```NIRF'21 ENGINEERING RANKING ANALYSIS```

NAMAN KUMAR

5TH JULY, 2020

ABHYAZ MTAB DATA ANALYST INTERN CANDIDATE

OUTLINE

- Executive Summary
- Introduction
- Methodology
  - Approach to Data
  - Analytical process
  - Tools used
- Results
  - Visualization
- Discussion
  - Innovative Insights
- Conclusion

EXECUTIVE SUMMARY

Summary of Tools & Methodologies:

- Data Collection
  - Using Web Scrapping
- Data Wrangling & Pre processing
  - numpy & pandas
- Data Analysis
  - Using Data Visualization
    - Plotly & Seaborn

Summary of Results:

- Exploratory Data Analysis
- Static Analytics

INTRODUCTION

PROBLEM STATEMENT:

Data Science: https://www.nirfindia.org/2021/EngineeringRanking.html
given is the link to NIRF ranked institutes in 2021.

Pick up the ranking of top 5 institutes, the related data for these institutes is available.

Perform an analysis of these 5 institutes.

Data points of interest are:

- Intake / Graduating students
- Placements / Higher studies
- Funding received
- Spending pattern

This isn't restricted to only this analysis - you can also provided further analysis.
The data should be represented visually with relevant references.

Here, I assumed the role of a Data Analyst to provide relevant insights to the given problem statement.

METHODOLOGY

We started with a General analysis of the given Institutes

- Using get requests we created a response object from the given HTML link

```
response = requests.get(static_url)
```

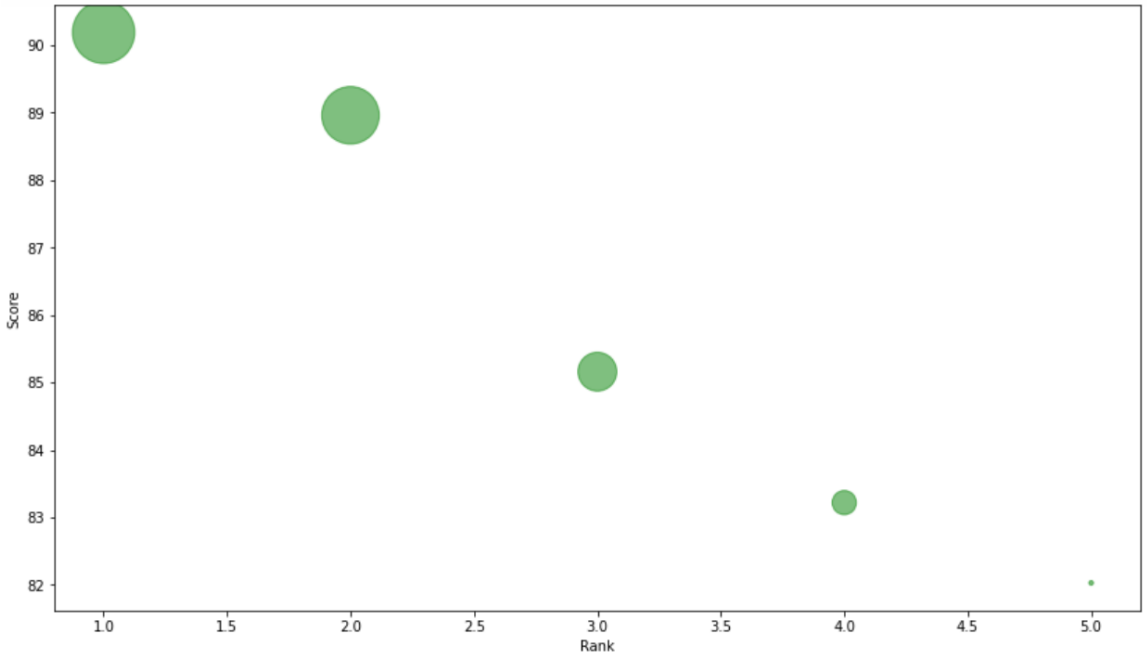
- Then we created a traversable HTML tree, soup object, using Beautiful Soup constructer

```
soup = BeautifulSoup(response.content, "html.parser")
```

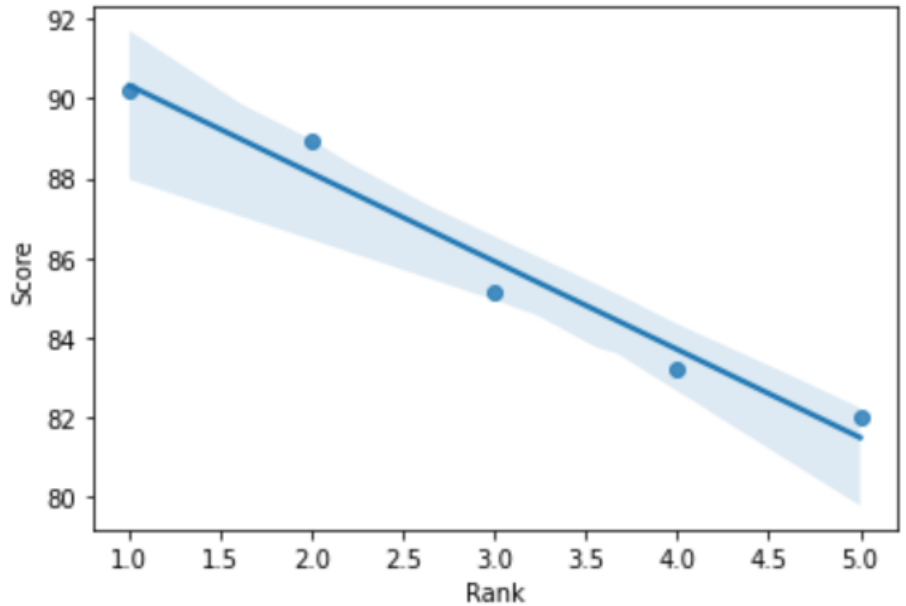
- We then created top 5 institute data frame.

	Institute ID	Name	City	State	Score	Rank
0	IR-E-U-0456	Indian Institute of Technology MadrasMore Deta...	Chennai	Tamil Nadu	90.19	1
1	IR-E-I-1074	Indian Institute of Technology DelhiMore Detai...	New Delhi	Delhi	88.96	2
2	IR-E-U-0306	Indian Institute of Technology BombayMore Deta...	Mumbai	Maharashtra	85.16	3
3	IR-E-I-1075	Indian Institute of Technology KanpurMore Deta...	Kanpur	Uttar Pradesh	83.22	4
4	IR-E-U-0573	Indian Institute of Technology KharagpurMore D...	Kharagpur	West Bengal	82.03	5

- We put a scatter plot of Score vs Rank



- And then tried to train a linear regression model to predict what score would a certain rank holder have



Now, moving onto individual Institute data

1. **EDA on `ins_id[0]` i.e. IIT MADRAS**

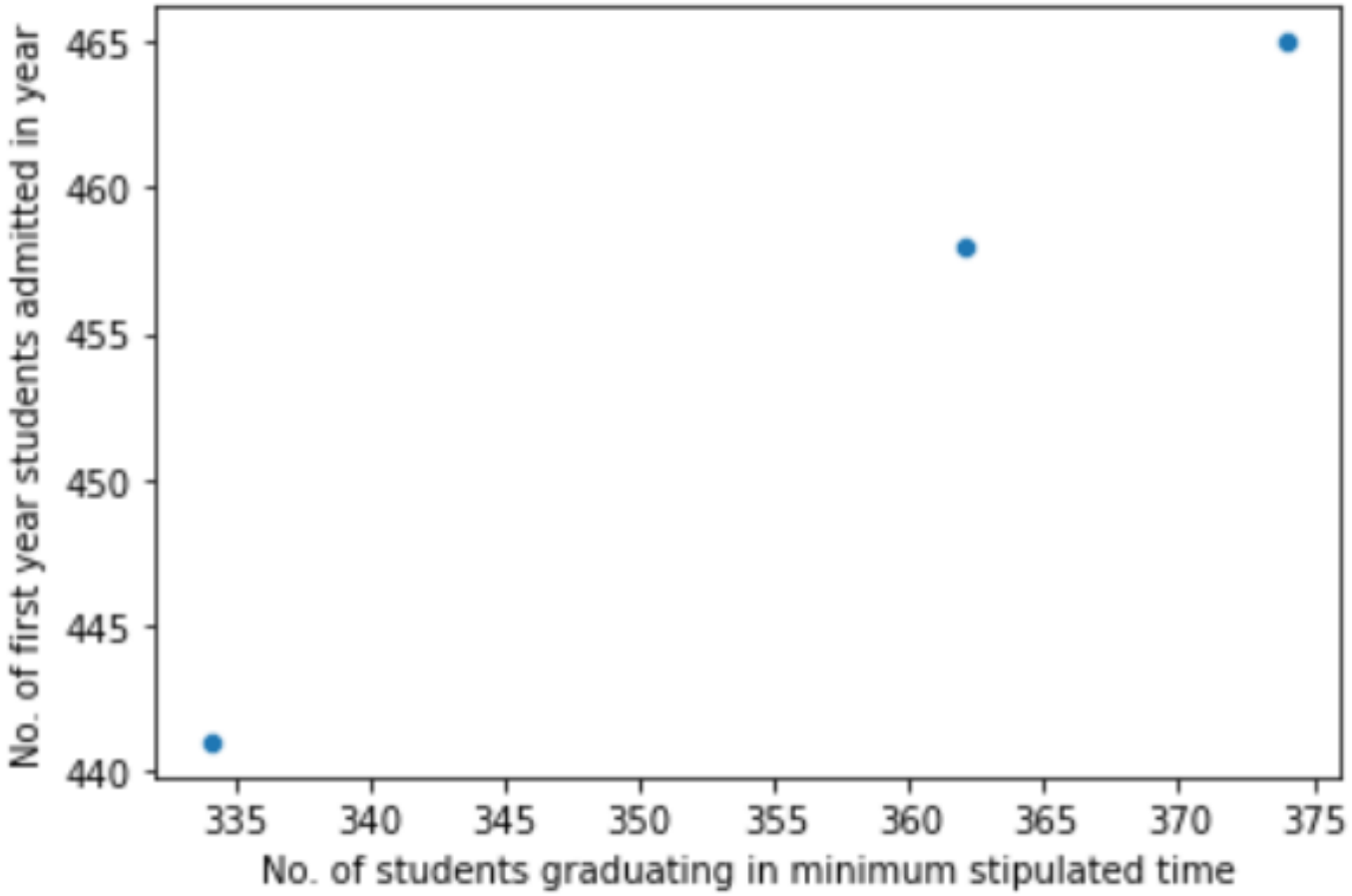
- We first used tabula-py to read remote PDFs as a list of tables, which gets converted to pandas dataframe automatically.

	Academic Year	No. of first year\rstudents intake in the\ryear	No. of first year\rstudents admitted in\rthe year	Academic Year.1	No. of students\radmitted through\rLateral entry	Academic Year.2	No. of students\rgraduating in\rminimum stipulated\rtime	No. of students\rplaced	Median salary of\rplaced\rgraduates(Amount in\rRs.)
0	2014-15	466	441	2015-16	0	2017-18	334	283	1306000(Thirteen\rlakhs six thousand)
1	2015-16	466	458	2016-17	0	2018-19	362	268	1360000(Thirteen\rLakhs Sixty thousand)
2	2016-17	466	465	2017-18	0	2019-20	374	286	1500000(Fifteen lakhs)

- Data Cleaning(using pandas)
  - Dropping irrelevant columns & rows
  - Changing column names to make them more comprehensible
  - Set & reset indices
  - Achieve final clean data ready to visualize

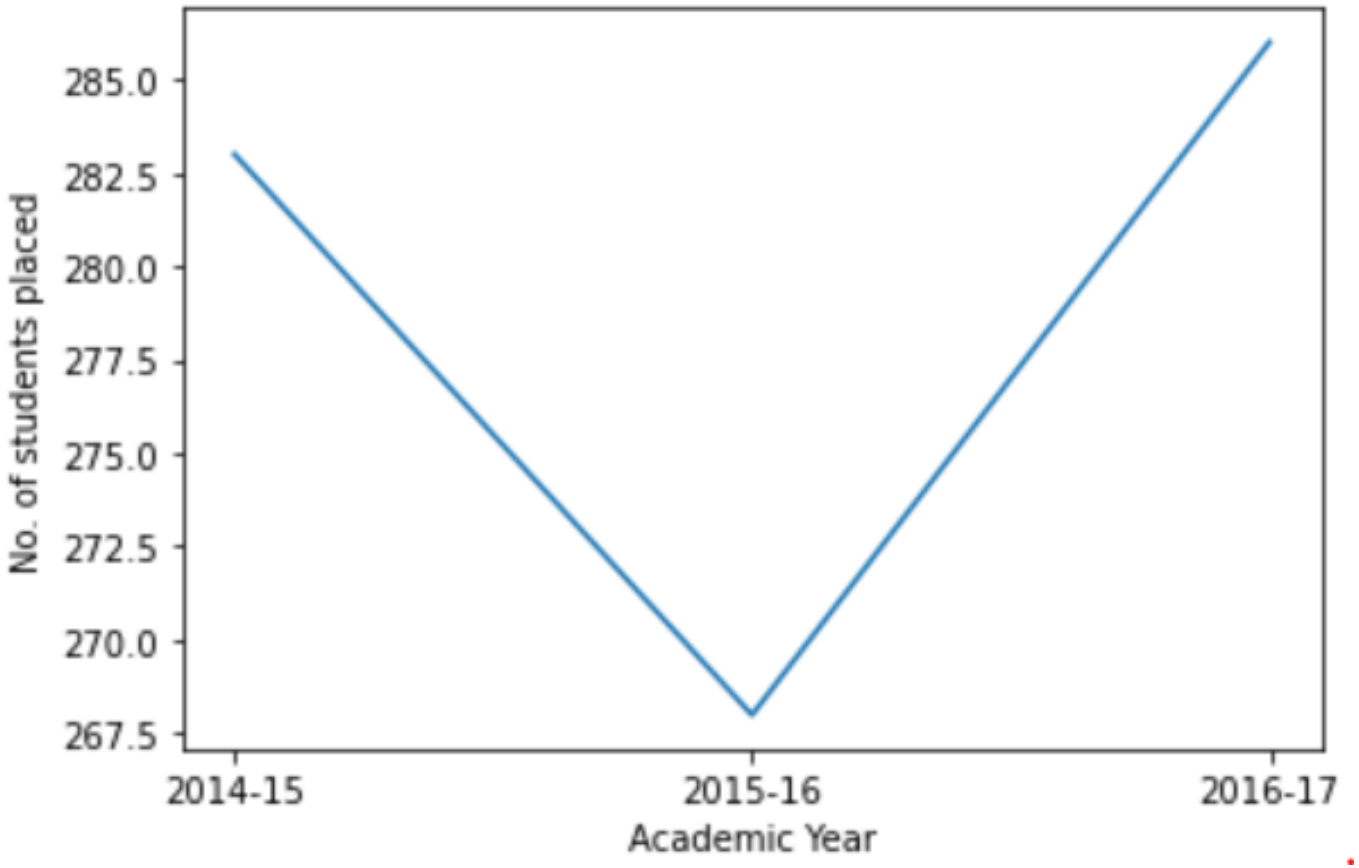
- Then we visualize & analyze this clean data using seaborn and pyplot

1. No. of first year students admitted in year VS No. of students graduating in minimum stipulated time



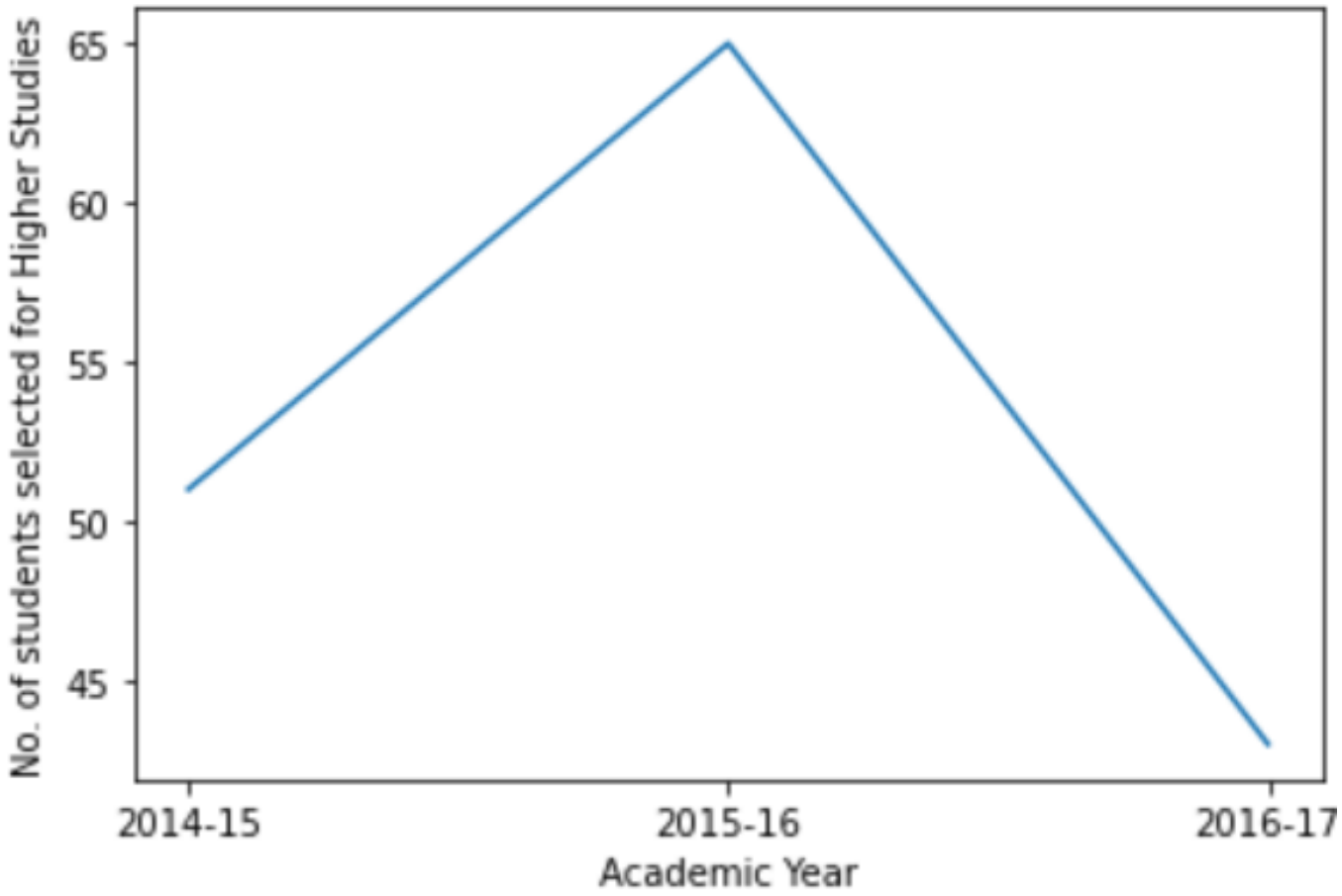
Result: We can clearly see that there is a positive collinearity between the above two attributes That is with increasing No. of first year students admitted in year, No. of students graduating in minimum stipulated time also increased, which is to be expected as well.

2. No. of students placed VS Academic Year



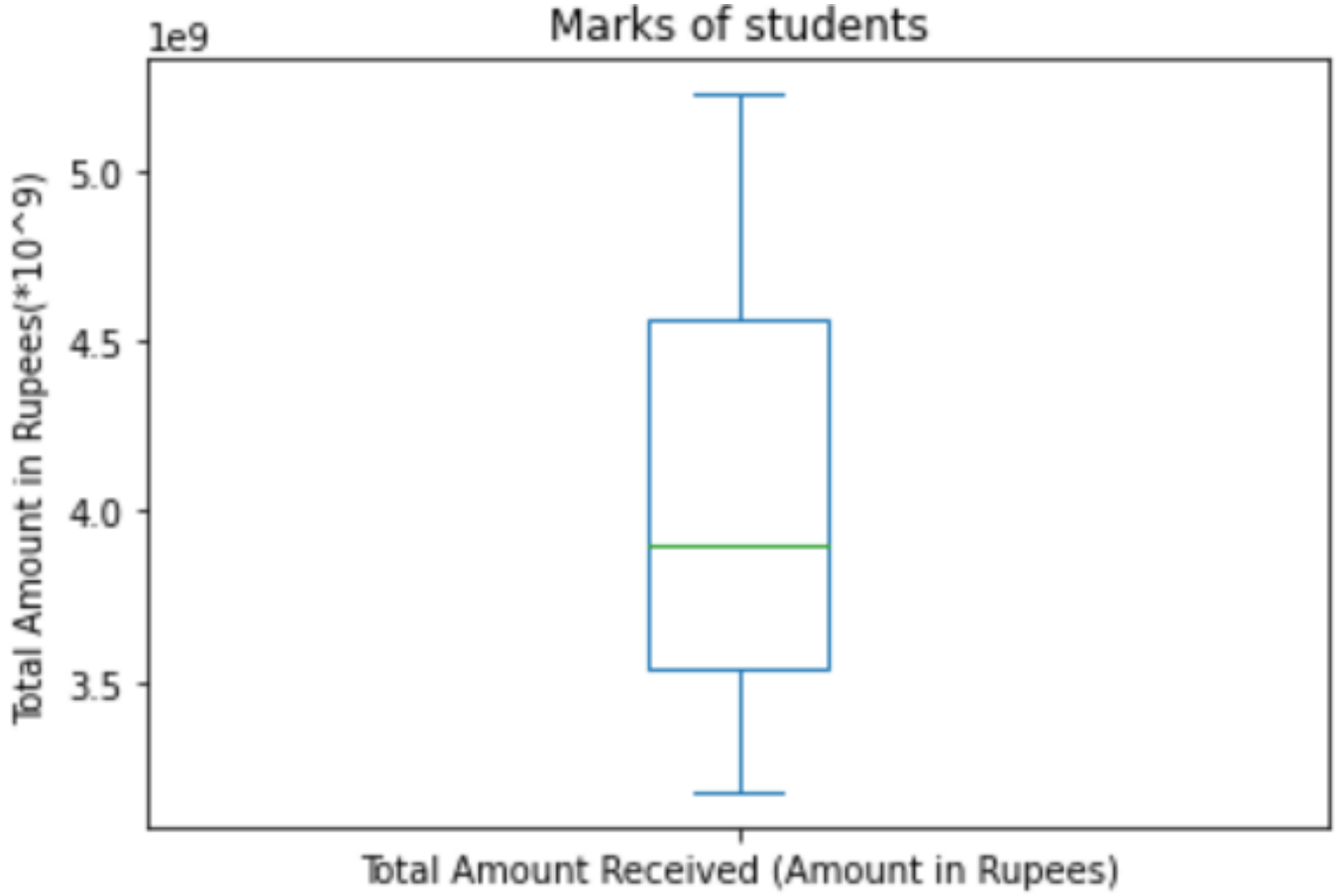
Result: As observed above, there is no linear relation between the two attributes
But what we can see is that there was dip in placements during the year 2015-16, which may be due to some underlying reason *(to which we have an answer in the below visualization).*

3. No. of students selected for Higher Studies VS Academic Year



Result: This has an inverse trend as compared to the above
we can see here that in year 2015-16 No. of students selected for Higher Studies were far more than other two academic years. Hence, this resulted in a dip in placements that year also.(answer to above analysis query)

4. Total Amount Received



Result:

- Median: 3903460211
- Maximum = 5223395988
- Minimum =3177083996

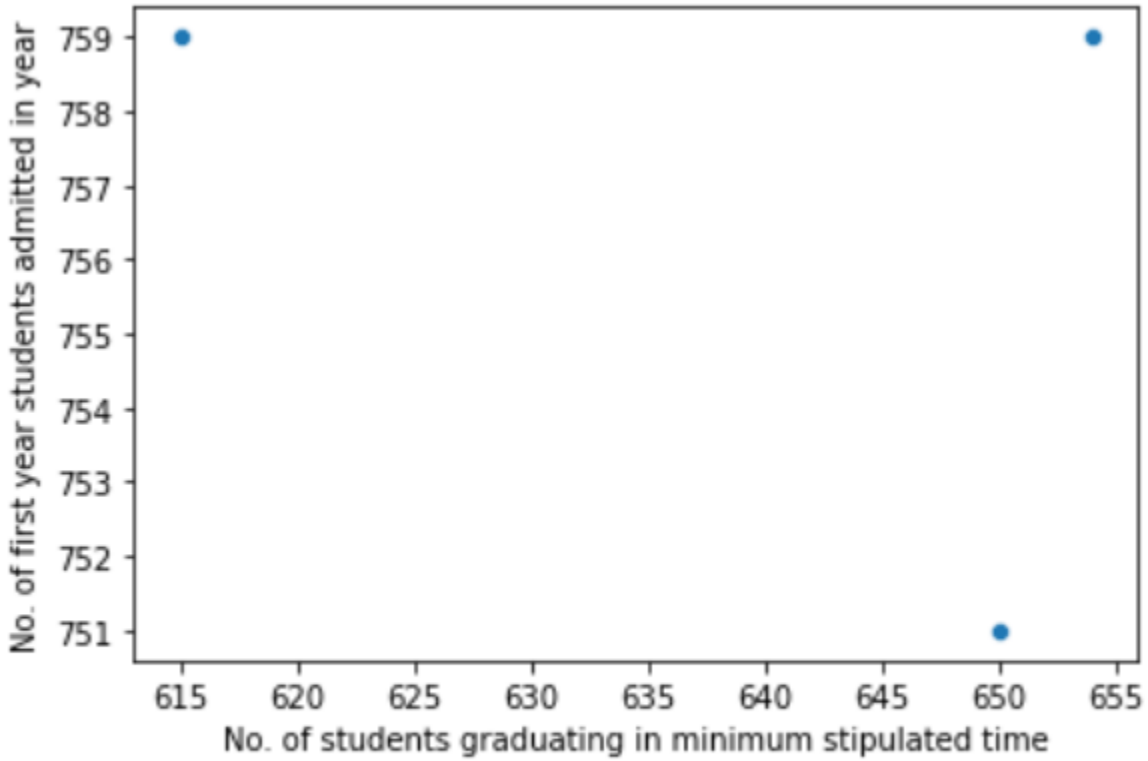
5. Then, we analyzed 'Financial Resources: Utilized Amount for the Capital expenditure for previous 3 years' table to understand spending patterns

- Transpose the Data Frame
- set columns & reset indices
- Drop irrelevant rows
- Add financial Column back
- Remove numeric data from financial column using str.replace()

Using the same methodology as above, we visually analyzed the other 4 data as well,

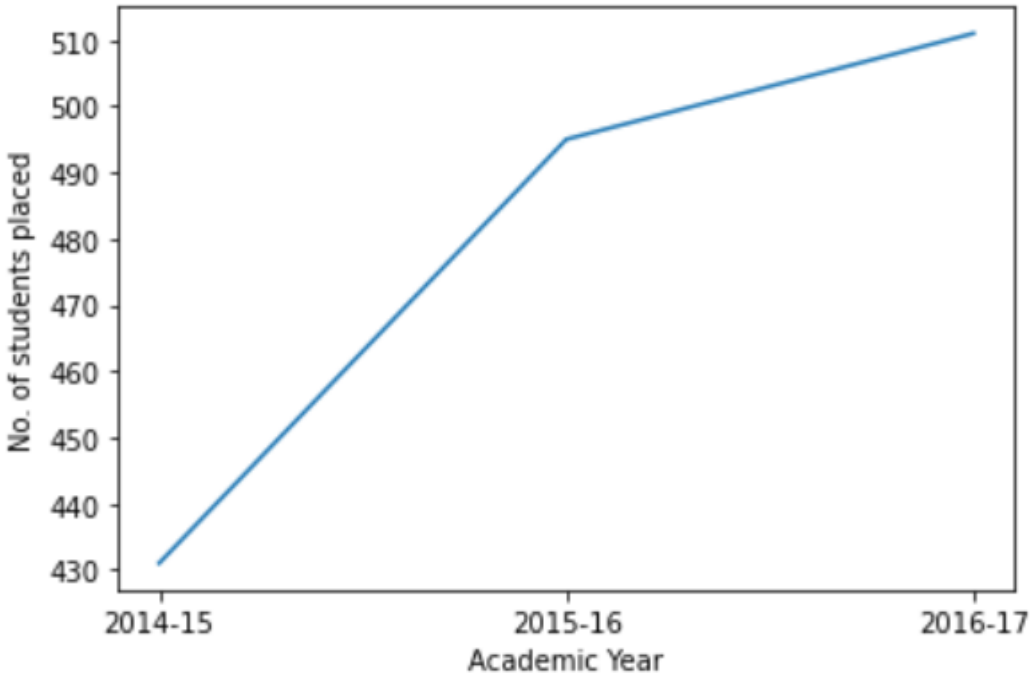
2. EDA on `ins_id[1]` i.e. IIT DELHI

- **No. of first year students admitted in year VS No. of students graduating in minimum stipulated time**



There Does not seem to be any relation among them

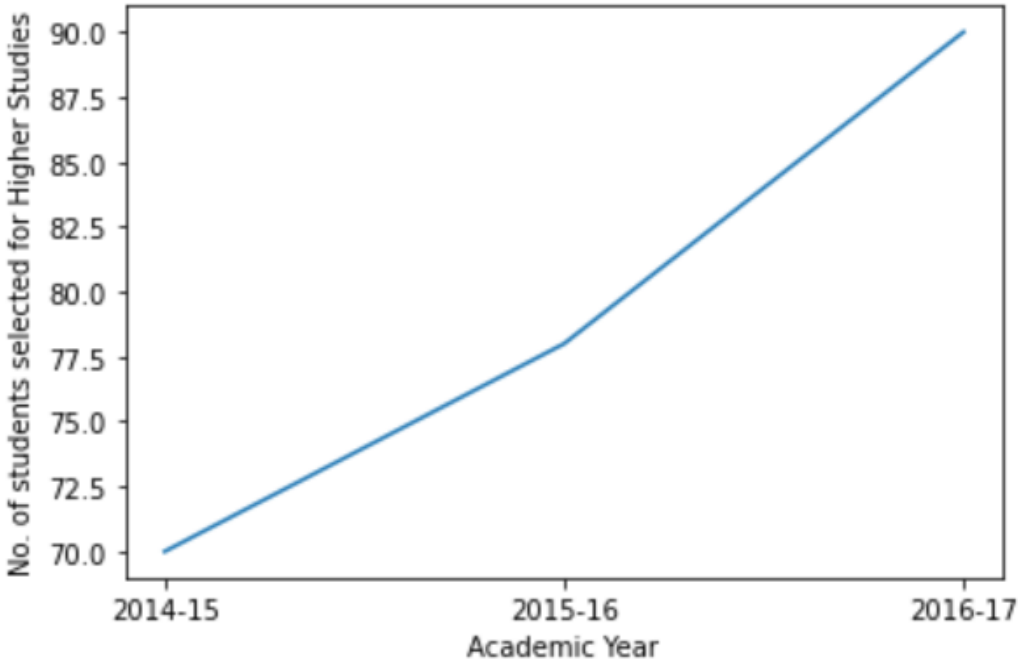
No. of students placed VS Academic Year



As observed above, there seems to be positive linear relation between the two attributes

But what we can see is that there was spike in placements during the year 2015-16, which may be due to some underlying reason *(to which we have an answer in the below visualization).*

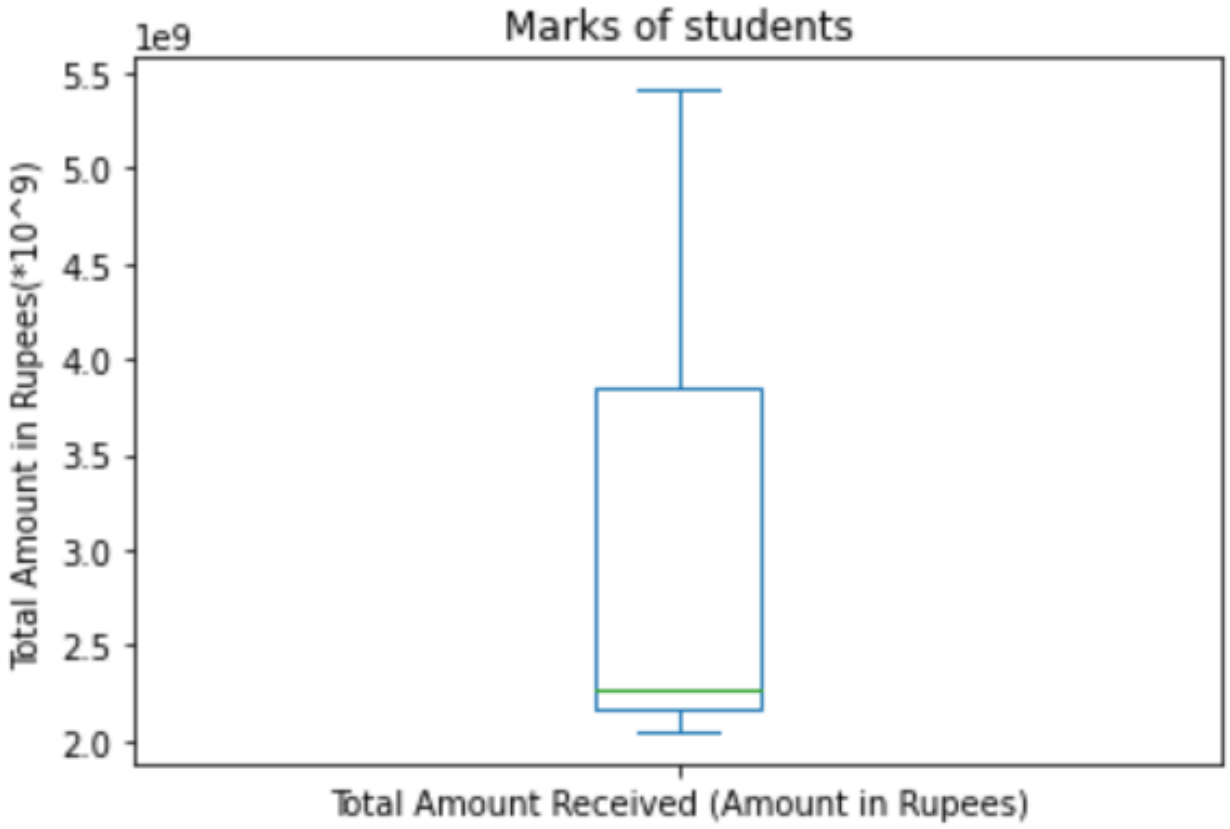
No. of students selected for Higher Studies VS Academic Year



This has an inverse trend as compared to the above

we can see here that in year 2015-16 No. of students selected for Higher Studies were less than other two academic years. Hence, this resulted in a spike in placements that year also.(answer to above analysis query)

Total Amount Received



From the above box plot, we can observe the following:

- Median: 2048965952
- Maximum = 5411417468
- Minimum = 2048965952

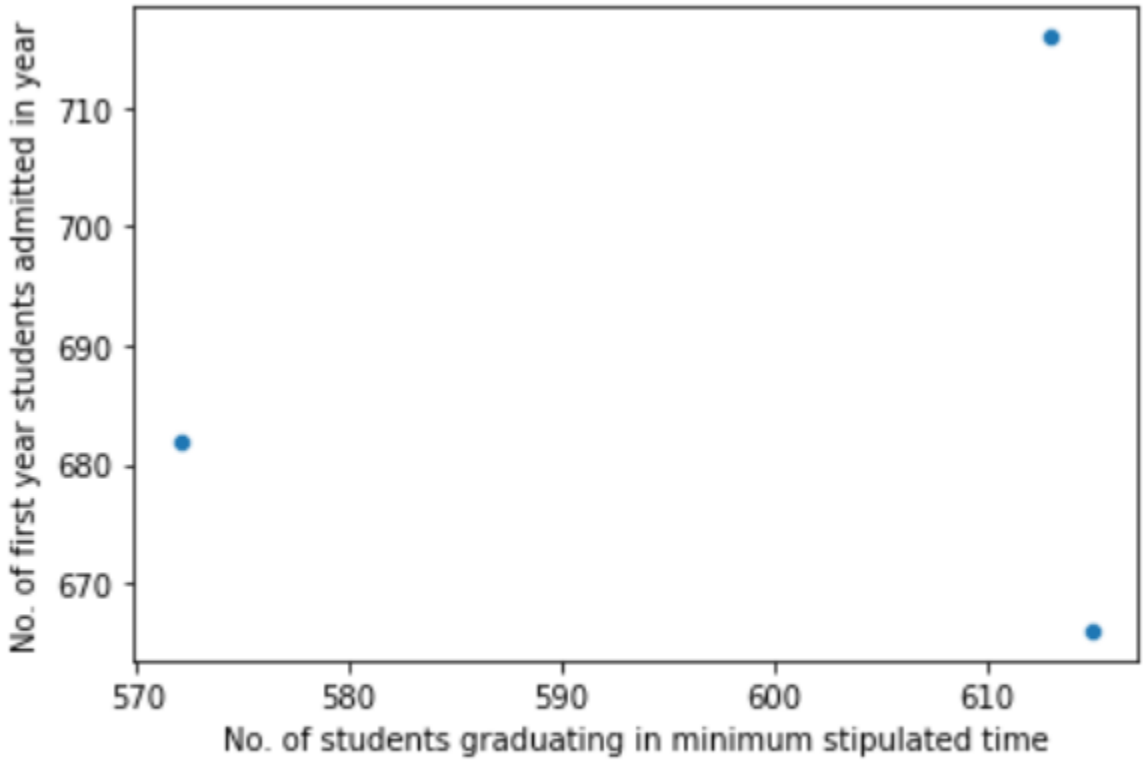
Now, let's analyze 'Financial Resources: Utilized Amount for the Capital expenditure for previous 3 years' table to understand spending patterns

	Library	Software & Equipment	Eng Workshop	Others	Financial Year
1	(Seventeen Crore Eighty Six Thousand One\rHun...	(One Hundred Eighty Two Crore Seventy Lakh\rT...	(Forty Lakh Fifty Eight Thousand Six Hundred ...	(Seventy Four Crore Fifty Three Lakh Eighty O...	2019-20
2	(Sixteen Crore Two Lakh Seventy One Thousand\...	(One Hundred Seventy Nine Crore Thirty Seven\...	(Forty Three lakh Two Thousand One Hundred\rS...	(Seventy Six Crore Forty five Lakh Thirty Eig...	2018-19
3	(Fourteen Crore Sixty One Lakh Twenty Five\rT...	(One Hundred Forty Two Crore Forty Six Lakh\r...	(Fifty Three Lakh Ninety Nine Thousand Only)	(Eighty Nine Crore Forty Eight Lakh Twenty Se...	2017-18

This is the final Cleaned data ready to analyze Spending Patterns.

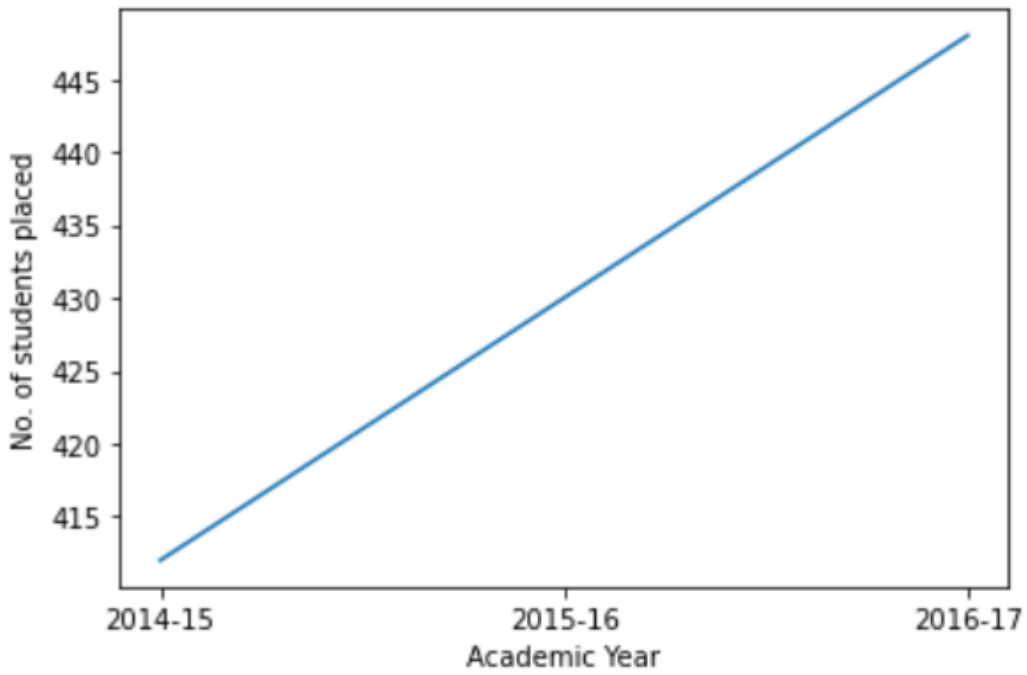
3. EDA on `ins_id[2]` i.e. IIT BOMBAY

1. No. of first year students admitted in year VS No. of students graduating in minimum stipulated time



There Does not seem to be any relation among them

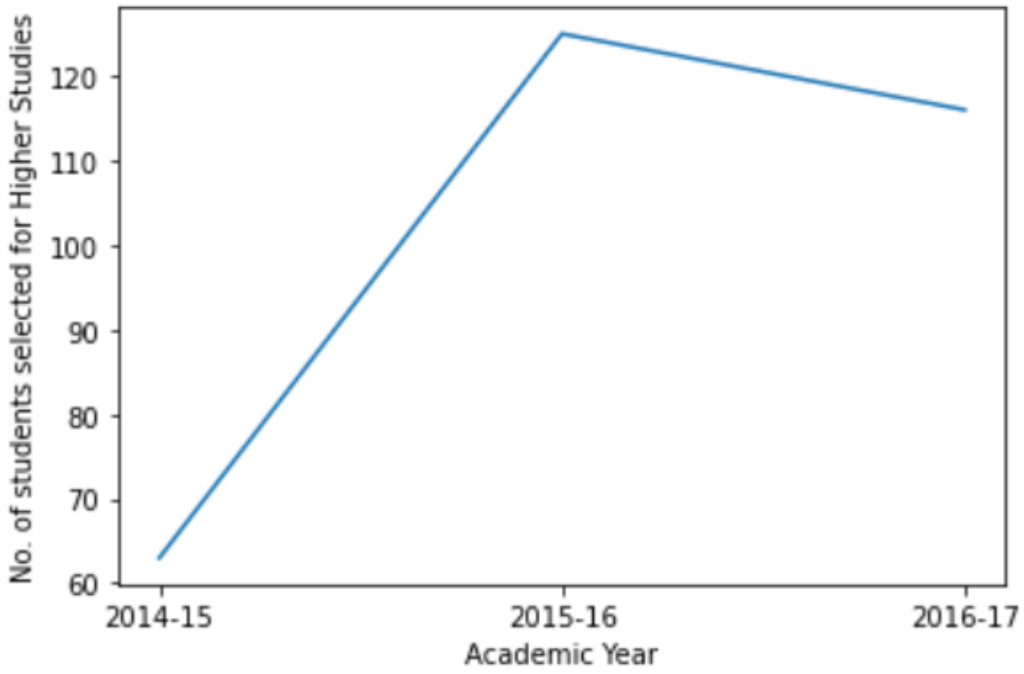
No. of students placed VS Academic Year



As observed above, there seems to be positive linear relation between the two attributes

But what we can see is the line is absolutely straigt with no spikes or dips, indicating there was a constast uniform increase in placed students throught the year

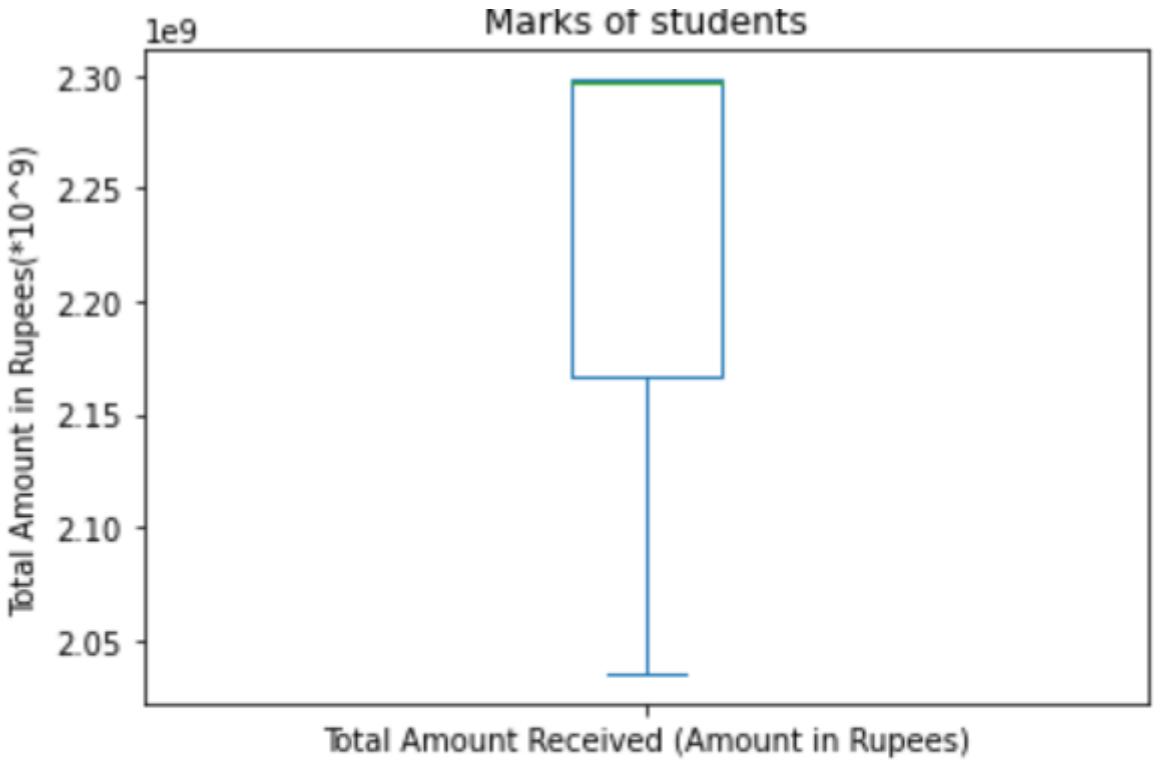
No. of students selected for Higher Studies VS Academic Year



There is sharpe increase in Students selected for higher studies in year 2015-16

If we were working with a hugh amount of data, this could have acted as an outlier and steered our model the other way. Best to remove such observations.

Total Amount Received



From the above box plot, we can observe the following:

- Median: 2297748211
- Maximum = 2298187443
- Minimum = 2035039984

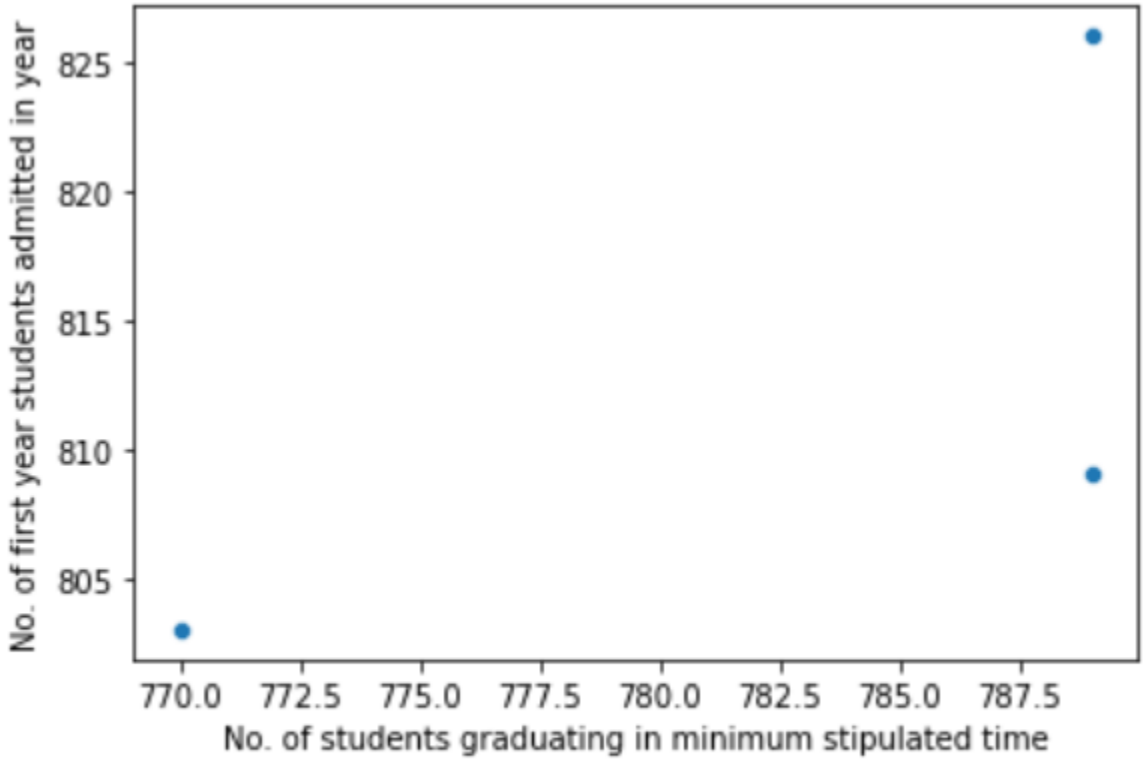
Now, let's analyze 'Financial Resources: Utilized Amount for the Capital expenditure for previous 3 years' table to understand spending patterns

	Library	Software & Equipment	Eng Workshop	Others
1	NaN	(Twenty Three Crores Thirty Three Lakhs Eleve...	(One Hundred Thirty Five Crores Eighty Two\rL...	(Seventy Eight Lakhs Four Hundred and Fourtee...
2	NaN	(Eighteen Crores Sixty Five Lakhs Twenty Nine...	(One Hundred Twenty Crore Thirty Six Lakhs\rN...	(One Crore Twenty One Lakhs Sixty Two Thousan...
3	NaN	(Sixteen Crores Sixty Seven Lakh Fourteen\rTh...	(Eight Crore Sixty One Lakh Thirty Seven Thou...	(Two Crores Forty Two Lakhs Twenty Eight\rTho...

This is the final Cleaned data ready to analyze Spending Patterns.

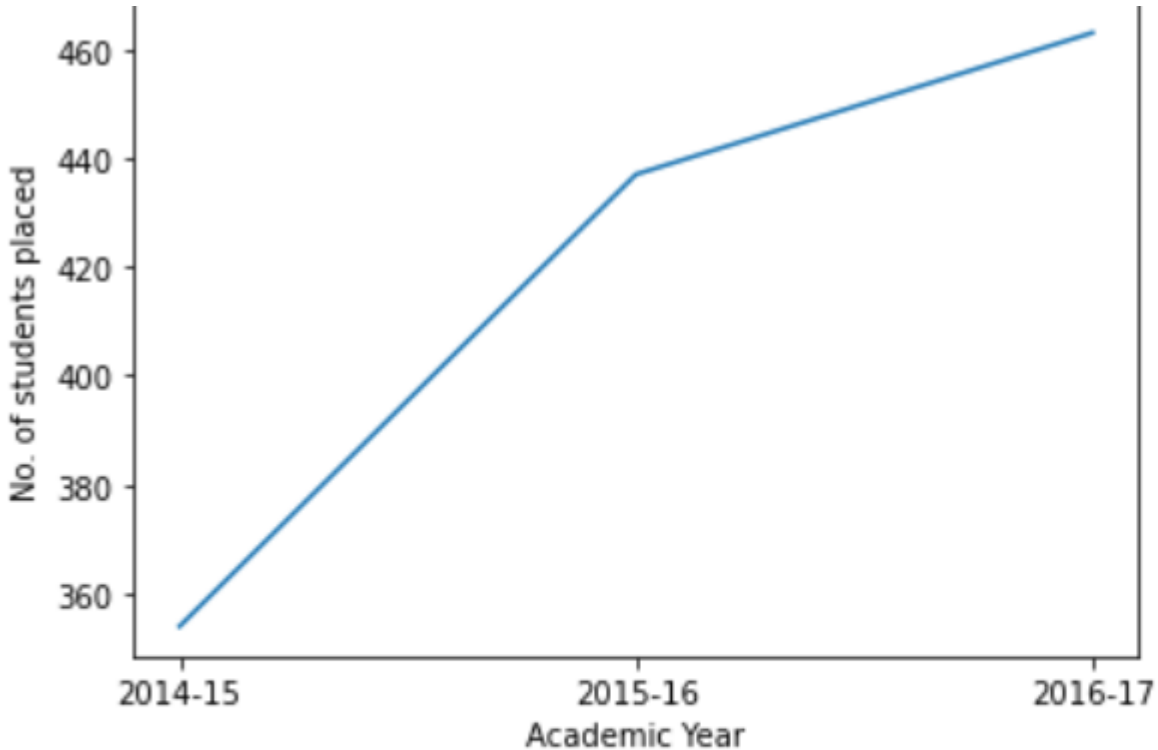
4. EDA on `ins_id[3]` i.e. IIT KANPUR

1. No. of first year students admitted in year VS No. of students graduating in minimum stipulated time



There Does not seem to be any relation among them

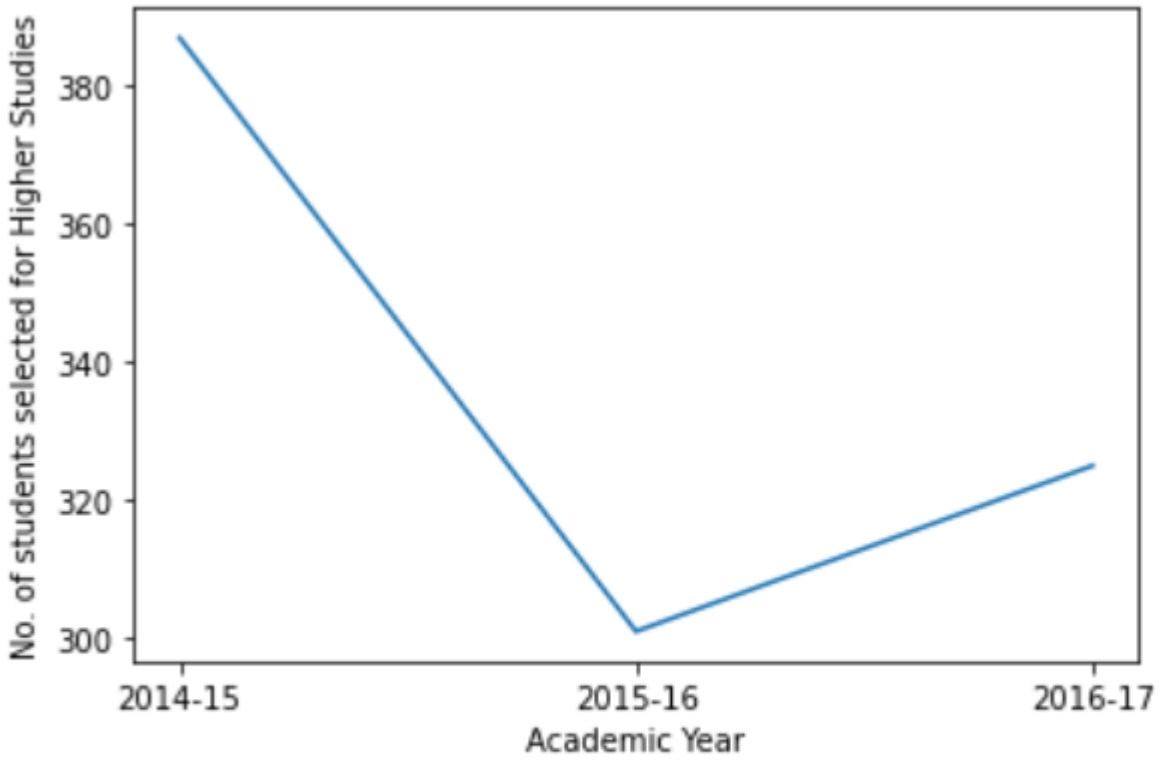
No. of students placed VS Academic Year



As observed above, there seems to be positive linear relation between the two attributes

There is a slight spike at year 2015-16 for students getting placed.

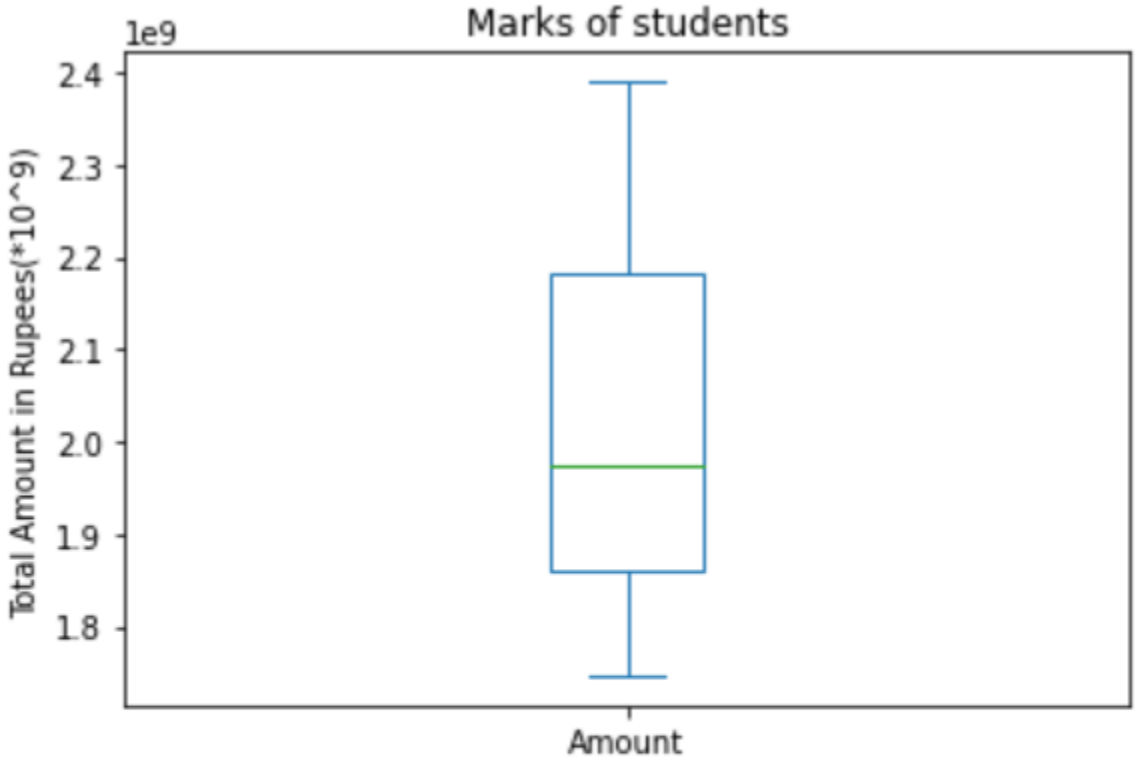
No. of students selected for Higher Studies VS Academic Year



There is sharpe decrease in Students selected for higher studies in year 2015-16

This could be the reason for spike in students getting placed in year 2015-16 above.

Total Amount Received



From the above box plot, we can observe the following:

- Median: 1974913396
- Maximum = 2389745850
- Minimum = 1747013587

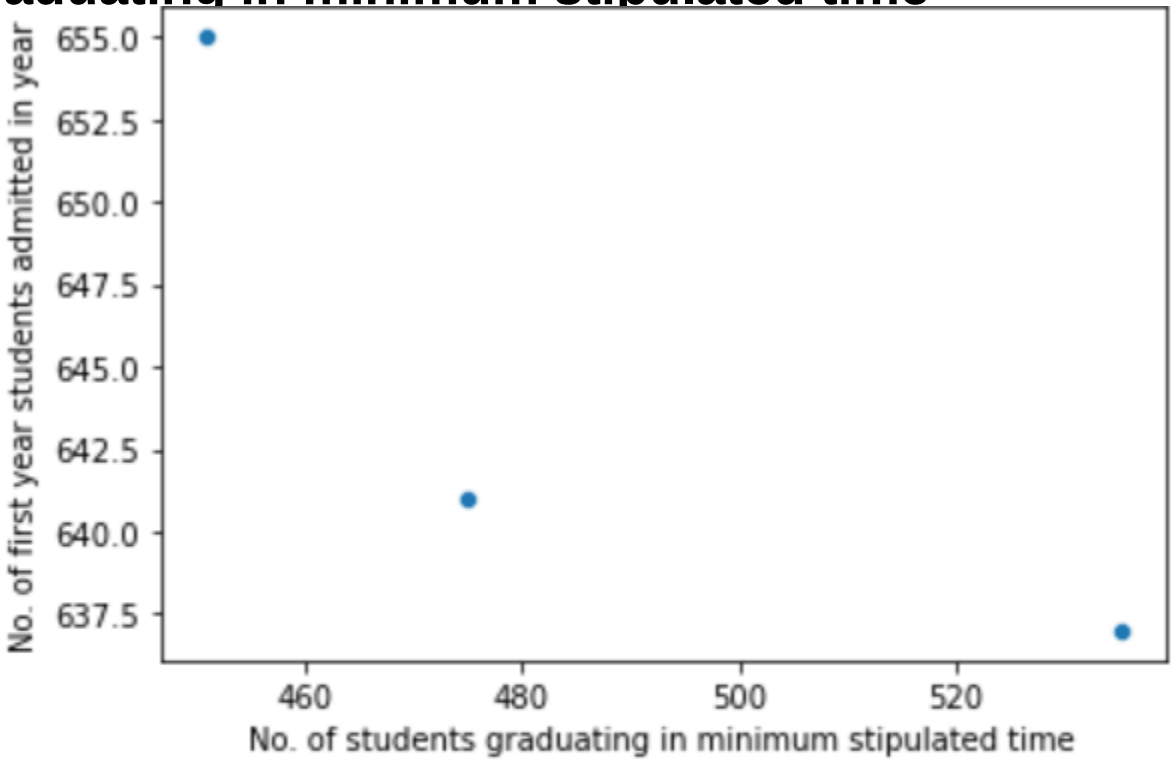
Now, let's analyze 'Financial Resources: Utilized Amount for the Capital expenditure for previous 3 years' table to understand spending patterns

	Library	Software & Equipment	Eng Workshop	Others
1	(Sixteen Crore Eighty Lakh)	(Sixty Eight Crore Fifty Seven Lakh Sixty Nin...	(Nine Crore Eighty Five Lakh)	(Seven Crore Thirty Four Lakh)
2	(Fifteen Crore Thirty Lakh)	(Sixty Seven Crore Forty Five Lakh Eighty Six...	(Four Crore Forty Four Lakh)	(Four Crore Sixty Seven Lakh)
3	(Twelve Crore Ninety Six Lakh)	(Sixty Two Crore Eleven Lakh Ninety Four\rTho...	(Nine Crore Sixty Four Lakh)	(One Crore Fifty Lakh)

This is the final Cleaned data ready to analyze Spending Patterns.

5. EDA on `ins_id[4]` i.e. IIT KHARAGPUR

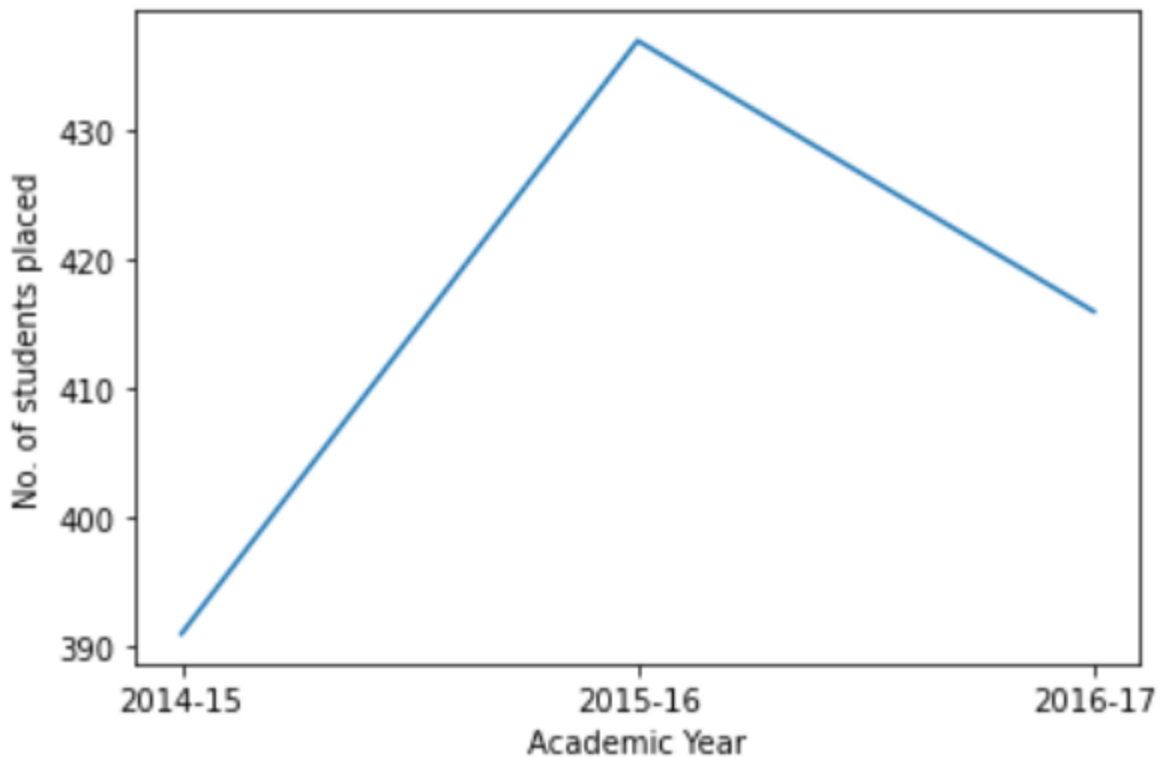
1. No. of first year students admitted in year VS No. of students graduating in minimum stipulated time



There seems to be a negative correlation b/w them

with increase in no. of students graduating, no. of students admitted seems to decrease.

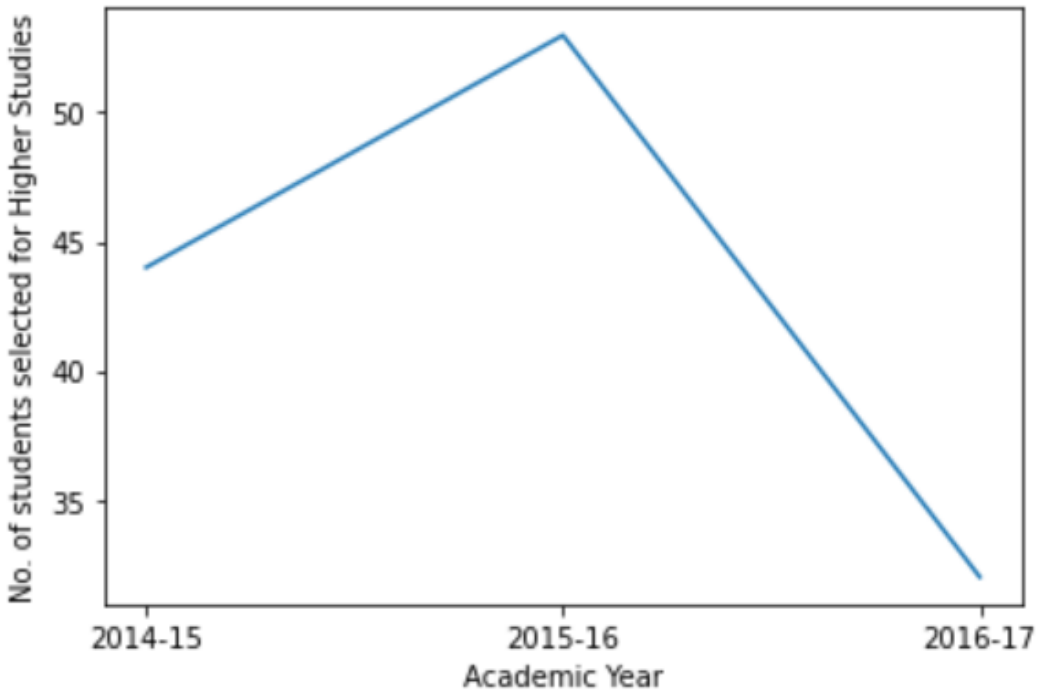
No. of students placed VS Academic Year



As observed above, there does not seem to be any linear relation here

There is a Huge spike at year 2015-16 for students getting placed.

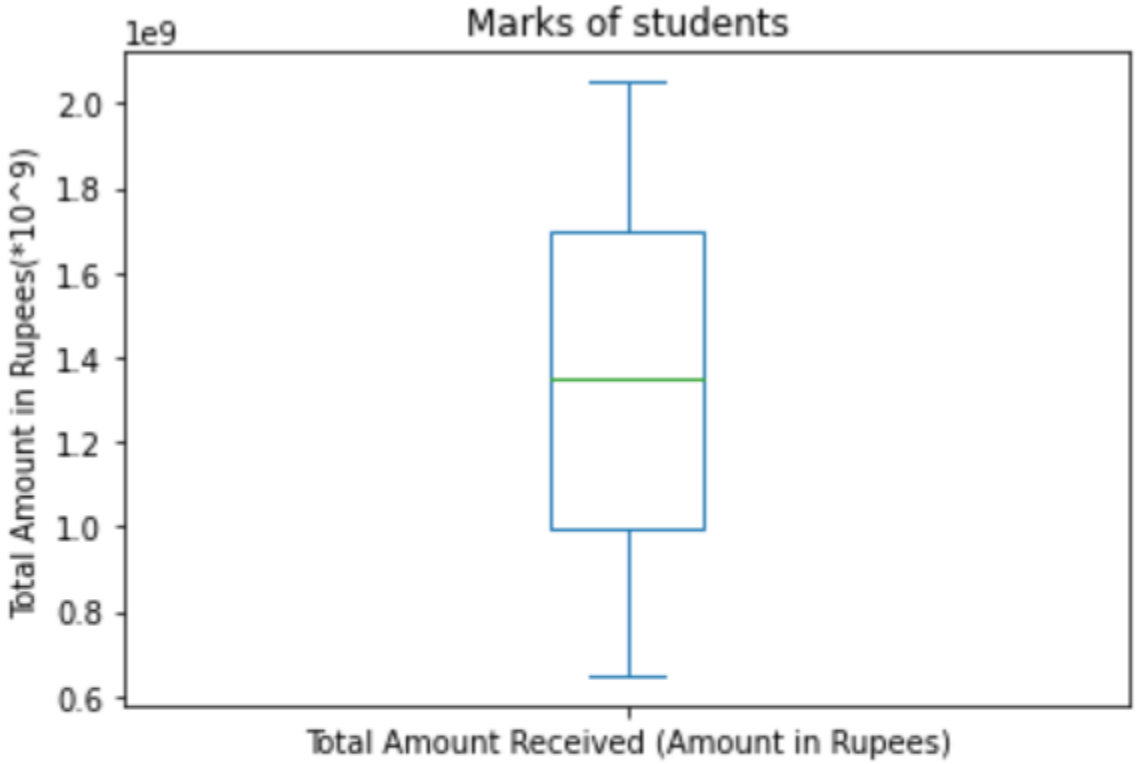
No. of students selected for Higher Studies VS Academic Year



There is sharpe spike again in Students selected for higher studies in year 2015-16

Both placements and students getting selected got spiked during this year, which checks out with the data and could be due to some underlying cause.

Total Amount Received



From the above box plot, we can observe the following:

- Median: 1349147118
- Maximum = 2051250000
- Minimum = 648432016

Now, let's analyze 'Financial Resources: Utilized Amount for the Capital expenditure for previous 3 years' table to understand spending patterns

	Library	Software & Equipment	Eng Workshop	Others	Financial Year
1	(Twenty Two Crores Seventy Three Lakhs Twenty...	(Twenty Two Crores Forty Two Lakhs Thirty Eig...	(Zero)	(Twenty Seven Crores Fifty Five Lakhs Sixty S...	2019-20
2	(twenty two crore three lakh ninety seven tho...	(eleven crore seventy three lakh sixteen thou...	(zero)	(fifty eight crore ninety one lakh eight two ...	2018-19
3	(seventeen crore fifty two lakh thirty thousa...	(twenty eight crore forty five lakh forty thr...	(zero)	(thirty five crore sixty five lakh forty six ...	2017-18

This is the final Cleaned data ready to analyze Spending Patterns.

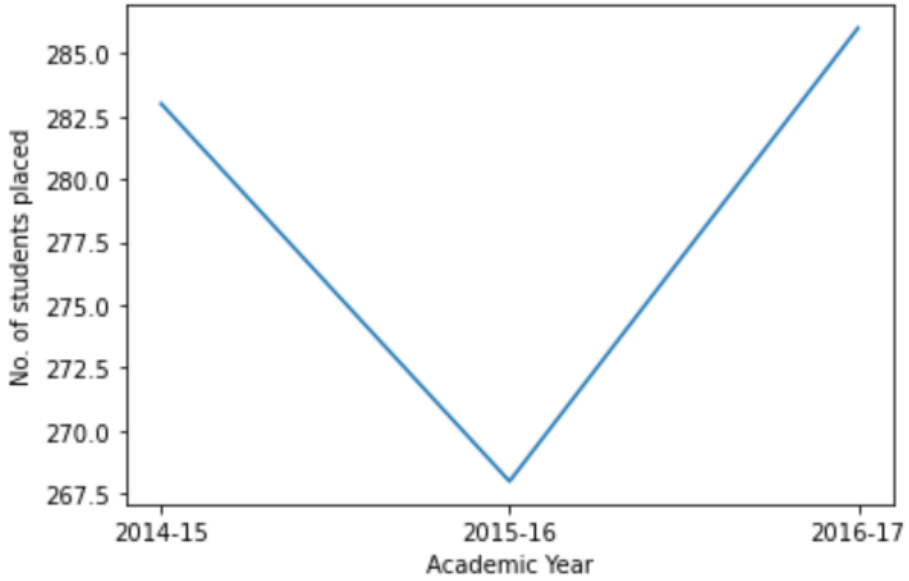
INNOVATIVE INSIGHT

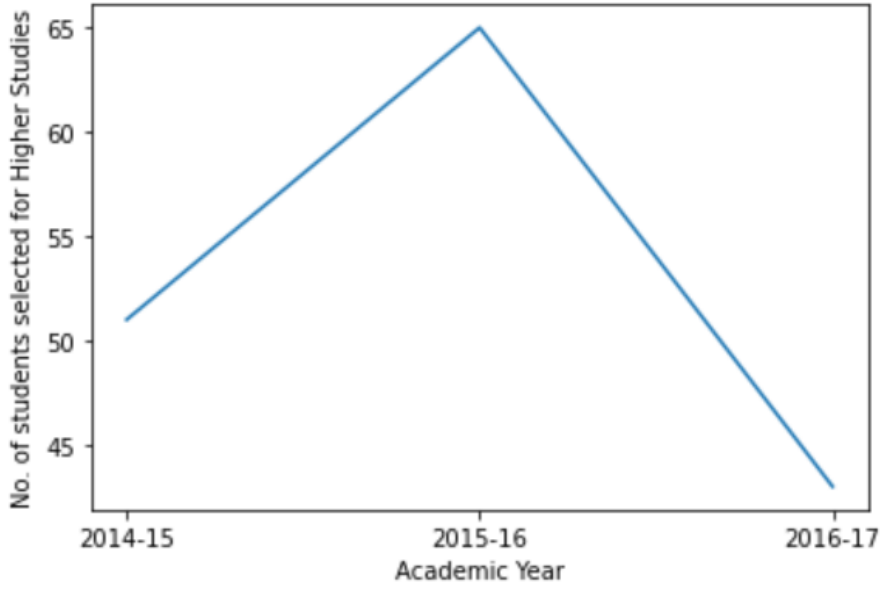
Whenever there was an unexpected fluctuation in Data trend,

There was some other explainable underlying change in attribute to compensate for it.

For e.g. In IIT Madras, in academic year 2015-15

Dip in No. of students placed can be explained due to **Spike in No. of students selected for Higher studies**





CONCLUSION

Data analytics & visuals

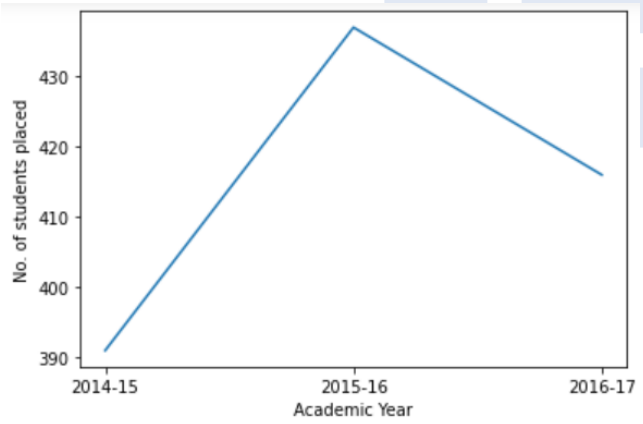
- help perform EDA,
- communicate Data clearly,
- share unbiased representation of data,
- use them to support insights to different stakeholder.

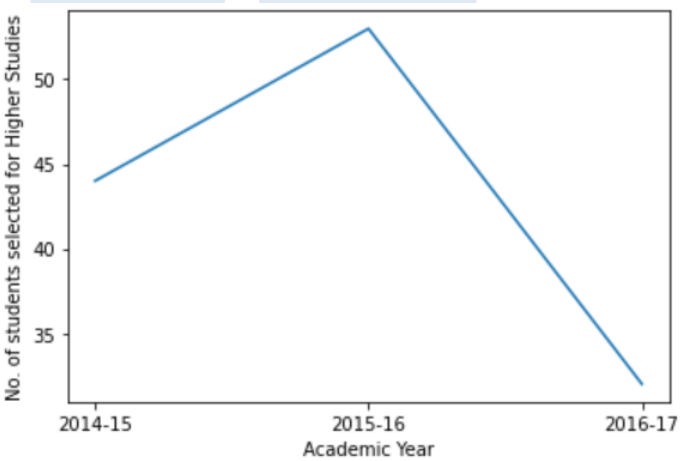
Also, data observation in IIT Kharagpur, for number of students getting placed & number of students going for higher studies, both were higher in academic year 2015-16, unlike other institutes.

This can reveal some new,

Previously unknown knowledge

& insights.





Mail