Improving Admissions Count and Student Performance through Data Analysis

Final Report for the BDM Capstone Project

Submitted by

Name: Krishna Varshney

Roll number: 22F3000605

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary and Title

R G Institute of IT & Engg. is one of the leading educational institute located in Shivaji Nagar, Jhansi. This institute runs under an organization named SSSGM Society. It has been in the business for around 14 years and has excelled in providing top notch education at cheaper fees to students of various backgrounds. It has successfully created a reputed position in the educational field and have attracted the attention of students not only from its own city but also from several nearby cities as well. R G Institute of IT & Engg. consists of modern facilities, latest technologies, well-equipped labs, experienced faculties and accreditation from various universities and government organizations.

Even after having such a good business model, the institute has faced some challenges in the past year. The institute aims to increase the number of enrollments in its various courses by around 50%. Also, the current passing rate of students is 76%, the institute aims to increase the performance of students to 85%. This project provides a data driven analytical method to address these challenges and provide effective solutions.

This is the final report of the project and it covers all the aspects needed for the objectives. The first part of this report contains the detailed explanation of the analysis, which includes the journey of how I approached the project. It contains the beginning steps, which are, problem understanding, data gathering, data cleaning and preprocessing. It will also explain the analysis approach I used for my project.

The second part of the report is the detailed analysis which is divided into two sections based on the two objectives. First section is Analysis on Students Enrollments Data and second section is on Students Performance Data. This part of the report contains all the important graphs along with the insights and proper explanation of observing the trends.

After doing a proper analysis, the results were shown to the respective authorities and discussion was done on it. The last part of report highlights the major keypoints about what can be the reasons that the institute is facing the challenges, that were discussed with the institute's staff. It also contains the easily implementable and effective recommendations that the institute can follow. By following this data analytics approach the institute would be able to achieve their expected objectives to a very great extent.

2 Detailed Explanation of Analysis Process/Method

Gathering the data and information

- The first step I followed was to understand about the business, their model, processes and other domain aspects. For this I visited the institute and interacted with the owner, institute's staff and some students.
- After clearly understanding the institute's objectives, I figured out what data I need for the project and collected the same. I collected institute's past year enrollments data and students' performance data. Along with this, I also gathered the courses information that the institute offer.

Data Cleaning and Preprocessing

- After gathering the relevant data, I worked on the preprocessing part. The data provided was in the form of excel tables. I got the course-wise data of students' enrollments and performance data was available for two courses only (as the institute intended to improve their result for those two courses only, i.e., CCC and 'O' Level).
- **Preprocessing Steps:**
  i. First, I removed the columns which were of institute's personal use (example- fee payment details, certificate issue details etc.), as they were non-significant for my analysis. The data did not have any missing values and outliers as such.
  ii. For students' enrollment data, I merged all the different course-wise tables into one, and created a new column to fetch student's age from its date of birth. For this operation, I used formula
  =INT(YEARFRAC(F2,TODAY(), 1)).
  Also, there is a field name Address which contained the residential address of the student. From the analysis point of view, the whole address was not significant, however the city of the student was needed for geographical analysis, so I extracted city from the address using the Text to Columns feature of Excel.
  iii. For students' performance data, I had two different tables to work with. The first table, which contain results of the student had columns- Roll No., and Grade of student and another table contain the detail of students. Both tables had Roll No. as a common column which I used to merged the tables together with the help of VLOOKUP function using formula:
  =VLOOKUP(A1, Sheet2!A:B, 2, FALSE).

Also, I felt the need to convert the grades (A, B, C, D, F, and Absent) into three categories i.e., Pass, Fail and Absent. In order to achieve this, I used the following formula:

=IF(OR(L2="S", L2="A", L2="B", L2="C", L2="D"), "Pass", IF(L2="F", "Fail", IF(L2="ABS", "Absent", "Invalid Grade")))

iv. After the following preparations, data was ready to be analyzed and gain insights. For insights I have used Pivot Table feature of Excel, and NumPy, Pandas libraries of Python.

Samples of Final Data Used

Sr No	Programme Name	Type	Category	Qualification	Duration
1	CCC	Certificate	Govt. Course	NIL	3 Months
2	Tally ERP 9	Certificate	Institute	NIL	5 Months
3	Desktop Publishing	Certificate	Govt. Project	10th Pass	5 Months
4	PMKVY	Certificate	Govt. Project	10th Pass	4 Months
5	Web Designing	Certificate	Institute	NIL	5 Months
6	DCA	Diploma	Institute	12th Pass	1 Year
7	O Level	Diploma	Govt. Course	12th Pass	1 Year
8	BCA	UG	University	12th Pass	3 Year
9	MBA	PG	University	BBA Pass	2 Year
10	MCA	PG	University	BCA Pass	2 Year

Figure 1: Course Details Data

#	Candidate Name	Father/Guardian Name	Gender	DOB	Age	Category	City	Course
1	Abhay Dwivedi	Chandra Prakash Dwivedi	Male	20-11-1994	29	Gen	Jhansi	MCA
2	Abhishek Pandey	Avinash Kumar	Male	29/11/1995	28	General	Jhansi	MBA
3	Abhishek Samadhiya	Om Prakash Samadhiya	Male	27/02/1993	31	OBC	Mahoba	MCA
4	Abhishek Shrivastava	Ramesh Shrivastava	Male	13/09/2000	23	General	Jhansi	Web Designing
5	Abhishek Yadav	Brijesh Yadav	Male	04/05/2005	19	OBC	Mahoba	Tally ERP 9
6	Aditya Pandey	Rajendra Pandey	Male	27/10/1999	24	General	Jhansi	Web Designing
7	Ajay Sachan	Vinod Sachan	Male	03/11/1996	27	OBC	Lalitpur	MCA
8	Akanksha Rai	Rajendra Kumar Rai	Female	02-10-2000	23	OBC	Jhansi	MBA
9	Akhilesh Singh	Radhe Shyam Singh	Male	19/08/1992	31	OBC	Mahoba	MCA
10	Alok Sen	Rajkumar Sen	Male	25/06/2000	23	OBC	Jhansi	Web Designing

Figure 2: Sample of Students' Enrollments Data

#	Roll No.	Candidate Name	Father/Guardian Name	Gender	Date of Birth	Age	Category	City	Course	Month	Result	Status
1	GO2301012723	Aditya Pratap Singh	Hari Singh Rajawat	M	10-01-2004	20	General	Jhansi	CCC	Jan	B	Pass
2	GO2301012927	Jagbhan	Asau	M	15-06-1997	26	OBC	Jhansi	CCC	Jan	C	Pass
3	GO2301012396	Abhishek Patel	Ravindra Patel	M	15-06-2005	18	General	Jhansi	CCC	Jan	ABS	Absent
4	GO2301012779	Shivam Rajpoot	Ashok Rajpoot	M	14-08-2002	21	OBC	Jhansi	CCC	Jan	A	Pass
5	GO2301012938	Nripat Singh	Ambika Prasad	M	27-10-2003	20	OBC	Jhansi	CCC	Jan	ABS	Absent
6	GO2301012778	Krishna Bhanupratap Khushwaha	Bhanupratap Kushwaha	M	27-08-2002	21	OBC	Lalitpur	CCC	Jan	C	Pass
7	GO2301012476	Rishabh Patel	Munnalal Patel	M	04-09-2003	20	OBC	Lalitpur	CCC	Jan	C	Pass
8	GO2301012413	Dhirendra Chaurasiya	Prabhu Dayal Chaurasiya	M	25-11-1984	39	OBC	Jhansi	CCC	Jan	A	Pass
9	GO2301012580	Pradeep Chaurasiya	Prabhu Dayal Chaurasiya	M	25-03-1986	38	OBC	Lalitpur	CCC	Jan	C	Pass
10	GO2301012965	Surendra Kumar	Suresh Kumar	M	20-06-1996	27	SC	Jhansi	CCC	Jan	A	Pass

Figure 3: Sample of Students' Performance Data

Initial Analysis and Feedback

- For initials, I did some basic analysis on the data, which included some descriptive statistics, basic trends and patterns in the data. It was done in order to get a better understanding of the business and gaining some important patterns in the data.
- I made a course pareto chart, to get the visual representation of distribution of enrollments across various courses offered by the institute.
- In addition to that, this analysis also included the distribution based on city, age groups, and gender.
- Also, I conducted the analysis on overall performance of the students in the past year.
- After the analysis was done, the findings along with the graphs were shown to the authorities. A proper discussion was done with them, they told which factors has the more scope for improvement and where they would want to lay emphasis on, and I got a direction for further advanced analysis.

Final Analysis and Suggestions

- After getting a clarity on how to proceed, I went for my final analysis. The final analysis comprises of many advanced analyses. This included correlations between various factors and their nature.
- These analyses show how different factors are affecting the enrollments distribution and performance grades of the students.
- For this, I first converted various ages of students into different bins of 8, i.e., into 4 groups- 13-20, 21-28, 29-36, 37-44. This helped in simplifying the analysis which would have otherwise have been quite bulky.

- I have divided my analysis into two sections, separately for the two objectives the institute have. First section comprises of Analysis for Students Enrollments Data and the second section has the Analysis for Students Performance Data.
- For the analysis of students' enrollments distribution, I did distributive analysis, category vs city analysis, age vs course analysis, course vs city analysis, and age vs city analysis.
- For the analysis of students' performance data, I did grade distribution analysis, monthly trend analysis, grade vs month analysis, grade vs age analysis, and grade vs city analysis.
- After performing all the analysis, the graphs along with observations shown to the authorities. We discussed about the major keypoints the institute can focus on and then possible recommendations on the objective was provided.

Tools Used

- Pivot tables to analyze student enrollment data and identify trends and patterns.
- Graphs like bar graphs, pie charts, scatter plots, histograms, stacked charts etc. for data visualizations.
- Excel functions and formulas for calculation and analysis, such as averages or rates and also for data cleaning and preprocessing steps.

Challenges faced during the project

i. The biggest challenge was with the time of admission. Different courses and project in the institute has different timeline for admission, which made it much difficult to observe any descriptive statistics, mean, deviation etc. in terms of month wise enrollments. Also, getting any time trends was difficult, as there were occurrences where there were very less enrollments (40-50) in a month to around 100-200 enrollments in another.

ii. Institute didn't have any record keeping system before the previous year; therefore, insights comparison was done only by the insights told by the staff verbally which were not accurate but rather an approximation.

iii. The data given by the institute was mostly categorical, so it became quite difficult to understand and get insights about the correlation between the various factors affecting both the enrollments and performance of students.

3 Results and Findings

Analysis on Student Enrollments Data

A. Distributive Analysis

The purpose of this analysis was to get an initial understanding about various factors effecting course enrollments. This included enrollments distribution city-wise, age-wise, gender-wise, course-wise.

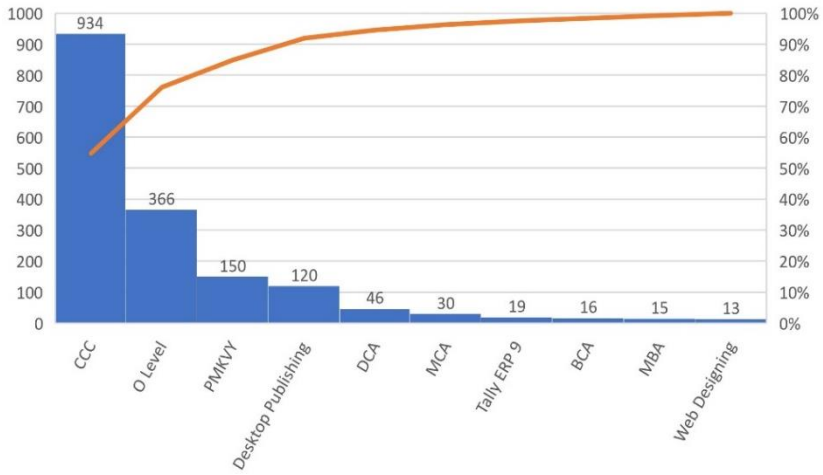

Figure 4: Students' Enrollments Course-wise Pareto Chart

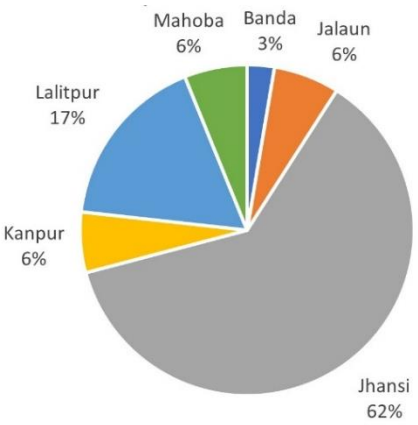

Figure 5: City-wise Enrollments Distribution

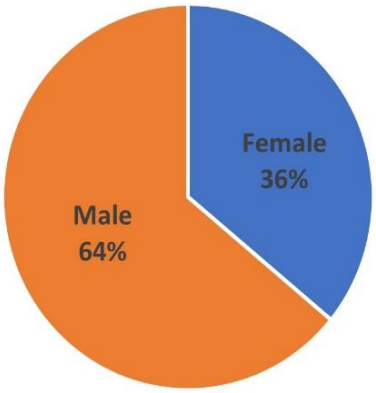

Figure 6: Gender-wise Enrollments Distribution

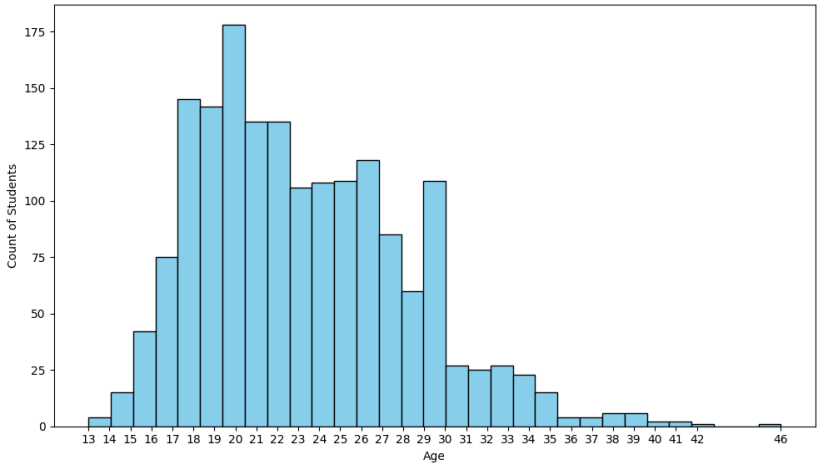

Figure 7: Age-wise Enrollments Distribution

Observations from the graph:

- Students are majorly interested in enrolling in CCC and 'O' Level Courses, which contributes to around 80% of all enrollments.
- Institute gets 62% of enrollments from its own city, second most enrollments are from Lalitpur (17%) as the city is closest to Jhansi among others.
- Proportion of female students is very less (36%) as compared to boys (64%).
- Students varying from age of 13 to 46 studies in the institute. However, most students lie in the age group of 18-24.

B. Category vs City Analysis

On talking with the staff, I got to know that category of students (General, OBC, SC, ST) is one of the important factors for the enrollments. This was because the government of Uttar Pradesh keeps introducing several education projects for category specific students, (for example- Free O Level course for OBC category). So, I found the need to understand the city wise category trends in past enrollments.

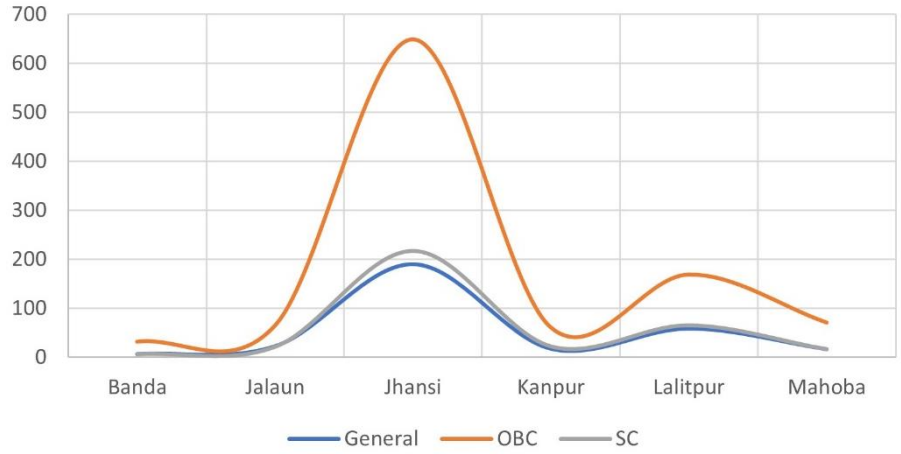

Figure 8: Category vs City Enrollments Distribution

Observations from the graph with possible explanations:

- The institute mainly has admissions from OBC Category, as there are more training and projects introduced by government for them.
- The distribution of General and SC students is almost similar.
- Also, institute has no ST student enrolled.

C. Age vs Course Analysis

Students of different age tend to do different type of courses; also, various courses have the age eligibility criteria. So, I found a need to understand the age proportion of students for different course, hence, I used 100% stacked chart as it helps in understanding and comparing the percentage as a whole.

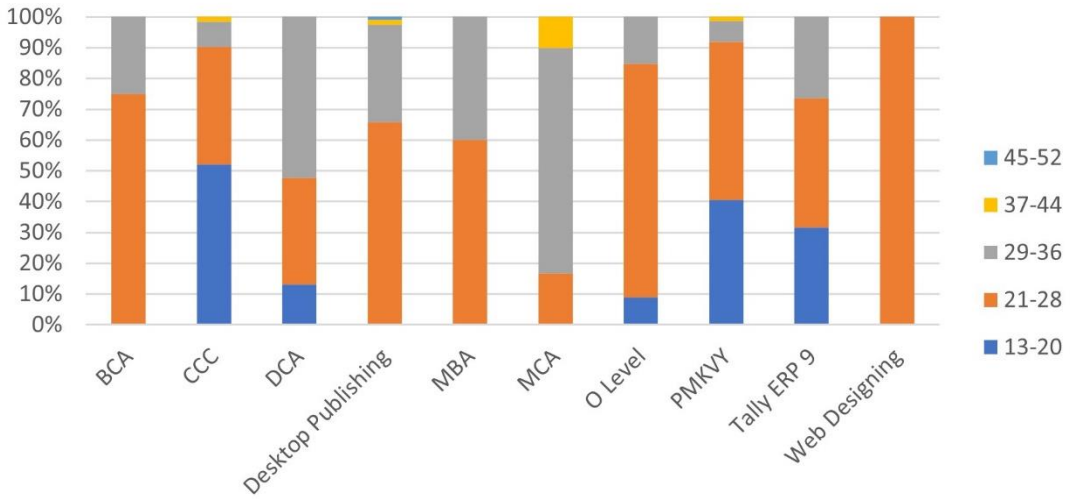

Figure 9: Age vs Course Enrollments Distribution

Observations from the graph with possible explanations:

- The students of age group 13-20 mostly enroll for CCC, PMKVY, Tally, as the duration of these courses is short, so students can do them as a supplement with their studies.
- Students of age group 29-36 are mostly enrolled for Master's degree.
- Student of age group 21-28 are enrolled across all courses offered by the institute.
- Institute also have a small proportion of admissions in age group of 37-44 and 45-52, however, they also take the short-term courses for their interest in studies.

D. Course vs City Analysis

The institute witnessed a hike in students' engagement from other cities than years before, therefore, it was important to analyze the pattern in this increasing interest of students from other cities. This analysis was done to detect the relationship existing between enrollments in different courses with respect to the cities. A 100% stacked chart is used for this analysis, as the purpose here is not to analyze the number of enrollments, I intend to analyze the pattern in the various courses' percentage of enrollments across different cities.

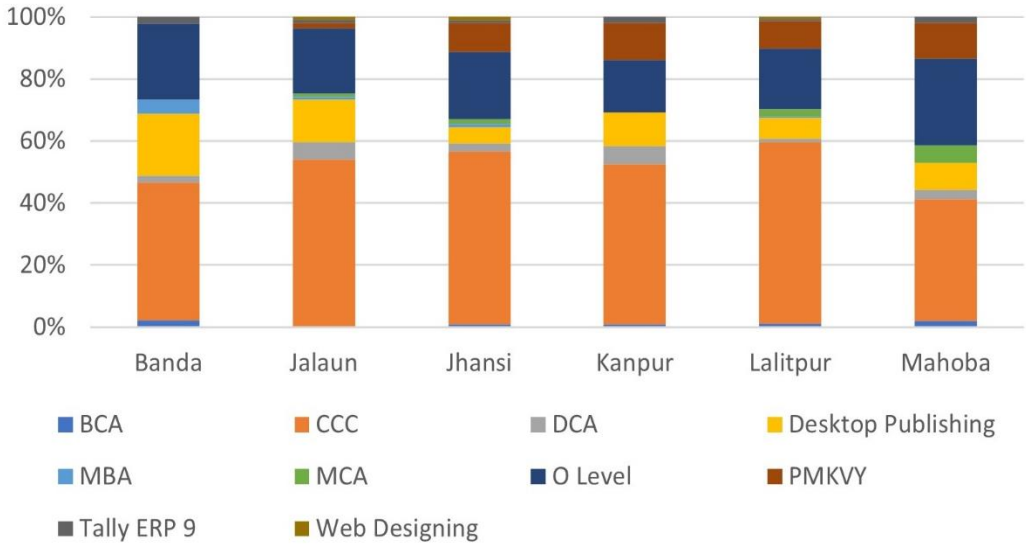

Figure 10: Course vs City Enrollments Distribution

Observations from the graph with possible explanations:

- The students from other cities are showing major interests in CCC and O Level. This is because these two are government courses which are necessary to be done if one wants to get job in various government sectors.
- Also, a good interest is shown in Desktop Publishing and PMKVY. As these are government projects, the training is free of cost and students also get placed after successful completion of the course.

E. Age vs City Analysis

The purpose of this analysis was to understand whether there exists any correlation for the enrollments between the student's age group and city to which the student belongs to.

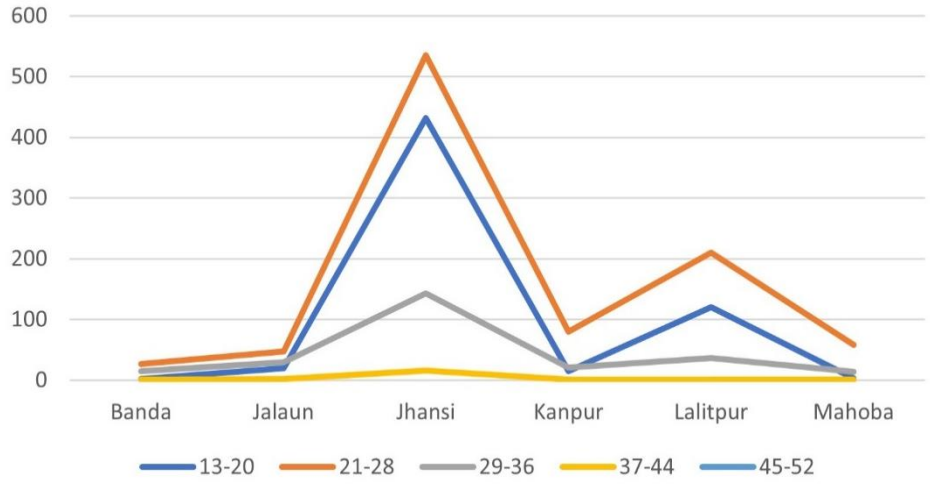

Figure 11: Age vs City Enrollments Distribution

Observations from the graph with possible explanations:

- For the age group of 13-20, it has been observed that institute have more enrollments from their own city, and very less from other cities. Apart from the local, institute also get enrollments in this age group from Lalitpur, as Lalitpur is comparatively near to Jhansi.
- The students of age group 21-28 and 29-36 are more tending to come to the institute for their enrollments.

Analysis on Students' Performance Data

- The institute provided the students scores data for two courses CCC and O Level, as they intend to optimize the performance of students for these two courses.
- Last year performance of institute was, 76% students got passed while 18% failed and 6% did not appear in the final examination, in aggregate for both CCC and O Level.
- The grades are in 6 categories- S, A, B, C, D, Fail, and Absent.
- The forthcoming analysis shows the various factors and their relationship with the grades of students.
- Some of the below analysis has been done as aggregate for both CCC and O Level, as they are both are very similar courses. While, some of the analysis has been done separately for both courses also.

A. Grade Distribution Analysis

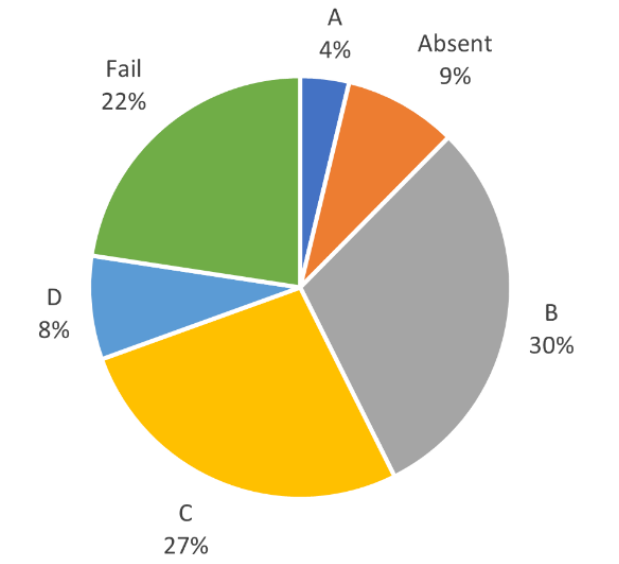

Figure 12: CCC Grade Distribution

Observations from the graph:

- Most students get B and C grades. Both B and C grades share equal proportion of the result, i.e., 30% and 27% respectively.
- Students getting A grades are very low (4%).

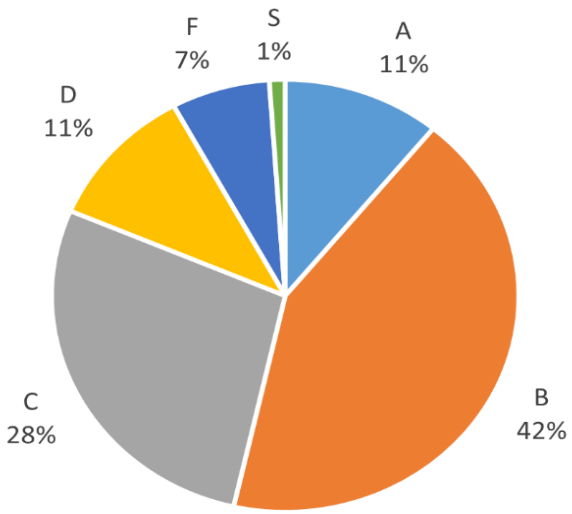

Figure 13: 'O' Level Grade Distribution

Observations from the graph:

- The average student grade is B.
- Proportion of S and A grades is very low.
- No proportion of students is absent in the examination as this course fees is high; students do not afford to miss exams.

B. Monthly Enrollments Trend Analysis

The purpose of this analysis is to see the pattern in enrollments month-wise. CCC admissions occurs every month, however 'O' level admissions are done only in January and July month. Therefore, this analysis has only been done for CCC.

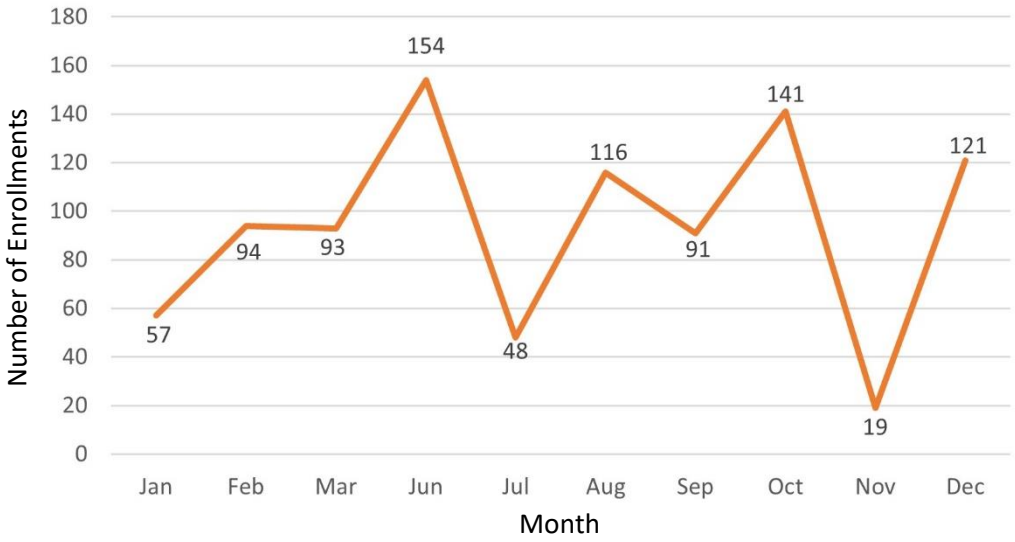

Figure 14: CCC Monthly Enrollments

Observations from the graph with possible explanation:

- Average enrollments in the course are 78.
- The greatest number of admissions happen in the month June, that is when most students are passed out of their current academic session and advance to further studies.
- January, July and November have the least number of admissions. CCC is a course which is done in side with some other main course, and as there are exams in universities and colleges in these three months, therefore students enroll less in these months.

C. Grade vs Month Analysis

After understanding the monthly enrollments of CCC, I felt the need to get the patterns in the grades of students' month-wise. For this, I used a stacked bar chart, which helps in understanding the enrollments distribution as well the proportion of grades for each month.

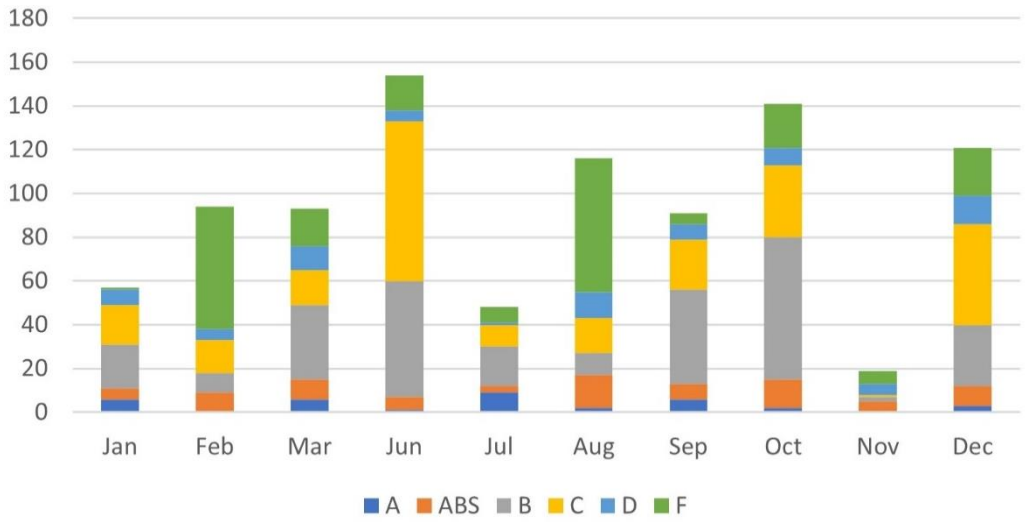

Figure 15: CCC Grade vs Month Distribution

Observations from the graph with possible explanation:

- February, August, and December have a large percentage of students getting failed. The reason for this is that students usually have other exams, a month a prior to these months, making it difficult for them to study for the exam.

D. Grade vs Age Analysis

This analysis is done by using both CCC and O Level performance data. The purpose of this analysis is to understand the correlation between grades and different age-groups.

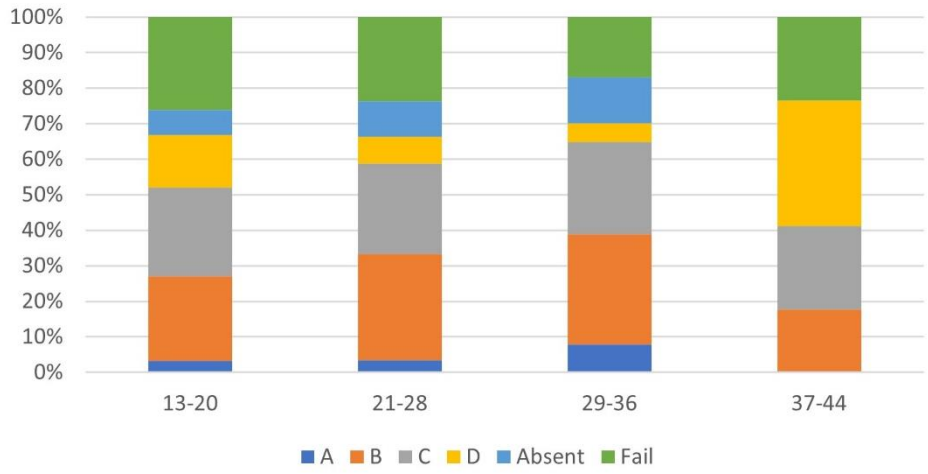

Figure 16: Grade vs Age Distribution

Observations from the graph with possible explanation:

- The proportion of students getting failed or lower grades are more in the age group of 13-20 and 37-44. In the age group of 37-44, no student got A grade, also a large percentage of students got D grade or fail. For the age group of 13-20, there is a noticeable percentage of students getting failed.

E. Grade vs City Analysis

This analysis is done by using both CCC and O Level performance data. The purpose of this analysis is to understand the correlation between grades and different age-groups.

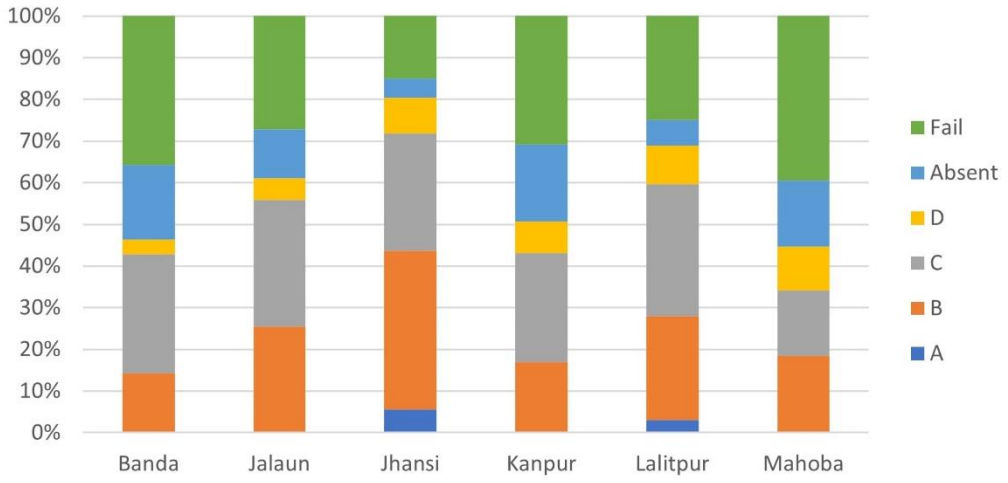

Figure 17: Grade vs City Distribution

Observations from the graph with possible explanation:

- The proportion of students getting failed or absent are more from other cities than the local.
- Banda, Kanpur and Mahoba have more percentage of students getting failed or absent, as these cities are located far away from Jhansi.

4 Interpretation of Results and Recommendation

Major Key Points or Interpretations of the Analysis

- ❖ Institute gets major admissions from two courses CCC and O Level contributing around 80% of total enrollments. This is because of the value of these certificates in government sectors jobs.
- ❖ The institute has a quite low proportion of female students compared to male students which the institute should lay their focus on.
- ❖ The students of age group 18-24 are more present as they pursue further education immediately after their schools.
- ❖ Students of lower age group tend to take short term courses however students with middle and upper age group tend to take more of degree courses.
- ❖ The institute has remarkable amount of admission in OBC category, due to the government projects offering free education and placement. However, the admissions for General, SC and ST are not much.

- The government projects and short-term courses are majorly drawing the attention of students from other nearby cities to get enrolled as the projects are free of cost and short-term courses like CCC are necessary for a good government job.
- Students coming from other cities are majorly in the age group of 21-28 or 29-36. The reason of this can be the regular up and down to the city for classes or getting a residential facility in Jhansi itself.
- For CCC, the major enrollments are done in the month of June when students proceed for the further studies. Also, the month of January, July and November has the lowest enrollments because of other exams being held in universities and colleges.
- The percentage of students getting failed or lower grades are in the month of February, August, and December as students could not study properly due to their other exams being held one month prior.
- Students coming for far distant cities have more percentage of failed or absent in the examination, due to residential issue they face.
- Students in the age group of 13-20 and 37-44 are more tend to get bad grades. The possible reason for this can be the adjustability with the new technologies. Students lying within these groups are not much familiar with the technologies used today for studies making it difficult for them.

Recommendations

- **Promote Gender Related Initiatives:**

  Implement targeted initiatives to increase female enrollment by addressing barriers such as societal norms, accessibility issues, and tailored marketing strategies to attract female students.

- **Customized Learning for different Age Groups**

  Develop customized educational programs aligned to the specific needs and preferences of different age groups, ensuring relevant curriculum content.

- **Outreach Programs for Different Categories**

  Implement targeted outreach programs to attract students from categories such as General, SC, and ST, where admissions are comparatively low, offering scholarships or incentives to encourage enrollment.

❖ **Introduce Residential Facilities for Distant Students**

- Introducing residential facilities or provide alternative accommodation options for students coming from far distant cities to address issues related to accommodation and residential challenges.
- Providing transportation facilities within a radius of at least 100 kms from the institute.

❖ **Implement Special Academic Support Programs**

Introduce academic support programs such as tutoring, mentoring, or remedial classes to assist students in improving their academic performance and cover the topics they left out due to their exams, especially during challenging months like February, August, and December.

❖ **Provide Technological Training and Support**

- Offer training programs or workshops to familiarize students in the age groups of 13-20 and 37-44 with modern technologies used for studies, ensuring they are adequately prepared to navigate digital learning environments.
- Also, implement peer mentoring programs where senior students can provide guidance and support to their juniors.

❖ **Enhanced Student Engagement**

- Introducing interactive teaching methods, project-based learning, and real-world applications to increase student engagement and motivation.
- Also, offering extracurricular activities, clubs, and events to promote a sense of community and belonging among students.

❖ **Continuous Monitoring and Evaluation**

- Establish a robust system for monitoring student progress, academic performance, and satisfaction levels to identify areas for improvement and measure the impact of interventions.
- Regularly review and update enrollment and admission policies based on feedback from students, faculty, and stakeholders to ensure inclusivity and equity.

❖ **Infrastructure and Technology Upgrades**

- Invest in modern technology infrastructure, including high-speed internet, multimedia classrooms, and digital learning resources, to enhance the learning experience.
- Ensure accessibility and usability of technology platforms for students from diverse backgrounds and levels of digital literacy.

❖ **Partnerships and Collaborations**

- Forge partnerships with local businesses, industries, and government agencies to create internship opportunities, industry-relevant projects, and job placement programs for students.
- Collaborate with educational institutions, non-profit organizations, and community groups to share resources and best practices in student support and engagement.