Project Title:

Predicting Student Performance and Risk of Failure using Data Analytics and Machine Learning
_____

1. Background & Business Objective

Educational institutions generate large volumes of data from student demographics, attendance records, exam results, teacher performance, extracurricular participation, and disciplinary actions.

However, most schools fail to leverage this data to:

- Identify struggling students early
- Improve teaching quality
- Optimize resource allocation
- Enhance student engagement & retention

This project focuses on building a data-driven decision-support system using SQL, Power BI, and Python to address these challenges.

2. Business Problem

- Student Performance Analysis
  - What is the overall pass/fail rates across grades and subjects?
  - How do demographics (gender, family income, parental education) affect performance?
- Attendance Insights
  - How does attendance % affect exam performance?
  - Which students are at risk due to chronic absenteeism?
- Teacher Effectiveness
  - Does teacher experience and qualification correlate with student outcomes?
  - Which teachers/subjects show consistently high/low performance?
- Extracurricular Activities
  - Do students who participate in activities perform better academically?
  - Which types of activities (sports, arts, STEM) correlate most with success?
- Disciplinary Impact
  - Do students with disciplinary records perform worse academically?
  - What types of incidents correlate most with underperformance?
- Predictive Modeling
  - Can we predict whether a student will pass/fail using available features (attendance, scores, family background, activities, discipline, teacher)?
  - Can we identify students most likely to drop out?

3. Goal

- Build a centralized relational database (SQL) for school data.
- Perform detailed analysis using SQL queries.
- Develop an interactive Power BI dashboard for administrators & teachers.
- Build a Python predictive model to identify at-risk students.
- Provide data-driven recommendations for policy and interventions.

4. Tools & Technologies

- **SQL (Snowflake)**: Data modeling, cleaning, transformations
- **Python (Pandas, NumPy, Seaborn, Scikit-learn)**: EDA, feature engineering, logistic regression modeling
- **Power BI**: Interactive dashboards and storytelling
- **Jupyter Notebook**: Data analysis and modeling workflow
- **GitHub**: Code and documentation repository

5. Deliverables

- SQL Scripts → insights & KPIs
- Power BI Dashboard → interactive storytelling
- Python Notebook → predictive modeling
- Final Report → insights, recommendations, visualizations
- Project Documentation → Proposal + PACE Strategy + Execution Summary

Data Analyst & Consultant: Shijin Ramesh

Date: 31st August, 2025