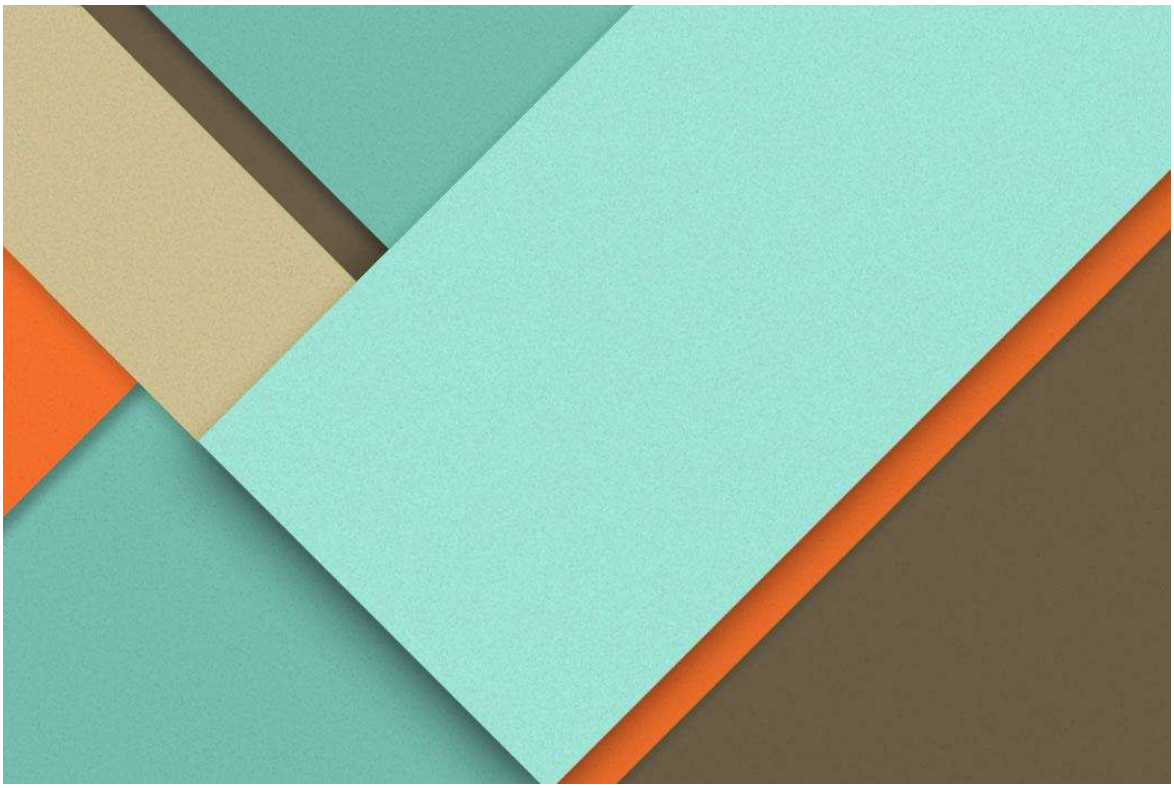


Exploratory Data Analysis

On Praxis' PGPDS Course Parameters and Students' Opinions

02.12.2021

—

Adarsh Raj - A21001

Mohammad Shaan - A21018

DS July 2021

STS-I Project

Motivation behind the project

The aim of this project study is to explore the distribution of variables and responses through exploratory data analysis and to find the most appropriate ways of utilizing those factors to gain insights on the course and student reviews concerned with Praxis' PGPDS course. Furthermore, based on the insights and reviews, the course structure and few of the parameters related to it can be upgraded to cater to the changing dynamics of the students' comfortability & opinions.

Praxis went ahead and launched its first full-time Business Analytics course which was later rechristened as Data Science. This course was offered full-time and with in-campus mode but ever since Covid hit in 2020, the course shifted to online mode and now it is being offered in dual-learning mode i.e., both offline and online. But, as we know, this mode has its own set of advantages and disadvantages. So, this study aims at taking a deeper look into the opinions of students around how the course is delivered and how their experience has been after taking the course.

Introduction

1) What Data Science is all about?

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply machine learning algorithms to numbers, text, images, video, audio, and more to produce artificial intelligence (AI) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate insights which analysts and business users can translate into tangible business value.

2) What Praxis is all about and how is it changing the data science industry?

Praxis has been the forefront in India's data science stratosphere transformation and has been leading the cause for the last decade. It has enabled a wide number of students gain technical and practical expertise and helped them transition into the career of Data Science & Machine Learning.

3) Dataset description

The dataset contains responses from 63 students from Praxis Business School's PGPDS course, which contains opinions on a varied number of parameters related to the course and how the course is delivered in the current dual learning environment.

Most of the questions are based on Likert Scale and other information like respective batch of student and campus location has been collected to segment the results for better understanding.

Exploratory Data Analysis

EDA is one of the crucial steps in data science that allows us to achieve certain insights and statistical measures that are quite essential. EDA is generally cross classified in two ways. First, each method is either non-graphical or graphical.

And second, each method is either univariate or multivariate (usually just bivariate).

Non-graphical methods generally involve calculation of summary statistics, while graphical methods obviously summarize the data in a diagrammatic or pictorial way.

Below, we'll explore each of the variables in the dataset one by one and try to have an extended understanding and distribution of it across other variable features.

1) Distribution of responses

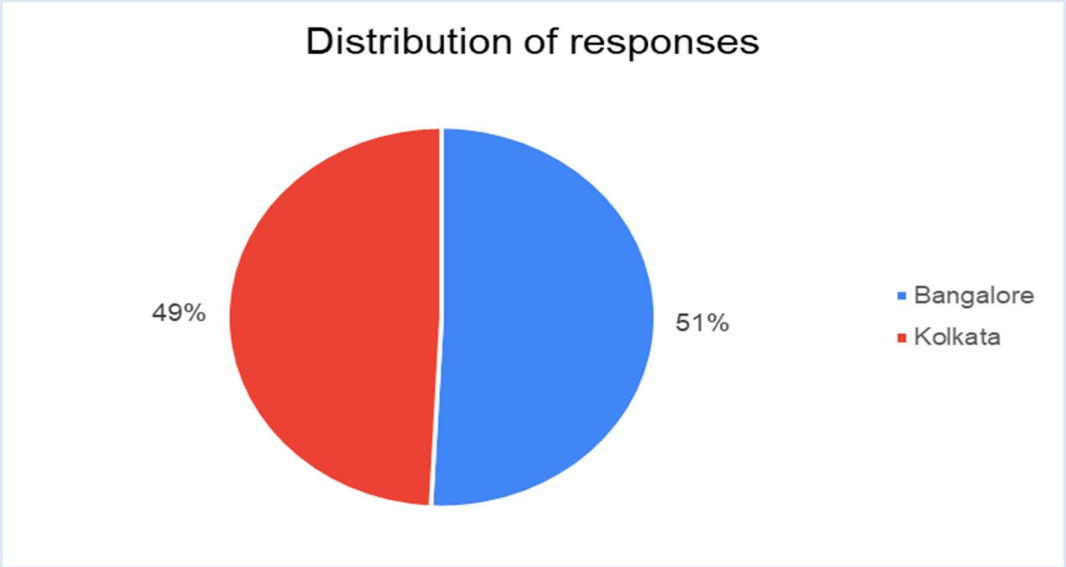
The survey generated responses from 63 students from different batches of the data science course offered by Praxis Business school. It also incorporates both the campuses of Praxis Business School i.e., Bangalore and Kolkata.

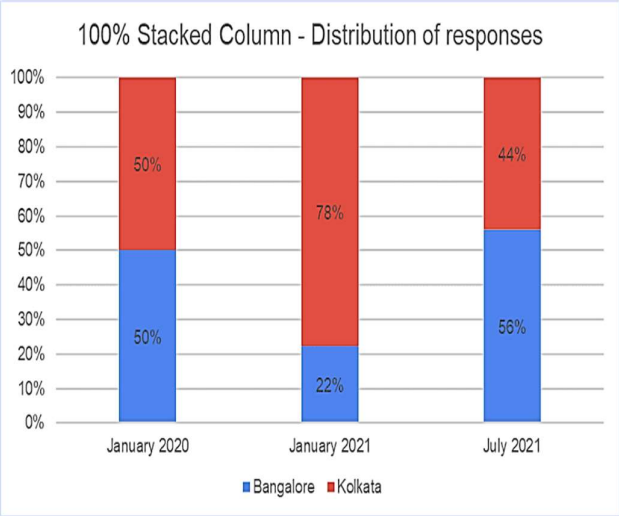
Campus location	January 2020	January 2021	July 2021	Total
Bangalore	2	2	28	32
Kolkata	2	7	22	31
Total	4	9	50	**63**

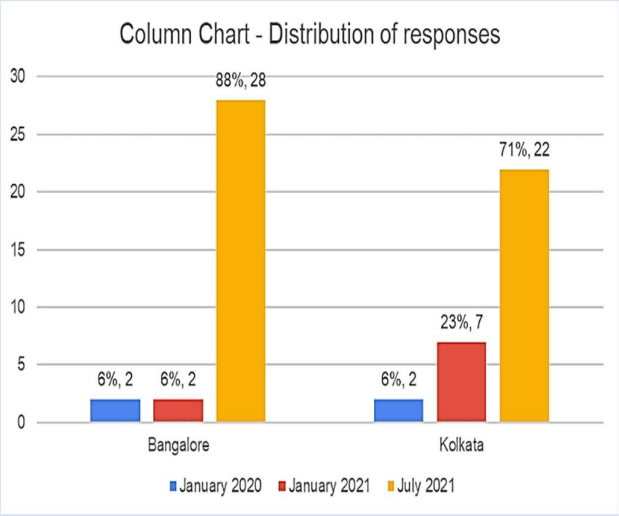
With help of the above table, we can see the distribution of responses across different batches and both campus locations.

Graphically, we can see –

- Distribution of responses across both campus locations
- Distribution of responses across different batches of both campus locations







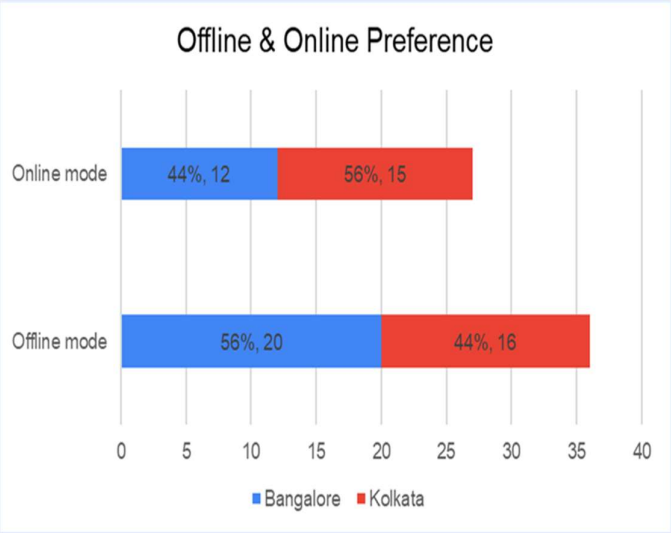
Insights –

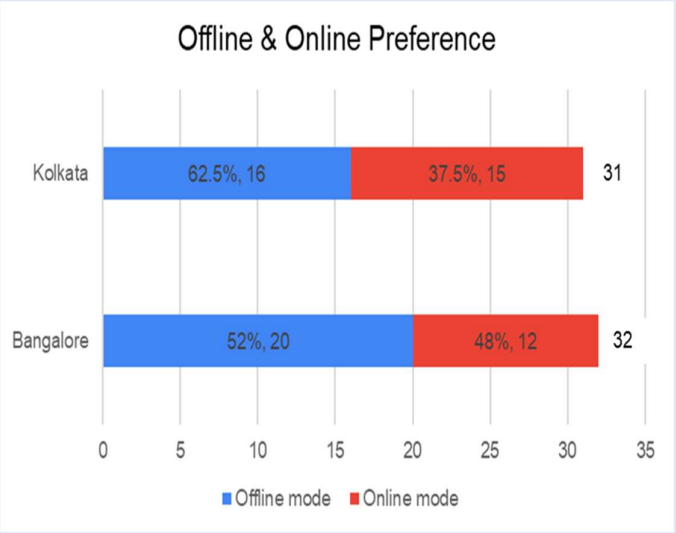
- From the table, we can see that we received 32 responses for Bangalore campus and 31 responses for Kolkata campus. The distribution of responses across both the campus location is pretty much the same.
- For both the campus locations, majority of responses came from the newly started batch i.e., July 2021, which amounts to 88% for Bangalore campus and 71% for Kolkata campus. Overall including both the campuses, responses from July 2021 batch has been at 79%.

2) Preferred mode of learning

For the preferred mode of learning by students, we plot the tabular representation of data and get to understand that for students who prefer offline mode are more in numbers compared to students who prefer online mode.

Preferred mode of learning	Bangalore	Kolkata	Total
Offline mode	20	16	36
Online mode	12	15	27
Total	32	31	**63**



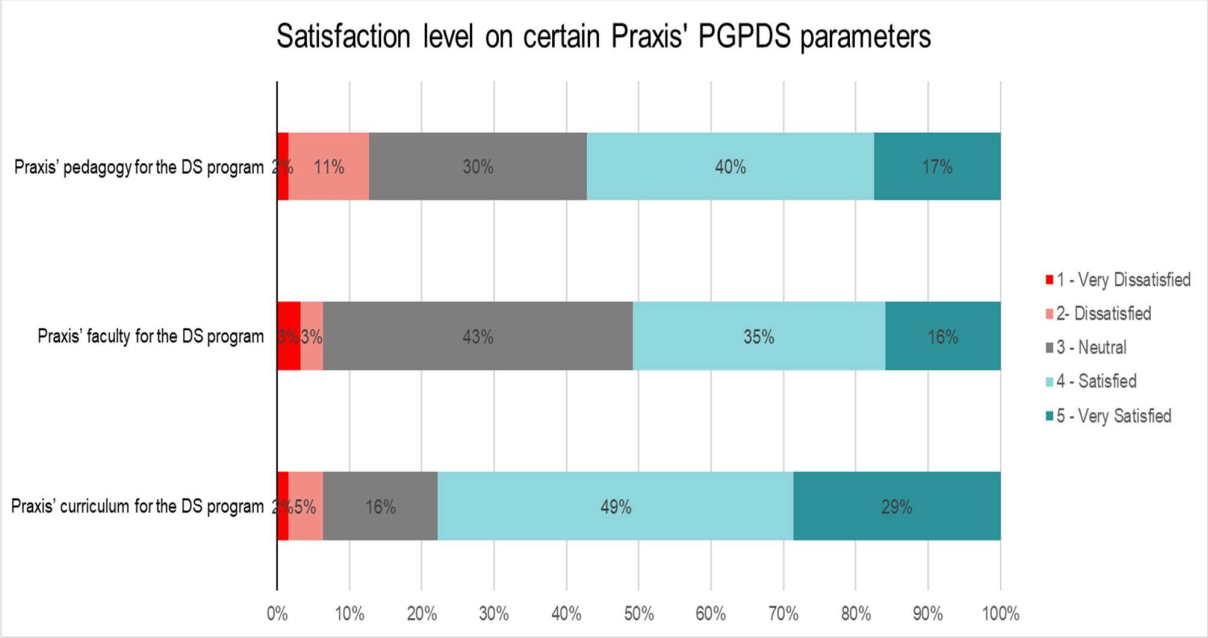


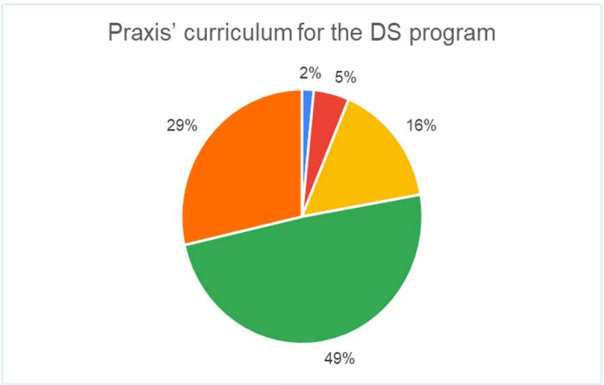
Taking a deeper dive into the proportions and segmenting it based on location, we get the following insights –

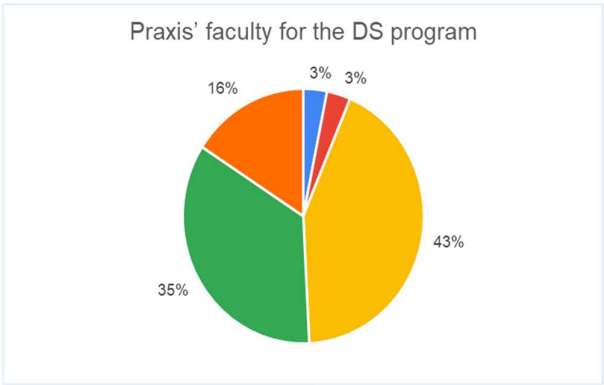
- Out of 63 responses, we have 57% respondents opting for offline mode while, for online mode is preferred by 43% of respondents.
- For Bangalore campus, 52% of students prefer offline mode while for Kolkata campus, approximately 63% of students prefer offline mode.
- From the above, we get to know that for Kolkata batch, students are more interested in going with offline mode and this could be due to the reason that Kolkata campus of Praxis Business School hosts a large & proper college infrastructure compared to Bangalore campus.

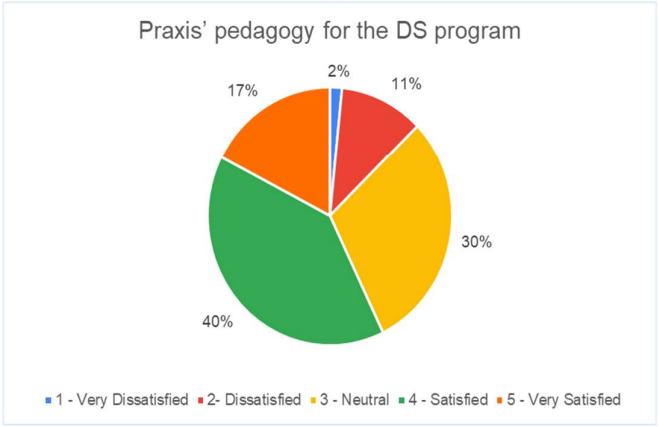
3) Praxis' Curriculum, Faculty & Pedagogy

For the above three parameters related to Praxis' PGPDS course, we obtained responses from students based on Likert scale and plotted it w.r.t different bins collected –









From the help of above chart, we get to understand that with –

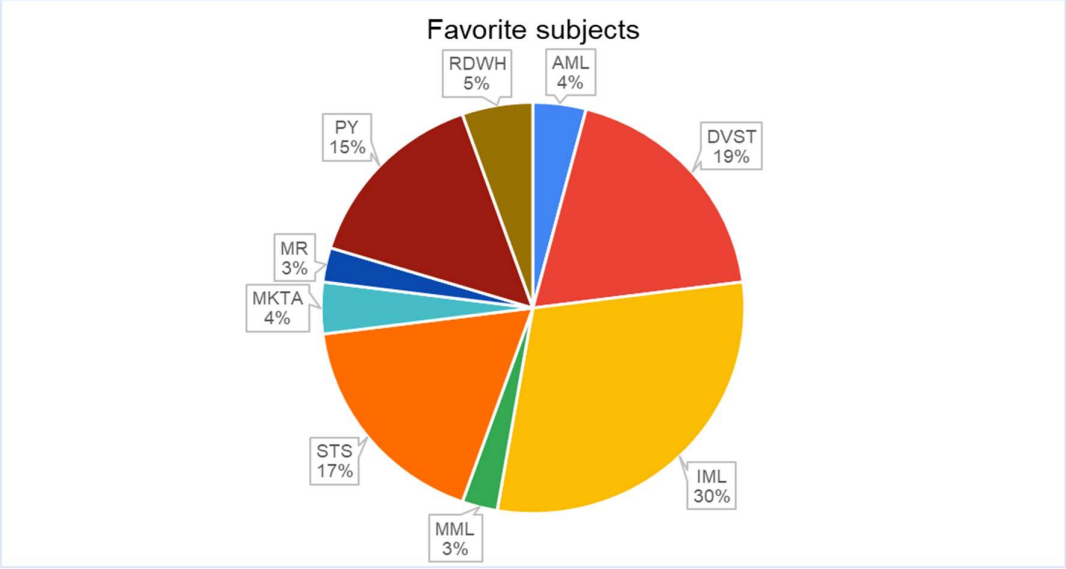
- Praxis' pedagogy for the DS program, only 57% of the respondents are either "Satisfied" or "Highly Satisfied"
- Praxis' faculty for the DS program, only 51% of the respondents are either "Satisfied" or "Highly Satisfied"
- Praxis' curriculum for the DS program, a greater proportion i.e., 78% of respondents are either "Satisfied" or "Highly Satisfied"

Using the three pointers above, we can conclude that for all three parameters – Pedagogy, Curriculum and Faculty, on overall basis, students are satisfied with the approach from Praxis Business School.

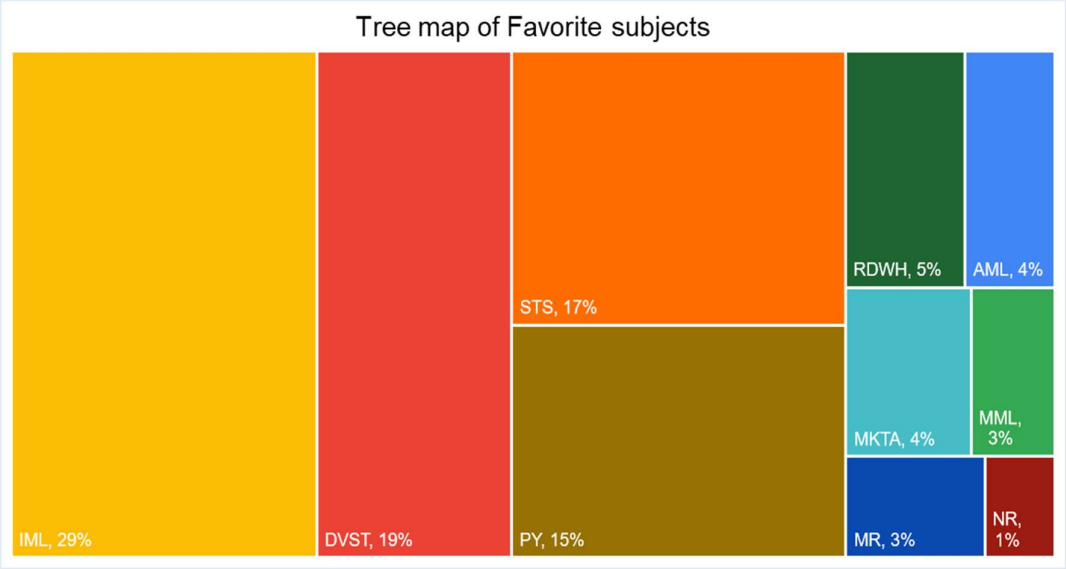
4) Favorite subjects

For the favorite subjects, data collected was converted from plain text to categorical feature and then it was plotted against the number of respondents. Note that, in this selection the respondent was allowed to pick more than one favorite subject.

NR in the below visual demonstrates that there was no response from that percent of students. Other subject codes stand their usual meaning.



Another way of representing the above visual is with the help of Tree map and the same is shown below.

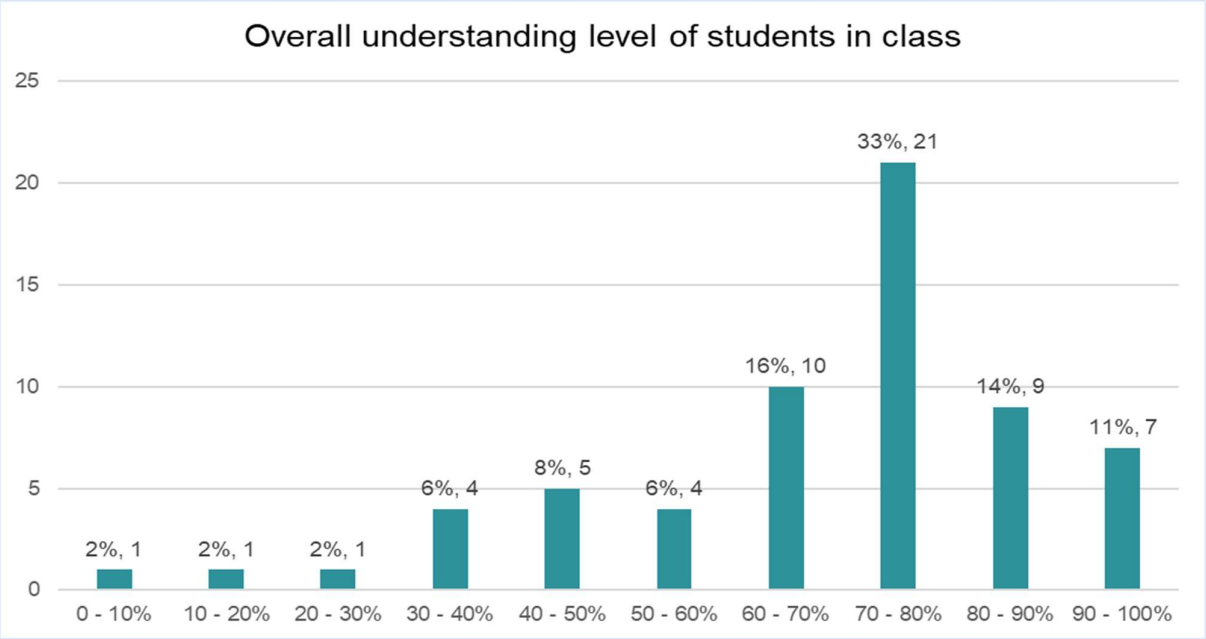


From the help of above pie chart and tree diagram visuals, we can understand the favorite subjects of the students currently. Further insights –

- Approximately 23% of students are more interested in Machine learning subject
- Subjects like RDWH & MML have very low likability (combined 8%) but these 2 subjects are quite major for holistic development of a student's data science and machine learning concepts

5) Understanding level in the class

Responses were collected from students based on how much are they able to understand during live lectures and then same was plotted in a column chart.



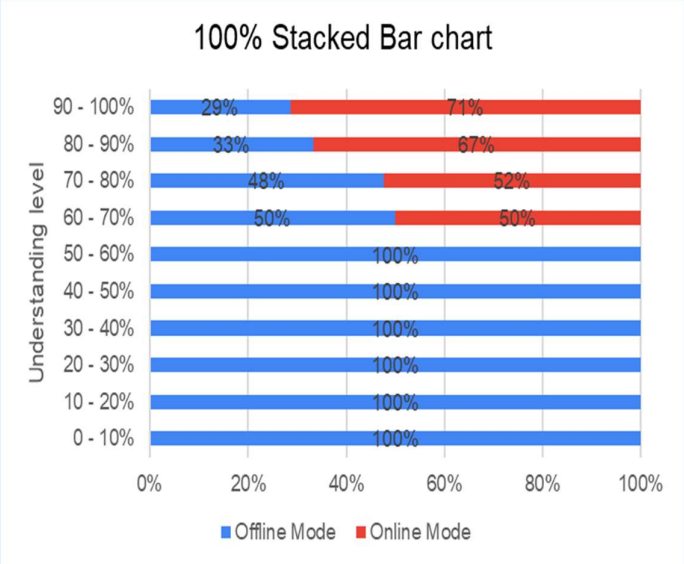
Here, we would like to define three terms in relevance to this understanding level in live class –

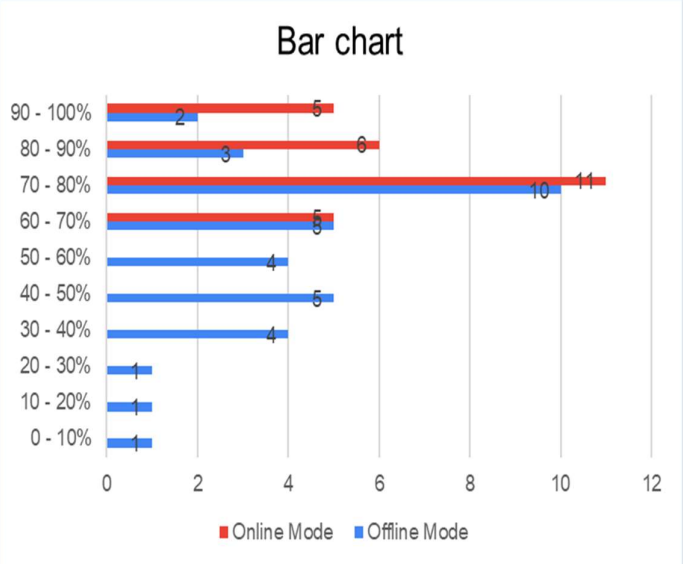
1. High absorption – Understanding level more than 70% in class
2. Medium absorption – Understanding level between 30% and 70%
3. Low absorption – Understanding level of below 30% in class

With the help of above graph, we can conclude that –

- More than half number of students (58%) can better understand the concepts taught in class and have high absorption rate
- A significant percent of students i.e., 36% can understand concept with medium level of absorption. Feedback could be taken from them to understand why they are not able to understand the most and accordingly actions could be taken to cater to their needs, if possible
- A very low proportion of students i.e., 6% of students are not able to understand significant number of concepts and probably end up wasting their time by revisiting lecture videos or learn through other platforms. Similar actions like above can be taken for them to cater to what is missing and ensure these students also benefit the most from live lectures

a) Understanding level w.r.t preferred mode of learning

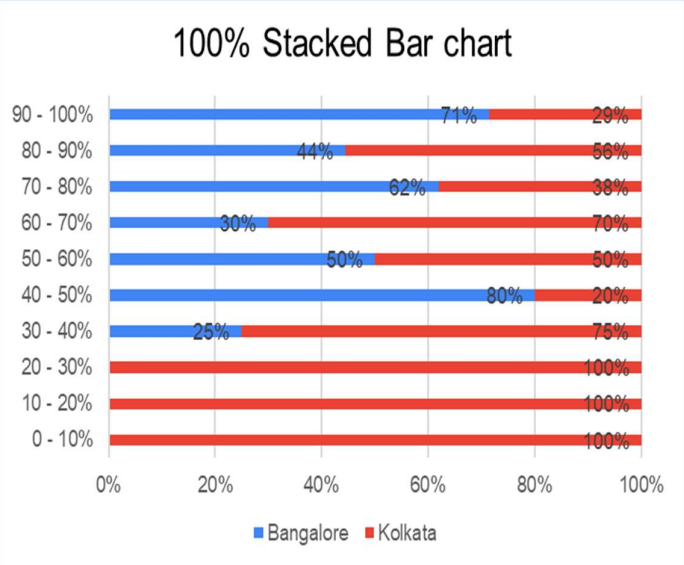


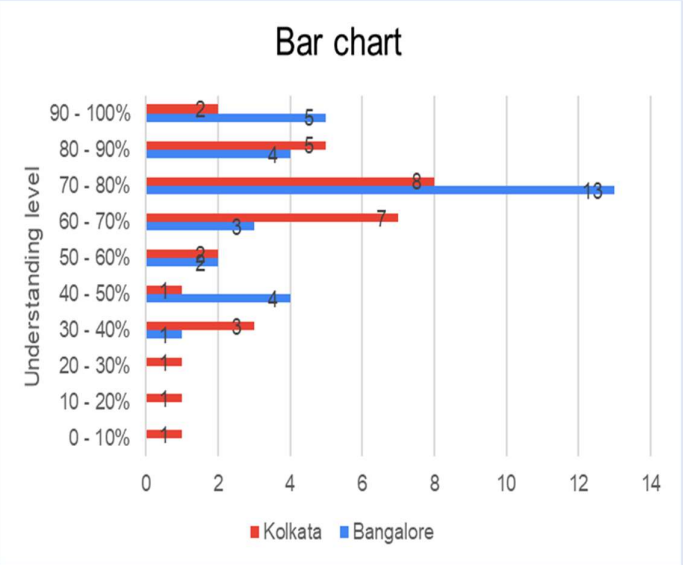


Above we have shown 2 plots for understanding level w.r.t preferred mode of learning – Bar chart and 100% Stacked Bar chart. With the help of above visual, we can conclude that –

- All the respondents who prefer Online mode can understand the concepts taught in class with almost high level of absorption
- Respondents who prefer offline mode are the ones who are not able to understand the concepts with high level of absorption

b) Understanding level w.r.t Campus location





Above we have plotted understanding level w.r.t campus location and similarly 2 charts have been plotted for the same – Bar and 100% Stacked Bar. Insights which we can derive are –

- In low absorption rate of understanding level, all the respondents are from Kolkata campus
- Overall, Bangalore campus has shown better understanding level in class compared Kolkata campus in medium and high absorption category

6) Attention level in the class

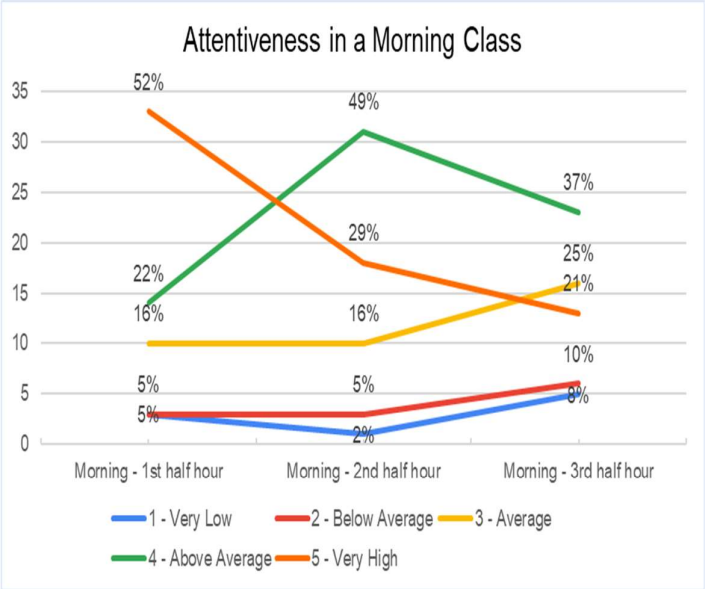
Study of student's attention level in both morning and afternoon class. Later on, this will be further described by other parameters like preferred mode of learning of students and campus location.

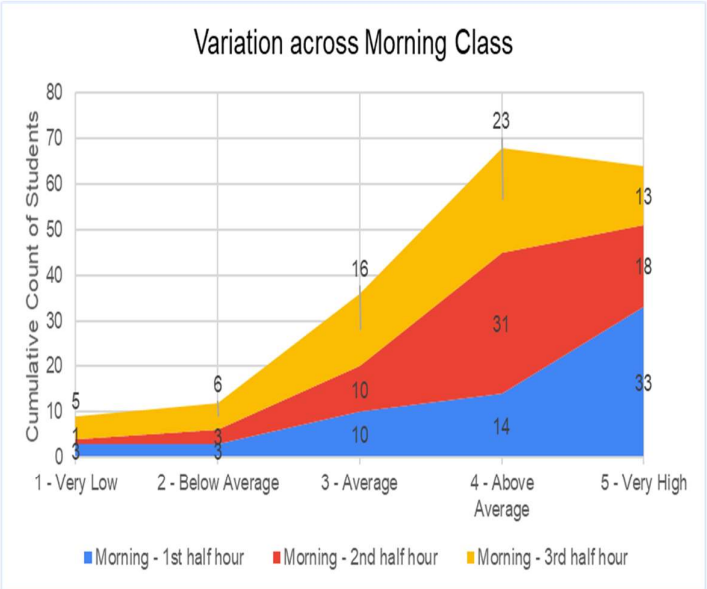
a) Morning Class

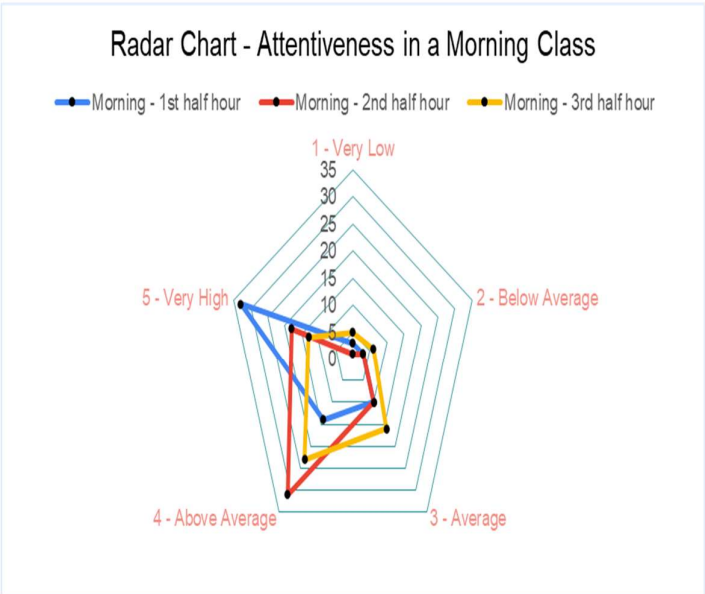
Please take note of the below tabular representation of the responses –

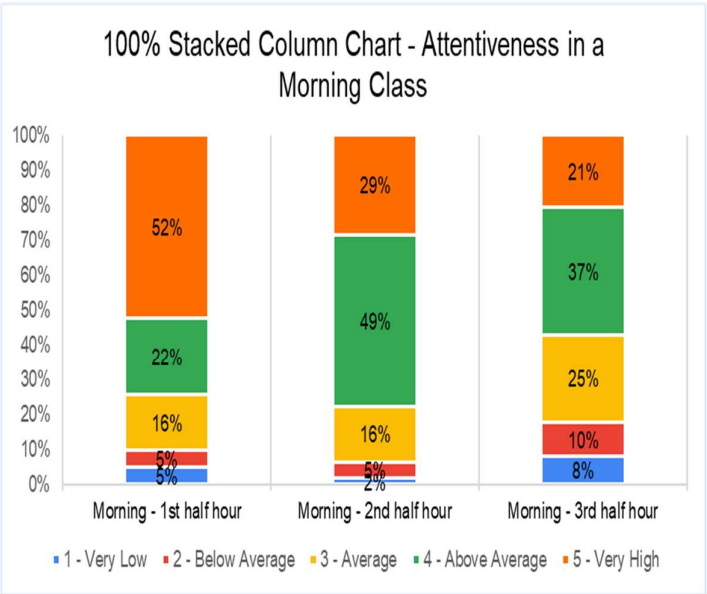
Instances	1 - Very Low	2 - Below Average	3 - Average	4 - Above Average	5 - Very High	Total
Morning - 1st half hour	3	3	10	14	33	63
Morning - 2nd half hour	1	3	10	31	18	63
Morning - 3rd half hour	5	6	16	23	13	63

The survey form related to this section takes in response for two different timelines – morning class and afternoon class.
Respondents were required to rate in their attentiveness in class during the three stages of a class – first half hour, second half hour and third half hour, based on a scale from 1 to 5, where 1 means Very low and 5 means Very high. Other scale parameters can be seen in graph plotted below. For the below graph, "Above Average" & "Very High" attention level can be grouped under same branch and termed as High attention group and rest of the three scales i.e., "Very Low", "Below Average" & "Average" can be grouped under one and termed as Low attention group.









Above we have 4 different plots showing how attentiveness varies along with the progression of a morning class through three different stages of half hours. We can conclude that –

- In the first half hour of morning class, majority of students fall under “Very High” attention category – 52%, which falls progressively as the class goes on
- There isn’t much change in the proportion of students falling in “Average” or below attention category as the class commences

- For the high attention group (as defined above), the proportion of students remain similar for the first hour of the class, but it falls drastically in the last half hour of the class
- For “Above Average” attention category, it increases as the class moves from 1st half hour to 2nd half hour but then it decreases towards the class end
- For any half hour of the class (amongst the three it has been divided into), majority number of students fall under either “Above Average” or “Very High” attention category

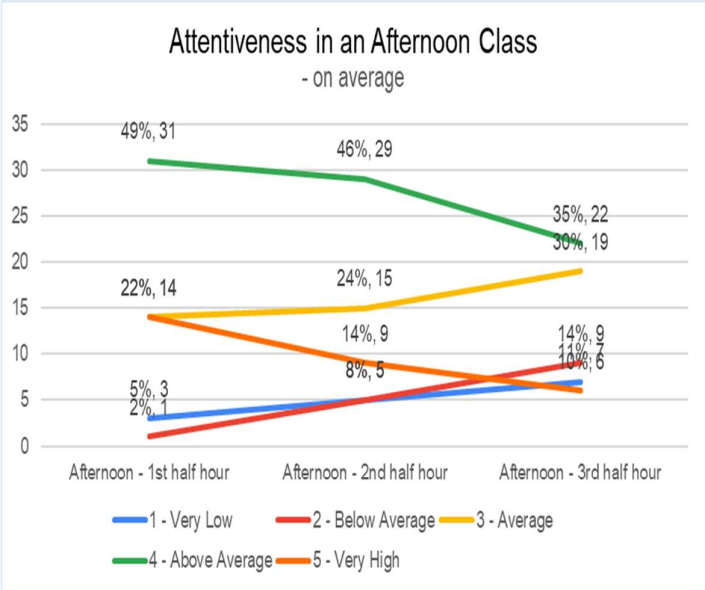
With the help of above 4 insights, we can fairly conclude that morning time is quite better for classes as large proportion of students are able to put their best attention to use during the class.

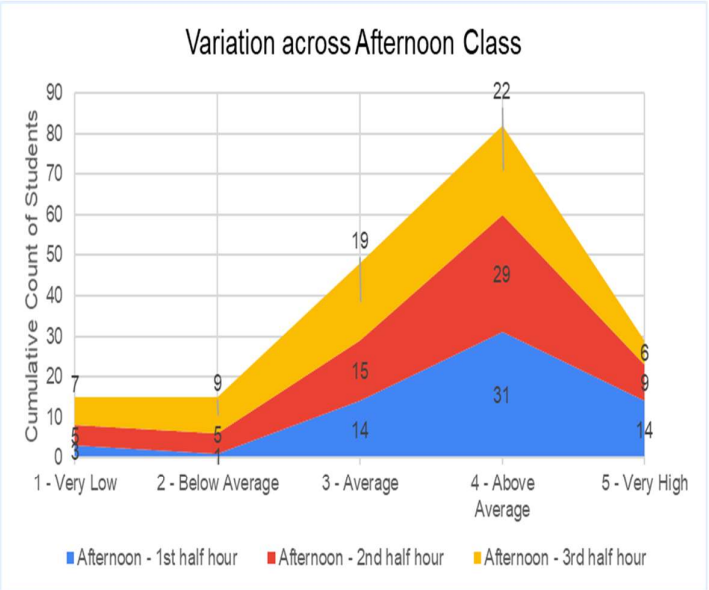
b) Afternoon Class

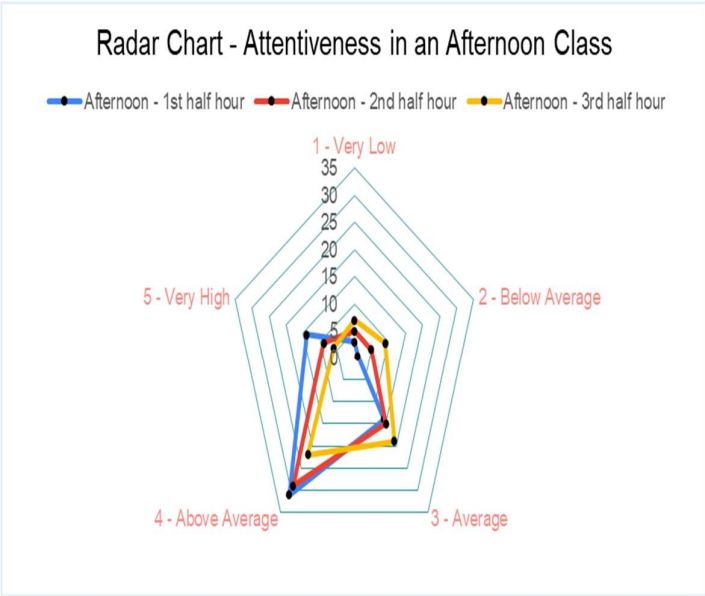
The responses collected and expressed in this section is like the previous section (morning class). Please refer to previous on morning class data type description to understand better. Below is the tabular representation of the responses –

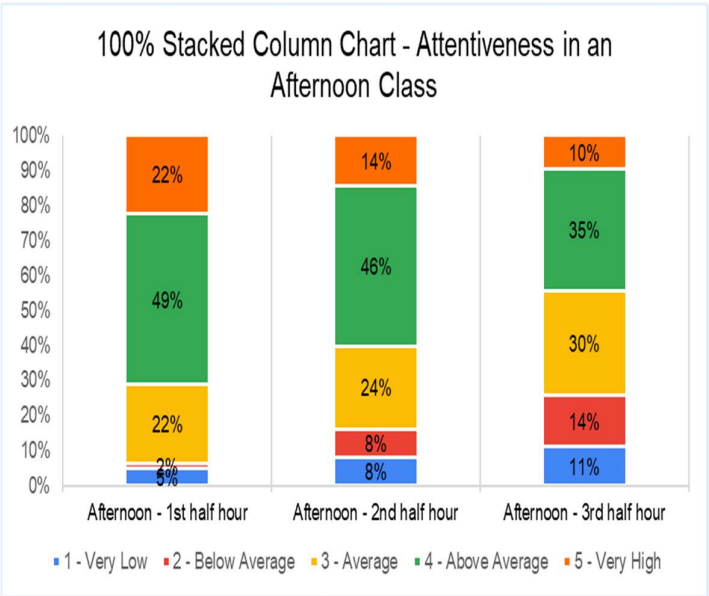
Instances	1 - Very Low	2 - Below Average	3 - Average	4 - Above Average	5 - Very High	Total
Afternoon - 1st half hour	3	1	14	31	14	63
Afternoon - 2nd half hour	5	5	15	29	9	63
Afternoon - 3rd half hour	7	9	19	22	6	63

Shown below is the graphical representation of the responses received under various categories pertaining to attentiveness scale in an afternoon class.









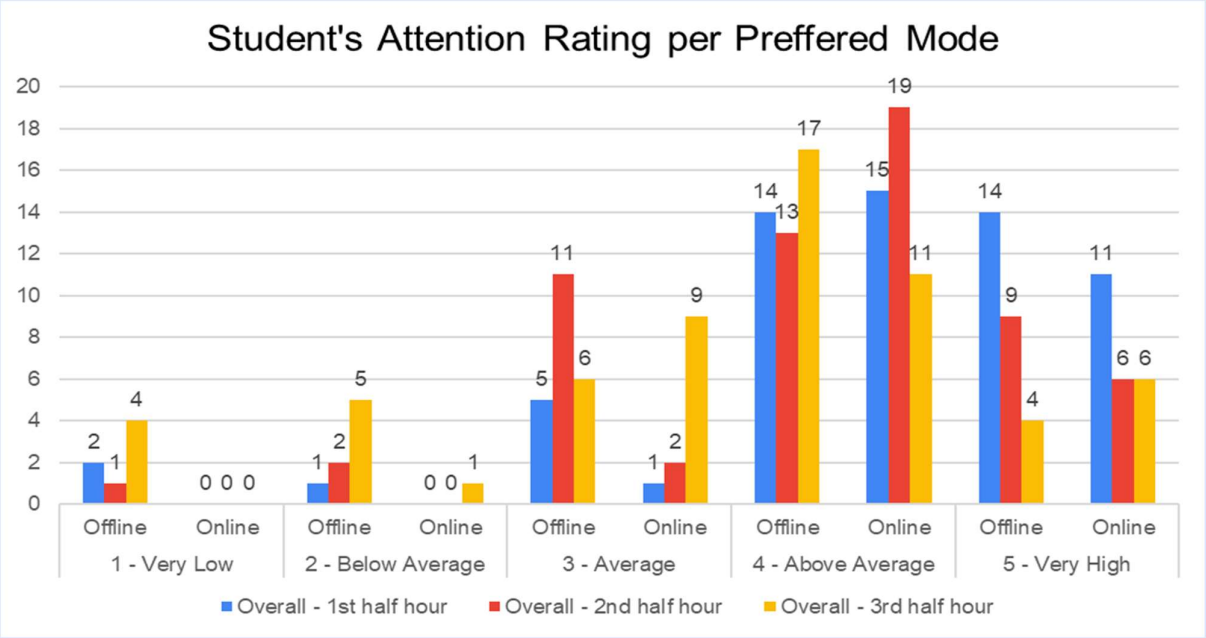
With the help of above graphical representations, we can conclude that –

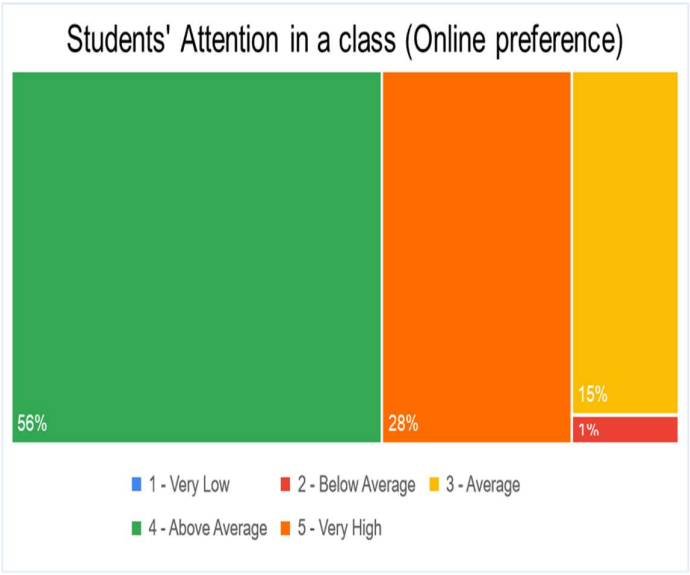
- Proportion of students with “Above Average” or “Very High” attention level keeps on decreasing as the class goes on
- Parallelly, the proportion of students with “Average” attention level or below tend to increase as the class goes on

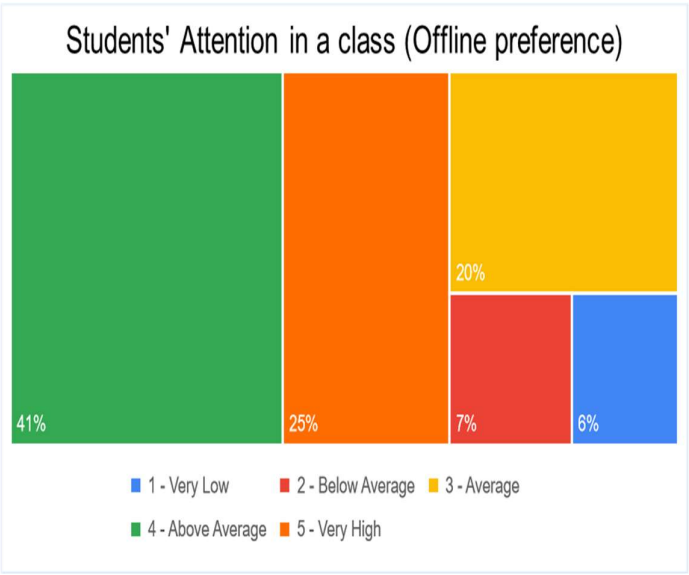
The above 2 pointers help us conclude that, for an afternoon session, as the class goes on, the attention level of the student in the class continues decreasing progressively which is quite different from the trend shown in the morning class as there the proportion of students falling under High attention group (very high &

above average combined) remains same for the first hour of the class but then falls drastically towards the end.

c) Attention level w.r.t preferred mode of learning – Overall class



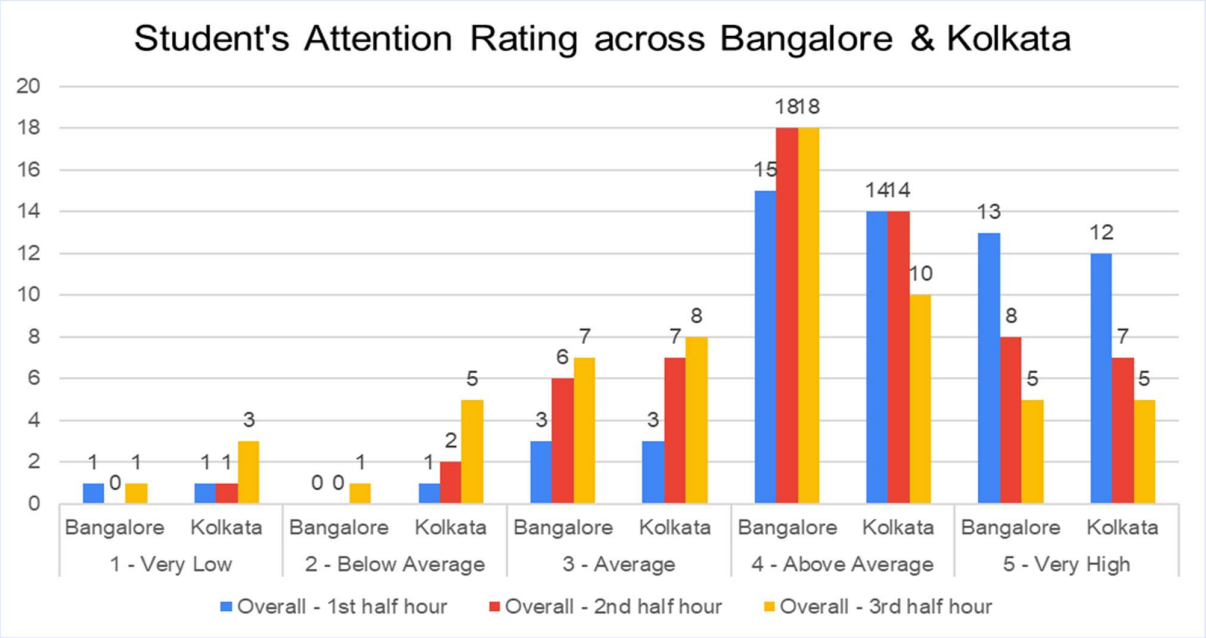


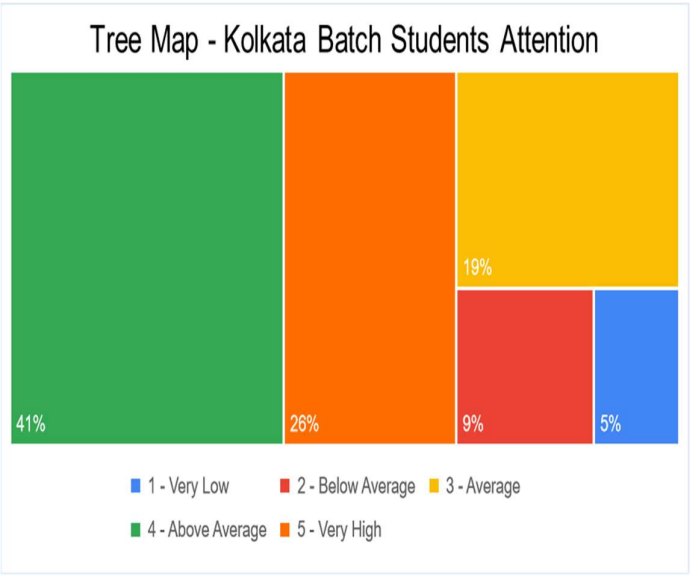


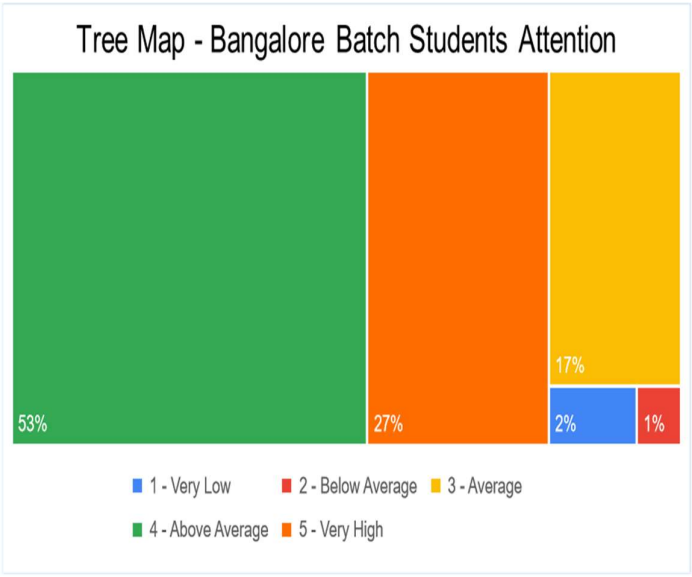
The above graphs help us understand the distribution of student's attention in a class, on overall basis (morning & afternoon combined) across their preferred mode of learning i.e., online, or offline. Inferences we can draw from above graph are –

- Amongst the students falling under Low attention group (very low, below average, and average combined), majority of students prefer for offline mode
- For “above average” and “very high” attention group, distribution of number of students between offline and online mode is pretty much the same
- For students preferring for online mode, no one falls under “very low” category
- There is approximately double the proportion of students in offline preferring mode who comes under “average” and below category compared to online mode

d) Attention level w.r.t campus location – Overall class



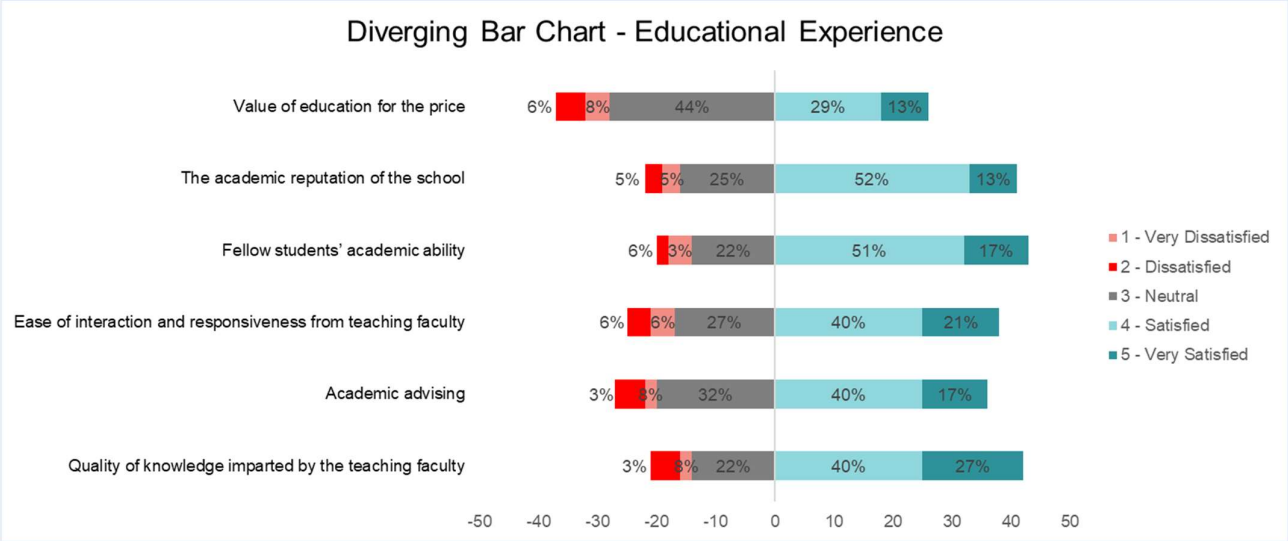


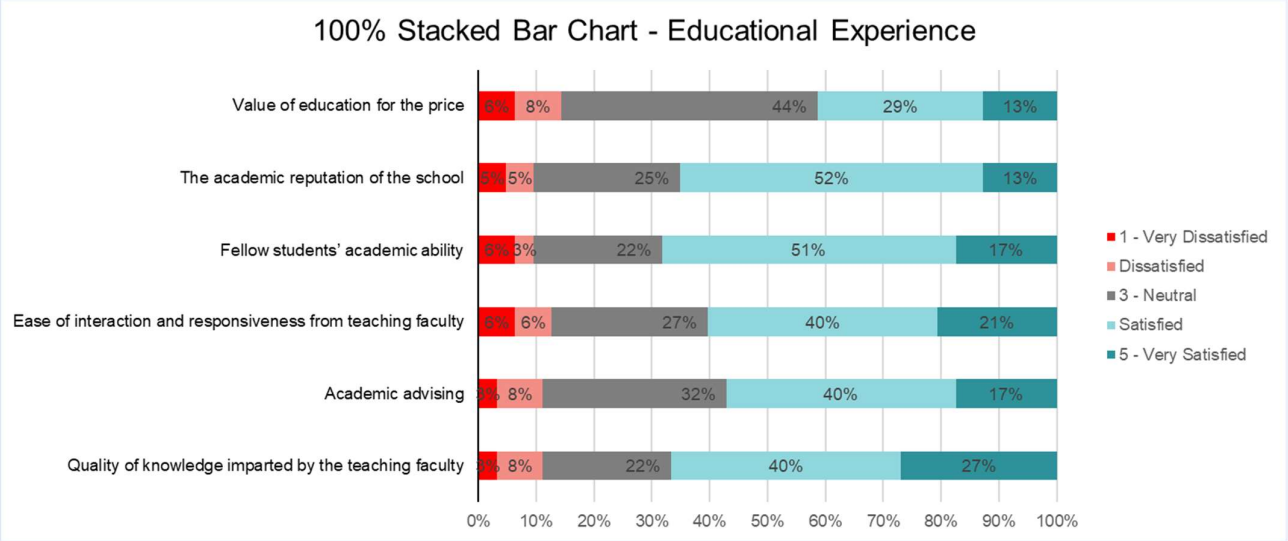


From the help of above charts, we can conclude that –

- Bangalore campus shows lesser number of students falling under “Average” category or lower, compared to Kolkata campus (33% for Kolkata campus compared to 20% for Bangalore). This shows that on overall basis, students from Bangalore campus tend to be more focused or attentive in class
- Parallelly, we can draw that student falling under “Above average” or “very high” category, tend to be more in number from Bangalore campus compared to Kolkata campus

7) Educational Experience of Students





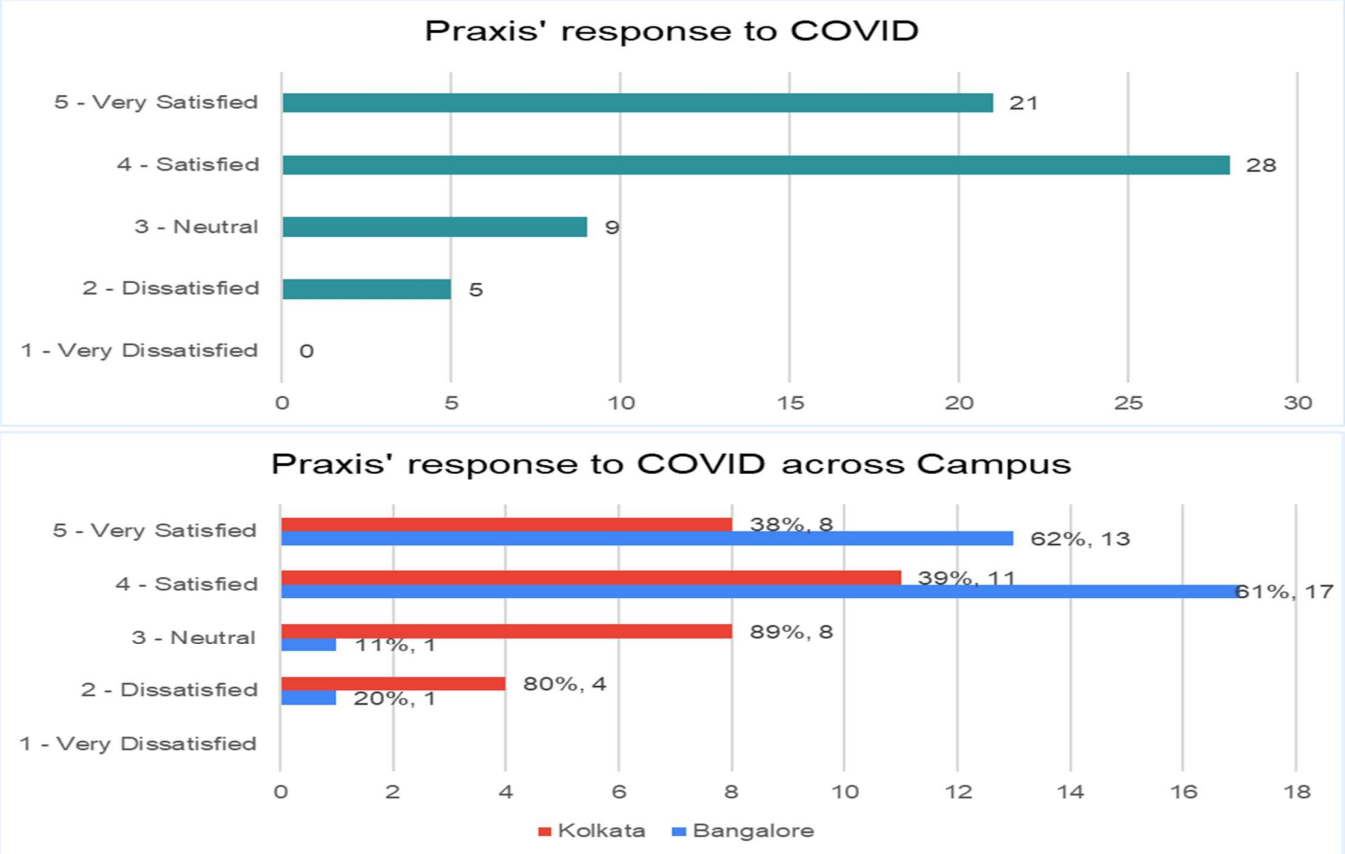
Within Educational experience of a student, the following parameters are evaluated –

1. Value of education for the price
2. The academic reputation of the school
3. Fellow students' academic ability
4. Ease of interaction and responsiveness from teaching faculty
5. Academic advising
6. Quality of knowledge imparted by the teaching faculty

Taking inferences with the help of above visuals, we can say that –

- For all 6 parameters combined on overall basis, more than half of number of respondents have been either “Satisfied” or “Highly Satisfied” with the educational experience
- For value of education of price, there seems to be many students (58%), who are either “Very Dissatisfied”, “Dissatisfied” or “Neutral”
- Quite a number of students are happy with fellow student’s academic capability (68%) and it shows that Praxis has managed to attract quite talented students for their data science batch
- Similarly, many students (67%) are happy with the quality of knowledge provided by the faculty which in turn is in concurrence with the proportion of students happy with the curriculum (78%), faculty (51%) and pedagogy (57%)

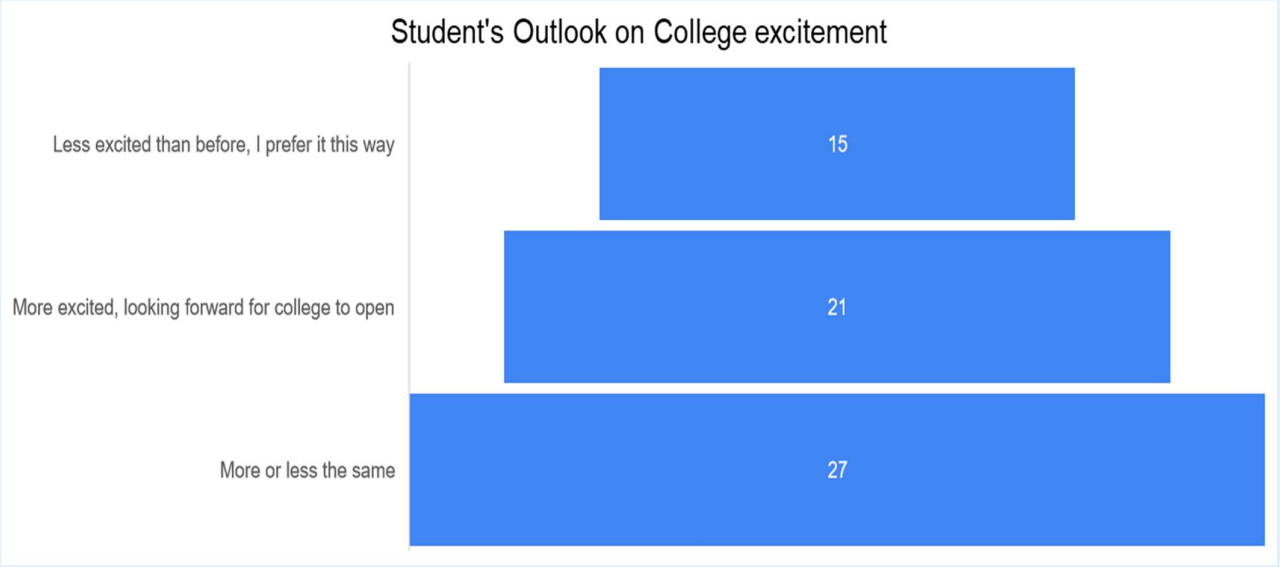
8) Praxis’ response to COVID

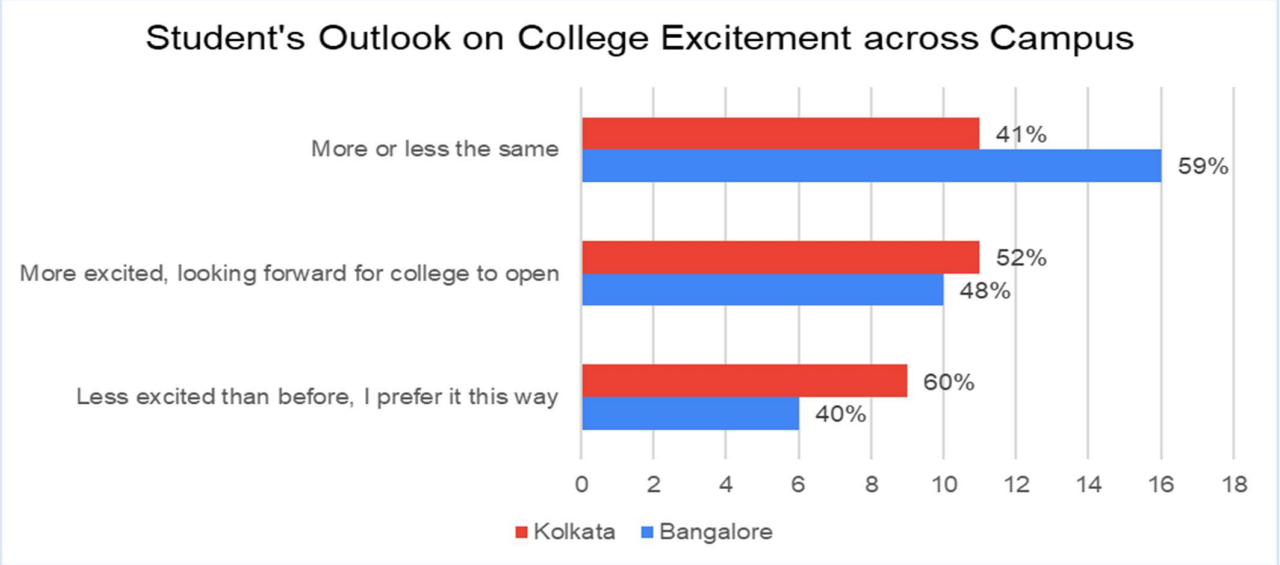


Above charts shows the distribution of satisfaction level of students from Praxis' response to COVID on their own educational front on overall basis and across both campuses. Inferences –

- Overall, quite a significant number of students are either "Satisfied" or "Very Satisfied" with Praxis' response to COVID, and the number stands at 78%
- When segmenting the responses across both campuses, we see that –
    - Quite a significant number of students (above 80%) are from Kolkata campus who are either "Dissatisfied" or "Neutral" with Praxis' response to COVID
    - Parallelly, maximum number of students who are either "Very Satisfied" or "Satisfied" with Praxis' response to COVID are from Bangalore campus

9) Student's Outlook on Back-to-College (B2C) Excitement





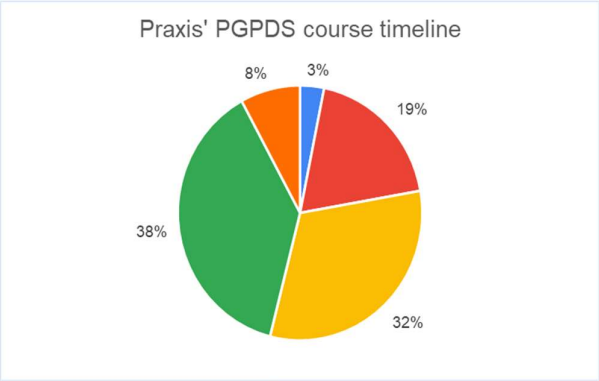
Above we have graphically represented the responses received from students regarding their excitement level for back-to-college scenario and the same has been plotted across campus as well. Inferences –

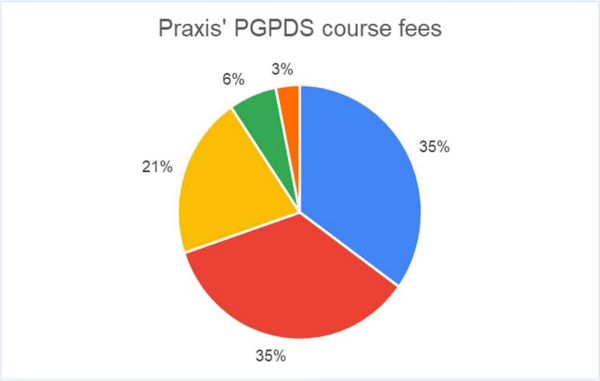
- Quite a significant number of students – 67%, are not excited with back to college scenario. Interestingly, this is in contrast with the proportion of students preferring for offline mode of learning, which stands at 57%
- For Kolkata campus, only 35% of students are excited about college to open compared to Bangalore campus where the number stands at 31%

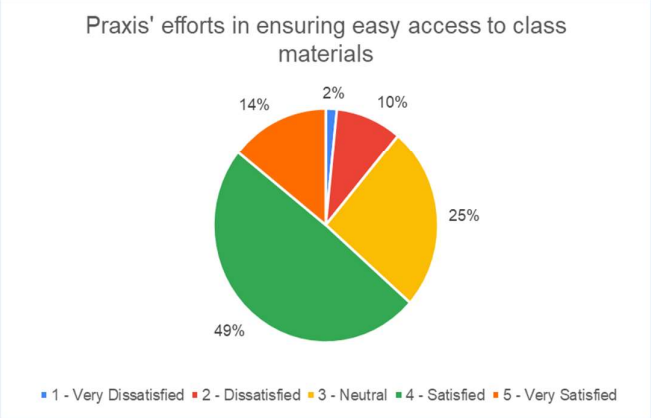
- For students falling under “less excited” or “more or less the same” category, the proportion is similar in Bangalore (69%) & Kolkata (65%) campus. Interestingly, Kolkata campus has better infrastructure compared to Bangalore and houses a proper college campus

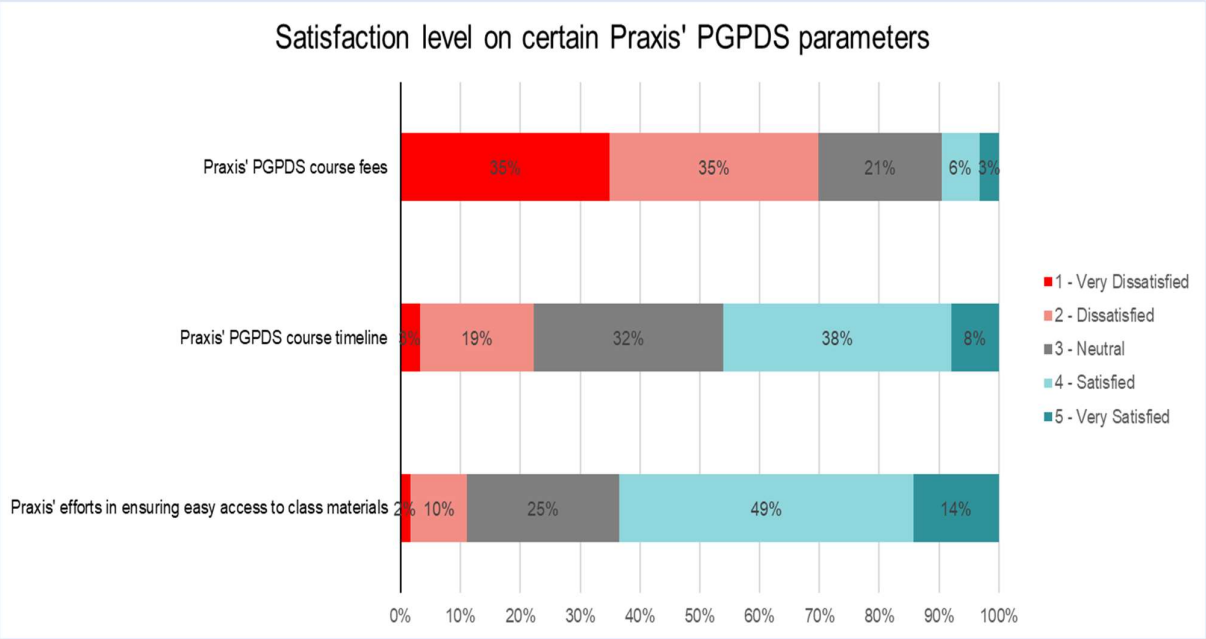
10) Praxis’ PGPDS course parameters

The parameters covered under this section are – Course fees, Course timeline, Praxis efforts in ensuring easy access to class materials i.e., recordings, slides etc.









Above we have plotted small multiple pie charts & 100% stacked bar chart to show satisfaction level of students on different course parameters which are stated above. The inferences we can draw here are –

- More than half of students – 63%, are either “Satisfied” or “Very Satisfied” with Praxis’ efforts in ensuring easy access to class materials
- There is a mixed response for the course timeline parameter from the students as approximately 48% students are either “Very Satisfied” or “Satisfied” and 32% students are “Neutral”
- For the course fees parameter, majority of students’ opinion – 70%, has been that they are either “Very dissatisfied” or “Dissatisfied” with course fees set by Praxis business school for their PGPDS course

Conclusion

With the help of insights generated in our report, we see that how exploratory data analysis can be essential in generating such insights & inferences which are not visible from a bird's eye view. One indeed needs to do a deep dive into the data to extract meaningful information and then work on the insights gained to build a more robust and holistic system of use.

For our use case here, we used the basics of statistics-I course to generate various charts and investigated relationship of different variables and the spread of a certain variable across different campus location and batches. The EDA here thus, gives us notions and information on how Praxis Business School can take future steps for the betterment and overall upgradability of its PGPDS course.