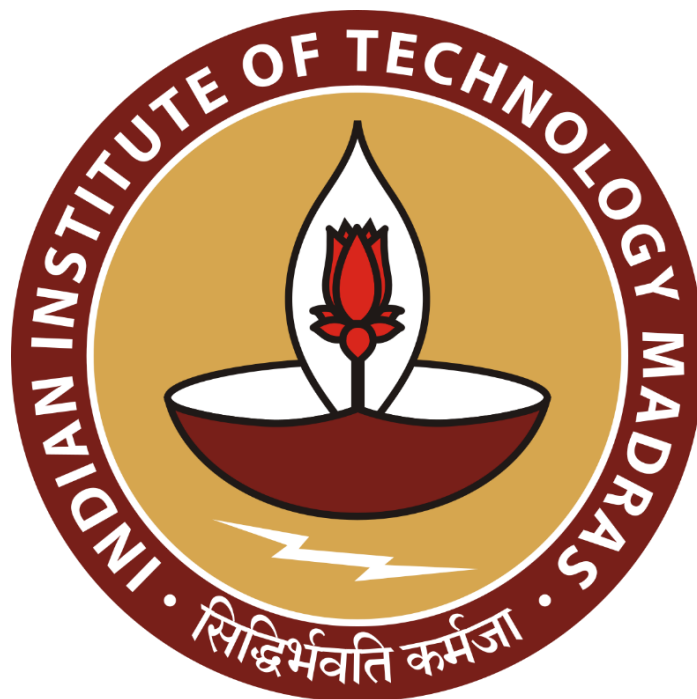**A Primary Data-Based Study of Improving Academic Performance and Business Enhancement in Small-Scale Educational Services**

A Proposal report for the BDM capstone Project

Submitted by

Name: Nirmit Desai
Roll Number: 23F3001502

IITM Online BS Degree Program,

Indian Institute of Technology, Madras,

Chennai Tamil Nadu, India, 600036

Content

1) Executive Summary

This project analyzes academic performance and business enhancement opportunities at Meghna Tutorials, a small-scale educational service catering to Grades 4–10. Using primary data collected from 50 students across three trimesters and eight subjects, the study aims to identify trends, risk factors for dropout, and the effectiveness of remedial interventions.

Descriptive statistical analysis was employed to summarize key metrics such as test scores, attendance, and feedback. The findings reveal that overall academic performance is strong, with most students scoring above 80 in total scores. Attendance is highest in lower grades (over 90% in Grade 4) but declines in higher grades, correlating with increased dropout risk and lower performance. Remedial class participation is concentrated in mathematics and science, particularly in Grades 7 and 8, indicating these subjects pose greater challenges. Although the dropout rate is low (3.3% of records), nearly all dropout cases are linked to low attendance and feedback, highlighting clear warning signs for early intervention.

Based on these insights, the report recommends targeted support for at-risk students, enhanced engagement strategies for higher grades, and regular monitoring of key performance indicators. These measures are expected to improve student outcomes and strengthen Meghna Tutorials' competitive position in the local education market. This data-driven approach provides a replicable model for similar tutorial centers seeking measurable academic and business improvements.

2) Proof of Originality

The google-drive link contains:
- Interview video link
- 3 images
- Letter from the organisation with the letterhead
- Dataset link

[Proof of Originality](#)

3) Metadata

	Student_Name	Grade	Trimester	Subject	Monthly_Test_Score	Term_End_Score	Total_Score	Attendance_Percentage	Feedback_Score	Dropout	Took_Remedial_Classes
2	Ishaan	4	1	Mathematics	15	65	80	90	8	0	0
3	Ishaan	4	1	Science	17	71	88	95	7	0	0
4	Ishaan	4	1	Social Studies	18	75	93	95	8	0	0
5	Ishaan	4	1	Hindi	16	59	75	95	8	0	0
6	Ishaan	4	1	English	15	66	81	97	9	0	0
7	Ishaan	4	2	Mathematics	20	79	99	98	10	0	0
8	Ishaan	4	2	Science	18	76	94	98	9	0	0
9	Ishaan	4	2	Social Studies	19	76	95	97	10	0	0
10	Ishaan	4	2	Hindi	16	57	73	96	9	0	1
11	Ishaan	4	2	English	19	74	93	97	8	0	0
12	Ishaan	4	3	Mathematics	17	79	96	98	9	0	0
13	Ishaan	4	3	Science	16	78	94	99	9	0	0
14	Ishaan	4	3	Social Studies	20	76	96	97	8	0	0
15	Ishaan	4	3	Hindi	17	69	86	96	9	0	0
16	Ishaan	4	3	English	18	64	82	98	10	0	0
17	Kavya	4	1	Mathematics	15	69	84	93	8	0	0
18	Kavya	4	1	Science	16	71	87	94	9	0	0
19	Kavya	4	1	Social Studies	18	68	86	92	9	0	0
20	Kavya	4	1	Hindi	14	59	73	98	9	0	1

Overview

1. Dataset Overview
- Title: Meghna Tutorials Academic Performance Dataset
- Source: Primary data collected from Meghna Tutorials, Hiranandani Estate, Thane
- Date Range: Academic years 2024–2025
- Population: Students from Grades 4 to 10
- Data Collection Method: Manual entry from internal records (test scores, attendance registers, feedback from students)
- Update Frequency: Trimester-wise

2. Data Dictionary

Field Name	Data Type	Description	Example Values	Allowed Values	Purpose/ Usage
Student_Name	String	Full name of the student	Ishaan	Text	Identifies individual students for tracking longitudinal progress
Grade	Integer	Academic grade level	4	4-10	Enables cohort analysis and grade-wise performance comparison
Trimester	Integer	Academic term within the year	3	1,2,3	Allows temporal analysis of academic progress and intervention timing
Subject	String	Subject name	English	Mathematics, Science, Social Studies, Hindi, English, Physics, Chemistry, Biology	Facilitates subject-wise diagnostic and remedial analysis
Monthly_Test_Score	Integer	Internal assessment test score	18	0-20	Measures ongoing performance and short-term learning outcomes
Term_End_Score	Integer	Final exam	76	0-80	Indicates

		score			cumulative knowledge and exam readiness
Total_Score	Integer	Internal assessment + final exam score	94	0-100	Used for overall performance and risk flagging
Attendance	Integer	Percentage of classes attended	89	0-100	Key predictor for academic risk and engagement
Feedback	Integer	Tutor's qualitative assessment	6	0-10	Captures behavioral and qualitative factors affecting performance
Dropout	Boolean	Whether the student dropped out	1	1=Yes 0=No	Outcome variable for retention analysis
Took_Remedial_Classes	Boolean	Whether remedial classes were attended	0	1=Yes 0=No	Used to assess intervention effectiveness

Table 1

3. Granularity and Structure
- Unit of Observation: Each row represents a unique combination of Student, Grade, Trimester, and Subject.
- Granularity: Data is granular at the subject and trimester level for each student, enabling detailed cohort, subject, and temporal analyses.
- Normalization: Data is structured in a flat, tabular format for ease of analysis.

3) Descriptive Analysis

The dataset contains 639 records representing 50 unique students across Grades 4–10 at Meghna Tutorials.

1) Summary Statistics

Metric	Monthly Test Score	Term End Score	Total Score
Mean	14.96	65.13	80.09

Median	16	68	83
Standard Deviation	3.87	14.44	17.44
Minimum	7	35	45
Maximum	20	80	100

Table 2

Interpretation: Most students score between 13–17 in monthly tests and 61–74 in term-end exams, with a few outliers scoring very low (possibly due to dropouts or absences).

2) Attendance
- Mean: 82.36%
- Median: 85%
- Range: 61% to 100%
- Distribution:
  - High (>90%): 194 records
  - Medium (75–90%): 338 records
  - Low (<75%): 86 records

Observation: The majority of students maintain medium to high attendance, but a significant minority have low attendance, often aligning with poor academic performance or dropout risk.

3) Feedback Score
- Mean: 7.79
- Median: 8
- Range: 6 to 10
- Distribution:
  - High (8–10): 402 records
  - Low (6–7): 216 records

4) Dropouts and Remedial Classes
- Number of Dropout Records: 21 (representing 5 students)
- Number of Remedial Class Instances: 116

5) Categorical Data Insights
- Unique Students: 50
- Subjects Covered: 8 (including PCMB subjects for higher grades)
- Grades Represented: 4 to 10

6) Breakdown by Subject

Subject	Mean Total Score	Mean Attendance (%)	Mean Feedback	Remedial Instances
Biology	82.3	83.61	7.83	22
Chemistry	82.68	83.84	7.71	22
English	83.04	90.58	8	4
Hindi	82.54	90.96	8.37	4
Mathematics	83.43	85.28	7.92	27
Physics	82.05	83.49	7.54	28
Science	83.31	88.88	7.70	4

Social Studies	85.35	90.88	7.62	5

Table 3

7) Breakdown by Trimester

Trimester	Mean Total Score	Mean Attendance (%)	Mean Feedback	Remedial Instances
1	82.92	85.61	8.07	36
2	82.04	85.06	7.99	48
3	83.54	84.79	8.10	32

Table 4

Observation: Scores and attendance remain fairly stable across trimesters, but remedial class participation peaks in the second trimester, possibly reflecting mid-year interventions.

8) Top and Bottom Performers
- Top 5 Students (by average total score):
  - Aarna (92.7)
  - Nisha (91.9)
  - Anwesha (91.3)
  - Aryan (91.1)
  - Anmol (90.5)
- Bottom 5 Students:
  - Hriday (70.7)
  - Sidharth (70.3)
  - Vivaan (69.6)
  - Neev (69.4)
  - Aryaman (68.8)

Note: The lowest performers often have lower attendance and feedback scores, and are more likely to have taken remedial classes.

4) Analysis Process/Method

Descriptive analysis has been selected as the primary analytical method for this mid-term report based on comprehensive evaluation of the research context, data characteristics, and business objectives. This methodological choice aligns with established educational research practices and provides the most suitable approach for the current phase of the project.

1. Research Phase Alignment
The mid-term report serves as an exploratory phase where the primary objective is to understand "what is happening" in the data rather than making predictions or establishing causal relationships. Descriptive analysis is specifically designed for this purpose, providing systematic methods to summarize and describe the features of educational datasets. This approach allows for the identification of patterns, trends, and anomalies that will inform deeper analysis in the final report.

2. Data Characteristics Justification

Mixed Data Types

The Meghna Tutorials dataset contains categorical variables (Student Names, Subjects, Grades), numerical variables (test scores, attendance), and binary variables (dropout status, remedial classes). Descriptive analysis is uniquely suited to handle this mixed-type data structure, providing appropriate summary measures for each variable type.

Hierarchical Structure

The data exhibits a nested hierarchical structure: Students → Grades → Trimesters → Subjects. With 50 students across 7 grades, 3 trimesters, and 8 subjects, descriptive analysis provides the most efficient method to summarize patterns at each level of this hierarchy.

Sample Size Considerations

With 639 records representing 50 students, the dataset is moderately sized but insufficient for complex statistical modeling that would require larger samples for statistical power. Descriptive analysis makes no assumptions about sample size and provides meaningful insights regardless of data volume.

3. Business Context Requirements

Stakeholder Communication

The primary stakeholders—parents, educators, and tutorial management—require clear, interpretable insights rather than complex statistical models. Descriptive statistics provide intuitive measures like averages, percentages, and ranges that are easily understood by non-technical audiences.

Actionable Insights

The business needs immediate, actionable insights to identify at-risk students, evaluate remedial programs, and inform parent communications. Descriptive analysis delivers these insights through straightforward metrics like mean attendance (82.4%) and dropout rate (3.3%) that directly inform decision-making.

4. Comparison with Alternative Methods

Inferential Statistics - Why Not Appropriate

- Sample Size Requirements: Inferential statistics require larger samples for adequate statistical power
- Distribution Assumptions: The data shows significant skewness (attendance: -3.27, total scores: -3.17), violating normality assumptions required for many inferential tests
- Research Questions: The mid-term focuses on description rather than hypothesis testing or population inference

Predictive Modeling - Why Not Suitable

- Data Volume: Limited data insufficient for robust training and validation of predictive models
- Time Horizon: Mid-term focuses on current state analysis rather than future predictions
- Interpretability: Complex models would be difficult for stakeholders to understand and act upon

Advanced Analytics (Machine Learning) - Why Not Justified

- Insufficient Data Volume: Deep learning approaches require substantially more data
- Black Box Nature: Conflicts with the need for transparent, interpretable results
- Scope Mismatch: Mid-term scope doesn't justify the complexity of advanced analytics

5. Specific Advantages of Descriptive Analysis

Immediate Interpretability

Descriptive analysis provides direct, interpretable results that stakeholders can immediately understand and act upon. For example, identifying that Grade 4 students have the highest mean attendance (93.5%) and lowest dropout rate provides clear, actionable insights for tutorial management.

Comprehensive Coverage

The method addresses multiple analytical needs simultaneously:

- Performance Analysis: Through means, medians, and distributions of test scores
- Engagement Assessment: Via attendance percentages and feedback scores
- Risk Identification: Through threshold analysis and outlier detection
- Comparative Analysis: Across grades, subjects, and trimesters

Foundation for Future Analysis

Descriptive analysis establishes baseline metrics and identifies variables for deeper investigation in the final report. It reveals data quality issues early and provides the groundwork for more advanced analytical approaches.

6. Methodological Rigor

Educational Research Standards

The approach aligns with Institute of Education Sciences (IES) guidelines for educational data analysis, which emphasize the importance of descriptive analysis as a foundation for educational research. This ensures the methodology meets established academic standards.

Systematic Approach

Descriptive analysis provides a systematic framework for data exploration that ensures comprehensive coverage of all relevant aspects of student performance and engagement.

7. Alignment with Project Objectives

The chosen method directly addresses the core project objectives:

- At-risk Student Identification: Through threshold analysis of attendance, scores, and feedback
- Remedial Program Evaluation: Via group comparisons between students who did and didn't take remedial classes
- Report Card Generation: Providing individual student metrics for personalized reporting
- Business Enhancement: Delivering clear performance indicators for strategic decision-making

8. Practical Considerations

Computational Efficiency

Descriptive analysis is computationally efficient and can be performed with readily available tools like Excel or Python, making it accessible and practical for implementation.

Minimal Assumptions

The method makes minimal assumptions about data distribution and is robust to outliers when appropriate measures (like median instead of mean) are used.

Scalability

The approach is scalable for future data collection, providing a consistent framework for ongoing analysis as the tutorial grows.

Conclusion

Descriptive analysis represents the most appropriate methodological choice for this mid-term report because it perfectly balances analytical rigor with practical applicability. It accommodates the data characteristics, meets stakeholder communication needs, provides actionable insights within project constraints, and establishes a solid foundation for advanced analysis in the final report. This approach follows educational research best practices while directly addressing the business objectives with interpretable, actionable results that can immediately benefit Meghna Tutorials' operations and student outcomes.

The methodology's emphasis on exploration, description, and foundation-building aligns perfectly with the mid-term phase of the project, ensuring that subsequent advanced analytics in the final report will be built on a solid understanding of the data's fundamental characteristics and patterns.
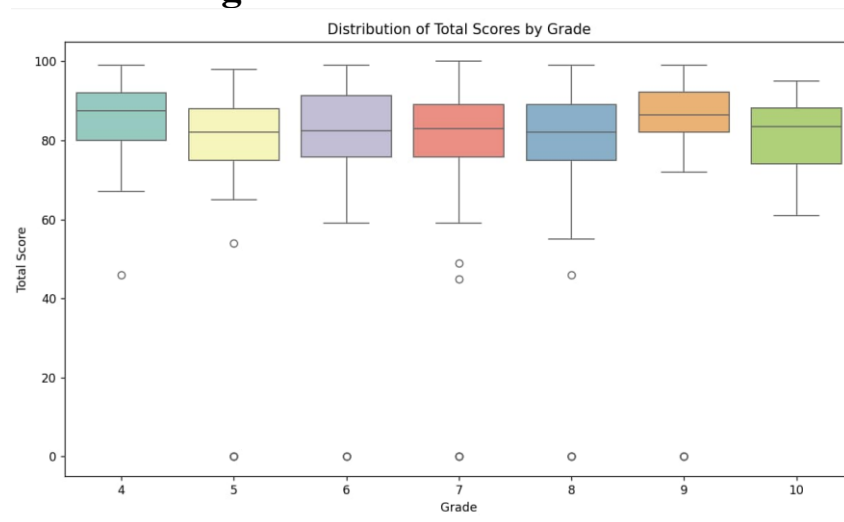
5) Results and Findings

Figure 1

The boxplot in Figure 1 illustrates the distribution of total scores across Grades 4 to 10. Most grades have a high median score (above 80), indicating strong overall academic performance. The boxes (interquartile ranges) are relatively tight, suggesting that the majority of students in each grade perform within a similar range. However, several grades show outlier points at the lower end (notably Grades 5, 6, 7, and 8), which represent students with significantly lower scores—these may correspond to students at risk of dropout or those with poor attendance. The presence of these outliers highlights the need for targeted interventions for lower-performing students.
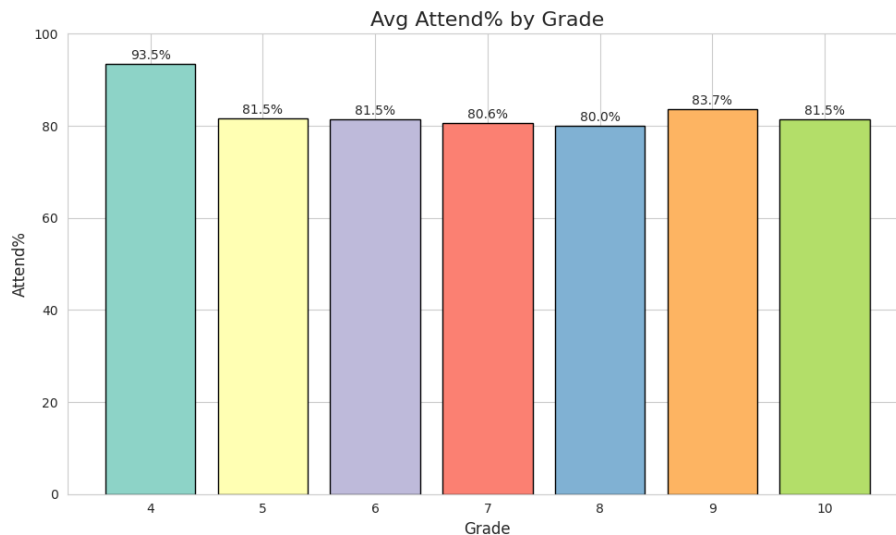
Figure 2

This bar chart in Figure 2 displays the average attendance percentage for each grade. Grade 4 stands out with the highest attendance (over 90%), while Grades 5 through 10 show a consistent but slightly lower average (around 80–83%). This suggests that younger students are more regular, while attendance tends to decrease in higher grades. The dip in attendance in middle and higher grades could be linked to increased academic pressure, competing commitments, or early signs of disengagement, which may warrant targeted interventions.
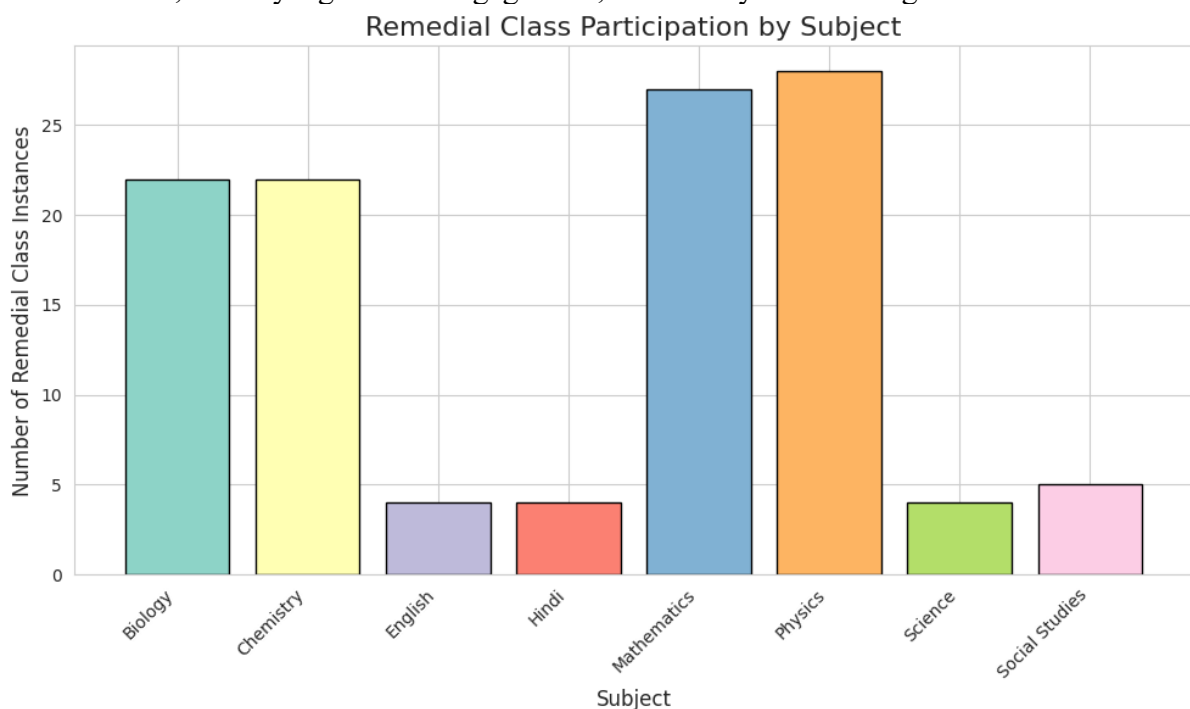

Figure 3

This bar chart in Figure 3 shows the number of remedial class instances for each subject. Mathematics and Physics have the highest remedial participation, followed by Biology and Chemistry. In contrast, English, Hindi, Science, and Social Studies have far fewer remedial sessions. This indicates that students find science and math subjects more challenging, and these areas may require additional instructional support or curriculum review to boost student confidence and performance.
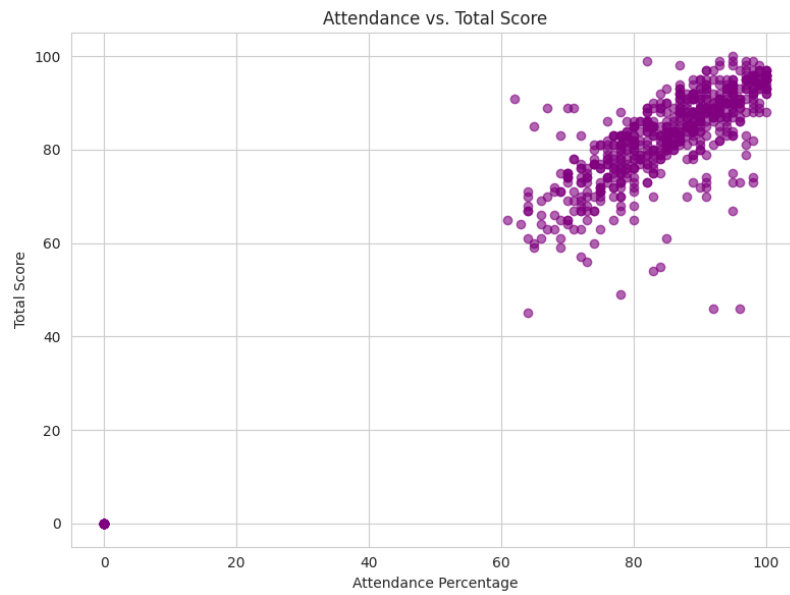
Figure 4

This scatter plot in Figure 4 shows the relationship between students' attendance percentage and their total academic scores.

There is a clear positive correlation; students with higher attendance tend to achieve higher total scores. Most data points cluster in the upper right, indicating that the majority of students with attendance above 75% score above 75 as well.