimbursed by various organizations etc.

- **Placement & Higher Studies** - This section contains multiple tables depending on the number of programs provided by the institute. There is one table for each program. Each table provides data for a program stating the performance of students in that particular programs. Each table consists of data pertaining to columns namely number of students graduated, number of students placed, students pursuing higher studies, median salary received by the students etc.

- **Ph.D Student Details** - States total number of full time and part time students and their year of admission for previous 3 years as counted from the year in which that report is being published.

- **Financial Resources** - States data about amount utilised by the institute for the Capital expenditure and Operational expenditure for previous 3 years as counted from the year in which that report is being published.

- **Sponsored Research Details** - This table consists of details of last 3 years pertaining to fields namely Number of Sponsored Projects, Number of Funding Agencies, and the Amount Received in lieu of such work.

- **Consultancy Project Details** - This table consists of details of last 3 years pertaining to fields namely Number of Consultancy Projects, Number of Client Organizations, and the Amount Received in lieu of such work.

- **Executive Development Program/Management Development Programs** - This table consists of details pertaining to fields like Total no. of Executive Development Programs/ Management Development Programs, Number of students in such program and the Amount received in lieu of such program. This data is provided for previous 3 years as counted from the year in which the report is being published.

- **PCS Facilities** - This table consists of details about the facilities provided to Physically Challenged Students. These faculties include having lifts/ramps in buildings, walking aids like wheelchair for transportation and facility for toilets for handicapped students.

- **Faculty Details** - This table contains the number of faculty members on roll in the institute.

3.3 Special Note Regarding Dataset

A special point to note is that due to the varying number of programs in different institutes, the number of tables, columns and rows is different in different PDF files.

4 Data Transformation

It was challenging to perform analytical analysis on the raw data given in the official NIRF website was in PDF format. Therefore we tried to extract the data using various packages available in Python. These were PyPDF, Tabula, Camelot. Using a combination of these packages, we extracted the relevant yet common information across all the top colleges across last three years and stored in the **csv** files. The files were then used for further processing and analysis.

- There were PDF files, one for each of the top colleges, for the year 2020-21. The same trend continued for years 2019-20, 2018-19 as well.

National Institutional Ranking Framework
Ministry of Education
Government of India

Data Submitted by Institution for India Rankings '2021'
Institute Name: Indian Institute of Technology Kanpur [IR-O-I-1075]

Sanctioned (Approved) Intake

Academic Year	2019-20	2018-19	2017-18	2016-17	2015-16	2014-15
UG [4 Years Program(s)]	907	910	827	827	-	-
PG [2 Year Program(s)]	1107	976	-	-	-	-

Total Actual Student Strength (Program(s) Offered by your Institution)

(All programs of all years)	No. of Male Students	No. of Female Students	Total Students	Within State (Including male & female)	Outside State (Including male & female)	Outside Country (Including male & female)	Economically Backward (Including male & female)	Socially Challenged (SC+ST+OBC Including male & female)	No. of students receiving full tuition fee reimbursement from the State and Central Government	No. of students receiving full tuition fee reimbursement from Institution Funds	No. of students receiving full tuition fee reimbursement from the Private Bodies	No. of students who are not receiving full tuition fee reimbursement
UG [4 Years Program(s)]	3100	407	3507	883	2622	2	398	1732	548	1541	0	41
PG [2 Year Program(s)]	1642	265	1907	455	1444	8	582	746	67	1151	0	110

Placement & Higher Studies

UG [4 Years Program(s)]: Placement & higher studies for previous 3 years

Academic Year	No. of first year students intake in the year	No. of first year students admitted in the year	Academic Year	No. of students admitted through Lateral entry	Academic Year	No. of students graduating in minimum stipulated time	No. of students placed	Median salary of placed graduates(Amount in Rs.)	No. of students selected for Higher Studies
2014-15	827	809	2015-16	0	2017-18	789	354	1400000(Fourteen Lakh)	387
2015-16	827	803	2016-17	0	2018-19	770	437	1500000(Fifteen Lakh)	301
2016-17	827	826	2017-18	0	2019-20	789	463	1700000(Seventeen Lakh)	325

PG [2 Years Program(s)]: Placement & higher studies for previous 3 years

Academic Year	No. of first year students intake in the year	No. of first year students admitted in the year	Academic Year	No. of students graduating in minimum stipulated time	No. of students placed	Median salary of placed graduates(Amount in Rs.)	No. of students selected for Higher Studies
2016-17	950	952	2017-18	780	499	1128650(Eleven Lakh Twenty Eight Thousand Six Hundred Fifty)	225
2017-18	910	906	2018-19	774	534	1088500(Ten Lakh Eighty Eight Thousand Five Hundred)	201

Figure 1: Snapshot of Page 1 of a NIRF PDF file

5 Data Pre-processing

This section is further divided into two subsections: Data Extraction and Data Cleaning.

5.1 Data Extraction:

The data was extracted from the PDF files available at the NIRF web page and were stored into csv in a suitable format.

5.1.1 Characteristics of Data:

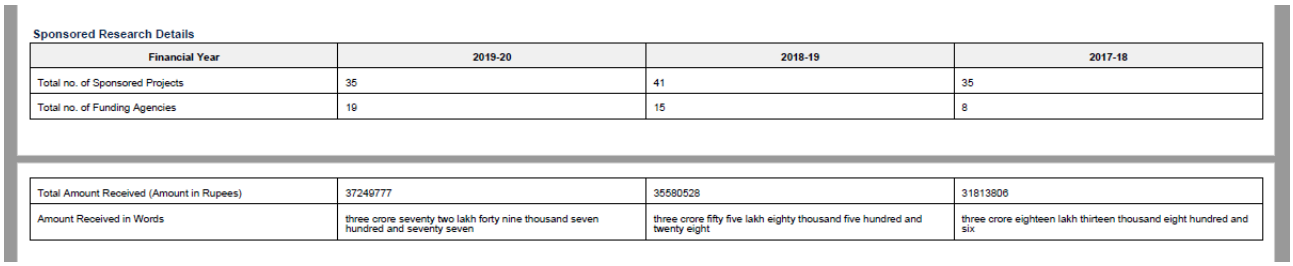
The data available at NIRF(National Institutional Ranking Framework) webpage has the following characteristics:

1. Each college has a separate PDF file. Therefore, there are PDF files for each year.
2. The number of tables vary in every PDF ranging from 9 to 19.

For a particular year, for example 2021 some of the PDF had 9 tables and some had more than 15. We analysed the PDFs to find the consistent tables and then went with the extraction of the same. Table 1, for all the colleges were extracted and stored in a single csv along with their college names. Storing in this format helped us in doing a comparative analysis of all the colleges on various important points. All the relevant information was extracted in similar format.

5.1.2 Challenges:

1. For the year 2021, each of the PDF file had a single lined table. A single line table cannot be extracted directly in tabular format using the available packages. These tables were extracted by reading texts from a page.
2. Some of the tables were split across two pages as shown in Figure 2. The split tables cannot be detected using the available packages. A custom parser was created by looking at the data files to resolve such issues.



Sponsored Research Details

Financial Year	2019-20	2018-19	2017-18
Total no. of Sponsored Projects	35	41	35
Total no. of Funding Agencies	19	15	8

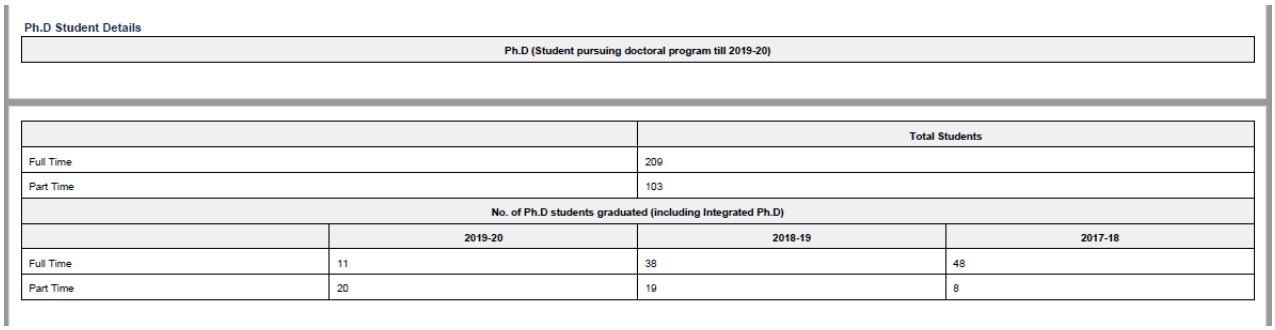
Total Amount Received (Amount in Rupees)	37249777	35580528	31813806
Amount Received in Words	three crore seventy two lakh forty nine thousand seven hundred and seventy seven	three crore fifty five lakh eighty thousand five hundred and twenty eight	three crore eighteen lakh thirteen thousand eight hundred and six

Figure 2: Snapshot of cross page table in a PDF file

3. PDF files of each year had a different format, so the parser had to be created separately for every year.

4. Some of the names of colleges had minor differences in institute name. For example: BITS Pilani and Birla Institute of Technology and Science Pilani. Such cases were handled separately.

5. Some of the cross page tables had mismatching number of columns as shown in Figure 3 because of the type of program in that institute.



Ph.D Student Details

Ph.D (Student pursuing doctoral program till 2019-20)			
	Total Students		
Full Time	209		
Part Time	103		
No. of Ph.D students graduated (including Integrated Ph.D)			
	2019-20	2018-19	2017-18
Full Time	11	38	48
Part Time	20	19	8

Figure 3: Snapshot of mismatched columns in cross page table

6. Many of the edge cases had to be handled manually, because of lack of uniformity in the data format as shown in Figure 4.

Institute	Ph.D (Student pursuing doctoral program till 2019-20)	a	b	c
Institute of Chemical Technology			Total Students	
Institute of Chemical Technology	Full Time		511	
Institute of Chemical Technology	Part Time		106	
Institute of Chemical Technology	No. of Ph.D students graduated (including Integrated Ph.D)			
Institute of Chemical Technology		2019-20	2018-19	2017-18
Institute of Chemical Technology	Full Time	140	135	137
Institute of Chemical Technology	Part Time	5	7	3
Jamia Millia Islamia			Total Students	
Jamia Millia Islamia	Full Time		1483	
Jamia Millia Islamia	Part Time		0	
Jamia Millia Islamia	No. of Ph.D students graduated (including Integrated Ph.D)			
Jamia Millia Islamia		2019-20	2018-19	2017-18
Jamia Millia Islamia	Full Time	345	360	309
Jamia Millia Islamia	Part Time	0	0	0
Kalasalingam Academy of Research and Higher Education			Total Students	
Kalasalingam Academy of Research and Higher Education	Full Time		303	
Kalasalingam Academy of Research and Higher Education	Part Time		137	
Kalasalingam Academy of Research and Higher Education	No. of Ph.D students graduated (including Integrated Ph.D)			
Kalasalingam Academy of Research and Higher Education		2019-20	2018-19	2017-18

Figure 4: Snapshot of a csv for a table extracted from the PDF

5.2 Cleaning:

- The spellings for names of the colleges varied over the years. Such cases were taken care of manually.
- The data-set contained NULL values for some of the tables. The NULL values were filled with 0 or ignored depending upon the requirement and feasibility.
- In order to make the data uniform, the colleges which were not in top for all the three years were ignored from analysis.

- Some of the college names had unnecessary spaces in the beginning and in the end. In such cases, we used strip function in Python to make it correct. The resultant csv after performing these steps is shown in Figure 5

Institute	Number of Faculties
Institute of Chemical Technology	168
Jamia Millia Islamia	742
Kalasalingam Academy of Research and Higher Education	526
Sri Venkateswara University	510
Indian Institute of Technology Guwahati	436
Gujarat University	381
Osmania University	428
Indian Institute of Science Education & Research Kolkata	127
Anna University	1024
Indian Institute of Technology Roorkee	574
Jamia Hamdard	473
University of Delhi	1060
Sardar Vallabhbhai National Institute of Technology	314
Indian Institute of Technology Patna	127
Amity University	2068

Figure 5: Snapshot of a csv after data cleaning and formatting

6 Data Processing & Analysis

In this section, we discuss in detail about the various processing that we performed and the analysis that we inferred from them using graphical and numerical methods.

6.1 Analysis of consistency of Institutes in NIRF Rankings

NIRF ranks the institutes and publishes the top hundred institutes each year along with their details. We have analysed the Top institutes as published by NIRF for the years 2019, 2020 and 2021. 86 of the institutes were consistent and have made into the Top for all the three years. The names of such institutes can be found in *Consistent_Institutes.csv*. Out of the rest 14 institutes in NIRF Top 100, 2021, 8 of the institutes managed to grab a rank within for the years 2020-21. The remaining six institutes ranked under only for the year 2021.

Insititute
Indian Institute of Technology Indore
Dr. B. R. Ambedkar National Institute of Technology
Sardar Vallabhbhai National Institute of Technology
Guru Gobind Singh Indraprastha University
Chandigarh University
Malaviya National Institute of Technology
Lovely Professional University
Visvesvaraya National Institute of Technology
Motilal Nehru National Institute of Technology
Gujarat University
Mumbai University
National Institute of Technology Silchar
Indian Institute of Technology Ropar
Kalasalingam Academy of Research and Higher Education

Figure 6: List of the institutes which do not appear in Top for the three consecutive years. *Institutes which were in Top for both the year 2020 and 2021 are marked with yellow, rest of the institutes managed a rank within top only for the year 2021.*

6.2 How frequently a program is conducted in the top institutes and which institutes are conducting rare programs.

In this section, we have found out what all programs are conducted in the top institutes across the country and counted their frequency. Figure 7 shows the graph representing the frequency of a particular program. This information was then used and the three least frequent programs were considered to be rare. A csv was then created which containing the institutes conducting the rare programs named as **rare_program.csv**.

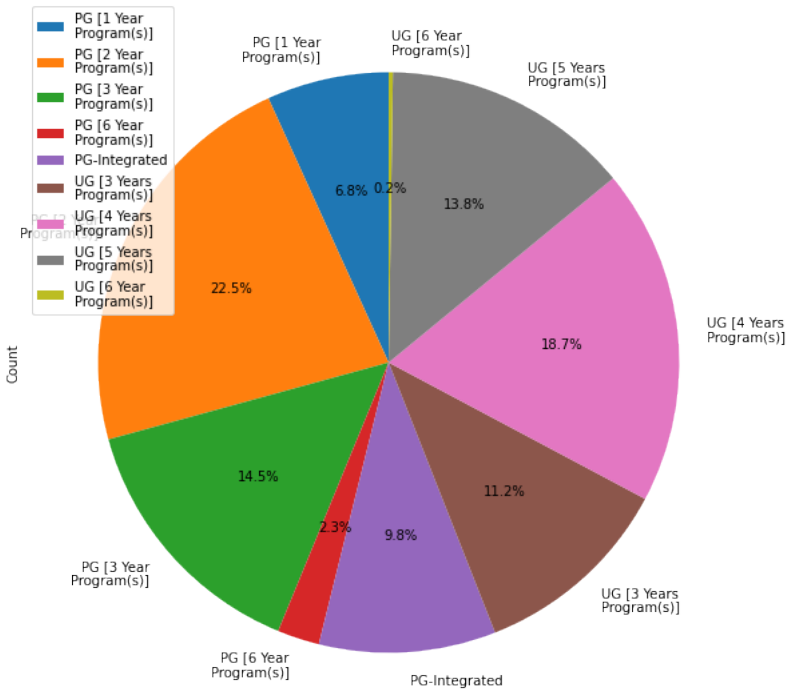


Figure 7: Frequency of every program taught in institutes

6.3 Program per institute that have the highest as well as the lowest intake for last three years.

For this section, the allowed number of students per year for each program were added and compared. Afterwards this data was compared to yield the programs that have the highest and the lowest intake. This was done for each of the institutes and the data was saved in the csv named as **min_max_program.csv**. Using this csv we have calculated how many times a particular program comes under the section "Maximum intake program" and "Minimum intake program". Figure 8 shows the graph representing the frequency of a each program coming under the category of "Maximum intake program" and "Minimum intake program".

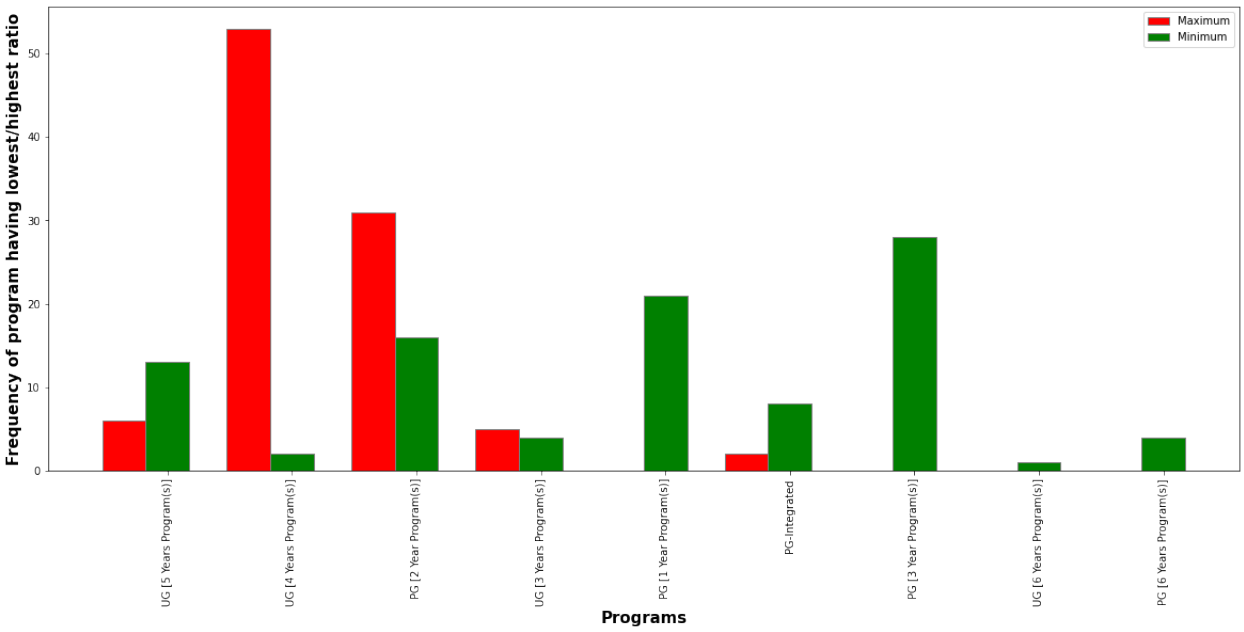


Figure 8: Frequency of every program coming under the category of "Maximum intake program" and "Minimum intake program"

6.4 Year in which the funding agencies brought in the maximum research based projects.

This section, to calculate the year in which the funding agency provided for the maximum number of research based projects, we calculated how many projects were brought by each of the agency(by calculating the ratio of projects to the number of agencies) and selecting the year with highest ratio. This was performed for every institute and data is stored in **max_project_year.csv**. This data was then plotted revealing with what frequency each of the year has most active funding agencies showed in Figure 9.

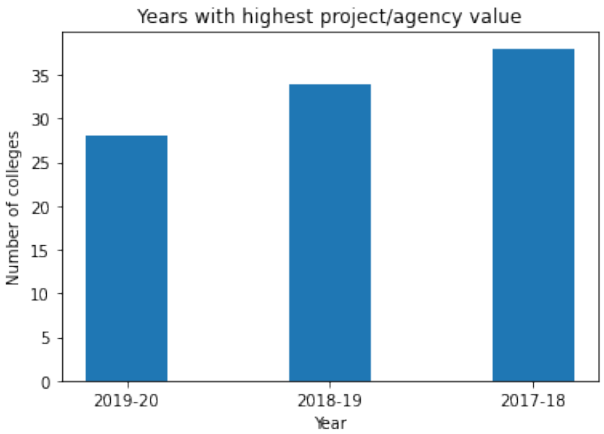


Figure 9: Frequency of every program coming under the category of "Maximum intake program" and "Minimum intake program"

6.5 The institutes that have been receiving the highest sponsorship for research projects.

In this section, the amount received by every institute by the sponsors have been observed for the past three years and the added. The amount reveals that how much the institute is being funded for conducting researches. The result is stored in **highest_sponsorship.csv**. Figure 10.

Institute	Total
Indian Institute of Science	1.29562e+10
Indian Institute of Technology Madras	1.23039e+10
Indian Institute of Technology Delhi	9.89643e+09
Indian Institute of Technology Bombay	8.36133e+09
Indian Institute of Technology Kanpur	6.29249e+09

Figure 10: The institutes having highest sponsorship in last three years.

6.6 Total intake of every institute in last years.

This section, helps us knowing how the number of students admitted to the top institutes of the country increases with time also indicating the growth of the institute in terms of facilities, infrastructure, faculty, better career option, etc. This leaves us with the formation of 6 csv files namely **2014-15_intake.csv**, **2015-16_intake.csv**, **2016-17_intake.csv**, **2017-18_intake.csv**, **2018-19_intake.csv** and **2019-20_intake.csv**. We have also created a overall file which shows the institutes having highest intake in past 6 years by the name **overall_intake.csv** Figure 11.

Institute	Total
S.R.M. Institute of Science and Technology	59147
Vellore Institute of Technology	41545
Amity University	40430
Lovely Professional University	32790
Chandigarh University	26501

Figure 11: The institutes with maximum intake in past six years.

6.7 Average number of PhD students over last 3 years

In the module, the average number of PhD students in each institute in Part Time and average number of PhD students in each institute in Full Time PhD program is calculated and stored in file named **phd.csv**. Here we show the top 5 institutes having highest

number of PhD students in Part Time PhD Figure 12 and similarly for Full Time PhD program Figure 13.

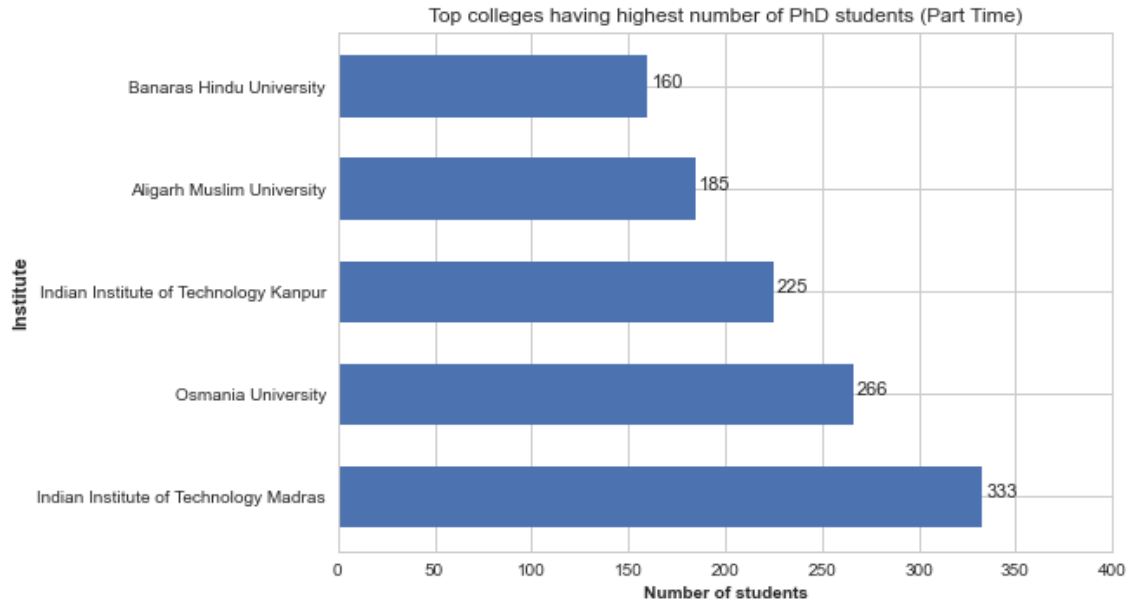


Figure 12: Part Time PhD students

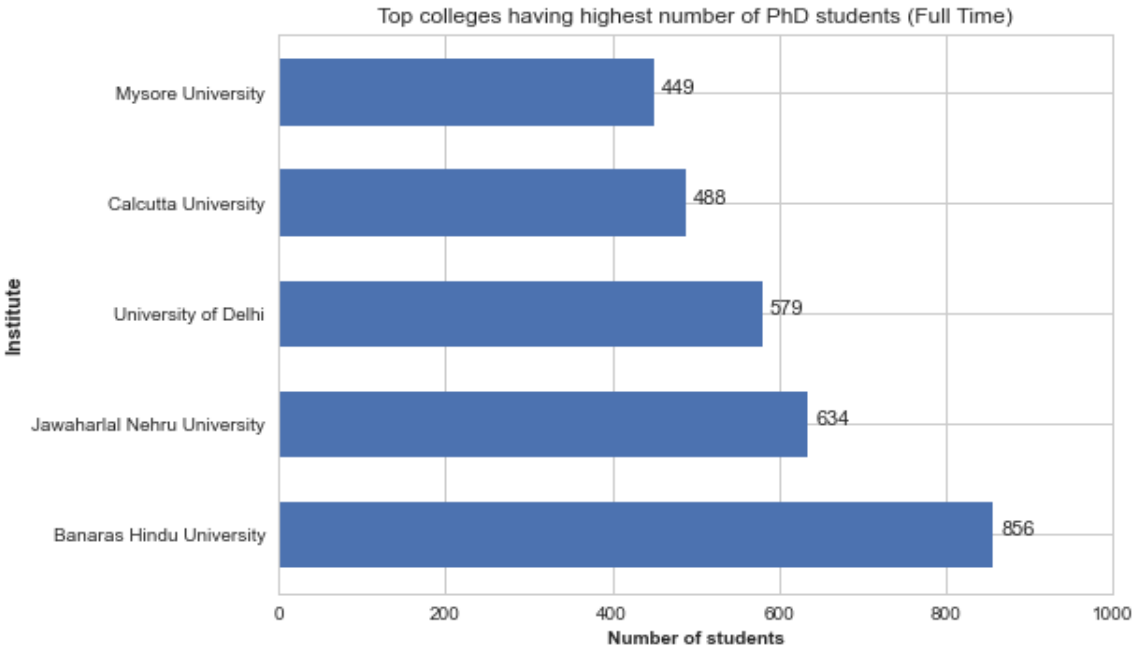


Figure 13: Full Time PhD students

6.8 Best Program in terms of percentage of students getting placed

In the module, the best program in each college in terms of percentage of students getting placed is calculated. To do this, we have used placement statistics for academic year 2015-

16 upto 2019-20. For each college, we have calculated the program which has highest percentage of students getting placed and the actual percentage of placements in that program. This result is stored in file named **best_program.csv**. Here we show a pie chart representing the number of colleges having a particular program as best in terms of placements Figure 14.

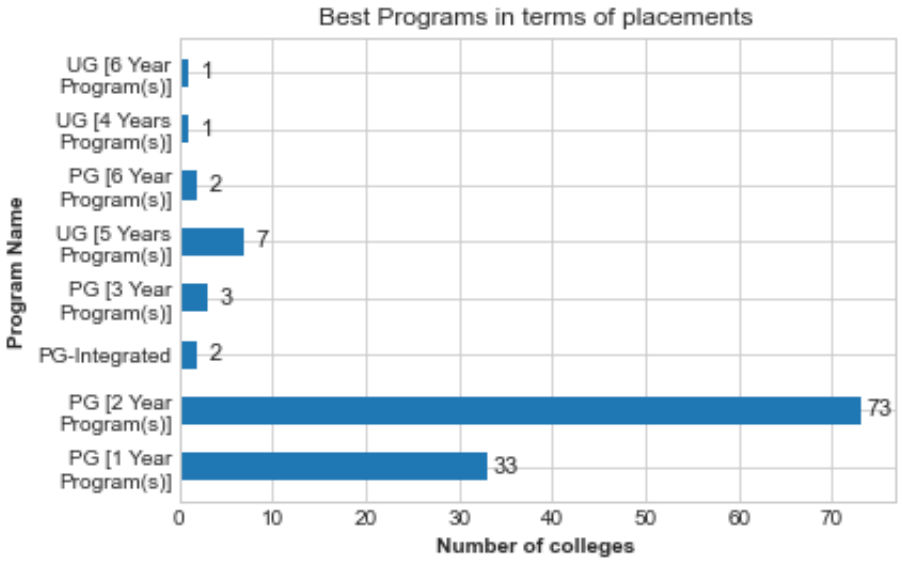


Figure 14: Best programs in college

6.9 Percentage of students going for higher studies

In the module, the best and the worst institute in each year in terms of percentage of students going for higher studies is calculated. To do this, we have used placement statistics for academic year 2015-16 to 2019-20. For each year, we have calculated the institute which has highest and the lowest percentage of students going for higher studies and the actual percentage of students going for higher studies. This result is stored in file named **higher_studies.csv**. This csv file could be used in future to see the performance of any institute over the years. Here we show graphs showing the best institute in each year Figure 15 and the worst institute in each year Figure 16.

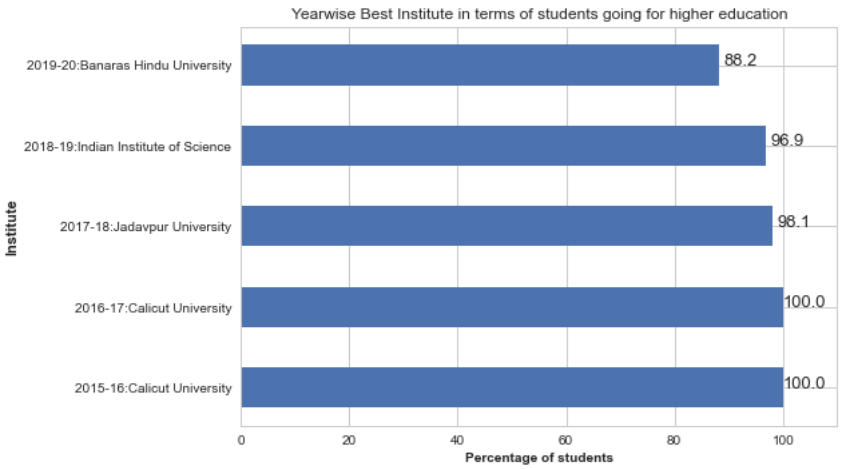


Figure 15: Institute having highest percentage of students going for higher studies

Academic Year	Institute	% Students going for higher studies
2015-16	Alagappa University	0.0
2016-17	Bharath Institute of Higher Education & Research	0.0
2017-18	Bharath Institute of Higher Education & Research	0.0
2018-19	Bharath Institute of Higher Education & Research	0.0
2019-20	Cochin University of Science and Technology	0.0

Figure 16: Institute having lowest percentage of students going for higher studies

6.10 Percentage of students unplaced

In the module, the best and the worst institute in each year in terms of percentage of students unplaced is calculated. To do this, we have used placement statistics for academic year 2015-16 to 2019-20. For each year, we have calculated the institute which has highest and the lowest percentage of students unplaced and the actual percentage of students unplaced. This result is stored in file named **unplaced.csv**. This csv file could be used in future to see the performance of any institute over the years. Here we show graphs showing the best institute in each year Figure 17 and the worst institute in each year Figure 18.

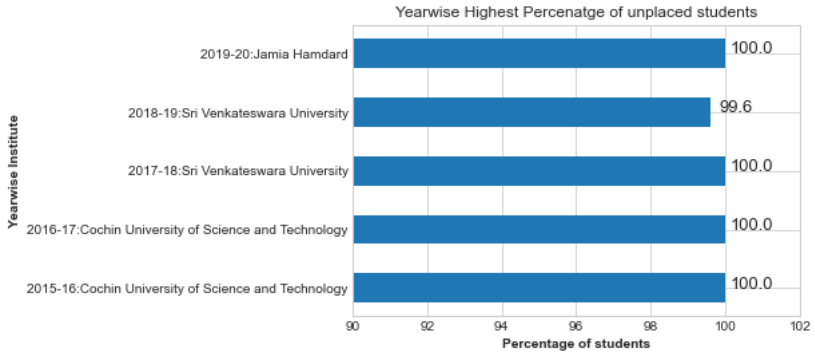


Figure 17: Institute having highest percentage of unplaced students

Academic Year	Institute	Percentage of unplaced
2015-16	Alagappa University	0.0
2016-17	Bharath Institute of Higher Education & Research	0.0
2017-18	Bharath Institute of Higher Education & Research	0.0
2018-19	Andhra University	0.0
2019-20	Bharath Institute of Higher Education & Research	0.0

Figure 18: Institute having lowest percentage of unplaced students

6.11 Analysis of colleges based on the number of migrants

The elite colleges of India attract a lot of students from other states as well as from foreign countries. Here we are showing top 5 colleges with highest(Figure 19) as well as lowest(Figure 20) percentage of migrants to their total number of students respectively. The corresponding output file containing all the information is: **Top institues with highest number of migrants.csv**.

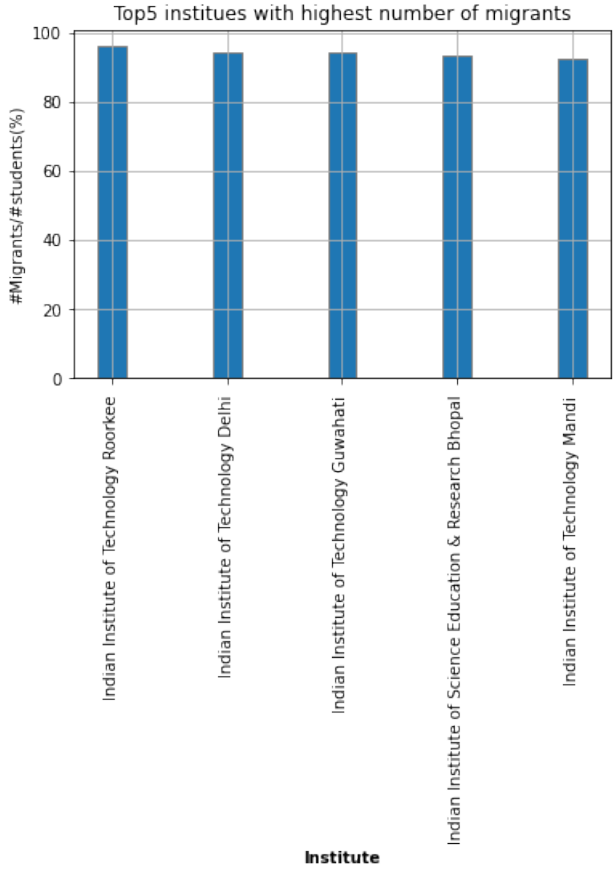


Figure 19: Top5 institutes with highest percentage of migrants-graph

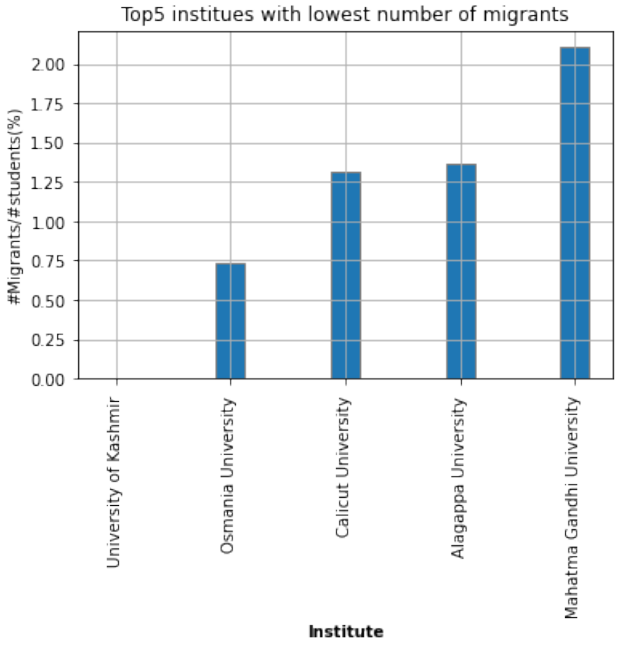


Figure 20: Top5 institutes with lowest percentage of migrants-graph

6.12 Analysis on Physically Challenged People data

In this module, the percentage of institutes having less than 80% of their buildings with lift/ramp, toilet, aided movement facilities for physically challenged people. To do this we have used the PCS Facilities data of all institutes in the ranking.The result is stored in the following tables **Lift Facilities.csv, Toilet Facilities.csv, Movement Facilities.csv**. The Pie charts below show the percentage of colleges containing these facilities in more than 80% of their buildings.

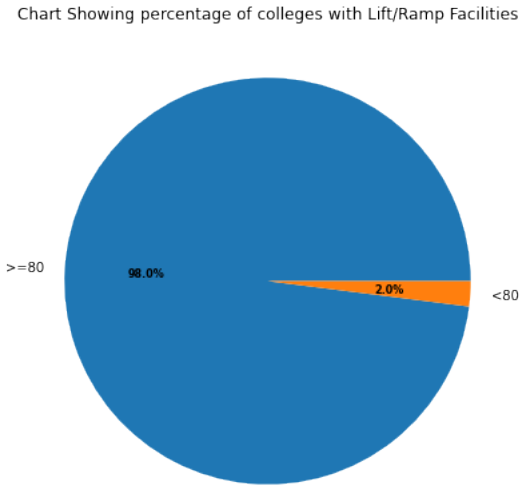


Figure 21: Percentage of Institutes with lift facilities for Physically Challenged People

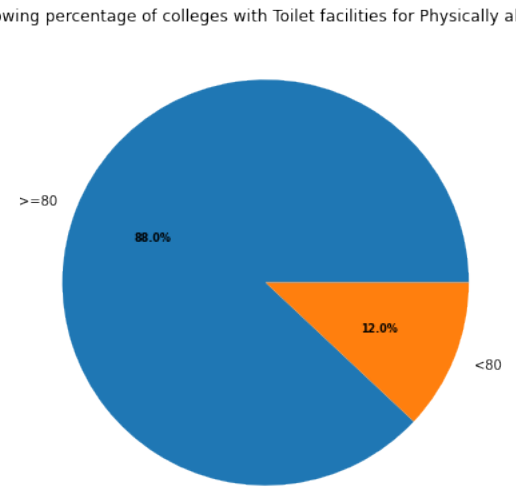


Figure 22: Percentage of Institutes with toilet facilities for Physically Challenged People

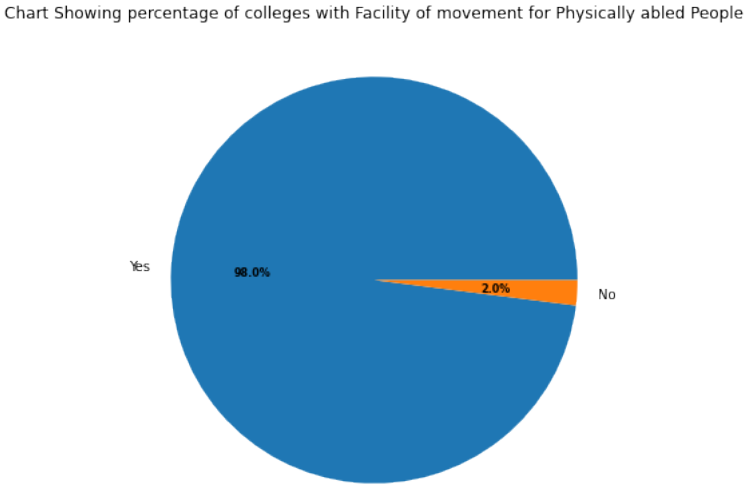


Figure 23: Percentage of Institutes with movement facilities between buildings for Physically Challenged People

6.13 Analysis of Student to Faculty Ratio

In this module, the best and the worst institutes in terms of ratio of number of student's to the number of faculty members in the institute is calculated. To do this, we have used Faculty details data. We have calculated the institute which has highest and the lowest value of this ratio .The results are stored in files named **Top 5 student to faculty ratio.csv, Worst 5 student to faculty ratio.csv**. These csv files could be used in future to see the performance of any institute over the years. Here we show graph Figure 24 showing the best and worst institutes with their student to faculty ratios.

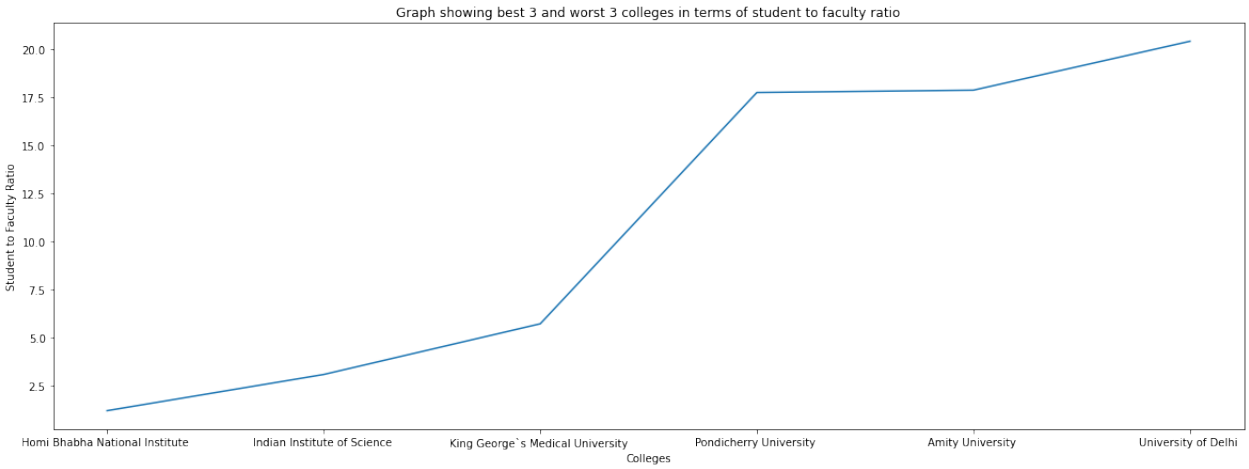


Figure 24: Bottom 3 and Top 3 Institutes in terms of student to faculty ratio

6.14 Analysis of Executive Development Program Earnings

In this module, Top 5 institutes receiving the highest amount share of Executive Development Program(EDP) Earnings for the last 3 years is calculated. The results are stored in files named **2017-18 EDP Earnings.csv,2018-19 EDP Earnings.csv,2019-20 EDP Earnings.csv, Total EDP Earnings.csv**. These csv files could be used in future to see the performance of any institute over the years. Here are the Pie charts Figure 25, Figure 26, Figure 27, Figure 28 showing the percent share of EDP earnings of top5 and all other institutes for the last 3 years {2018, 2019, 2020} and in total of all years respectively.

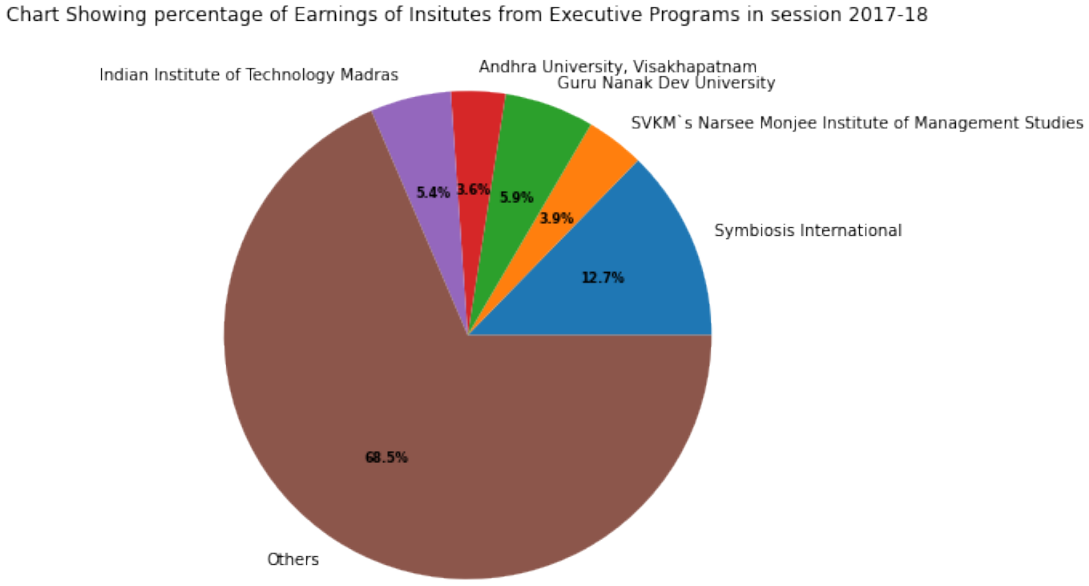


Figure 25: Percentage of EDP Earnings of Top 5 Institutes in 2017-18

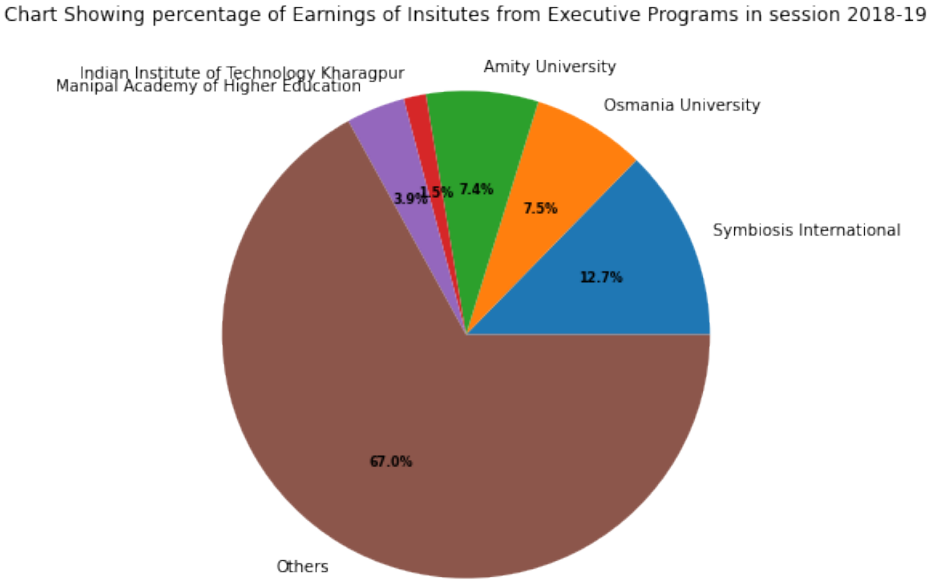


Figure 26: Percentage of EDP Earnings of Top 5 Institutes in 2018-19

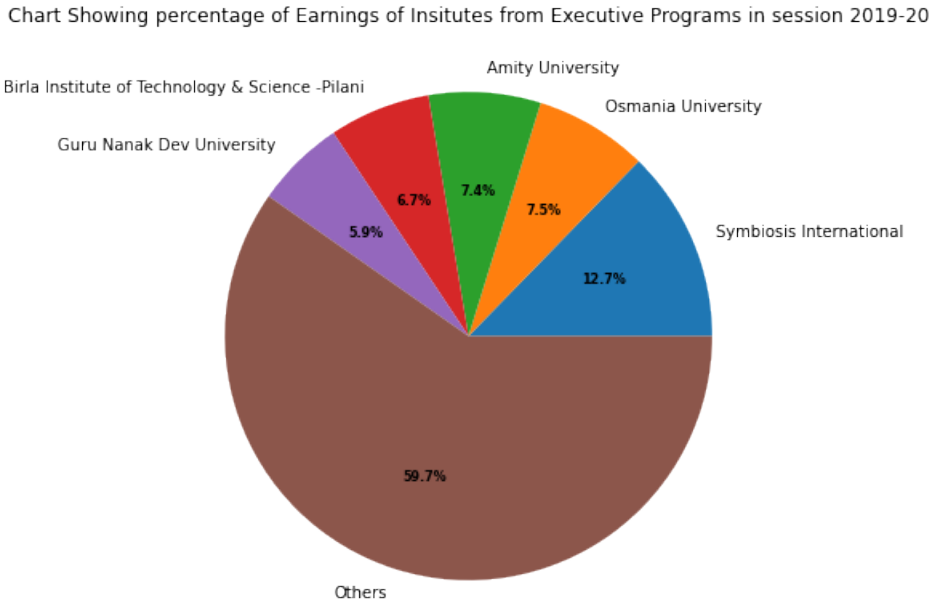


Figure 27: Percentage of EDP Earnings of Top 5 Institutes in 2019-20

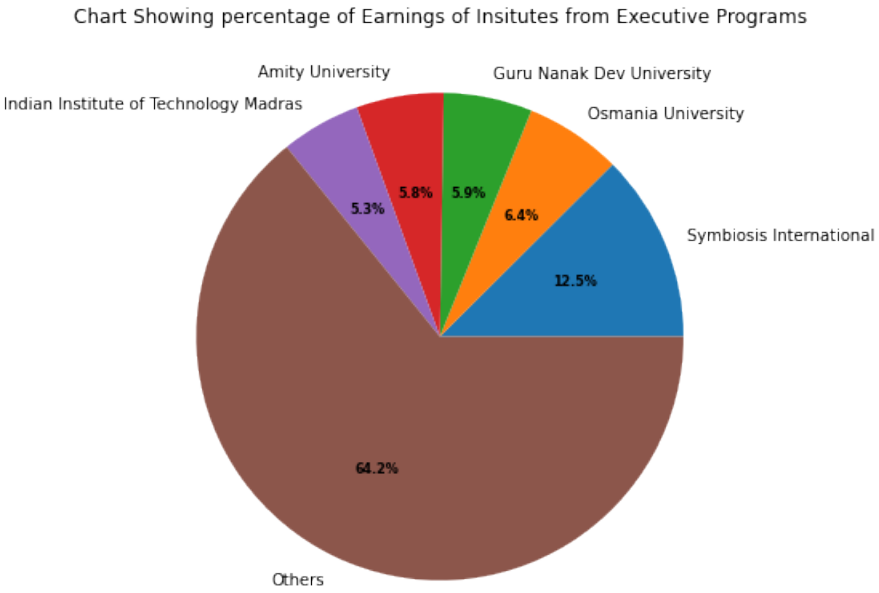
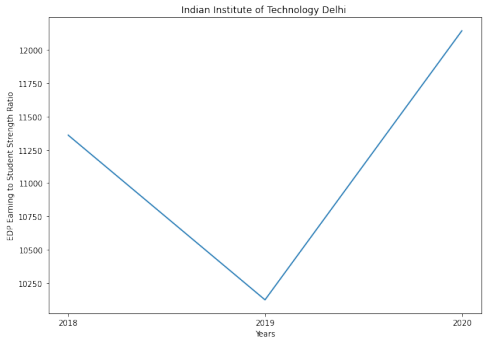
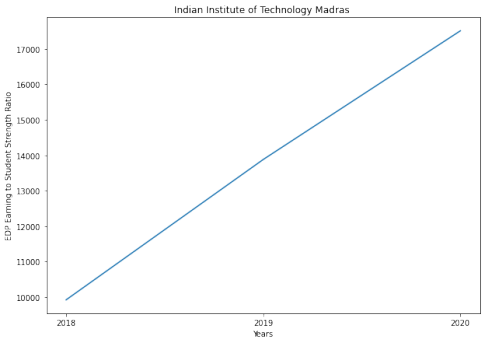


Figure 28: Percentage of EDP Earnings of Top 5 Institutes in Total

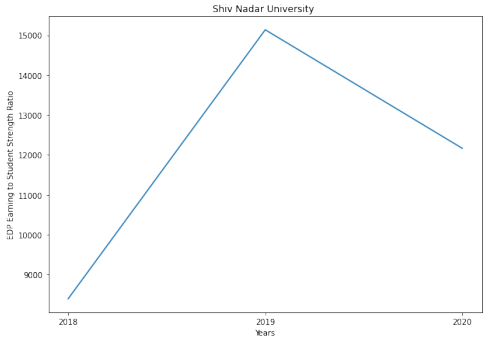
In this module, we have also found out the best institutes in terms of highest ratio of amount received from Executive Development Program(EDP) to total actual student strength of all institutes for the last 3 years and in total of last 3 years.The results are stored in files named **EDP to Strength ratio 2018.csv,EDP to Strength ratio 2019.csv,EDP to Strength ratio 2020.csv,EDP to Strength ratio Total.csv**. Here are the graphs showing the yearly ratio of EDP Earnings to actual student strength of the top 3 institutes of last 3 years. The tables (below graphs) show that some institutes have constantly been on the top for last years in terms of EDP Earnings to Student Ratio.



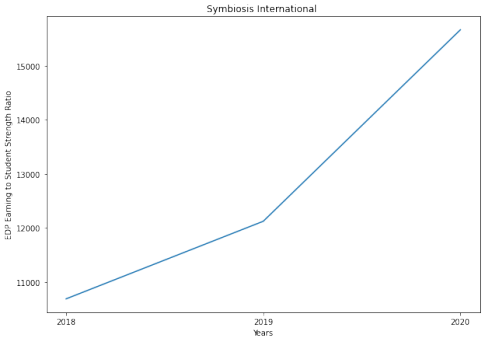
(a) Indian Institute of Technology Delhi



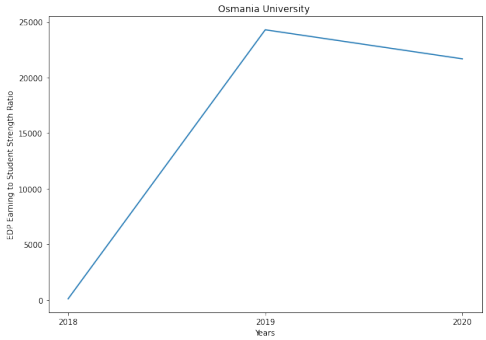
(b) Indian Institute of Technology Madras



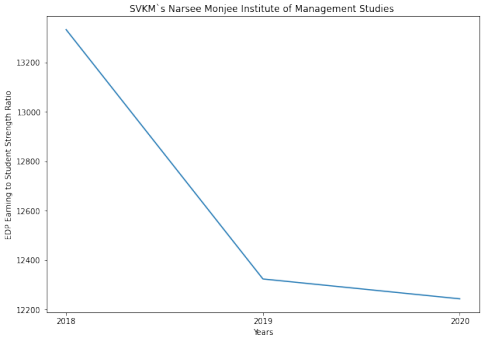
(a) Shiv Nadar University



(b) Symbiosis International



(c) Osmania University



(d) SVKM's Narsee Monjee Institute of Management Studies

Institute	2018_Ratio
SVKM`s Narsee Monjee Institute of Management Studies	13333.53037
Indian Institute of Technology Delhi	11359.87738
Symbiosis International	10690.36756
Indian Institute of Technology Madras	9924.207103
Guru Nanak Dev University	9770.363968

(e) Top 5 Institutes 2018

Institute	2019_Ratio
Osmania University	24296.27465
Shiv Nadar University	15135.49098
Indian Institute of Technology Madras	13880.50015
SVKM`s Narsee Monjee Institute of Management	12324.06851
Symbiosis International	12124.64314

(f) Top 5 Institutes 2019

Institute	2020_Ratio
Osmania University	21680.93645
Indian Institute of Technology Madras	17507.99713
Symbiosis International	15665.78139
Indian Institute of Technology Hyderabad	13451.02506
Guru Nanak Dev University	13138.05248

(g) Top 5 Institutes 2020

Institute	Total_Ratio
Osmania University	14930.24417
Indian Institute of Technology Madras	13824.88842
Symbiosis International	12943.72565
SVKM`s Narsee Monjee Institute of Management Studies	12634.51087
Shiv Nadar University	11790.92126

(h) Top 5 Institutes in Total of past 3 years

Figure 30: Yearly EDP Earnings to Student Ratio of top Institutes of last 3 years

6.15 Analysis of reimbursement provided by various institutes

Here we are trying to analyse which are the colleges that provide fees waivers to maximum(Figure 31) as well as minimum(Figure 32) number of students. These waivers include funds provided by **private bodies, state and central governments** and from **institute funds** as well. The corresponding output file containing the complete set of information is **Top programs and colleges which provide reimbursement to maximum number of students.csv**.

Institute	percent reimbursement
Osmania University	81.4716
Jamia Millia Islamia	79.775
Calcutta University	70.31
Indian Institute of Technology (Banaras Hindu University) Varanasi	68.8063
Calicut University	67.9625

Figure 31: Top5 institutes which provide reimbursement to highest number of students

Institute	percent reimbursement
Jawaharlal Nehru University	0
Dr. D. Y. Patil Vidyapeeth	0.203528
Birla Institute of Technology & Science - Pilani	0.387214
Vellore Institute of Technology	0.654759
SVKM`s Narsee Monjee Institute of Management Studies	0.704122

Figure 32: Top5 institutes which provide reimbursement to lowest number of students

6.16 Analysis of economical and socially backward people admitted in the top institutes

Here, we have calculated the proportion of entire student population, who either belong to socially or economically backward classes. This analysis can further be used if someone wants to get involved in the welfare of such people. The plots below give details of the institutes with highest as well as lowest ratios of economically and socially backward people. The corresponding output file containing the complete set of information are:

- **Top institutes with highest economically backward-total ratio.csv**
- **Top institutes with highest socially backward-total ratio.csv**

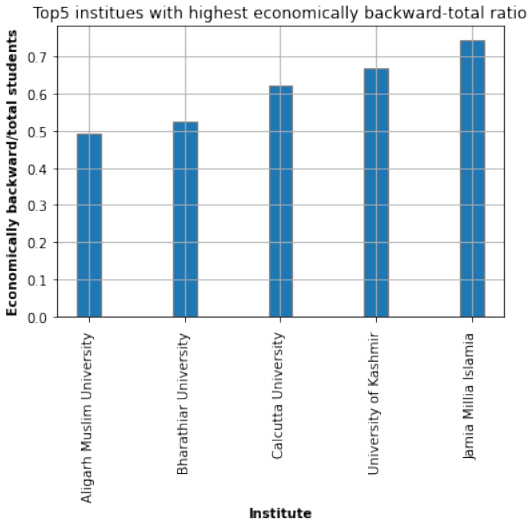


Figure 33: Top5 institutes with highest economically backward-total ratio

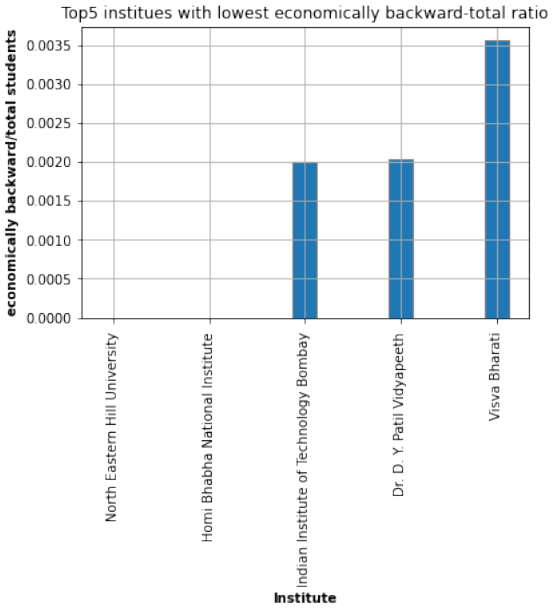


Figure 34: Top5 institutes with lowest economically backward-total ratio

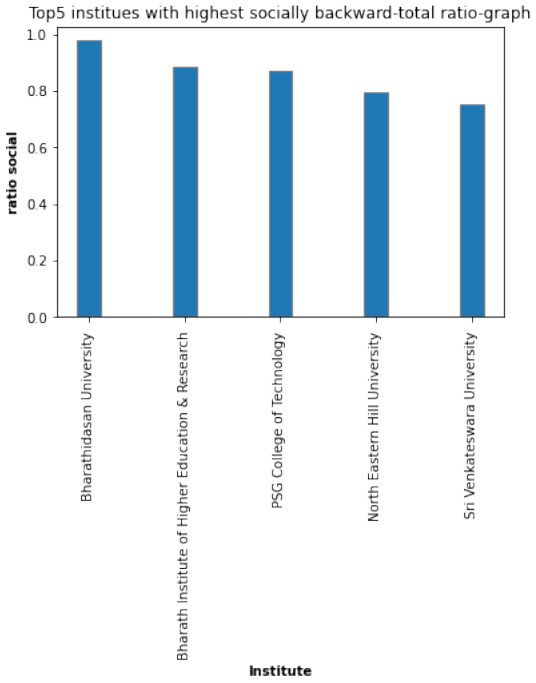


Figure 35: Top5 institutes with highest socially backward-total ratio

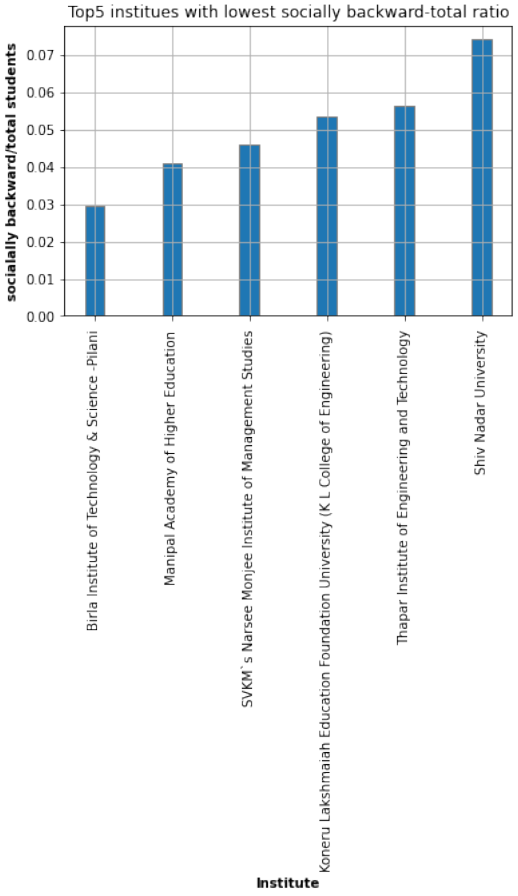


Figure 36: Top5 institutes with lowest socially backward-total ratio

6.17 Analysis of funds received through consultancy projects

The institutes receives some funds from client organizations for consultancy projects. In this section, we perform an analysis about the distribution of such funds among the top 85 colleges over the past three years. In order to get some idea about the distribution, we consider the summation of the funds received by the 85 colleges as the total funding received and then calculate the percentage of total funds received by each of the college.

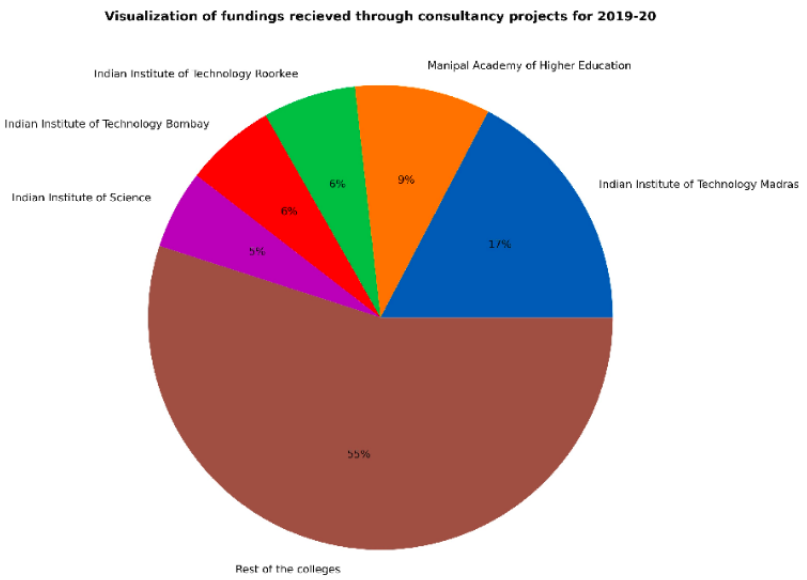


Figure 37: Percentage of total funds received by the top 5 colleges for the session 2019-20. It is evident from the above chart that 41% of the total funds is received by the Top 5 colleges.

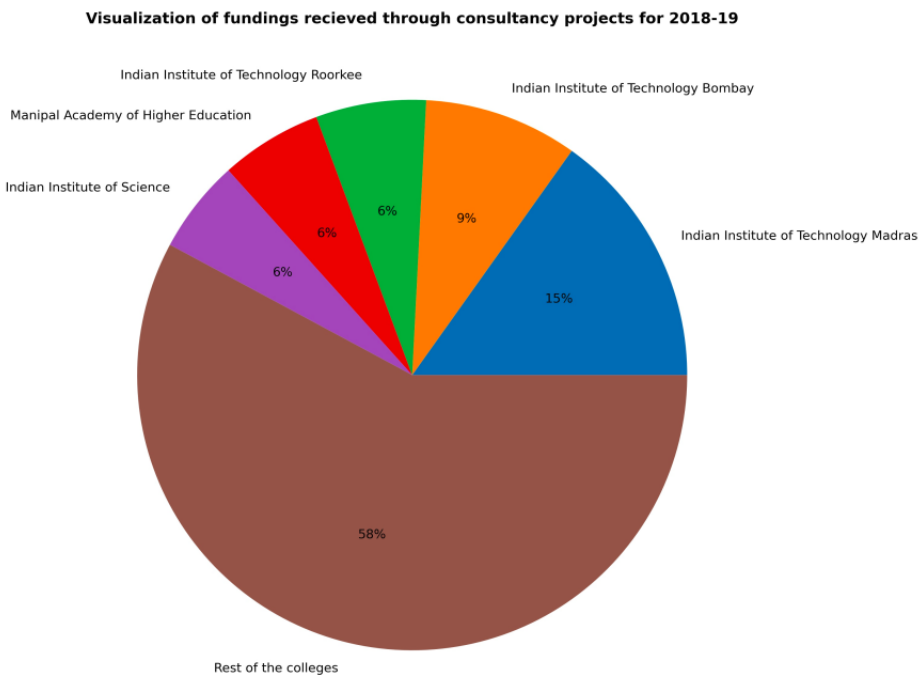


Figure 38: Percentage of total funds received by the top 5 colleges for the session 2018-19. It is evident from the above chart that 42% of the total funds is received by the Top 5 colleges.

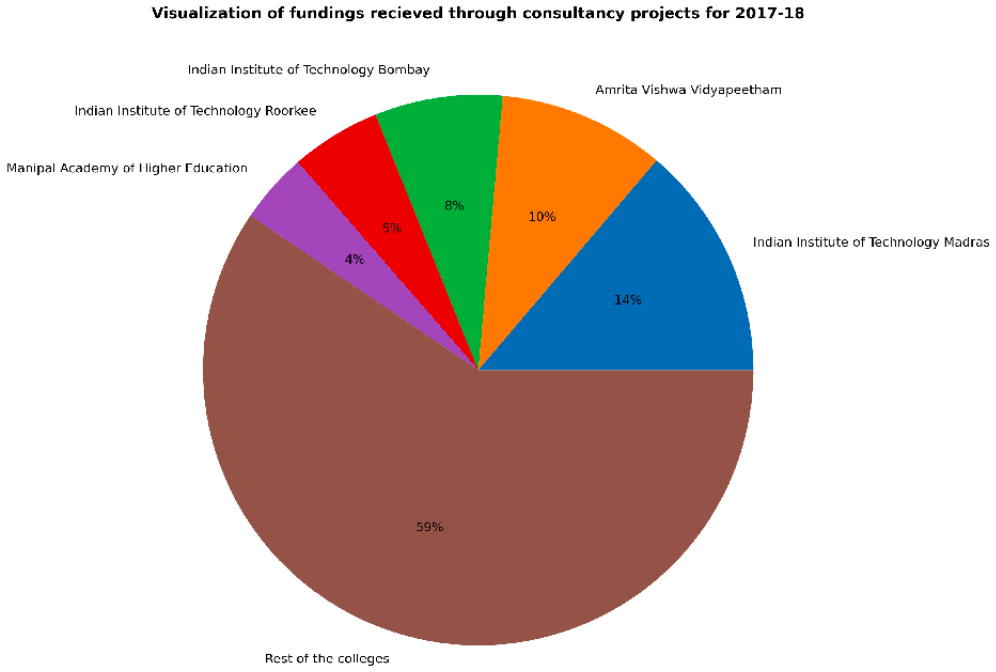


Figure 39: Percentage of total funds received by the top 5 colleges for the session 2017-18. It is evident from the above chart that 41% of the total funds is received by the Top 5 colleges.

6.18 Analysis of number of consultancy projects in the institutes over the last three years

The number of consultancy projects received by an institute is different for each year. We found the total number of consultancy projects for each of the institute over the past three years and analysed the trends of the Top 5 institutes.

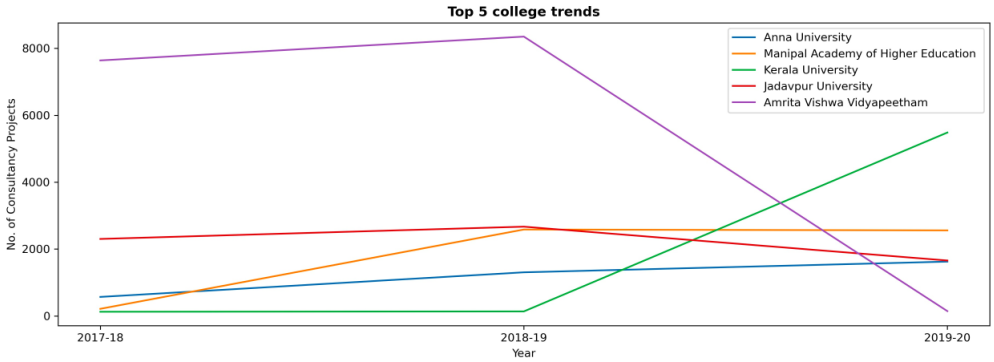


Figure 40: Trends in number of consultancy projects received by the top 5 institutes.

For the session 2019-20, Amrita Vishwa Vidyapeetham, had a steep decline in number of consultancy projects whereas Kerala University has shown a significant increase in the same. Jadavpur University has been consistent with the number of consultancy projects over the the last three 3 years.
Institutes with most number of consultancy projects for the last three sessions is available in greater detail in *consultancy_projects_top5.csv*

6.19 Analysis of number of client projects received and number of client organisations

We have analysed the number of projects a college gets on average from a client organisation by calculating the ratio of number of client projects to number of client organisations for three years. Surprisingly, the ratio takes values greater than 50 to even less than 1 for some of the institutes. The detailed ratios for every institute for three years is available in *consultancy-project-to-client-ratio.csv*

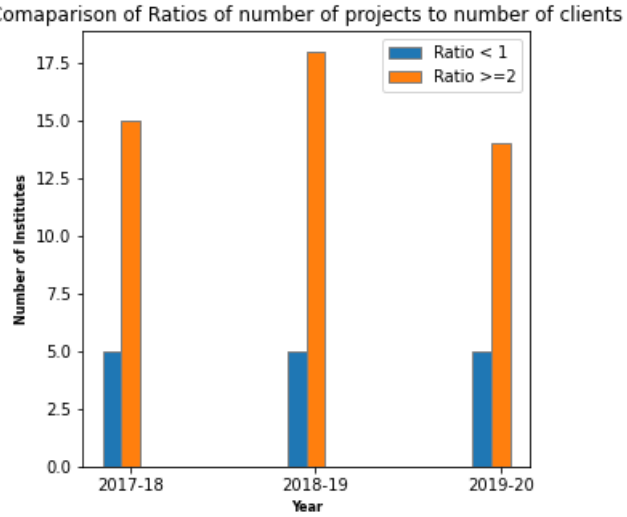


Figure 41: Comparison of number of institutes having the ratio between number of consultancy projects and number of client organisations greater than equal to 2 and less than 1

The year for which this ratio is maximum for each of the institute is available in *Maximum-project-to-client-ratio-year.csv*

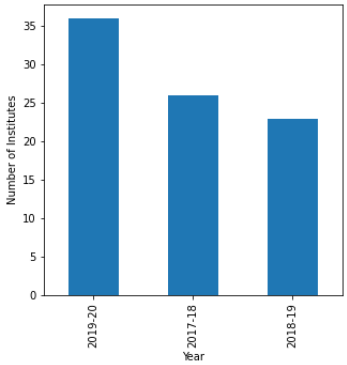


Figure 42: The above bar graph shows count of institutes which had the maximum number of projects to number of client ratio for a particular year

36 out of the top 86 institutes had the highest number of consultancy projects to number of client organisations ratio in the year 2019-20.

6.20 Analysis of sponsored research projects and consultancy projects for the Top 86 institutes

There are two types of projects in a institute, sponsored research projects and consultancy projects from clients. We have analysed the percentage of these two types of projects for all institutes for three years. The percentages for each institute is available in *Sponsored Research projects and consultancy projects.csv*

	A	B	C	D	E	F	G
	Institute	Reasearch Project %(2017-18)	Consultancy Project%(2017-18)	Reasearch Project %(2018-19)	Consultancy Project%(2018-19)	Reasearch Project %(2019-20)	Consultancy Project%(2019-20)
	Alagappa University	88.37	11.63	86.57	13.43	94.74	5.26
	Aligarh Muslim University	72.00	28.00	70.97	29.03	53.29	46.71
	Amity University	95.36	4.64	92.86	7.14	94.90	5.10
	Amrita Vishwa Vidyapeetham	1.77	98.23	1.70	98.30	55.06	44.94
	Andhra University, Visakhapatnam	30.00	70.00	21.58	78.42	21.54	78.46
	Anna University	21.25	78.75	7.65	92.35	5.31	94.69
	Banaras Hindu University	99.48	0.52	99.47	0.53	99.47	0.53

Figure 43: A snapshot from *Sponsored Research projects and consultancy projects.csv*

Most of the colleges either have many sponsored research projects or have many client projects. There are only a handful of colleges which have comparable numbers of sponsored research projects and client projects.

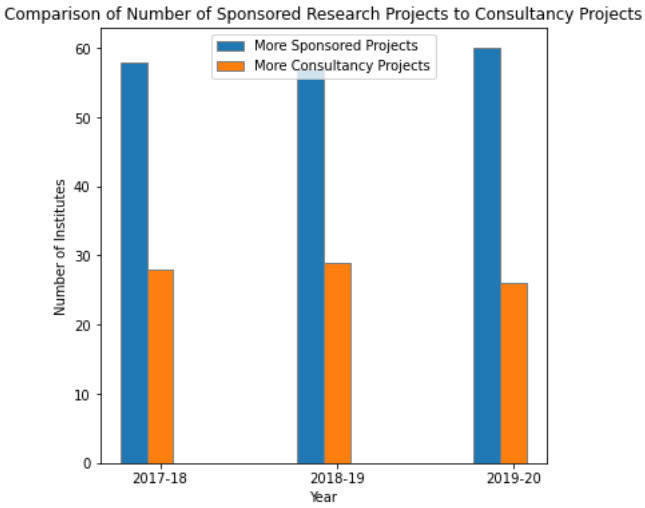


Figure 44: Comparison of number of research projects and consultancy projects for three years

Around 56 out of the top 86 institutes, have greater number of sponsored research projects when compared to number of consultancy projects.

6.21 Analysis of programs in various colleges which either overfill or have vacant seats

Every program in every college is assigned a fixed number of seats which they can accommodate. They either overfill, or a few seats are left vacant. We have shown those 5 programs and their corresponding institutes which overfill(Figure 45) and those which have vacant seats(Figure 46). The corresponding output file containing the complete set of information are:

- **Top institues with highest economically backward-total ratio.csv**
- **Top institutes with highest number of vacancies.csv**

Institute	Programs	Overfill-percent
Indian Institute of Technology Ropar	PG [3 Years Program(s)]	inf
National Institute of Technology Durgapur	PG [3 Years Program(s)]	inf
Indian Institute of Engineering Science and Technology	PG-Integrated	inf
Sri Venkateswara University	UG [3 Years Program(s)]	inf
Tezpur University	UG [3 Years Program(s)]	inf

Figure 45: Top5 programs with highest overfill percentage

Institute	Programs	Vacancies-Percent
Tezpur University	PG [1 Years Program(s)]	58.7302
Banasthali Vidyapith	PG [1 Years Program(s)]	53.3333
Visva Bharati	PG [1 Years Program(s)]	46.0526
Aligarh Muslim University	PG [1 Years Program(s)]	43.3333
Indian Institute of Technology Indore	PG [2 Years Program(s)]	34.0426

Figure 46: Top5 institutes with highest percentage of vacancies

6.22 Analysis of sex-ratio for the top institutes

Here we are trying to analyse the sex ratio of various colleges. We have shown colleges having highest(Figure 47) as well as lowest(Figure 48) **females/males** ratio. The corresponding output file containing the complete set of information is **Top institutes with highest female-male ratio.csv**.

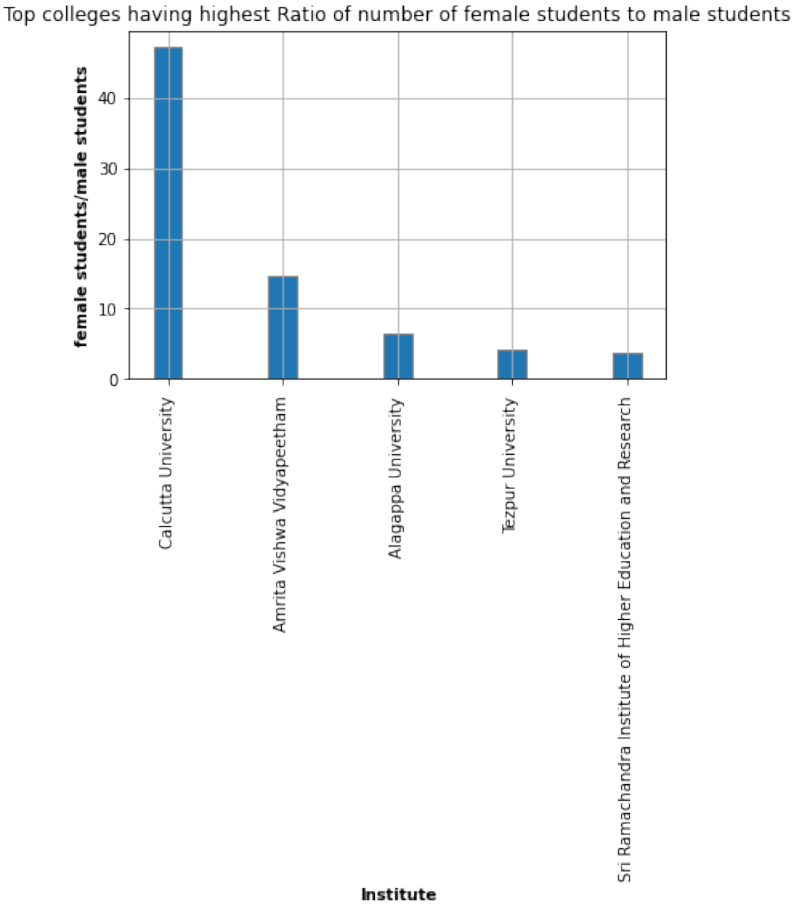


Figure 47: Top5 institutes with highest female-male ratio

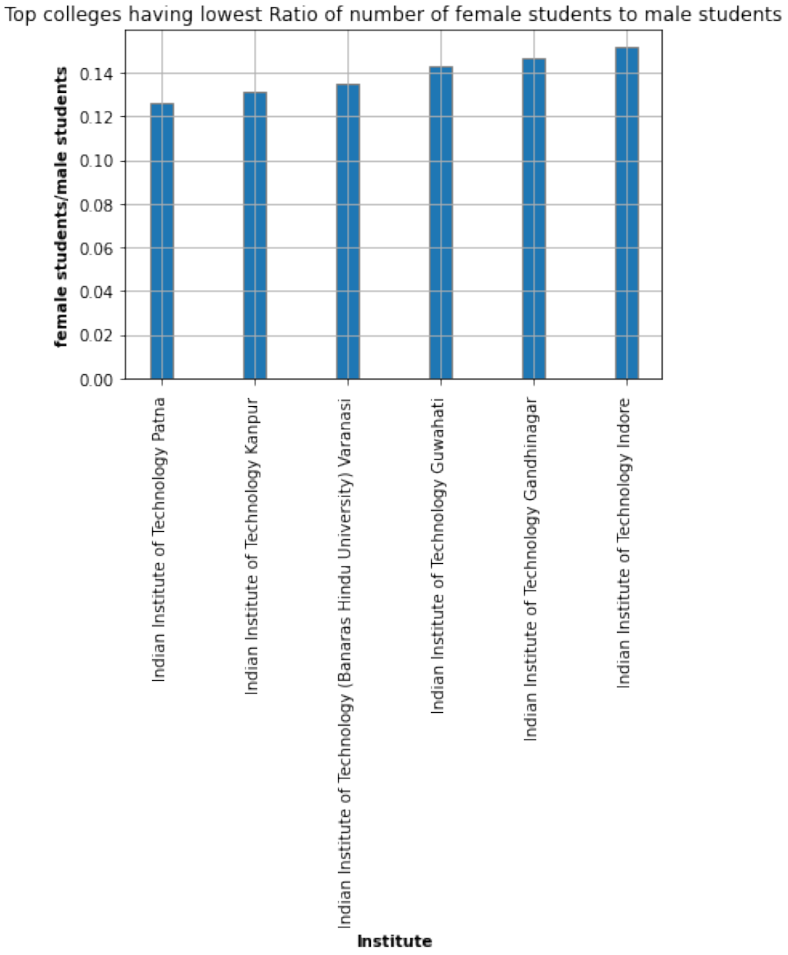


Figure 48: Top5 institutes with lowest female-male ratio

7 Conclusion and Future Work

In this project we have analysed the data of the institutes present in the NIRF rankings of past 3 years to draw interesting patterns and useful insights into how the institutes actually performed on various aspects of data submitted by the institutes to the NIRF. All the insights that we were able to draw from the data is presented in this report in the form of various visualizations like piechart's, line graphs, etc. The name's of the resulting csv's are also shared in every section of the analysis. These results can be used further for a lot of future work like to do further research for finding the reasons why some institutes are doing well consistently for the past three years and why others are not or to suggest actions that institutes can take on the areas where they are lacking behind.

8 Important Links and References

1. NIRF 2021 Data: https://www.nirfindia.org/2021/OverallRanking.html
2. NIRF 2020 Data: https://www.nirfindia.org/2020/OverallRanking.html
3. NIRF 2019 Data: https://www.nirfindia.org/2019/OverallRanking.html

4. Github Project Link: https://github.com/pranshu27/NIRF-Data-Analysis