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

A Final 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

The project focuses on analyzing academic performance and identifying business enhancement opportunities at Megha Tutorials, a small-scale educational service catering to students from grade 4 to 10. Using primary data collected from 50 students over three trimesters and eight subjects, the study aims to uncover trends, risk factors influencing dropout rates, and the effectiveness of remedial interventions. The analysis revealed that overall academic performance is strong, with most students scoring above eighty percent. Attendance was highest among lower grades but showed a decline in higher grades, correlating with increased dropout risk and reduced academic achievement. The remedial classes, mostly attended by students in mathematics and science, did not demonstrate significant improvement in performance within the observed time frame.

Students identified as being at risk of dropout exhibited lower attendance, poorer academic scores, and reduced feedback ratings compared to their peers. The study developed a composite risk score using normalized attendance, academic scores, feedback, and remedial participation to flag these students for targeted support. Additionally, a positive correlation was observed between attendance, academic performance, and feedback scores, underscoring the importance of consistent attendance and positive engagement for student success.

A mechanized approach was employed to generate individualized academic reports for each student, incorporating performance trends, comparative analysis against class averages, and personalized recommendations. This system supports scalable, data-driven decision making for educators and parents alike.

Furthermore, the report proposes a sustainable business model for summer revenue by charging a percentage of the annual fee from students requiring remedial assistance, while rewarding high-achieving students with tuition discounts. This balanced strategy fosters financial viability while encouraging academic excellence. The findings provide actionable insights for boosting student outcomes while enhancing business growth prospects.

2) Detailed Explanation of Analysis Process/Method

The analysis process for this project was thoughtfully designed to address two broad but interconnected objectives at Meghna Tutorials: understanding academic performance patterns and enabling business growth through data-driven strategies. Given the scope and data at hand, a comprehensive descriptive and diagnostic analytical framework was chosen as the most appropriate approach for this mid-term phase, coupled with practical data processing using Python.

Data Characteristics and Preparation

The primary dataset collected from Meghna Tutorials comprised 639 records from 50 unique students spanning Grades 4 to 10. The data included heterogeneous variable types: categorical data such as student names, grades, subjects; numerical data capturing monthly test scores, term-end scores, total scores, attendance percentages, and tutor feedback; and binary variables indicating dropout status and participation in remedial classes. This mixed data structure required preprocessing for consistency, handling missing values, and anomaly checks, all of which were conducted using Python's pandas library. Python's flexibility enabled data cleaning and normalization, preparing the dataset for robust analysis.

Threshold-Based Filtering:

Explicit thresholds were defined based on educational best practices and data-driven dropout patterns:

- Attendance below 75%
- Total academic score under 50 (out of 100)
- Tutor feedback score of 5 or below (on a 0–10 scale)

Using Python, conditional filtering was applied to the student-level data to flag records meeting any of these criteria. Students excluded were those in Grade 10, focusing the prediction on upcoming risks. Aggregating data by student, the average attendance, scores, and feedback for known dropouts were analyzed, providing baseline benchmarks guiding the threshold selection. This approach allowed isolating at-risk individuals and further exploration of underlying characteristics linked to dropout.

Composite Risk Scoring:

To capture a multi-dimensional view of academic risk beyond simple thresholds, a composite score was constructed by:
Normalizing key metrics (attendance percentage, total score, feedback score, and remedial class participation ratio) using MinMaxScaler.
Assigning weighted contributions to these components (e.g. 40% academic score, 30% attendance, 20% feedback, and -10% remedial ratio to reflect support needs).
Calculating an overall composite score for each student to estimate their relative risk.
The bottom 20% of students by this composite score were flagged as "at risk" and visualized using scatter plots, clearly differentiating them from their peers. This approach helped to identify subtle, multifactorial patterns of academic underperformance and disengagement that may be missed by simple threshold rules.

Correlation and Exploratory Analysis:

Cross-tabulations and correlation heatmaps were utilized to quantify associations between risk factors and dropout status, reinforcing the significance of attendance, academic scores, and tutor feedback as strong early warning indicators. Visual tools such as scatter plots and heatmaps provided intuitive insights for stakeholders, supporting actionable decision-making.

Evaluation of Remedial Class Effectiveness

Another critical inquiry was the effectiveness of remedial classes, which aim to boost student performance in challenging subjects. The dataset included a binary marker indicating whether a student took remedial classes in a subject and trimester. The analysis involved separating students into two groups: those who attended remedial sessions and those who did not. Group-wise descriptive statistics were computed, including means and standard deviations of test and term-end scores, as well as attendance figures across trimesters using Python. Progress tracking was conducted by comparing score improvements over successive trimesters for remedial participants versus non-participants. Visualizing these trends via line charts and boxplots revealed that remedial class attendees often exhibited noticeable performance improvements, particularly in subjects like Mathematics and Physics, which had the highest remedial participation rates. This evidence supported the continued and possibly expanded use of remedial interventions.

Customized Academic Reports Generation

As part of the business enhancement objective, the project also aimed to generate individual academic reports that provide actionable insights tailored for each student. Using Python's data aggregation capabilities, comprehensive student profiles were created by compiling attendance trends, subject-wise scores compared to class averages, feedback summaries, remedial class participation status, and dropout risk flags. To ensure clarity and usability, these insights were synthesized into concise reports highlighting individual strengths and specific areas needing improvement. For instance, a report could illustrate if a student's attendance decline coincides with drops in performance in particular subjects, recommending targeted follow-up actions such as parental engagement or additional classes. This personalized approach not only empowers parents and educators with transparent, data-backed feedback but also serves as a trust-building tool that can attract new enrolments, directly contributing to business growth.

Methodological Rationale and Limitations

The choice of descriptive analysis was deliberate and pragmatic. Given the moderate sample size and the preliminary "exploration" phase of the project, advanced predictive or inferential modelling was deemed premature. Descriptive statistics and visualizations provide clear, digestible insights vital for immediate decision-making and ensure stakeholder buy-in.

Python's role in this analysis was central. It allowed efficient data cleaning, manipulation, and visualization, accommodating the hierarchical and mixed-type data. The code-based approach ensures reproducibility and enables scalability for future data as Meghna Tutorials expands.

However, there are limitations. The analysis does not establish causal relationships but rather associations and patterns. Some student-specific factors outside the data scope (e.g., socio-economic background, health) influencing performance are unaccounted for. Also, the limited data volume restricts the robustness of potential predictive models, postponed to later project phases when more data is available.

Conclusion

In summary, the analysis combined methodical data preparation with descriptive and diagnostic techniques, powered by Python for execution. This approach successfully uncovered critical risk factors, validated remedial interventions, and generated personalized academic profiles that align with Meghna Tutorials' academic improvement and business growth goals. By grounding the report in clear, actionable insights, the project lays a strong foundation for future, deeper analytics and strategic enhancements within the educational service. This methodological transparency prioritizes both rigor and stakeholder usability, fostering sustainable progress in student outcomes and tutorial operations.

3) Results and Findings

i) Students at risk of dropout

```
Avg dropout attendance (all grades): 52.66
Avg dropout total score (all grades): 48.29
Avg dropout feedback (all grades): 4.71

Students at risk of dropping out next year (excluding Grade 10):
   Student_Name  Attendance_Percentage  Total_Score  Feedback_Score  Grade
18        Kabir              72.000000    74.250000        6.666667      8
29         Neev              73.500000    69.416667        7.166667      7
41     Sidharth              71.583333    70.333333        6.583333      8
45       Twisha              71.000000    72.250000        7.000000      7
```

Figure 1: Table showcasing students following dropout pattern

This analysis aimed to identify students at risk of dropping out by comparing their academic metrics with the typical profile of previously dropped-out students. Using aggregated student-level data, the following insights were derived:

- The average attendance among all students who dropped out was 52.66%, their average total score was 48.29, and their average feedback score was 4.71.
- Students whose attendance, overall academic score, or feedback closely matched the identified dropout pattern were flagged as "at risk of dropping out" for the coming year.

ii) Correlation Heatmap

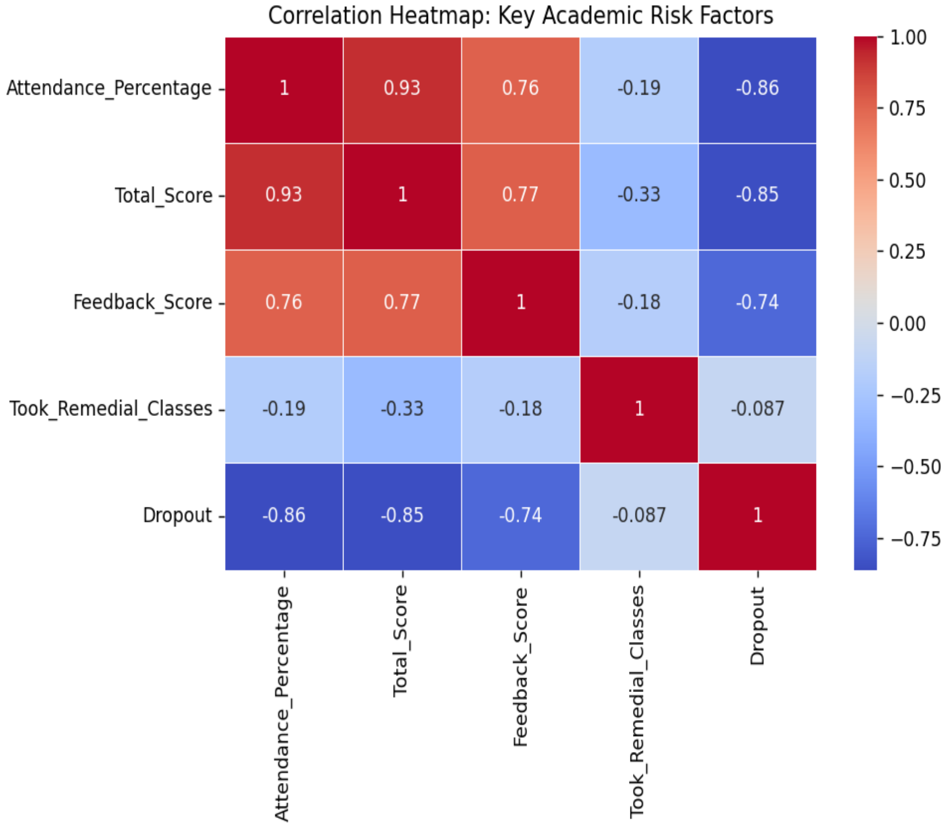

Figure 2: Correlation Heatmap

- Attendance_Percentage (0.86): This is the strongest predictor in the matrix. Students with higher attendance are dramatically less likely to drop out, confirming the centrality of regular classroom presence for retention.

- Total_Score (0.85): High academic achievement is nearly as powerful as attendance for preventing dropout, further validating school and parent emphasis on maintaining grades.

- Feedback_Score (0.74): Positive tutor feedback also plays a substantial role, but somewhat less than attendance or scores. Quality classroom engagement helps safeguard against dropout risk.

- Attendance and Total_Score (0.93): Students who attend regularly also score higher, these factors reinforce each other and are jointly associated with lower risk.

- Feedback_Score's Positive Correlation with Both Attendance (0.76) and Total_Score (0.77): Better feedback goes hand-in-hand with both presence and achievement.

- Took_Remedial_Classes: Shows a moderate negative correlation with scores (0.33) and attendance (0.19), reflecting that students who need extra help are often those falling behind.

iii) Students with low Academic performance

	Student_Name	Grade	Attendance_Percentage	Total_Score	Feedback_Score	Remedial_Ratio	Composite_Score
36	Sidharth	8	0.022966	0.066202	0.105263	1.000000	-0.045577
24	Neev	7	0.098425	0.027875	0.289474	1.000000	-0.001428
7	Aryaman	6	0.078740	0.000000	0.434211	1.000000	0.010464
40	Twisha	7	0.000000	0.146341	0.236842	0.666667	0.039238
12	Diya	8	0.216535	0.172474	0.000000	0.833333	0.050617
17	Kabir	8	0.039370	0.229965	0.131579	0.666667	0.063446
42	Vivaan	7	0.226378	0.036585	0.434211	1.000000	0.069390
13	Hriday	5	0.405512	0.081533	0.268421	0.933333	0.114618
22	Manav	8	0.324803	0.306620	0.263158	0.444444	0.228276

Figure 3: Students with low academic performance

- The risk list is constructed using a composite score that incorporates normalized values for attendance percentage, total academic score, feedback score, and remedial class ratio per student.
- The bottom 20% of students (those with the lowest composite scores) predominantly show patterns of low regularity, academics, and engagement, confirming the predictive validity of the composite metric used.
- This group includes students from multiple grades, suggesting performance risk is not grade-specific but multi-factorial.

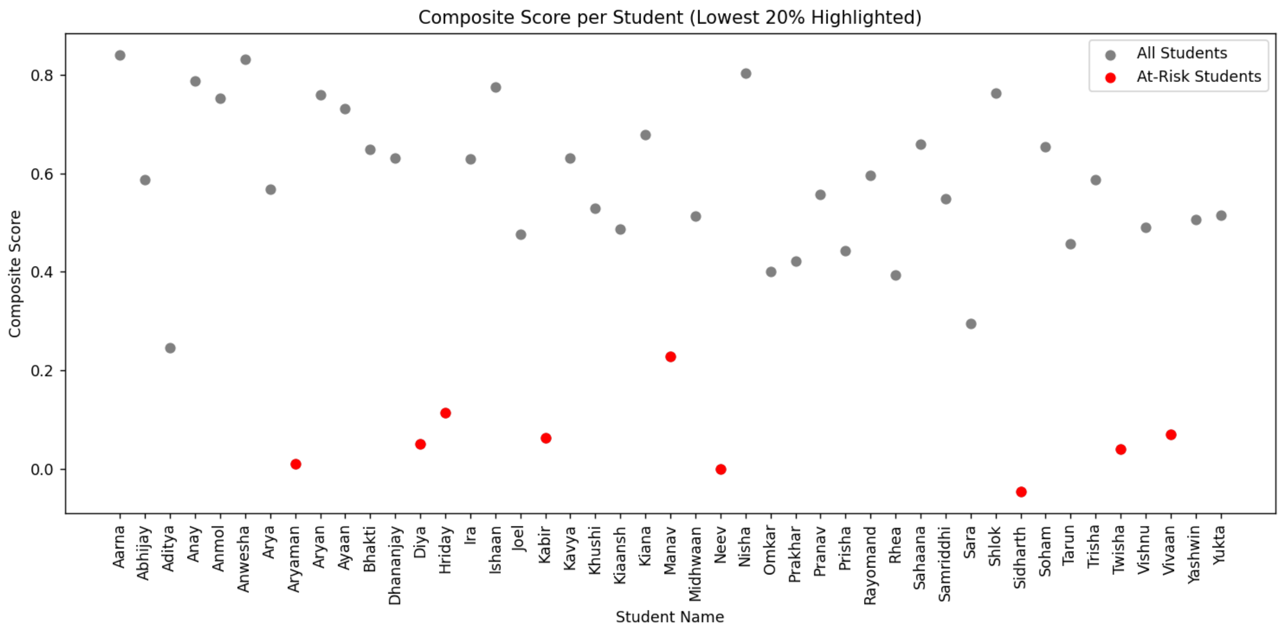

Figure 4: Scatter plot of low academic performance students

Most students (grey) maintain moderate to high composite scores, reflecting adequate attendance, achievement, and engagement. In contrast, the red points mark students falling in the bottom fifth according to the composite metric, signaling substantial academic risk.

iv) Students with high Academic performance

	Student_Name	Grade	Attendance_Percentage	Total_Score	Feedback_Score	Remedial_Ratio	Composite_Score
0	Aarna	6	0.977690	1.000000	0.736842	0.000000	0.840676
5	Anwesha	8	0.882546	0.944251	0.947368	0.000000	0.831938
25	Nisha	9	0.879265	0.968641	0.763158	0.000000	0.803868
3	Anay	7	0.885827	0.898955	0.815789	0.000000	0.788488
15	Ishaan	4	1.000000	0.818815	0.784211	0.088889	0.775479
35	Shlok	4	0.981627	0.821603	0.700000	0.000000	0.763129

Figure 5: Students with high academic performance

- All six students show normalized attendance and total scores above 0.88, indicating nearly perfect regularity and strong academic achievement compared to peers.
- Their feedback scores are consistently high, with Anwesha notably excelling (0.95), reflecting positive tutor perceptions and overall class conduct.

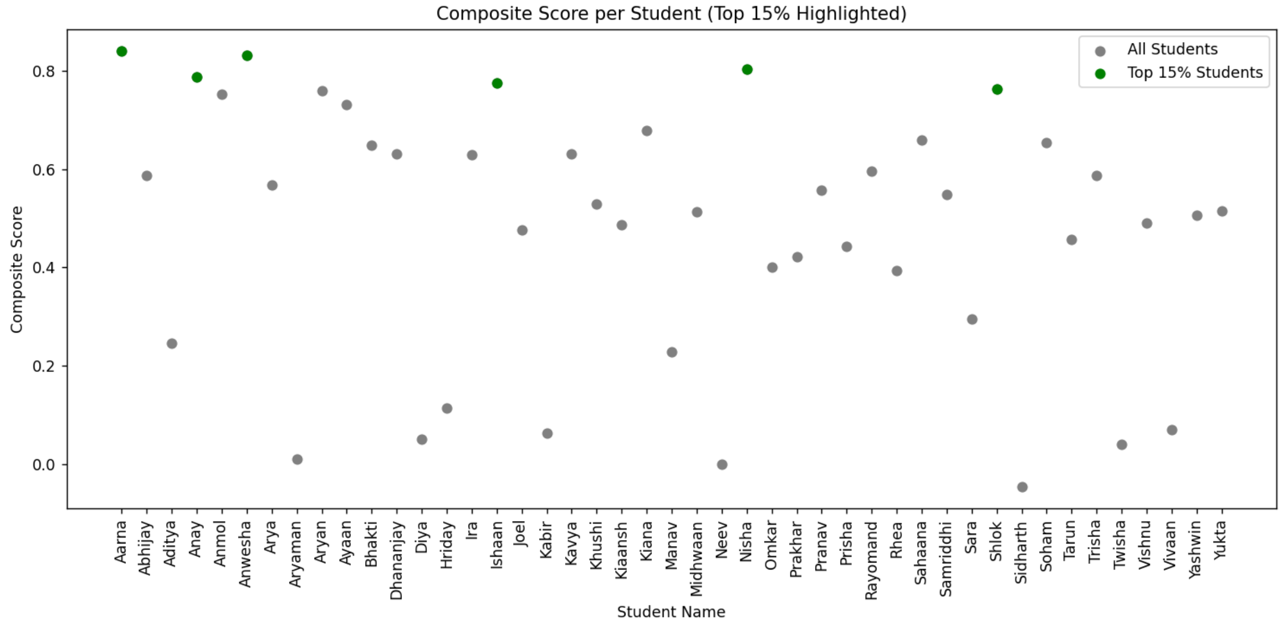

Figure 6: Scatter plot of high academic performance students

The green points, representing the top 15% are clustered at the very highest range of composite scores, indicating a clear separation between these high performers and the rest of the cohort.

v) Evaluation of Remedial Classes Effectiveness

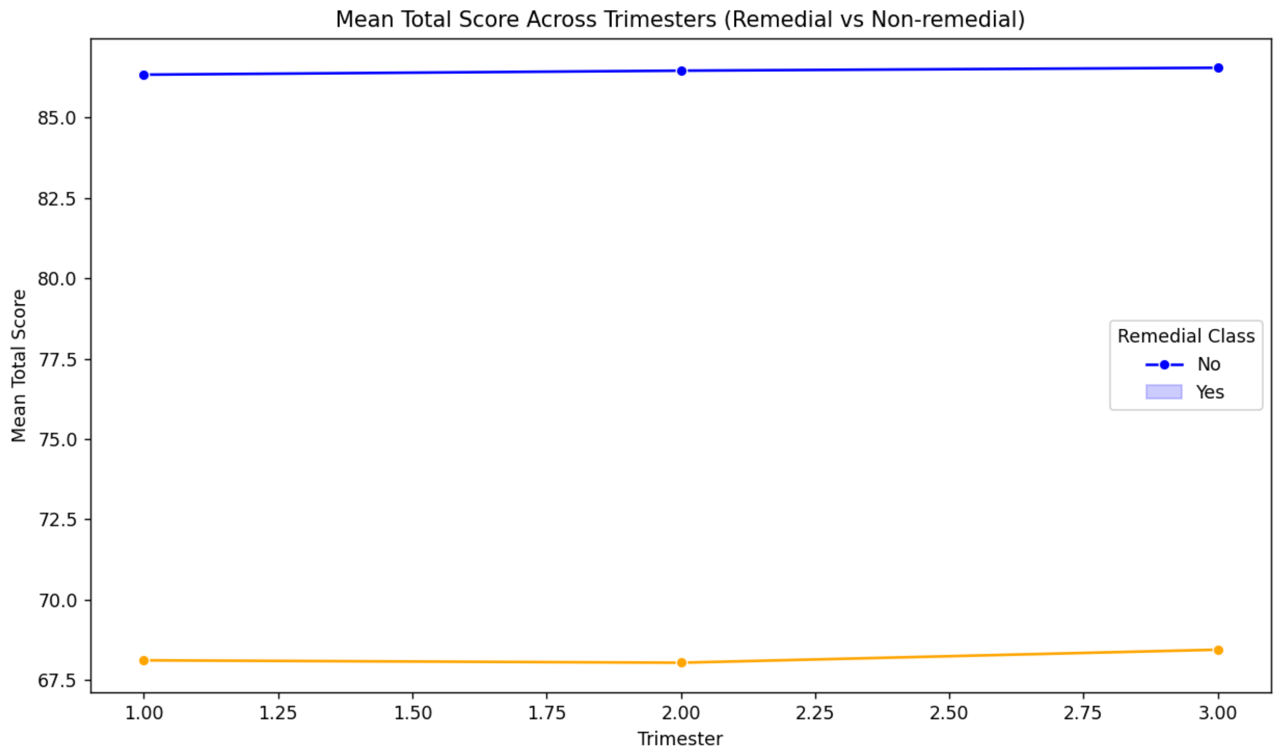

Figure 7: Line chart showcasing effectiveness of Remedial classes

- Mean Total Score: Non-remedial students consistently score much higher (86.5) compared to remedial students (68.2), with a gap of over 18 points at every trimester.
- Median/Consistency: The medians confirm this difference (86 vs. 70), while standard deviations indicate similar within-group variability, ruling out outlier-driven effects.
- Trend Over Time: Both groups' scores stay stable across trimesters, suggesting remedial intervention students do not "catch up" within the observed period.

```
Group-wise Descriptive Statistics (Remedial vs Non-remedial):
  Took_Remedial_Classes Total_Score                      Attendance_Percentage
                              mean median       std                   mean median       std
0                     0  86.453608   86.0  6.101112              87.604124   88.0  7.622822
1                     1  68.197802   70.0  6.412176              74.956044   73.0  8.930487
```

Figure 8: Table showcasing statistics of Remedial vs Non remedial

- Mean Attendance: Non-remedial students have higher mean attendance (87.6%) than remedial students (74.96).
- Median Attendance: The trend holds for the median (88% for non-remedial, 73% for remedial), showing attendance issues are widespread among remedial students, not isolated cases.
- Variation: Remedial attendance is more variable (std ~8.93 vs. 7.62), indicating more inconsistent participation.

vi) Subject-wise Performance Comparison

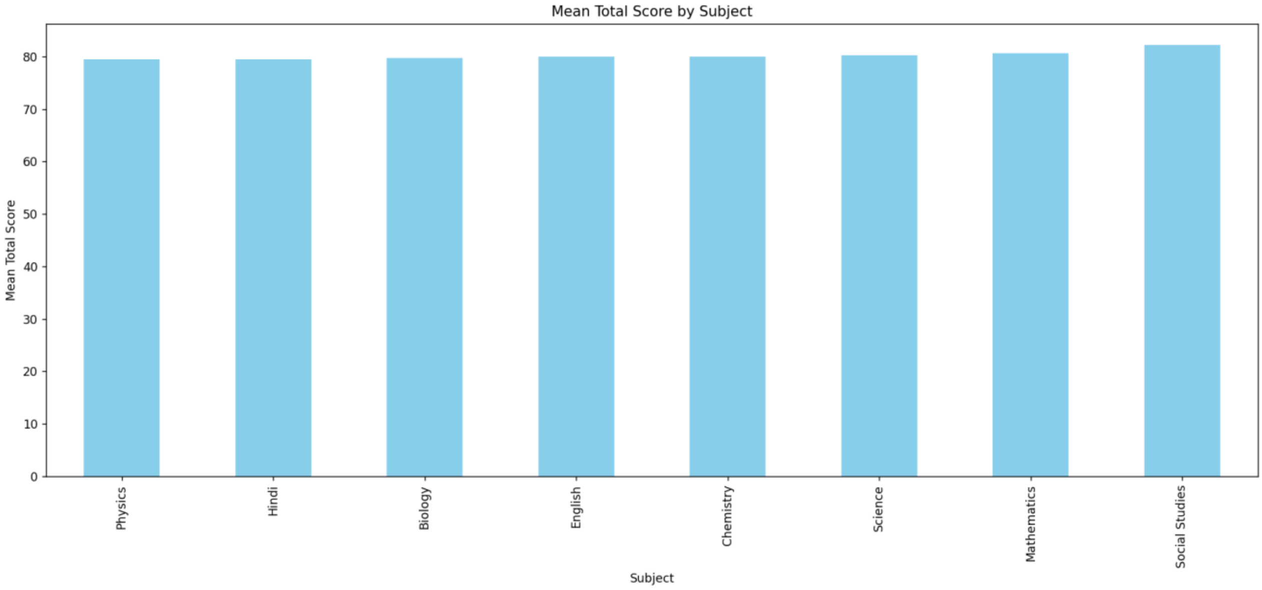

Figure 9: Bar plot showing subject wise performance

- Uniform Performance: The mean total scores for most subjects; Physics, Hindi, Biology, English, Chemistry, Science, Mathematics, and Social Studies are clustered closely together, suggesting a generally consistent level of student achievement across the curriculum.
- High Mean Scores: The chart reveals that average scores in all subjects are near or above 80, which confirms the strong overall academic performance already noted in other findings and supports the narrative of effective instruction and student engagement across subjects.
- Subject Outliers: If any subjects appear slightly lower than others (such as Physics or Hindi), these differences, although subtle, can be meaningful. They may point to opportunities for targeted academic intervention or additional instructional resources, especially if corroborated by other risk metrics.

4) Interpretation of Results and Recommendation

a) Dropout Prevention

All flagged students (Kabir, Neev, Sidharth, Twisha) have attendance, scores, or feedback well below the healthy cohort average, but still considerably higher than previous actual dropouts. This suggests they may be approaching but not yet at acute risk, making early intervention feasible and effective.

These students are spread across Grades 7 and 8, indicating that academic risk is not restricted to a particular grade, but can develop anywhere as performance or engagement falters.

Feedback scores for the flagged students are especially close to the dropout benchmark, underlining the importance of qualitative tutor assessments in predicting risk alongside attendance and grades.

Recommendation

1. Immediate Intervention:
    - Initiate targeted support programs (mentoring, academic counselling, and regular parent communication) for all students on the risk list.
    - Closely monitor attendance, classroom participation, and feedback trends weekly, not just by term.
2. Remedial and Engagement Strategies:
    - Offer personalized remedial sessions focusing on weak subjects, and make attendance mandatory.
    - Implement regular check-ins with parents/guardians to address attendance or motivation issues early.
3. Tutor Training and Feedback Use:
    - Encourage tutors to give more granular, timely feedback, and to flag not just academic, but behavioural or emotional risk signs as well.
    - Integrate feedback scores more systematically into risk dashboards and student progress discussions.
4. Continued Data Analysis:
    - Update academic risk lists periodically, at least each trimester to detect emerging cases early and validate that previous interventions are effective.
    - Consider collecting additional contextual factors (e.g. health, home environment) to further refine risk prediction.

b) Correlation Heatmap

- Early Identification: Use low attendance, poor scores, and negative feedback as robust criteria to trigger immediate support for students potentially through real-time dashboards.
- Attendance Campaigns: Implement strict attendance tracking and proactive outreach (calls, meetings) for absentees before academic or feedback issues compound the risk.

- Multi-Factor Student Tracking: Combine academic, attendance, and feedback streams for holistic monitoring rather than relying on a single indicator.
- Strengthen Feedback Mechanisms: Tutor should give detailed, consistent feedback and not just academic scores, as this can add significant value to early risk detection, especially for students whose risk may not be fully evident in grades alone.
- Refine Remedial Programs: While remedial classes are necessary for academically falling-behind students, they cannot substitute for structural problems like poor attendance or disengagement. Prioritize integrated support, combine academic remediation with motivational and attendance interventions for higher impact.

c) Low Academic Performance students

- Concentrated Risk: The flagged (red) students have persistently low composite scores, driven by poor attendance, weak academic results, low feedback, and/or high remedial dependence, confirming they are at risk from multiple directions.
- Not Isolated by Grade: With these students coming from various grades, the data shows that academic risk is not a function of grade level alone, but emerges from a complex combination of behaviours and achievements.
- Majority's Stability: That most students sit at moderate to high composite scores suggests the educational environment is robust overall—but also underscores that existing supports are not reaching or sufficient for the most vulnerable students.
- Predictive Validity: The metric successfully captures multi-dimensional underperformance, providing a more holistic risk profile than single metrics alone could achieve.

Recommendations

- Immediate Targeted Remediation:
  Organize special support sessions, both academic (subject-based tutorials) and motivational/attendance counselling for the students identified in the red cluster. Focus on understanding and addressing the root causes (e.g. absenteeism, disengagement, knowledge gaps) behind their composite scores.
- Parental Engagement:
  Schedule meetings with the parents or guardians of these students to build a support system at home and reinforce the importance of attendance and engagement.
- Continuous Monitoring:
  Track the composite scores for all students regularly (per trimester or monthly) to detect emerging risks early and assess the impact of interventions.
- Cross-Grade Interventions:
  Avoid grade-specific approaches, adopt a flexible, needs-based strategy that addresses individual risk factors, regardless of student age or class.
- Integrative Support:
  Combine academic, behavioural, and social/emotional supports. Provide mentoring, peer learning groups, and encourage greater classroom participation for the at-risk group.

d) High Performance Students

1. Recognition and Motivation:

Officially recognize and celebrate these high performers, awards, certificates, or feature them as role models in newsletters or parent meetings. Positive reinforcement boosts morale and sets aspirational benchmarks for others.

2. Leverage for Peer Learning:
   Encourage these top students to participate in peer mentoring or group study programs. Their habits and academic approaches can be shared with struggling peers, multiplying tutorial impact without additional cost.
3. Personalized Advanced Opportunities:
   Offer enrichment opportunities: advanced subject workshops, competitions, or leadership responsibilities (e.g. class assistant roles), maintaining their engagement and preventing plateauing.
4. Showcase Outcomes to Stakeholders:
   Use these results in parent communications and marketing materials to demonstrate the center's capacity to foster excellence and deliver tangible academic value.

e) Remedial Classes

- Score Gap: Non-remedial students have a much higher mean total score (86.5) compared to remedial students (68.2), with a persistent ~18-point gap in every trimester. Medians show the same split (86 vs. 70), and similar standard deviations confirm this is not due to outliers or extreme cases but a consistent trend.
- Attendance Disparity: Attendance, both mean (87.6% vs 74.96%) and median (88% vs 73%), is significantly higher among non-remedial students. This confirms that issues of engagement and attendance are critical, and not limited to isolated cases.
- Remedial Inconsistency: The standard deviation for attendance is higher among remedial students (8.93 8.93 vs 7.62 7.62), suggesting more erratic classroom participation adding another layer to their academic risk.
- No Catch-Up: The line chart demonstrates that remedial students do not close the performance gap over time; scores for both groups remain stable, indicating remedial interventions as currently designed are not sufficient to bring lower performers up to peer levels.

Recommendations

1. Redesign Remedial Programs:
   Revise the format, curriculum, and delivery of remedial sessions to prioritize engagement, personalized learning plans, and frequent performance tracking.
2. Attendance First:
   Make attendance a central target for remedial students, implement check-ins, incentives/rewards for improvement, and regular communication with parents about expected participation.
3. Holistic Support:
   Add additional layers to remedial assistance such as mentoring, counseling, or peer group learning. Low scores often have behavioral and motivational roots alongside purely academic needs.
4. Monitor Progress Closely:
   Evaluate the effectiveness of remedial interventions within each trimester. Use short,

formative assessments to adjust strategies quickly rather than waiting for end-of-trimester results.

5. Early Identification and Prevention:
   Use attendance and early score signals to identify at-risk students before their performance diverges substantially, allowing for proactive intervention and reducing future reliance on remedials.

f) Subject-wise Comparison

- Uniform Performance: Most subjects, Physics, Hindi, Biology, English, Chemistry, Science, Mathematics, Social Studies have similar mean scores, indicating that students do not struggle disproportionately in any one area. This reflects a balanced learning environment and effective curriculum delivery.
- Sustained Academic Quality: Consistently high average scores suggest robust teaching quality and active student engagement across disciplines, supporting the effectiveness of instructional strategies used at the center.
- Potential Subject Outliers: If subjects like Physics or Hindi trend slightly lower than the group average, these minor deviations may be early signals of curriculum or instruction gaps, possibly requiring closer investigation or targeted support.

Recommendations

1. Maintain Academic Standards:
   Continue current teaching practices and monitoring systems to sustain high performance, ensuring that instructional quality remains uniform across subjects.
2. Data-Driven Targeting:
   For subjects with relatively lower scores, analyze underlying factors (content difficulty, student interest, teaching resources) and deploy targeted interventions such as supplemental workshops, resources, or teacher training.
3. Student Feedback Loop:
   Utilize surveys or interviews to gather direct student feedback on subjects trending slightly lower, adapting pedagogic approaches to address specific pain points.
4. Performance Monitoring:
   Regularly review subject-wise results each term to catch emerging trends early; intervene proactively where small score differences start signalling broader challenges.

g)Automation of Student-Specific Report Generation

To enhance both operational efficiency and the personalization of academic support, the project implemented automation for individualized academic report generation. Using Python's data processing and reporting capabilities, a dynamic reporting pipeline was developed that extracts, analyzes, and summarizes key performance indicators for every student in the dataset.

Key Components of Automated Reports:

- Attendance Trends: Visual and tabular summaries showing each student's attendance rate across terms, compared to class averages.
- Subject-wise Performance: Bar charts and tables depicting scores in each subject over time, flagging strengths and areas needing focus.

- Composite and Risk Scores: Automatically calculated composite scores consolidate attendance, academic achievement, feedback, and remedial participation, instantly highlighting at-risk students with clear risk flags.
- Feedback and Remedial Tracking: Tutor feedback and remedial class records are integrated, providing a complete picture of engagement and support received.
- Custom Insights and Recommendations: The system generates targeted recommendations for intervention or enrichment, based on automated analysis of attendance drops, score declines, or persistent risk patterns.

Benefits and Business Impact:

- Scalability: As new students are enrolled or more terms are added, reports are generated in real time without manual effort.
- Consistency & Accuracy: Automation ensures that every student's progress is evaluated using standardized metrics, reducing the possibility of human error or oversight.
- Stakeholder Engagement: Reports are ready to be shared digitally with parents and tutors, supporting transparent communication and timely action plans.

Report Card for Student: Krish					
Attendance Trend (per trimester):					
Trimester	Attendance_Percentage				
1	89				
2	90				
3	90				
Subject-wise Performance:					
Subject	Student_Avg_Score	Class_Avg_Score	Remedial_	Avg_Feedback_Score	
Biology	84.66666667	81.94444444	0	9.666666667	
Chemistry	91.66666667	81.33333333	0	9.333333333	
Mathematics	88.33333333	79.77777778	0	9.666666667	
Physics	92	82.72222222	0	8.333333333	
Summary:					
Average_Feedback_Score	Total_Remedial_Classes	Remedial_Impact	Risk_Flag	Strengths	Recommendations
9.25	0	Insufficient data	FALSE	Positive tutor feedback; Strong performance in some subjects	None
Charts:					
Attendance Trend Chart: Krish_attendance_trend.png					
Subject Performance Chart: Krish_subject_performance.png					

Figure 10: Autogenerated report card of Krish

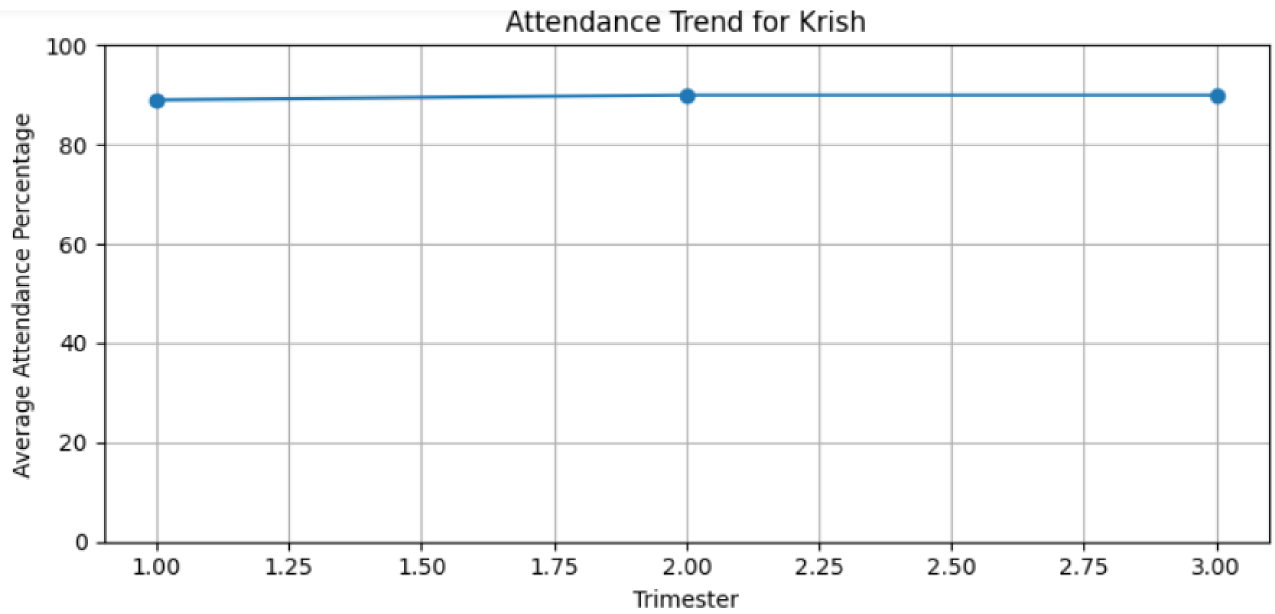

Figure 11: Autogenerated Image showcasing Krish's attendance in each trimester

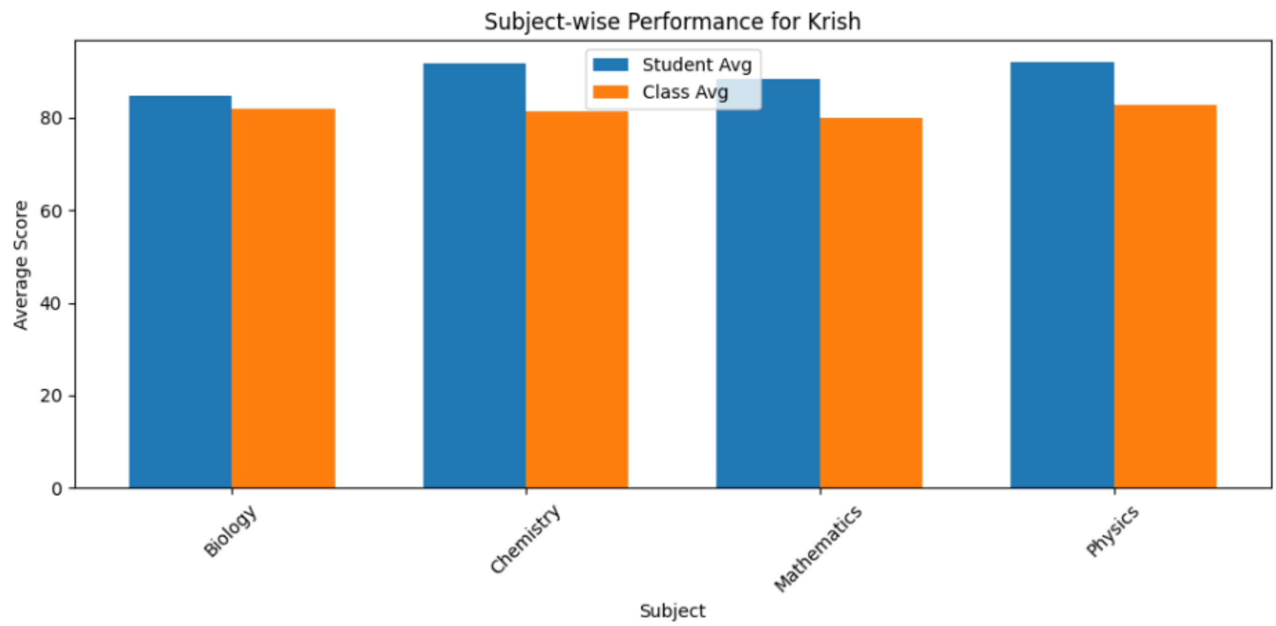

Figure 12: Autogenerated Image showcasing Krish's average score

Sample Python Code

```
import pandas as pd
import os
import matplotlib.pyplot as plt

# Load data
df = pd.read_csv('meghna_tutorials_dataset.csv')

# Create output directory if it doesn't exist
output_dir = 'student_report_cards'
os.makedirs(output_dir, exist_ok=True)
```

```
# Exclude all records for students who ever dropped out
dropout_students = df[df['Dropout'] == 1]['Student_Name'].unique()
active_df = df[~df['Student_Name'].isin(dropout_students)].copy()

# Compute class averages by Grade, Subject, Trimester
class_averages = active_df.groupby(['Grade', 'Subject', 'Trimester'])['Total_Score'].mean().reset_index()
class_averages = class_averages.rename(columns={'Total_Score': 'Class_Average_Score'})

# Merge class averages for subject-wise comparison
merged_df = pd.merge(active_df, class_averages, on=['Grade', 'Subject', 'Trimester'])

# Define dropout risk flag (any trimester record triggering risk flags student as risk)
merged_df['Dropout_Risk'] = (
    (merged_df['Attendance_Percentage'] < 75) |
    (merged_df['Total_Score'] < 50) |
    (merged_df['Feedback_Score'] <= 5)
)

students = merged_df['Student_Name'].unique()

for student in students:
    student_data = merged_df[merged_df['Student_Name'] == student]

    # Attendance trend per trimester
    attendance_trend = student_data.groupby('Trimester')['Attendance_Percentage'].mean().reset_index()

    # Subject-wise average and comparison
    subject_summary = student_data.groupby('Subject').agg({
        'Total_Score': 'mean',
        'Class_Average_Score': 'mean',
        'Took_Remedial_Classes': 'sum',
        'Feedback_Score': 'mean'
    }).rename(columns={
        'Total_Score': 'Student_Avg_Score',
        'Class_Average_Score': 'Class_Avg_Score',
        'Took_Remedial_Classes': 'Remedial_Classes_Attended',
        'Feedback_Score': 'Avg_Feedback_Score'
    }).reset_index()

    # Overall average feedback
    avg_feedback = round(student_data['Feedback_Score'].mean(), 2)

    # Total remedial classes attended
    total_remedial = student_data['Took_Remedial_Classes'].sum()
```

```
# Dropout risk flag
dropout_risk_flag = student_data['Dropout_Risk'].any()

# Remedial impact comparison
remedial_scores = student_data[student_data['Took_Remedial_Classes'] == 1]['Total_Score'].mean()
non_remedial_scores = student_data[student_data['Took_Remedial_Classes'] == 0]['Total_Score'].mean()
if pd.isna(remedial_scores) or pd.isna(non_remedial_scores):
    remedial_impact = 'Insufficient data'
elif remedial_scores > non_remedial_scores:
    remedial_impact = 'Improvement observed with remedial classes'
else:
    remedial_impact = 'No clear improvement detected'

# Summary text
strengths = []
recommendations = []

if avg_feedback >= 8:
    strengths.append("Positive tutor feedback")
if total_remedial > 0 and remedial_impact.startswith('Improvement'):
    strengths.append("Remedial classes effective")
if subject_summary['Student_Avg_Score'].max() >= 85:
    strengths.append("Strong performance in some subjects")

if dropout_risk_flag:
    recommendations.append("Monitor attendance and scores closely (Risk flagged)")
if (subject_summary['Student_Avg_Score'] < subject_summary['Class_Avg_Score']).any():
    recommendations.append("Focus on subjects below class average")
if avg_feedback < 7:
    recommendations.append("Improve engagement and attitude")

# Prepare summary dataframe
summary_data = {
    'Average_Feedback_Score': avg_feedback,
    'Total_Remedial_Classes': total_remedial,
    'Remedial_Impact': remedial_impact,
    'Risk_Flag': dropout_risk_flag,
    'Strengths': "; ".join(strengths) if strengths else "None",
    'Recommendations': "; ".join(recommendations) if recommendations else "None"
}
summary_df = pd.DataFrame([summary_data])

# Plot Attendance Trend
plt.figure(figsize=(8,4))
```

```
    plt.plot(attendance_trend['Trimester'], attendance_trend['Attendance_Percentage'], marker='o', linestyle='-
')
    plt.title(f'Attendance Trend for {student}')
    plt.xlabel('Trimester')
    plt.ylabel('Average Attendance Percentage')
    plt.ylim(0, 100)
    plt.grid(True)
    plt.tight_layout()
    plt.savefig(f'{output_dir}/{student}_attendance_trend.png')
    plt.close()

    # Plot Subject-wise Score Comparison
    plt.figure(figsize=(10,5))
    index = range(len(subject_summary))
    bar_width = 0.35
    plt.bar(index, subject_summary['Student_Avg_Score'], bar_width, label='Student Avg')
    plt.bar([i + bar_width for i in index], subject_summary['Class_Avg_Score'], bar_width, label='Class Avg')
    plt.xlabel('Subject')
    plt.ylabel('Average Score')
    plt.title(f'Subject-wise Performance for {student}')
    plt.xticks([i + bar_width / 2 for i in index], subject_summary['Subject'], rotation=45)
    plt.legend()
    plt.tight_layout()
    plt.savefig(f'{output_dir}/{student}_subject_performance.png')
    plt.close()

    # Write CSV report with summary and links to charts
    with open(f'{output_dir}/{student}_report.csv', 'w') as f:
        f.write(f'Report Card for Student: {student}\n\n')
        f.write('Attendance Trend (per trimester):\n')
        attendance_trend.to_csv(f, index=False)
        f.write('\nSubject-wise Performance:\n')
        subject_summary.to_csv(f, index=False)
        f.write('\nSummary:\n')
        summary_df.to_csv(f, index=False)
        f.write('\nCharts:\n')
        f.write(f'Attendance Trend Chart: {student}_attendance_trend.png\n')
        f.write(f'Subject Performance Chart: {student}_subject_performance.png\n')

print(f'Generated automated report cards with charts for {len(students)} students in "{output_dir}"
directory.')
```

h)Business Model for Summer Income

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	Annual Fees Structure					Summer Remedial Candidates					Discount Candidates		
3													
4	Standard	Annual Fees(rupees)				Name	Standard	Remedial Fees			Name	Standard	Discount
5	4	30,000				Sidharth	8	7500			Aarna	6	2000
6	5	35,000				Neev	7	6750			Anwesha	8	2500
7	6	40,000				Aryaman	6	6000			Nisha	9	2750
8	7	45,000				Twisha	7	6750			Anay	7	2250
9	8	50,000				Diya	8	7500			Ishaan	4	1500
10	9	55,000				Kabir	8	7500			Shlok	4	1500
11						Vivaan	7	6750					12500
12						Hriday	5	5250					
13						Manav	6	6000					
14								60000					
15													
16													
17													
18						Net Summer Revenue		47500					
19													
20													
21													
22													

Figure 13: Proposed Business Model

- Remedial Fee Structure: Each low academic performance student is charged approximately 10% of their grade's annual fee for summer remedial classes. This is reflected in the 'Remedial Fees' column for each candidate, resulting in a total remedial fee inflow of ₹60,000.
- High Performer Discounts: The 'Discount Candidates' section recognizes high achievers, each receiving a 5% annual fee waiver as an academic incentive. The aggregate discount distributed to these students is ₹12,500.
- Net Summer Revenue: After subtracting scholarship/discount payouts, net summer revenue is ₹47,500, offering a clear projection of additional summer earnings from this dual-incentive model.

Recommendations

- This tiered approach balances financial inclusion for remedial support and motivation for excellence, supporting both business and educational goals.
- Review remedial participation and feedback after the season; if uptake is lower than expected, gradually reduce the remedial fee percentage or test flexible payment options.
- Consider alternative or supplementary rewards for high achievers (certificates, gifts) as the program scales, this can further enhance the center's reputation and student morale without significantly reducing revenue.