Student Academic Performance Analysis(G11): Semester Long Assignment

Prepared By-

Ritik Agnihotri
(524410011)

Saksham Chanouria
(524410013)

Anshul Singh Chauhan
(524110020)

Ashutosh Dwivedi
(524410016)

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)

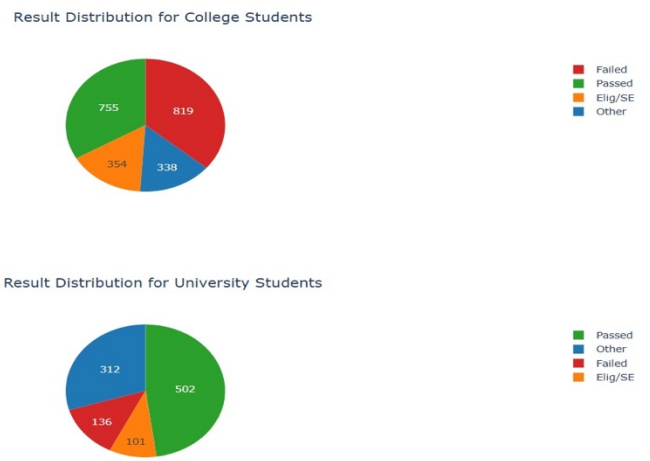
Objectives

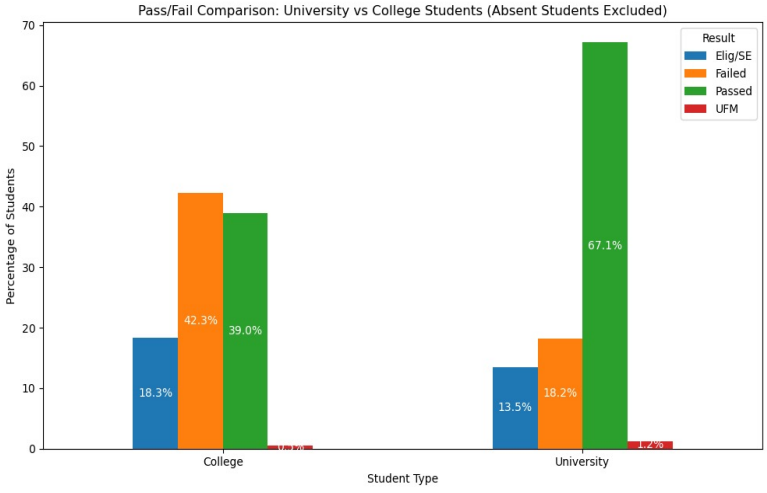
The academic performance of students is influenced by a wide range of factors—ranging from the type of institution they attend, to the subjects they choose, and even the presence of practical components in their curriculum. Understanding these influences is crucial not only for evaluating the effectiveness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

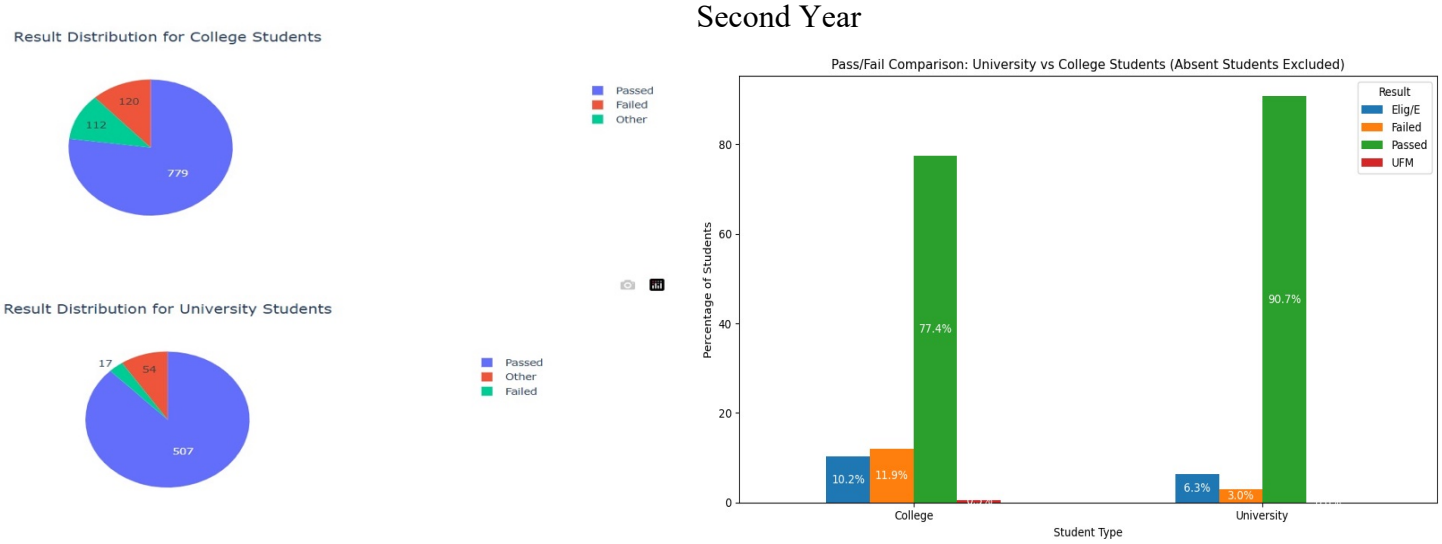
Some questions answered from this exploration includes:

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

First Year

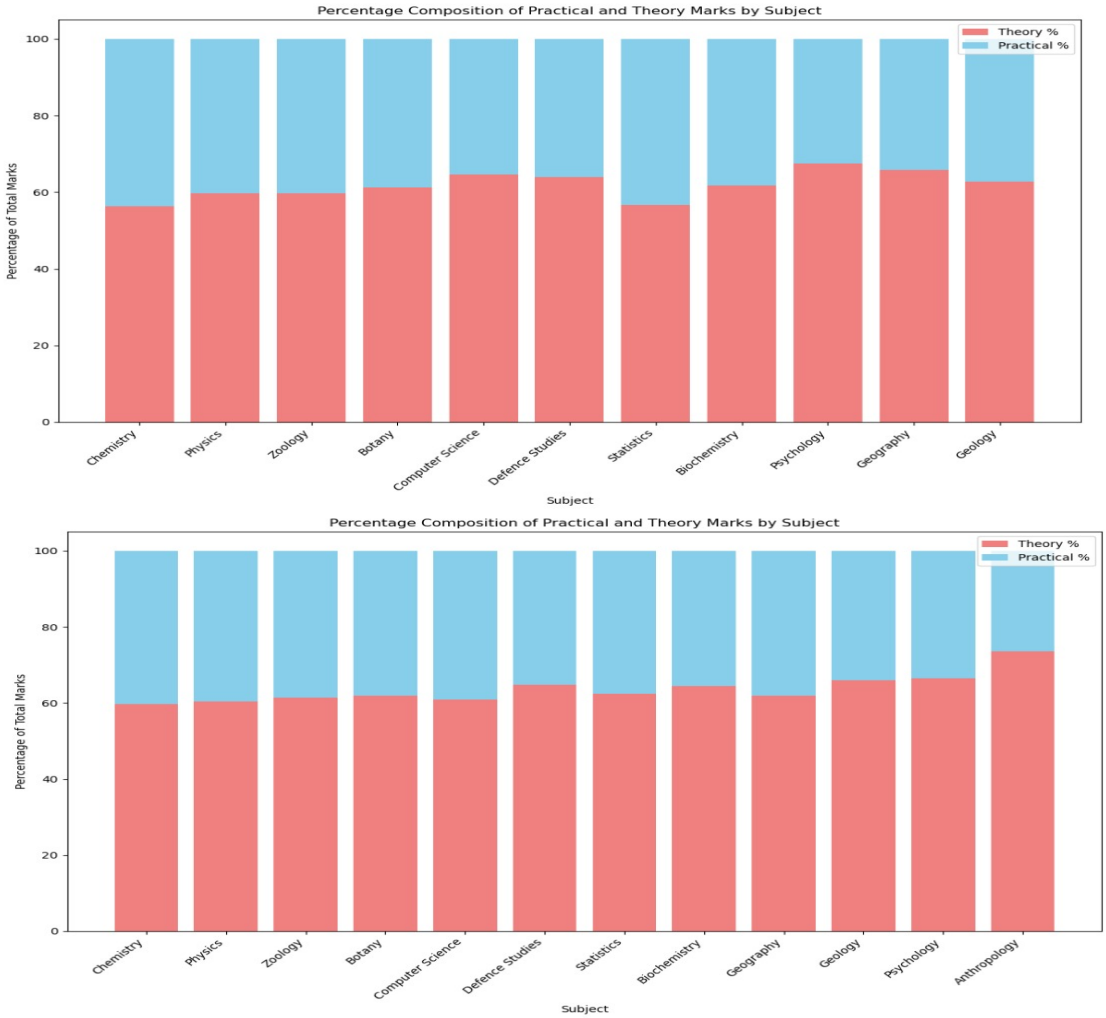






The above graphs show the number of passed, failed and Elig/SE students distribution among University and Affiliated colleges. From the above graph, it can be seen that the proportion of passed students in the university is significantly greater than number of passed students in affiliated colleges.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

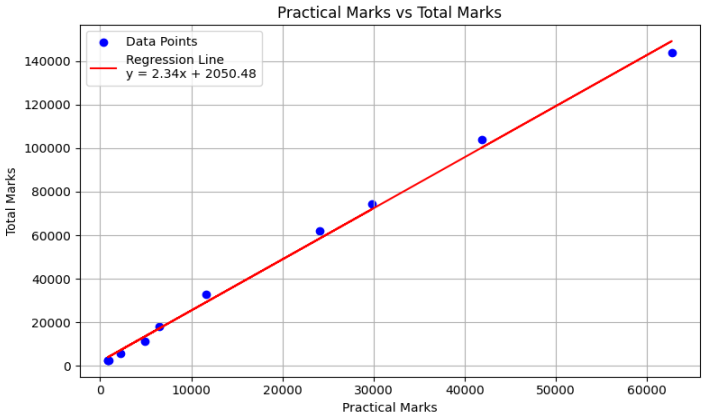


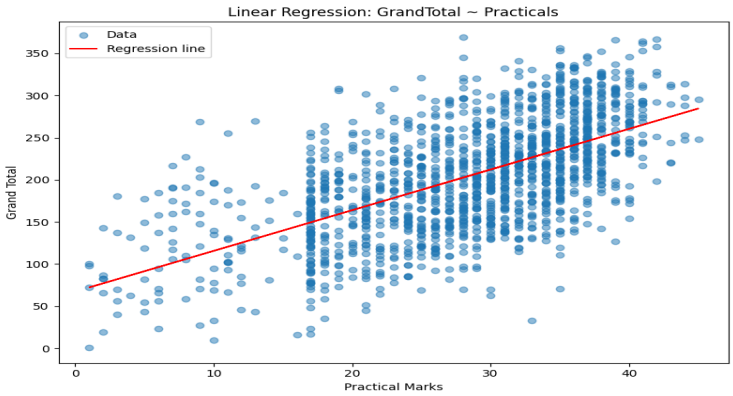
It is clear from the above graphs that even though the practical exams' weightage is less(50/150) , still it significantly has more contribution in overall result as compared to theoretical exams.
For instance, the weightage of practical is about 33%, still in most of the exams, they are contributing nearly 40% of total marks.

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

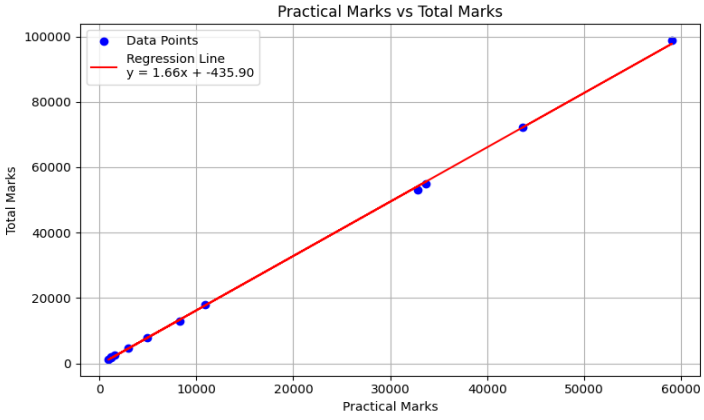
Subject	Practical Marks	Non-Practical Marks	Total marks	Practical %
Physics	41879	62097	103976	40.28
Chemistry	62729	81074	143803	43.62
Computer Science	11585	21123	32708	35.42
Psychology	881	1824	2705	32.57
Defence Studies	6480	11483	17963	36.07
Statistics	4890	6406	11296	43.29
Biochemistry	2234	3610	5844	38.23
Botany	24110	38031	62141	38.8
Zoology	29844	44373	74217	40.21
Geography	866	1669	2535	34.16
Geology	909	1531	2440	37.25

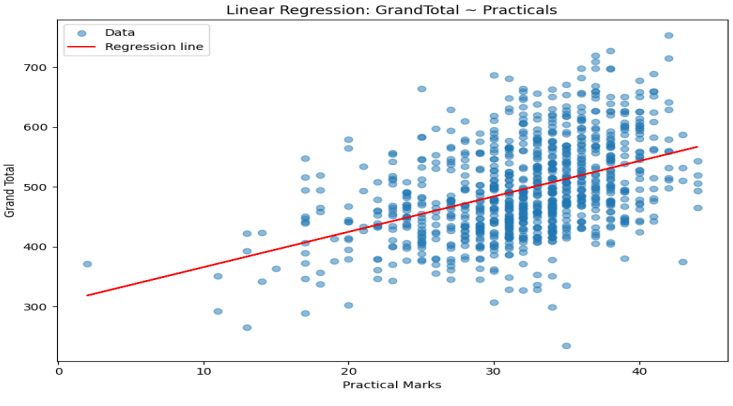
Subject	Practical Marks	Non-Practical Marks	Total Marks	Practical %
Physics	28533	43639	72172	39.53
Chemistry	39856	59060	98916	40.29
Computer Science	7010	10925	17935	39.09
Psychology	604	1200	1804	33.48
Defence Studies	4507	8301	12808	35.19
Statistics	2993	4975	7968	37.56
Anthropology	335	930	1265	26.48
Biochemistry	1647	2993	4640	35.5
Botany	20135	32826	52961	38.02
Zoology	21140	33704	54844	38.55
Geography	975	1590	2565	38.01
Geology	643	1249	1892	33.99





Pearson Correlation Coefficient: 0.9981
📊 P-value: 0.0000
✅ There is a statistically significant relationship between practical and total marks.

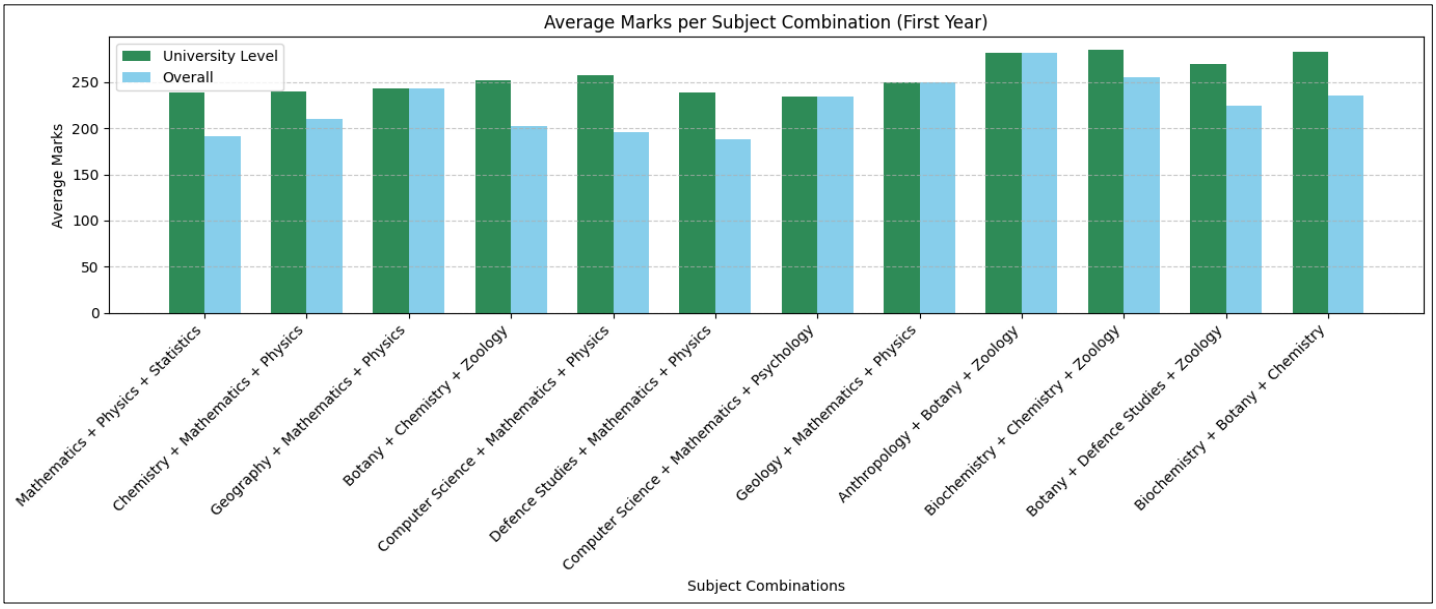


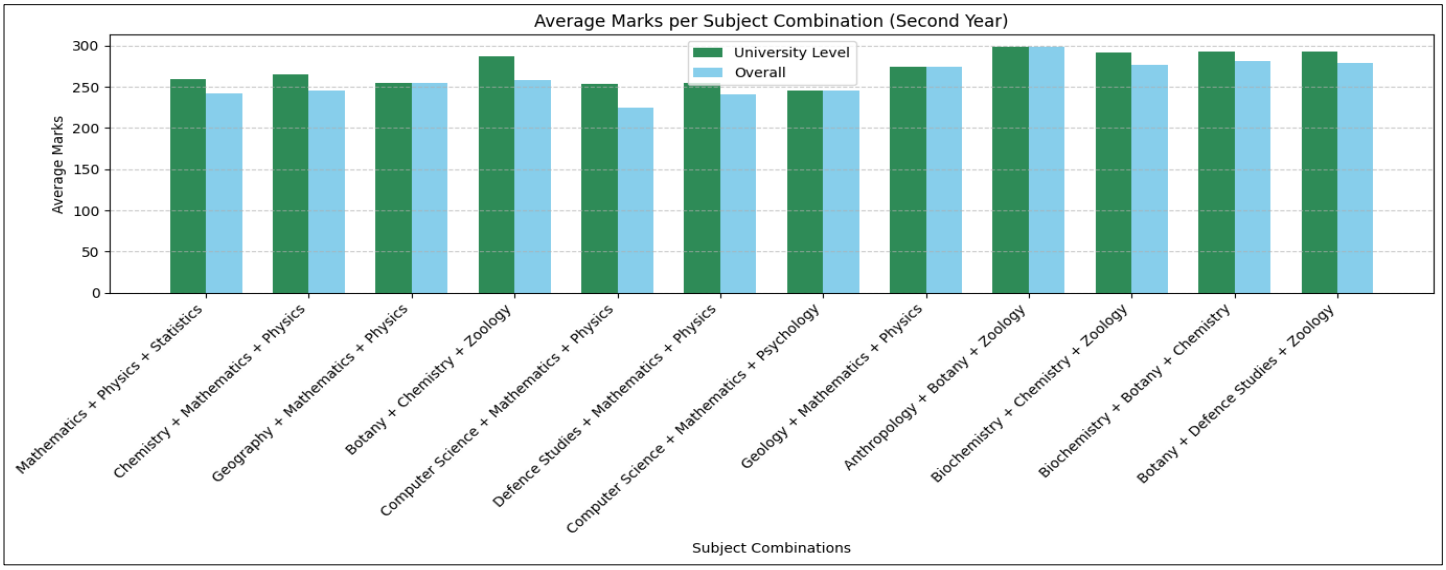


```
📈 Pearson Correlation Coefficient: 0.9998
📊 P-value: 0.0000
✅ There is a statistically significant relationship between practical and total
marks.
```

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

			Overall Students			University Students		
Subject 1	Subject 2	Subject 3	Total Marks	Count	Average	Total Marks	Count	Average
Mathematics	Physics	Statistics	28185	147	191.7347	7873	33	238.5758
Chemistry	Mathematics	Physics	158674	756	209.8862	66923	279	239.8674
Geography	Mathematics	Physics	6553	27	242.7037	6553	27	242.7037
Botany	Chemistry	Zoology	194413	960	202.5135	46618	185	251.9892
Computer Science	Mathematics	Physics	67263	344	195.532	11853	46	257.6739
Defence Studies	Mathematics	Physics	35343	188	187.9947	8110	34	238.5294
Computer Science	Mathematics	Psychology	7037	30	234.5667	7037	30	234.5667
Geology	Mathematics	Physics	5990	24	249.5833	5990	24	249.5833
Anthropology	Botany	Zoology	3666	13	282	3666	13	282
Biochemistry	Chemistry	Zoology	10468	41	255.3171	5127	18	284.8333
Botany	Defence Studies	Zoology	7854	35	224.4	4858	18	269.8889
Biochemistry	Botany	Chemistry	4714	20	235.7	1983	7	283.2857

Second Year

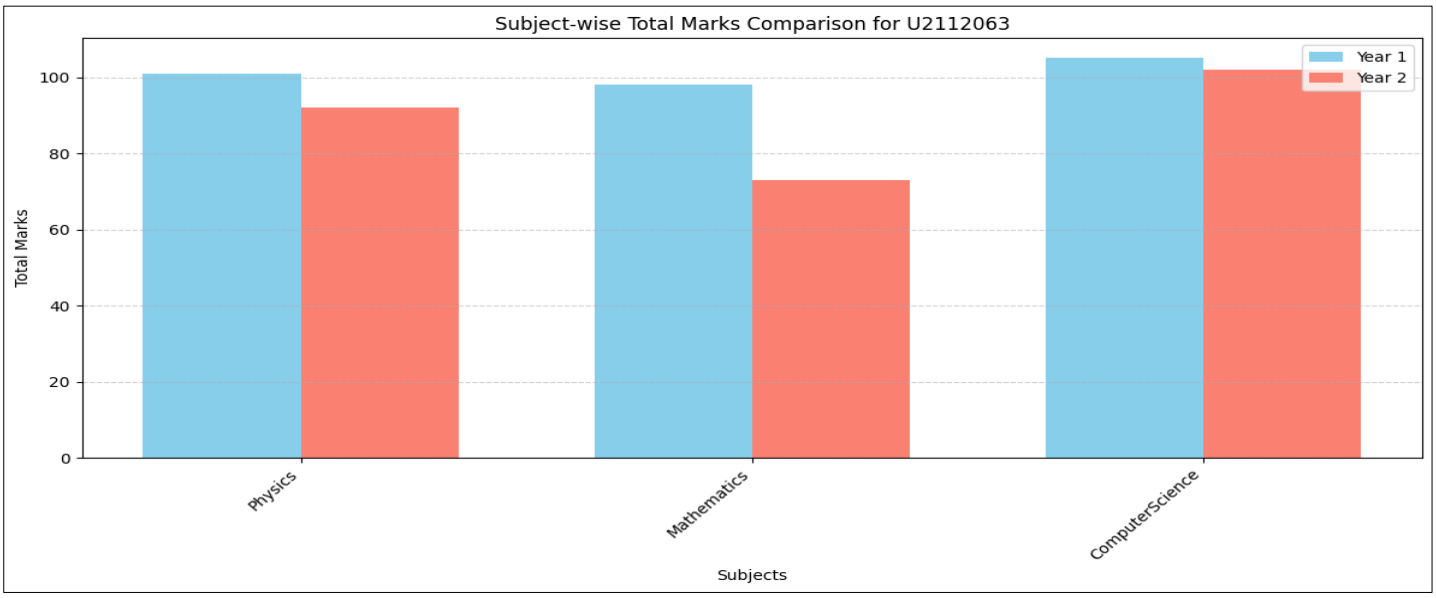
			Overall Students			University Students		
Subject 1	Subject 2	Subject 3	Total Marks	Count	Average	Total Marks	Count	Average
Mathematics	Physics	Statistics	20109	83	242.277108	6736	26	259.076923
Chemistry	Mathematics	Physics	117411	478	245.629707	53595	202	265.321782
Geography	Mathematics	Physics	6358	25	254.32	6358	25	254.32
Botany	Chemistry	Zoology	144761	560	258.501786	42771	149	287.053691
Computer Sci	Mathematics	Physics	42460	189	224.656085	9872	39	253.128205
Defence Studi	Mathematics	Physics	24577	102	240.95098	7640	30	254.666667
Computer Sci	Mathematics	Psychology	4659	19	245.210526	4659	19	245.210526
Geology	Mathematics	Physics	4932	18	274	4932	18	274
Anthropology	Botany	Zoology	3579	12	298.25	3579	12	298.25
Biochemistry	Chemistry	Zoology	8842	32	276.3125	4078	14	291.285714
Biochemistry	Botany	Chemistry	4212	15	280.8	2050	7	292.857143
Botany	Defence Studi	Zoology	7525	27	278.703704	4679	16	292.4375

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.

<u>Hypothesis Testing 1</u>: 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.

<u>Solution</u>: 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

```
Contingency Table:
Result_Status  Failed  Passed
Institution
College          1157    1109
University        448     603


Chi-Square Statistic: 20.1093
P-value: 0.0000
Degrees of Freedom: 1

Expected Frequencies:
Result_Status       Failed       Passed
Institution
College        1096.451613  1169.548387
University      508.548387   542.451613

Conclusion: Reject the null hypothesis (H₀). There is a significant association between institution type and pass/fail rates.
```

Second Year

```
Contingency Table:
Result_Status  Failed  Passed
Institution
College           232     779
University         71     507


Chi-Square Statistic: 26.4120
P-value: 0.0000
Degrees of Freedom: 1

Expected Frequencies:
Result_Status      Failed      Passed
Institution
College        192.783512  818.216488
University     110.216488  467.783512

Conclusion: Reject the null hypothesis (H₀). There is a significant association between institution type and pass/fail rates.
```

<u>Hypothesis Testing 2</u>: Verifying that a student's result aggregate is affected by his/her subject combination choice.

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

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

<u>Solution</u>: 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

```
Contingency Table:
Result                                  Elig/SE  Failed  Passed
Subject_Combination
Anthropology + Botany + Zoology               2       0      11
Biochemistry + Botany + Chemistry             2       0       5
Biochemistry + Chemistry + Zoology            2       3      13
Botany + Chemistry + Zoology                 15      26     144
Botany + Defence Studies + Zoology            0       2      16
Chemistry + Mathematics + Physics            41      58     180
Computer Science + Mathematics + Physics      5       6      35
Computer Science + Mathematics + Psychology  10       4      16
Defence Studies + Mathematics + Physics       8       3      23
Geography + Mathematics + Physics             6       2      19
Geology + Mathematics + Physics               4       4      16
Mathematics + Physics + Statistics            4       5      24


Chi-Square Statistic: 37.0360
P-value: 0.0234
Degrees of Freedom: 22

Expected Frequencies:
Result                                    Elig/SE     Failed      Passed
Subject_Combination
Anthropology + Botany + Zoology          1.802521   2.057423    9.140056
Biochemistry + Botany + Chemistry        0.970588   1.107843    4.921569
Biochemistry + Chemistry + Zoology       2.495798   2.848739   12.655462
Botany + Chemistry + Zoology            25.651261  29.278711  130.070028
Botany + Defence Studies + Zoology       2.495798   2.848739   12.655462
Chemistry + Mathematics + Physics       38.684874  44.155462  196.159664
Computer Science + Mathematics + Physics 6.378151   7.280112   32.341737
Computer Science + Mathematics + Psychology 4.159664 4.747899  21.092437
Defence Studies + Mathematics + Physics  4.714286   5.380952   23.904762
Geography + Mathematics + Physics        3.743697   4.273109   18.983193
Geology + Mathematics + Physics          3.327731   3.798319   16.873950
Mathematics + Physics + Statistics       4.575630   5.222689   23.201681

Conclusion: Reject H₀. Results depend on subject combinations.
```

Second Year

Contingency Table:

Result	Elig/SE	Failed	Passed
Subject_Combination			
Anthropology + Botany + Zoology	0	0	12
Biochemistry + Botany + Chemistry	1	0	6
Biochemistry + Chemistry + Zoology	1	0	13
Botany + Chemistry + Zoology	5	4	140
Botany + Defence Studies + Zoology	0	0	16
Chemistry + Mathematics + Physics	14	9	179
Computer Science + Mathematics + Physics	2	0	37
Computer Science + Mathematics + Psychology	2	0	17
Defence Studies + Mathematics + Physics	5	0	25
Geography + Mathematics + Physics	1	1	23
Geology + Mathematics + Physics	1	0	17
Mathematics + Physics + Statistics	3	1	22

Chi-Square Statistic: 19.5382
P-value: 0.6119
Degrees of Freedom: 22

Expected Frequencies:

Result	Elig/SE	Failed	Passed
Subject_Combination			
Anthropology + Botany + Zoology	0.754039	0.323160	10.922801
Biochemistry + Botany + Chemistry	0.439856	0.188510	6.371634
Biochemistry + Chemistry + Zoology	0.879713	0.377020	12.743268
Botany + Chemistry + Zoology	9.362657	4.012567	135.624776
Botany + Defence Studies + Zoology	1.005386	0.430880	14.563734
Chemistry + Mathematics + Physics	12.692998	5.439856	183.867145
Computer Science + Mathematics + Physics	2.450628	1.050269	35.499102
Computer Science + Mathematics + Psychology	1.193896	0.511670	17.294434
Defence Studies + Mathematics + Physics	1.885099	0.807899	27.307002
Geography + Mathematics + Physics	1.570916	0.673250	22.755835
Geology + Mathematics + Physics	1.131059	0.484740	16.384201
Mathematics + Physics + Statistics	1.633752	0.700180	23.666068

Conclusion: Fail to reject H_0. Results are independent of subject combinations.

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.