# Student Academic Performance Analysis(G11): Semester Long Assignment

Prepared By-



Saksham Chanouria

(524410013)

Ashutosh Dwivedi

(524410016)

Anshul Singh Chauhan

(524110020)

Ritik Agnihotri

(524410011)

# Introduction

Understanding the factors that influence student academic performance is essential for educational institutions aiming to enhance learning outcomes and support student success. This report presents a comprehensive data-driven analysis of undergraduate science students at the Central University of Allahabad, focusing on their academic performance across the first and second year.

This project, titled “Student Academic Performance Analysis” provides an in-depth exploration of academic outcomes by leveraging detailed datasets containing individual scores, subject combinations and examination results and employ a range of statistical tests including correlation analysis, hypothesis testing and comparative assessments- to uncover patterns and relation within academic data, and visualizations through boxplots and regression plots. Tools like Python, Seaborn, Matplotlib and Power BI were integrated to produce both analytical insight to produce both analytical insight and clear visual storytelling.

Using data collected from undergraduate students over two academic years, this analysis examines correlations between theory and practical marks, compares the performance of university-enrolled students versus those from affiliated colleges, and investigates how subject choices such as Physics–Maths–Chemistry (PMC), Physics–Maths–Computer (PMComp), and others affect overall academic success.

This report aims to present these findings in a concise and actionable manner, offering data-driven recommendations that can support academic strategy and student support interventions.and highlight key areas for academic intervention and policy development.

# Data Description

The dataset contains academic records of University of Allahabad BSc 1st Year and 2nd Year students of the 2021-2024 batch, covering performance across multiple subjects. It includes Roll Numbers, Enrollment Numbers, Subject-wise Marks, Practical Scores, Total Score and Final Results (Passed/Failed/Eligible for Supplementary Exams/Absent)

**Time Frame**: Two academic years (2021-2022 and 2022-2023)

**Source**: <https://allduniv.ac.in/files/result/NET_BSc2_2022-23.pdf>

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **Description** | **Data Type** | **Possible Values/Ranges** |
| **RollNo** | Unique identifier for each student | Numeric | Integer values |
| **EnrolNo** | Enrollment Number (Same in entire course) | Text | Alphanumeric string, starts with 'U' for university students(U2112626-U2110241) and other alphabet for college students |
| **Part1Total** | Total marks in Part 1 (First Year) | Numeric | Range: 0 to Maximum Possible Marks(450), includes 'AA' (converted to NaN) |
| **Part2Total** | Total marks in Part 2 (Second Year) | Numeric | Range: 0 to Maximum Possible Marks(450), Includes 'AA' (converted to NaN) |
| **GrandTotal** | Sum of Part1Total and Part2Total | Numeric | Calculated field, sum of Part1Total(First Year) and Part2Total(Second Year). (Range: 0 to 900) |
| **Result** | Final result of the student | Categorical | 'Passed', 'Failed', 'Elig/SE' (Eligible for Supplementary Exam), 'Absent', 'AA' |
| **Subject Columns** | Marks in individual subjects (Physics, Mathematics, Chemistry, etc.) | Numeric | Range: 0 to Maximum Possible Marks for each subject.  For Practical Subjects(Range: 0-50 for Practical and 0-33 for each theory exam)  For Non-Practical Subjects(Range: 0-50 for each Theory exam) |

**(1.1)** **Example of Student Data enrolled in a Practical subject**

RollNo. : 2202495

EnrolNo. : U2112063 (Second Year)

Part1Total : 304

Part2Total : 267

GrandTotal : 571

Result. : Passed

Subject and marks : 1. Physics

(Optics 🡪 27)

(Oscillations, Waves and Electromagnetism🡪 17)

(Atomic and Nuclear Physics🡪 12)

(Practical🡪 36)

2. Mathematics

(Linear Algebra 🡪 19)

(Real Analysis🡪 29)

(Mechanics🡪 25)

3. Computer Science

(Computer Org. and Architecture 🡪 14)

(Computer Networks 🡪 27)

(Data Structure 🡪 25)

(Practical 🡪 36)

**(1.2)** **Example of Student Data enrolled in a Practical subject**

RollNo. : 2202495

EnrolNo. : U2112063 (First Year)

Part1Total : 304

GrandTotal : 304

Result. : Passed

Subject and marks : 1. Physics

(Mechanics and Special Theory of Relativity 🡪 27)

(Thermal Physics🡪 14)

(Electronics🡪 24)

(Practical🡪 36)

2. Mathematics

(Geometry 🡪 40)

(Real Analysis🡪 29)

(Differential Equation🡪 29)

3. Computer Science

(Digital Electronics 🡪 25)

(Computer Electronics 🡪 19)

(C Programming 🡪 29)

(Practical 🡪 32)

# Objectives

The academic performance of students is influenced by a wide range of factors—ranging from the type of institution they attend, to the subjects they choose, and even the presence of practical components in their curriculum. Understanding these influences is crucial not only for evaluating the effectiveness of current educational practices but also for shaping future academic strategies and student guidance mechanisms.

Some questions answered from this exploration includes:

**Q1. Determine if there's a statistically significant association between the institution type (university vs. affiliated college) and student pass/fail status**

First Year

Second Year

The above graphs show the number of passed, failed and Elig/SE students distribution among University and Affiliated colleges. From the above graph, it can be seen that the proportion of passed students in the university is significantly greater than number of passed students in affiliated colleges. Similarly, other results can also be derived from the above graphs.

**Q2. Does the inclusion of practical subjects in a student's curriculum influence their overall academic performance?**

# It is clear from the above graphs that even though the practical exams’ weightage is less(50/150) , still it significantly has more contribution in overall result as compared to theoretical exams.

# For instance, the weightage of practical is about 33%, still in most of the exams, they are contributing nearly 40% of total marks.

# Q3. Is there a correlation between student performance in practical exams and their overall academic performance?

Pearson Correlation Coefficient: 0.9981

📊 P-value: 0.0000

✅ There is a statistically significant relationship between practical and total marks.

📈 Pearson Correlation Coefficient: 0.9998

📊 P-value: 0.0000

✅ There is a statistically significant relationship between practical and total marks.

In this analysis, we have used Karl Pearson’s Correlation coefficient between the practical marks and total marks on both processed data (subject wise) and the actual raw dataset (student - wise). After analysis, we found that in case of subject-wise processed data, the value of correlation coefficient was nearly 0.98 which is showing a strong relation between the two columns. However, in case of raw data (student wise), there was some noise in the graph because of irregularities in marks of student, but still we were able to prove a significant relationship between practical and total marks with value of correlation coefficient to be nearly 0.6.

After cross checking with actual dataset, it was observed that non-practical subjects (e.g: Mathematics) have comparatively lesser average than the practical subjects(e.g : Physics).

**Q4. Determine whether the subject combination impacts the overall academic performance of the students across both University and Affiliated Colleges.**

First Year

Second Year

In this, we are actually doing two type of analysis using a single Chart. Firstly, it determines whether any particular subject combination (e.g: Physics-Chemistry-Mathematics) has comparatively more average score than the other subject combinations. This can help the upcoming students to decide which particular subject combination can be suitable for them.

Secondly, we are also analyzing that whether there is a difference in the average score of subject combination across University and Affiliated colleges to determine whether the students in university get more marks than the Affiliated Colleges in that specific subject combination. As we can see from the above graph, for almost every subject combination, the average marks of the university students are more than other affiliated colleges.

**Q5. Which subject should a student drop in the third year based upon statistical analysis of the past two years?**

**Description:** In BSC, a student has to drop one subject from his curriculum in the third year, thus he has only two subjects in the third year which are 225 marks each. Thus with the help of descriptive analysis, we help a particular student to choose the subject combination so as to achieve better grades on the basis of his performance in the subjects in last two years.

The above graph shows the academic record of a student (Enrollment No. U2112063) in the past two years. Now, if he has to decide which subject to drop for the third year, it can be seen that he has been performing comparatively not as good in Mathematics as compared to other groups. So, according to stats, dropping Mathematics from third year can be academically more beneficial.

**Hypothesis Testing 1: Verifying that a student’s result whether Pass/Fail depends upon the institution in which he/she is studying.**

Null Hypothesis(Ho) : There is no relation between the result of student with the institute he is enrolled in.

Alternate Hypothesis(H1) : There is significant relation between the result of student with the institute he is enrolled in.

**Solution:** For this, Chi-Square Test can be a better option for Inferential Statistics because we had categorical data such as (College-Passed, College-Failed, University-Passed, University Failed)

First Year

Second Year

**Hypothesis Testing 2: Verifying that a student’s result aggregate is affected by his/her subject combination choice.**

Null Hypothesis(Ho) : There is no relation between the result of student with the subject combination he has choosen.

Alternate Hypothesis(H1): There is a significant relation between the result of student with the subject combination he has choosen

**Solution:** For this, Chi-Square Test can be a better option for Inferential Statistics because we had categorical data such as (‘Subject Combination’-‘Result Type’)

First Year

Second Year

**Conclusion:**

This project on *Student Academic Performance Analysis* has provided valuable insights into the academic trends and influencing factors among undergraduate science students of the Central University of Allahabad. By employing a combination of statistical techniques and data visualization tools, we were able to uncover significant relationships between practical exam performance and overall academic success, the influence of institution type (university vs. affiliated colleges) on pass rates, and the impact of subject combinations on student outcomes.

Our findings highlight the critical role that practical subjects play in boosting total scores, even when their weightage is relatively lower. The strong positive correlation between practical marks and total performance suggests that hands-on, application-oriented learning significantly benefits students. Additionally, students enrolled at the main university generally outperformed those from affiliated colleges, indicating a potential disparity in teaching quality, resources, or evaluation standards that merits further investigation.

The analysis also supports data-driven decision-making for students facing subject selection dilemmas—such as determining which subject to drop in the final year—by identifying weaker areas based on their academic history.

Overall, this project demonstrates how data analytics can be effectively used to support educational planning, student guidance, and institutional improvement. Educational institutions can leverage such analyses to tailor interventions, allocate resources strategically, and ultimately enhance student outcomes in a holistic manner.