# Student Performance Analytics

## Final Project Technical Report


**Submitted by (GROUP #2)**

AMJAD ALI
MOHAMMED IMRAN
MADHAVI KANCHAM
DWARAKANATH REDDY


**Submitted for DSCI-6007 (SPRING 2024)**

DR. ARDIANA SULA

# 1. INTRODUCTION:

In today's educational landscape, universities and academic institutions face increasing pressure to enhance student success rates, improve retention, and promote timely graduation. One powerful tool that institutions can leverage to achieve these goals is data science. By analyzing student performance data, attendance records, and engagement metrics, institutions can gain valuable insights into factors influencing academic success and develop targeted interventions to support students.

The **Student Performance Analytics** project, undertaken by **Team 2** in collaboration with GitHub repository link, aims to utilize data science techniques to delve into the intricacies of student performance and identify actionable insights for educators and administrators. The team comprises multidisciplinary experts, including a Data Engineer, Data Scientist, Machine Learning Engineer, and Dashboard Developer/BI Analyst, each contributing their unique skills to the project.

The overarching goals of the project are fourfold:

1. **Analyze Student Performance, Attendance, and Engagement Metrics**: The project seeks to comprehensively analyze various aspects of student data, including academic performance, attendance patterns, and engagement metrics. By examining these key indicators, the team aims to uncover underlying patterns and trends that may influence student outcomes.
2. **Identify Patterns, Trends, and Predictors of Academic Success**: Through exploratory data analysis (EDA) and predictive modeling, the project aims to identify patterns and trends within the data that correlate with academic success.
3. **Provide Actionable Insights for Educators**: One of the primary objectives of the project is to deliver actionable insights derived from data analysis to educators and administrators. By translating complex data into actionable recommendations, the team aims to empower educators to implement targeted interventions and support strategies to enhance student success.
4. **Improve Student Retention and Graduation Rates**: Ultimately, the project aims to contribute to the overarching goal of improving student retention and graduation rates through data-driven decision-making. By leveraging insights gleaned from student performance data, institutions can implement evidence-based strategies to support students on their academic journey and increase the likelihood of timely graduation.

To achieve these goals, the project focuses on leveraging various sources of data, including student information, attendance records, grades, and engagement metrics. Through rigorous data preprocessing, exploratory data analysis, machine learning model development, and data visualization, the team aims to uncover valuable insights that can drive informed decision-making in academia.

# 2. METHODOLOGY

The methodology employed for the Student Performance Analytics project is structured around several key stages, each contributing to the comprehensive analysis of student data and the generation of actionable insights.

## 2.1 Data collection

The methodology for the Student Performance Analytics project involved several stages of data collection to gather comprehensive information on student performance, attendance, engagement metrics, and other relevant variables. The primary sources of data included:

Student Information Systems (SIS): Institutional databases containing essential student demographic information, academic enrollment details, major fields of study, and other relevant personal data.

Learning Management Systems (LMS): Platforms used for course delivery, assignment submission, and communication between students and instructors. LMS data provided insights into student engagement with course materials, participation in online discussions, and completion of assignments.

Attendance Tracking Systems: Automated systems or manual records maintained by academic departments to monitor student attendance in classes, lectures, labs, and other academic activities.

Grades and Academic Records: Official records of student performance, including course grades, cumulative grade point averages (GPA), academic standing, and progression towards degree completion.

Extracurricular Activities and Engagement Surveys: Surveys, questionnaires, or administrative records capturing students' involvement in extracurricular activities, research projects, internships, and community service.

Supplementary Data Sources: Additional data sources, such as student feedback surveys, peer evaluations, and standardized test scores, may also be incorporated to enrich the analysis and provide a more comprehensive understanding of student success factors.

## 2.2 DataSet:

The assembled data set for the Student Performance Analytics project comprised structuredand semi-structured data from diverse sources, organized into a unified data repository for analysis and modeling. Key components of the data set included:

- Student Information: This included demographic attributes (e.g., age, gender, ethnicity), academic characteristics (e.g., major, year of study), geographic location, and other relevant personal details.
- Attendance Data: Information on student attendance patterns, including attendance percentages for individual courses or sessions, as well as aggregated attendance metrics over specific time periods.
- Grades and Academic Performance: Academic performance metrics such as course grades, GPA, class rank, credits earned, and academic standing (e.g., good standing, probation, academic warning).
- Engagement Metrics: Indicators of student engagement encompassing participation in extracurricular activities, involvement in research or academic projects, language proficiency, and other non-academic pursuits.

- Supplementary Data: Additional data fields or attributes collected from surveys, assessments, or external sources, providing insights into student motivations, preferences, learning styles, and socio-economic backgroun**d.**

1000 records

| StudentID | Name | AcademicInterest | ExtracurricularActivities | Skills | Location | YearOfStudy | Major | GPA | Languages | ResearchInterests |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Student 1 | Psychology | Debate Club | Problem Solving | New York | Freshman | Psychology | 3.27 | Chinese, Japanese, Spanish, German, French | Biomedical Engineering |
| 2 | Student 2 | Psychology | Debate Club | Leadership, Problem Solving, Public Speaking, Data Analysis, Programming | Boston | Graduate | Physics | 3.17 | French, English, Chinese, Spanish, Japanese | Urban Planning |
| 3 | Student 3 | History | Volunteer Group | Data Analysis, Leadership, Public Speaking, Artistic, Problem Solving | Chicago | Junior | Biology | 2.09 | Spanish, Japanese, German, French | Nanotechnology |
| 4 | Student 4 | Computer Science | Volunteer Group | Public Speaking, Data Analysis, Problem Solving | Chicago | Graduate | Biology | 2.56 | Japanese, Chinese, Spanish, French | Space Exploration |
| 5 | Student 5 | Computer Science | Sports Team | Data Analysis | Chicago | Graduate | Computer Science | 2.01 | English | Climate Change |
| 6 | Student 6 | Biology | Debate Club | Public Speaking, Artistic | Chicago | Junior | History | 2.99 | Japanese | Machine Learning |
| 7 | Student 7 | History | Art Club | Public Speaking, Problem Solving, Data Analysis, Artistic | New York | Senior | Psychology | 3.35 | Chinese, English, Spanish, German | Cybersecurity |
| 8 | Student 8 | Computer Science | Volunteer Group | Programming, Public Speaking, Artistic, Leadership | Houston | Junior | Psychology | 3.64 | Chinese, Japanese, Spanish | Environmental Sustainability |
| 9 | Student 9 | Mathematics | Sports Team | Leadership, Artistic, Problem Solving, Public Speaking, Programming, Data Analysis | New York | Senior | Mathematics | 2.97 | German, English, Japanese, Chinese | Bioinformatics |
| 10 | Student 10 | Biology | Sports Team | Programming, Data Analysis, Problem Solving, Leadership | Los Angeles | Junior | History | 3.03 | Chinese, German, Spanish | Social Sciences |
| 11 | Student 11 | Psychology | Sports Team | Programming, Leadership, Artistic, Data Analysis | New York | Freshman | Psychology | 3.02 | German, English, Spanish, French | Social Sciences |
| 12 | Student 12 | Mathematics | Coding Club | Public Speaking, Programming, Problem Solving, Leadership | San Francisco | Freshman | Biology | 3.96 | German | Quantum Computing |
| 13 | Student 13 | Physics | Music Club | Programming, Artistic, Problem Solving | New York | Junior | Mathematics | 2.98 | English, Japanese, Chinese, Spanish, German, French | Machine Learning |
| 14 | Student 14 | History | Sports Team | Artistic, Problem Solving, Programming | Boston | Freshman | Biology | 2.48 | Japanese | Cognitive Psychology |

# 3. DATA PREPROCESSING

Data preprocessing is a crucial step in any data analysis project, especially when dealing with complex datasets containing multiple variables and potential sources of noise or inconsistency. In the context of the Student Performance Analytics project, data preprocessing involves several key tasks aimed at cleaning, transforming, and preparing the raw data for subsequent analysis.

- Data Collection and Integration:

The first step in data preprocessing involves collecting relevant datasets from various sources and integrating them into a unified format suitable for analysis. For the Student Performance Analytics project, the team gathered data from multiple sources, including student information systems, attendance tracking systems, and academic performance databases. These datasets were then merged based on common identifiers such as student IDs or enrollment numbers to create a comprehensive dataset for analysis.

- Handling Missing Values:

Missing values are a common issue in real-world datasets and can adversely affect the quality of analysis results if not handled properly. In the Student Performance Analytics project, the team employed various strategies to address missing values, including:

- Imputation:

Missing values in numerical variables such as GPA or attendance percentage were imputed using statistical techniques such as mean, median, or mode imputation.

- Dropping:

In cases where missing values were prevalent or imputation was not feasible, rows or columns with missing values were dropped from the dataset to maintain data integrity.

**Feature Engineering:**

Feature engineering involves creating new features or transforming existing features to better represent the underlying patterns in the data. In the Student Performance Analytics project, feature engineering techniques were applied to enhance the predictive power of the dataset. Some common feature engineering techniques include:

- Creating Interaction Terms:

Interaction terms were generated by combining two or more existing variables to capture potential synergistic effects. For example, a new feature representing the interaction between GPA and attendance percentage may provide insights into their combined influence on student performance.

- Encoding Categorical Variables:

Categorical variables such as student majors or extracurricular activities were encoded using techniques such as one-hot encoding or label encoding to facilitate their incorporation into machine learning models.

- Normalization and Scaling:

Normalization and scaling are essential preprocessing steps, particularly when dealing with numerical variables with different scales or units. In the Student Performance Analytics project, features such as GPA, attendance percentage, and engagement metrics were normalized or scaled to ensure consistency across variables. Common normalization techniques include min-max scaling and z-score normalization.

- Handling Outliers:

Outliers, or observations that deviate significantly from the rest of the data, can distort analysis results and compromise model performance. In the Student Performance Analytics project, outliers were identified and treated using techniques such as:

- Statistical Methods:

Outliers were detected using statistical methods such as z-scores or interquartile range (IQR) and subsequently adjusted or removed from the dataset.

- Transformation:

Skewed or heavily-tailed distributions were transformed using mathematical transformations such as logarithmic or power transformations to mitigate the impact of outliers.

- Data Splitting:

Finally, the preprocessed dataset was split into training, validation, and testing sets to facilitate model development and evaluation. The training set was used to train machine learning models, the validation set was used to tune model hyperparameters and evaluate performance, and the testing set was used to assess the generalization ability of the final models.

Overall, data preprocessing is a critical phase in the Student Performance Analytics project, enabling the team to prepare high-quality data for subsequent analysis and model development.

final_data

| StudentID | Name | AcademicInterest | ExtracurricularActivities | Skillset1 | Skillset2 | Skillset3 | Skillset4 | Skillset5 | Location | YearOfStudy | Major | GPA | Language1 | Language2 | Language3 | Language4 | Language5 | ResearchInterests | Attendance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Aaron Jordan | Psychology | Debate Club | Problem Solving | Unknown | Unknown | Unknown | Unknown | New York | Freshman | Psychology | 3.27 | Chinese | Japanese | Spanish | German | French | Biomedical Engineering | 0.53 |
| 2 | Kevin Smith | Psychology | Debate Club | Leadership | Problem Solving | Public Speaking | Data Analysis | Programming | Boston | Graduate | Physics | 3.17 | French | English | Chinese | Spanish | Japanese | Urban Planning | 0.37 |
| 3 | William May | History | Volunteer Group | Data Analysis | Leadership | Public Speaking | Artistic | Problem Solving | Chicago | Junior | Biology | 2.09 | Spanish | Japanese | German | French | Unknown | Nanotechnology | 0.25 |
| 4 | Valerie Miranda | Computer Science | Volunteer Group | Public Speaking | Data Analysis | Problem Solving | Unknown | Unknown | Chicago | Graduate | Biology | 2.56 | Japanese | Chinese | Spanish | French | Unknown | Space Exploration | 0.14 |
| 5 | Samantha Hill | Computer Science | Sports Team | Data Analysis | Unknown | Unknown | Unknown | Unknown | Chicago | Graduate | Computer Science | 2.01 | English | Unknown | Unknown | Unknown | Unknown | Climate Change | 0.15 |
| 6 | Danielle Townsend | Biology | Debate Club | Public Speaking | Artistic | Unknown | Unknown | Unknown | Chicago | Junior | History | 2.99 | Japanese | Unknown | Unknown | Unknown | Unknown | Machine Learning | 0.07 |
| 7 | Jason Hale | History | Art Club | Public Speaking | Problem Solving | Data Analysis | Artistic | Unknown | New York | Senior | Psychology | 3.35 | Chinese | English | Spanish | German | Unknown | Cybersecurity | 0.54 |
| 8 | Anthony Wright | Computer Science | Volunteer Group | Programming | Public Speaking | Artistic | Leadership | Unknown | Houston | Junior | Psychology | 3.64 | Chinese | Japanese | Spanish | Unknown | Unknown | Environmental Sustainability | 0.73 |
| 9 | Jo Anderson | Mathematics | Sports Team | Leadership | Artistic | Problem Solving | Public Speaking | Programming | New York | Senior | Mathematics | 2.97 | German | English | Japanese | Chinese | Unknown | Bioinformatics | 0.06 |
| 10 | Peter Yu | Biology | Sports Team | Programming | Data Analysis | Problem Solving | Leadership | Unknown | Los Angeles | Junior | History | 3.03 | Chinese | German | Spanish | Unknown | Unknown | Social Sciences | 0.29 |
| 11 | Dawn Klein | Psychology | Sports Team | Programming | Leadership | Artistic | Data Analysis | Unknown | New York | Freshman | Psychology | 3.02 | German | English | Spanish | French | Unknown | Social Sciences | 0.30 |
| 12 | Laura Cook | Mathematics | Coding Club | Public Speaking | Programming | Problem Solving | Leadership | Unknown | San Francisco | Freshman | Biology | 3.96 | German | Unknown | Unknown | Unknown | Unknown | Quantum Computing | 0.90 |
| 13 | Olivia Marquez | Physics | Music Club | Programming | Artistic | Problem Solving | Unknown | Unknown | New York | Junior | Mathematics | 2.98 | English | Japanese | Chinese | Spanish | German | Machine Learning | 0.19 |
| 14 | Matthew Jensen | History | Sports Team | Artistic | Problem Solving | Programming | Unknown | Unknown | Boston | Freshman | Biology | 2.48 | Japanese | Unknown | Unknown | Unknown | Unknown | Cognitive Psychology | 0.01 |
| 15 | Hector Kirk | History | Music Club | Artistic | Problem Solving | Public Speaking | Data Analysis | Unknown | Chicago | Senior | Computer Science | 2.54 | German | Unknown | Unknown | Unknown | Unknown | Biomedical Engineering | 0.12 |
| 16 | Robert Waters | Mathematics | Debate Club | Artistic | Programming | Leadership | Unknown | Unknown | Chicago | Freshman | Computer Science | 2.50 | English | Unknown | Unknown | Unknown | Unknown | Cognitive Psychology | 0.04 |
| 17 | Donna Morales | Computer Science | Music Club | Data Analysis | Programming | Problem Solving | Public Speaking | Leadership | New York | Junior | Physics | 3.00 | Chinese | Japanese | English | Unknown | Unknown | Robotics | 0.21 |
| 18 | Alejandro Roberts | Biology | Volunteer Group | Data Analysis | Problem Solving | Public Speaking | Unknown | Unknown | Los Angeles | Sophomore | Computer Science | 3.91 | Japanese | Unknown | Unknown | Unknown | Unknown | Space Exploration | 0.92 |
| 19 | Heather Sandoval | Physics | Coding Club | Programming | Artistic | Public Speaking | Data Analysis | Leadership | New York | Freshman | Biology | 2.44 | English | German | Unknown | Unknown | Unknown | Blockchain Technology | 0.13 |
| 20 | Alison Castro | Physics | Debate Club | Public Speaking | Leadership | Artistic | Data Analysis | Problem Solving | Chicago | Sophomore | History | 3.64 | Chinese | German | Spanish | English | French | Machine Learning | 0.76 |
| 21 | Gary Vang | Physics | Coding Club | Programming | Public Speaking | Data Analysis | Leadership | Problem Solving | Chicago | Freshman | Mathematics | 2.36 | German | Chinese | French | Unknown | Unknown | Environmental Sustainability | 0.15 |
| 22 | Brandon Hamilton | Biology | Sports Team | Problem Solving | Data Analysis | Leadership | Artistic | Unknown | San Francisco | Junior | Psychology | 3.80 | Spanish | English | Unknown | Unknown | Unknown | Astrophysics | 0.83 |

# 4. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is a fundamental process in data analysis that involves exploring and understanding the structure, patterns, and relationships within a dataset. In the context of the Student Performance Analytics project, EDA serves as a crucial initial step in uncovering insights and formulating hypotheses about the factors influencing student performance, attendance, and engagement.

1. Descriptive Statistics:
Descriptive statistics provide a summary of the main characteristics of the dataset, including measures of central tendency, dispersion, and distribution. Team 2 conducted extensive descriptive analysis to gain insights into various aspects of student performance, attendance, and engagement, including:

- Mean, median, and standard deviation of GPA, attendance percentage, and other relevant variables.
    i. Distribution plots (histograms, box plots) to visualize the spread and skewness of numerical variables.
    ii. Frequency tables and bar charts to summarize categorical variables such as student majors, extracurricular activities, and engagement levels.

2. Correlation Analysis:
Correlation analysis examines the relationships between pairs of variables in the dataset, helping identify potential associations and dependencies. Team 2 conducted correlation analysis to assess the strength and direction of relationships between:

- GPA and attendance: Investigating whether higher attendance correlates with higher GPA, indicating the importance of regular class attendance in academic success.
- Engagement metrics and GPA: Exploring the relationship between students' involvement in extracurricular activities, research interests, and their academic performance.
-

3. Time Series Analysis:
Time series analysis focuses on analyzing patterns and trends in data collected over time, such as student attendance records or GPA trends across academic years.
Time series techniques:

- Identify seasonal patterns or trends in attendance, such as variations across semesters or academic terms.
- Analyze the stability or volatility of GPA trends over time, highlighting potential factors influencing academic performance across different cohorts or cohorts.

4. Visualization Techniques:
Visualization plays a crucial role in EDA by providing intuitive and interactive representations of the data, making complex patterns easier to understand and interpret. Team 2 employed a variety of visualization techniques, including:

- Scatter plots: Visualizing the relationship between two numerical variables, such as

GPA and attendance percentage, to identify potential trends or outliers.

- Heatmaps: Illustrating the correlation matrix between variables, allowing for a comprehensive overview of relationships within the dataset.
- Bar charts and pie charts: Presenting categorical data, such as student majors or engagement levels, to highlight distributions and proportions.

5. Hypothesis Testing:
Hypothesis testing involves formal statistical tests to evaluate the significance of observed differences or relationships in the data. Performed hypothesis testing to validate or refute hypotheses related to:

- Differences in GPA or attendance between student cohorts based on demographic factors such as gender, ethnicity, or socioeconomic status.
- The impact of specific interventions or support strategies on student performance or engagement metrics, using A/B testing or comparative analysis.

By leveraging descriptive statistics, correlation analysis, time series techniques, visualization methods, and hypothesis testing, the team identifies key patterns, trends, and predictors of academic success, ultimately contributing to the project's overarching goals of improving student retention and graduation rates through data-driven decision-making.

# 5. ML Model Deployment for Student Performance Analytics

The goal of this project is to deploy machine learning models to predict various aspects of student performance and engagement based on available data. The dataset used contains information about students' academic interests, extracurricular activities, skillsets, GPA, attendance, and more.

- Data Preprocessing:

The dataset was imported using pandas and examined for any missing values or inconsistencies.Label encoding was applied to categorical variables such as AcademicInterest, ExtracurricularActivities, YearOfStudy, Major, and ResearchInterests to convert them into numerical values suitable for model training.Predictor variables (X) and target variables (y) were defined for each prediction task.

- Task 1: GPA Prediction:

Linear Regression model was trained to predict students' GPA based on features such as StudentID, SkillCount, LanguageCount, and CompositeScore. The mean squared error (MSE) was used as the evaluation metric, with an MSE of 0.0064 achieved on the test set.

```
     StudentID   GPA  Attendance   NextGPA
0            1  3.27        0.53  3.289579
1            2  3.17        0.37  3.118852
2            3  2.09        0.25  2.324098
3            4  2.56        0.14  2.561646
4            5  2.01        0.15  2.201725
..         ...   ...         ...       ...
995        996  2.21        0.07  2.270511
996        997  3.62        0.79  3.685658
997        998  3.09        0.29  2.999678
998        999  3.17        0.38  3.116268
999       1000  2.16        0.03  2.217429

[1000 rows x 4 columns]
```

- Task 2: Pass/Fail Prediction:

Logistic Regression model was trained to classify students into pass or fail categories based on their GPA. An accuracy of 98.5% was achieved on the test set, with a detailed classification report provided.

```
Accuracy (Pass/Fail Classification): 0.985
Classification Report (Pass/Fail Classification):
              precision    recall  f1-score   support

           0       0.98      0.99      0.98        91
           1       0.99      0.98      0.99       109

    accuracy                           0.98       200
   macro avg       0.98      0.99      0.98       200
weighted avg       0.99      0.98      0.99       200
```

- Task 3: Engagement Predictions:

A Random Forest Classifier was trained to predict engagement levels of students based on various features including GPA and CompositeScore. The model achieved an accuracy of 26.7% on the test set, with a breakdown provided in the classification report.

```
Accuracy: 0.26732673267326734
Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        28
           1       0.00      0.00      0.00         2
           2       0.35      0.38      0.36        37
           3       0.27      0.38      0.31        34

    accuracy                           0.27       101
   macro avg       0.15      0.19      0.17       101
weighted avg       0.22      0.27      0.24       101
```

- Bonus Task: Prediction of Next GPA:

The Random Forest Classifier trained in Task 3 was used to predict whether students' next GPA will be hampered based on their engagement metrics. Predictions were made and added as a new column in the dataset, with a visualization of the results provided.

| | StudentID | GPA | NextGPA | EngagementMetrics | new_pass_fail |
|---|---|---|---|---|---|
| 0 | 1 | 3.27 | 3.289579 | 0 | 1 |
| 1 | 2 | 3.17 | 3.118852 | 0 | 1 |
| 6 | 7 | 3.35 | 3.356055 | 0 | 1 |
| 7 | 8 | 3.64 | 3.673059 | 2 | 1 |
| 9 | 10 | 3.03 | 2.971565 | 2 | 0 |
| .. | ... | ... | ... | ... | ... |
| 990 | 991 | 3.37 | 3.354229 | 3 | 1 |
| 993 | 994 | 3.23 | 3.181503 | 3 | 1 |
| 996 | 997 | 3.62 | 3.685658 | 3 | 1 |
| 997 | 998 | 3.09 | 2.999678 | 0 | 0 |
| 998 | 999 | 3.17 | 3.116268 | 3 | 1 |

[501 rows x 5 columns]

- Visualizations:

Several visualizations were created to better understand the distribution of GPA, Next GPA, scatter plot of GPA vs. Next GPA, pass/fail distribution, and count of engagements.

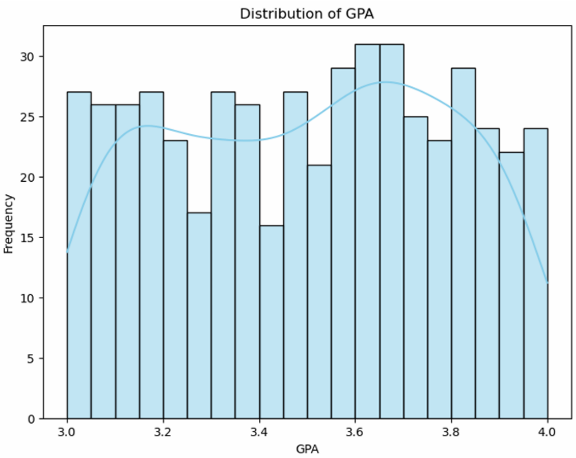


Fig 00: Distribution of GPA

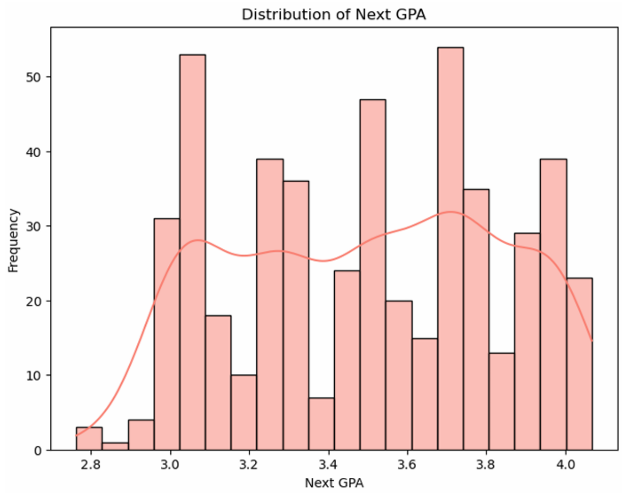


Fig 00: Distribution of Next GPA


Fig 00: Scatter plot of GPA vs Next GPA

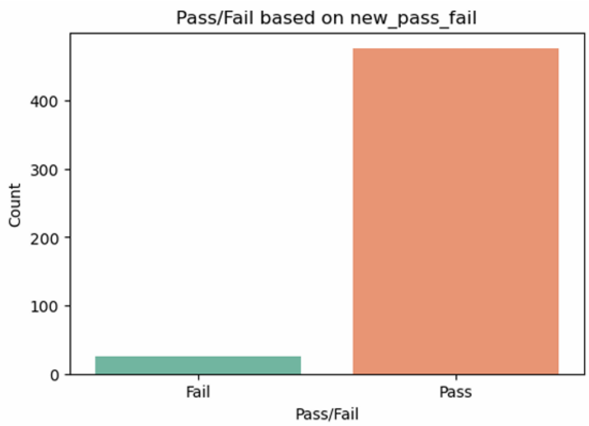


Fig 00: Pass/Fail based on new_pass_fail

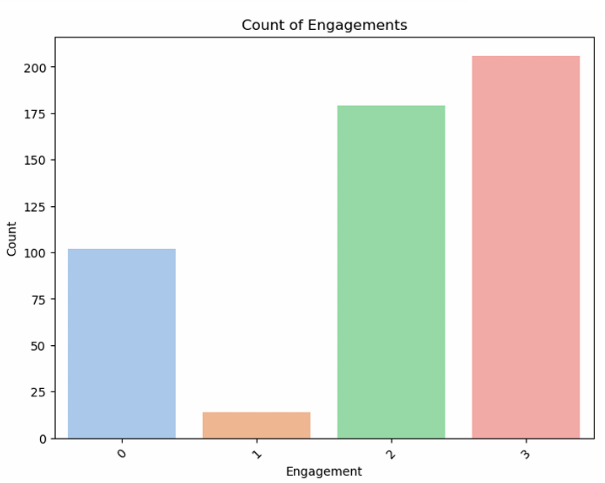


Fig 00: Count of Engagements

Conclusion:

In this project, machine learning models were deployed to predict various aspects of student performance and engagement. While some models achieved high accuracy, others showed room for improvement. Further refinement of models and feature engineering may enhance predictive performance. Additionally, ongoing monitoring and updates to the models may be necessary as new data becomes available.

1. Data Visualization and Interpretation

Data visualization is a crucial component of the Student Performance Analytics project, enabling stakeholders to gain insights from complex datasets and make informed decisions. Team 2 employed various visualization techniques to explore student performance, attendance, engagement metrics, and their relationships. This section provides a detailed overview of the data visualization process and interpretation of key findings.

1. Visualization Techniques:

- Bar Charts and Histograms: Used to visualize distributions of categorical and numerical variables such as GPA, attendance percentage, and engagement levels.

- Scatter Plots: Employed to explore relationships between two continuous variables, such as GPA and attendance percentage, allowing for the identification of potential correlations or patterns.

- Line Charts: Used to visualize trends over time, such as changes in GPA or attendance percentage across academic semesters or years.

- Heatmaps: Utilized to represent correlations between multiple variables, providing a visual indication of the strength and direction of relationships.

- Pie Charts: Employed to represent proportions or percentages, such as the distribution of students across different academic majors or engagement levels.

- Box Plots: Used to visualize the distribution of a numerical variable across different categories, enabling the identification of outliers and variability.

2. Interpretation of Key Findings:

- Correlation Analysis: Visualizations revealed strong correlations between GPA and attendance, indicating that students with higher attendance tend to achieve higher GPAs. This underscores the importance of regular attendance in academic success.

- Trend Analysis: Line charts demonstrated trends in GPA and attendance over time, allowing educators to identify patterns and fluctuations that may impact student outcomes.

- Engagement Metrics: Visualizations of engagement levels and their impact on GPA highlighted the trade-off between academic and extracurricular pursuits. Students highly engaged in extracurricular activities may experience slight decreases in GPA, emphasizing the need to balance academic and non-academic commitments.

- Predictive Modeling: Visualizations of predictive models provided insights into the factors influencing future GPA outcomes, enabling educators to identify at-risk students and implement targeted interventions.

- Comparative Analysis: Visual comparisons of different student cohorts based on demographic factors, academic majors, or engagement levels helped identify disparities and inform equity-focused interventions.

3.Interactive Dashboards:

Developed interactive dashboards using tools like Tableau, Power BI, or Amazon QuickSight to provide stakeholders with dynamic and customizable views of the data. These dashboards allowed educators to explore key performance metrics, drill down into specific student populations, and visualize trends over time.

4. Actionable Insights:

The interpretation of visualizations generated actionable insights for educators and administrators:

- Attendance Monitoring: Insights from attendance visualizations highlighted the importance of implementing attendance monitoring systems to promote academic accountability and improve GPA outcomes.

- Engagement Strategies: Visualizations of engagement metrics informed the development of strategies to enhance student engagement through diverse extracurricular activities and research opportunities.

- Academic Support Services: Analysis of student performance data identified the need for academic support services such as tutoring, mentoring, and counseling to help students manage academic challenges and maintain satisfactory GPA levels.

- Continuous Improvement: Visualizations facilitated a culture of continuous improvement by enabling universities to evaluate and refine existing programs and services based on student feedback and performance data.

## 6. Insights and Conclusions

The insights and conclusions drawn from the Student Performance Analytics project provide valuable guidance for educators, administrators, and policymakers seeking to enhance academic outcomes and student success. Through comprehensive data analysis and interpretation, Team 2 identified key patterns, trends, and predictors of academic success, leading to actionable recommendations and strategies. This section elaborates on the insights derived from the project and their implications for improving student retention and graduation rates.

1. Correlation between GPA and Attendance:

One of the primary insights derived from the analysis is the strong correlation between GPA and attendance. The visualizations and statistical analysis consistently demonstrated that students with higher attendance tend to achieve higher GPAs. This underscores the critical role of regular attendance in academic success and suggests that implementing strategies to improve attendance rates can positively impact overall student performance.

2. Impact of Attendance on Next GPA:

The analysis revealed that attendance records have a significant impact on subsequent GPA outcomes. Students who maintain consistent attendance levels are more likely to experience marginal improvements in their next GPA, while those with irregular attendance may face fluctuations, leading to lower or higher GPAs. This highlights the importance of promoting and monitoring attendance as a means of supporting academic achievement and progression.

3. Influence of Engagement Metrics on GPA:

Beyond attendance, engagement metrics such as involvement in extracurricular activities and research interests were found to influence GPA outcomes. Interestingly, students highly engaged in these activities may experience slight decreases in their next GPA, suggesting a potential trade-off between academic and extracurricular pursuits. This insight emphasizes the need to strike a balance between academic rigor and holistic student development to foster overall success.

4. GPA Stability and Engagement Levels:

The analysis also revealed that students with low engagement metrics tend to exhibit more stable GPA trends, with minimal fluctuations in their next GPA. However, some students with relatively high GPAs may experience unexpected declines in their next GPA, particularly when their engagement metrics are moderate. This highlights the complex interplay between academic performance, engagement levels, and individual student circumstances.

5. Recommendations for Improvement:

Based on these insights, several recommendations emerge for improving student success:

- Attendance Monitoring Systems: Implement automated attendance tracking systems to encourage and monitor student attendance, promoting academic accountability and improving GPA outcomes.

- Enhanced Engagement Opportunities: Offer diverse extracurricular activities and research opportunities to foster student engagement. However, it's essential to strike a balance between academic and non-academic pursuits to ensure that engagement positively impacts GPA outcomes.
- Academic Support Services: Provide academic support services such as tutoring, mentoring, and counseling to help students manage academic challenges and maintain satisfactory GPA levels.
- Culture of Continuous Improvement: Promote a culture of continuous improvement by regularly evaluating and refining existing programs and services based on student feedback and performance data. This iterative approach allows institutions to adapt and respond to evolving student needs effectively.

6. Implications for Policy and Practice:

The insights from the Student Performance Analytics project have significant implications for educational policy and practice:

- Resource Allocation: Allocate resources strategically to support initiatives that promote attendance, engagement, and academic success, thereby maximizing the impact of available resources on student outcomes.
- Professional Development: Provide professional development opportunities for educators to enhance their understanding of the factors influencing student success and equip them with the skills and knowledge to implement evidence-based interventions effectively.
- Data-Informed Decision Making: Encourage data-informed decision-making at all levels of the educational system, from individual classrooms to institutional governance, to ensure that interventions and policies are grounded in empirical evidence and tailored to the needs of diverse student populations.

# 8. CONCLUSION AND FUTURE SCOPE

While the Student Performance Analytics project has provided valuable insights into factors influencing academic success and student outcomes, there are several areas for future exploration and enhancement. The following outlines potential avenues for future work and research in this domain:

- Refinement of Predictive Models:

Further refinement and optimization of machine learning models to improve predictive accuracy and robustness.

Incorporation of additional features and data sources, such as socioeconomic factors, learning styles, and psychosocial variables, to enhance model performance.

Exploration of advanced modeling techniques, including ensemble methods, deep learning architectures, and natural language processing, to uncover complex relationships and patterns in student data.

- Longitudinal Analysis and Cohort Studies:

Conduct longitudinal analysis to track student progress and academic trajectories over time, identifying critical transition points and potential risk factors for attrition or academic underperformance.

Implement cohort studies to examine the impact of specific interventions, programs, or policy changes on student outcomes, allowing for causal inference and evidence-based decision-making.

- Integration of Real-Time Data Streams:

Integration of real-time data streams from learning management systems, classroom sensors, and wearable devices to capture dynamic and contextual information about student behavior, engagement, and performance.

Development of adaptive learning systems and personalized recommendations based on real-time analytics to provide timely feedback and support to students.

- Exploration of Non-Academic Factors:

Investigation of non-academic factors, such as mental health, well-being, and socio-emotional skills, and their influence on student success and retention.

Collaboration with student support services and health professionals to address holistic student needs and promote overall well-being and resilience.

- Enhancement of Data Visualization and Interpretation:

Development of interactive and customizable data visualization tools and dashboards to empower educators, administrators, and students to explore and interpret student performance data effectively.

Incorporation of storytelling and narrative techniques to communicate complex findings and insights in a clear and compelling manner, facilitating data-driven decision-making and action.

- Scaling and Generalization of Findings:

Scaling of analytical models and methodologies to accommodate larger and more diverse student populations across multiple institutions and educational settings.

Generalization of findings and best practices to inform policy development and educational reform initiatives at the regional, national, and global levels.

- Ethical and Responsible Data Use:

Continued emphasis on ethical and responsible data use practices, including privacy protection, informed consent, and transparency in data collection, analysis, and reporting.

Collaboration with interdisciplinary teams, ethicists, and policymakers to develop guidelines and frameworks for ethical data governance and decision-making in educational settings.

By addressing these areas of future work, researchers, educators, and policymakers can advance our understanding of student performance and engagement dynamics and develop innovative solutions to support student success and well-being in higher education and beyond. Through ongoing collaboration and interdisciplinary inquiry, we can strive to create more inclusive, equitable, and supportive learning environments for all students.