DatafestAfrica 2024 Datathon

09.10.2024

—

Team Rosh

Adebola Rabiu

Anuoluwapo Tenibiaje

Timilehin Olowolafe

1. Project Overview

Title:
Data-Driven Solution to Improve Student Performance in Ikorodu Community High School

Objective:
To design a data-driven system that helps improve students performance, focusing on preparing students for key standardized exams like JAMB and WASSCE. The solution will establish a robust data infrastructure, analyze key performance factors, and offer predictive insights and actionable recommendations.

Scope:
The solution will include data collection, storage, predictive modeling, and reporting, aiming to enhance performance through early interventions and data-driven decisions.

2. Problem Definition

Challenges Identified:

- Poor foundational education infrastructure.
- Limited data collection and utilization for improving academic outcomes.
- Low student performance in exams like JAMB (76% of students scored below 200 in 2024).
- Lack of insights into factors beyond academic scores, such as teacher impact, attendance, and resource allocation.

Goals:

- Understand key challenges affecting student performance.
- Develop a predictive model to identify students at risk of failing.
- Provide actionable insights to improve educational outcomes.

3. Solution Design

3.1. Data Collection Plan

Collection Methods:

- Digital surveys and forms for teachers and school administrators.
- Attendance tracking via digital or IoT systems.
- Periodic assessments and student engagement metrics.

Data Types:

Tables:

Academic Schema:

1. **Students**
    - **StudentID (PK, int)**
    - **FirstName (varchar(50))**
    - **LastName (varchar(50))**
    - **DateOfBirth (date)**
    - **Gender (varchar(10))**

- AdmissionDate (date)
- CurrentClass (varchar(10))

2. Classes
   - ClassID (PK, int)
   - ClassName (varchar(20))
   - ClassTeacherID (FK to Teachers, int)
3. Subjects
   - SubjectID (PK, int)
   - SubjectName (varchar(50))
   - SubjectCode (varchar(10))
4. ExamResults
   - ResultID (PK, int)
   - StudentID (FK to Students, int)
   - SubjectID (FK to Subjects, int)
   - Score (decimal(5,2))
   - Grade (varchar(2))
5. Attendance
   - AttendanceID (PK, int)
   - StudentID (FK to Students, int)
   - Date (date)
   - Status (varchar(10))

Administrative Schema:

6. Staff
   - StaffID (PK, int)
   - FirstName (varchar(50))
   - LastName (varchar(50))
   - Role (varchar(50))
   - Department (varchar(50))

Extracurricular Schema:

7. Clubs
   - ClubID (PK, int)
   - ClubName (varchar(50))
8. ClubMembership
   - MembershipID (PK, int)
   - StudentID (FK to Students, int)

- **ClubID (FK to Clubs, int)**
- **JoinDate (date)**

Facilities Schema:

9. Resources

- **ResourceID (PK, int)**
- **ResourceName (varchar(100))**
- **ResourceType (varchar(50))**
- **Quantity (int)**
- **LastMaintenanceDate (date)**

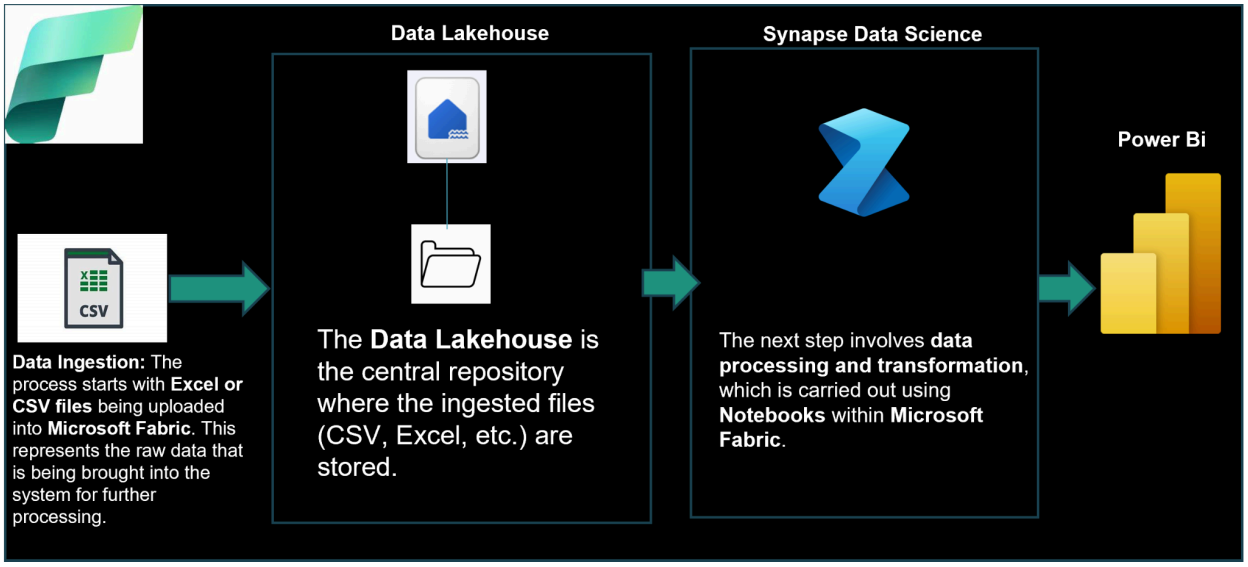
10. ResourceUsage

- **UsageID (PK, int)**
- **ResourceID (FK to Resources, int)**
- **UserID (int)**
- **UserType (varchar(20))**
- **UsageDate (date)**
- **Duration (int)**

- **Academic Schema::** Demographics, attendance, academic scores.
- **Facilities Schema:** Resources, facilities, internet access.
- **Extracurricular Activities:** Participation in non-academic programs.

All the data are collected in csv formats

Data Flow and Processing Pipeline



3.2. Data Lakehouse & Structuring

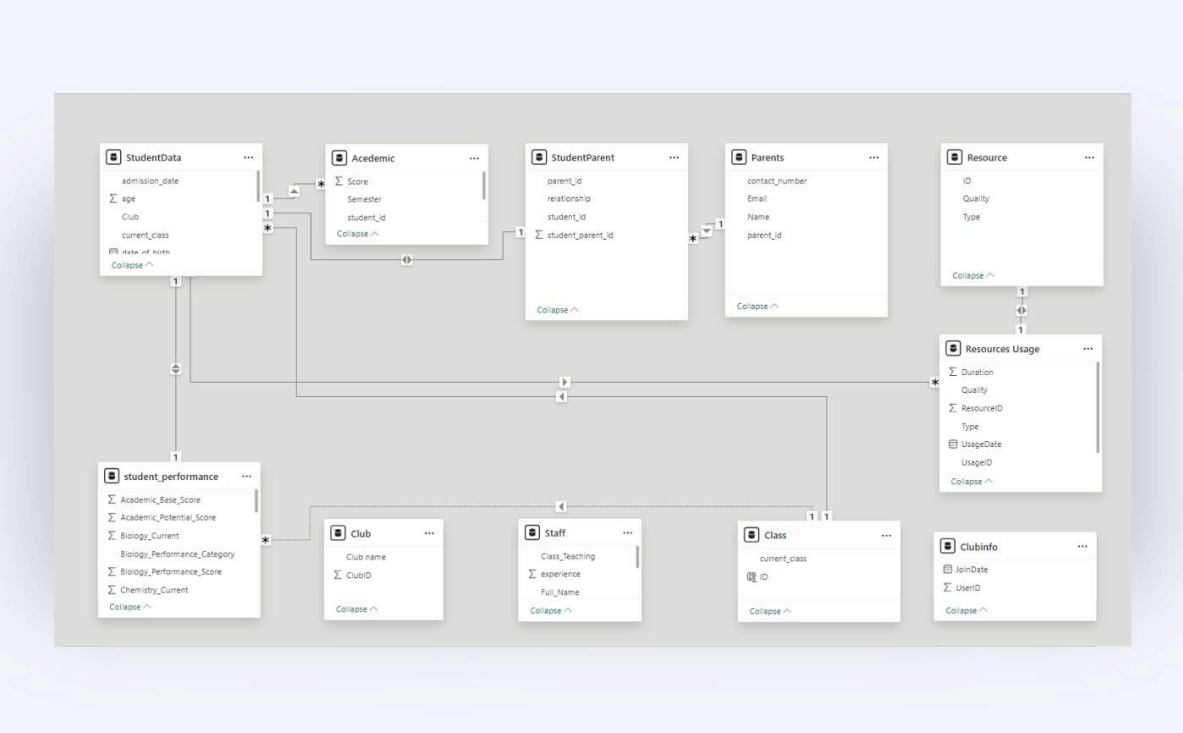
Ingesting the Excel/CSV File to the Lakehouse

We started by ingesting the CSV files, which contained both fact and dimension data, into Microsoft Fabric's Lakehouse which was used to store the data

Database Schema:

- **Tables:** StudentData,, Academics, Clubs,ClubsID, Resources, ResourcesUsage

- **Relationships:** Establish relationships between tables for a comprehensive view (e.g., students related to classes and teachers).



Data Storage:

- Centralized data lake designed to scale and integrate .
- Ensures secure and efficient access to data with role-based access controls.

Predictive Model:

- **Inputs:** Demographics, academic history, attendance, teacher performance, extracurricular participation.
- **Output:** Predicts likelihood of passing or failing upcoming exams.
- **Objective:** Early identification of at-risk students for targeted intervention.

After preparing the data, we transitioned to the data science phase:

Feature Engineering Process:

Academic Features

- Calculates current performance metrics per subject

- Generates progress metrics tracking improvement

- Measures consistency using score standard deviations
- Creates performance categories (Needs Improvement to Excellent)

Demographic Features

- Derives age categories
- Calculates socioeconomic scores
- Generates learning environment scores
- Creates engagement scores based on participation and resource usage

Performance Metric Calculation

The system generates three key metrics:

1. Academic Base Score
   - Normalized average of current subject scores

2. Academic Potential Score
   - Weighted combination of:
     - Academic base (40%)
     - Progress (20%)
     - Engagement (20%)
     - Learning environment (10%)
     - Consistency (10%)

3. Risk Factor Score
   - Identifies at-risk students using:
     - Academic performance
     - Learning disabilities
     - Parental education
     - Engagement levels

4. Clustering Analysis

- Processes data for each class level separately
- Uses multiple performance indicators for clustering
- Determines optimal cluster count (2-8 clusters)

Key Considerations

- Adjusts calculations based on class levels
- Handles missing data gracefully
- Uses class-specific thresholds
- Weights factors based on empirical importance
- Ensures fair comparison within class levels

Once the data science tasks were complete, we saved the processed data back into the Lakehouse as a CSV file making it accessible for further analysis or reporting.

---

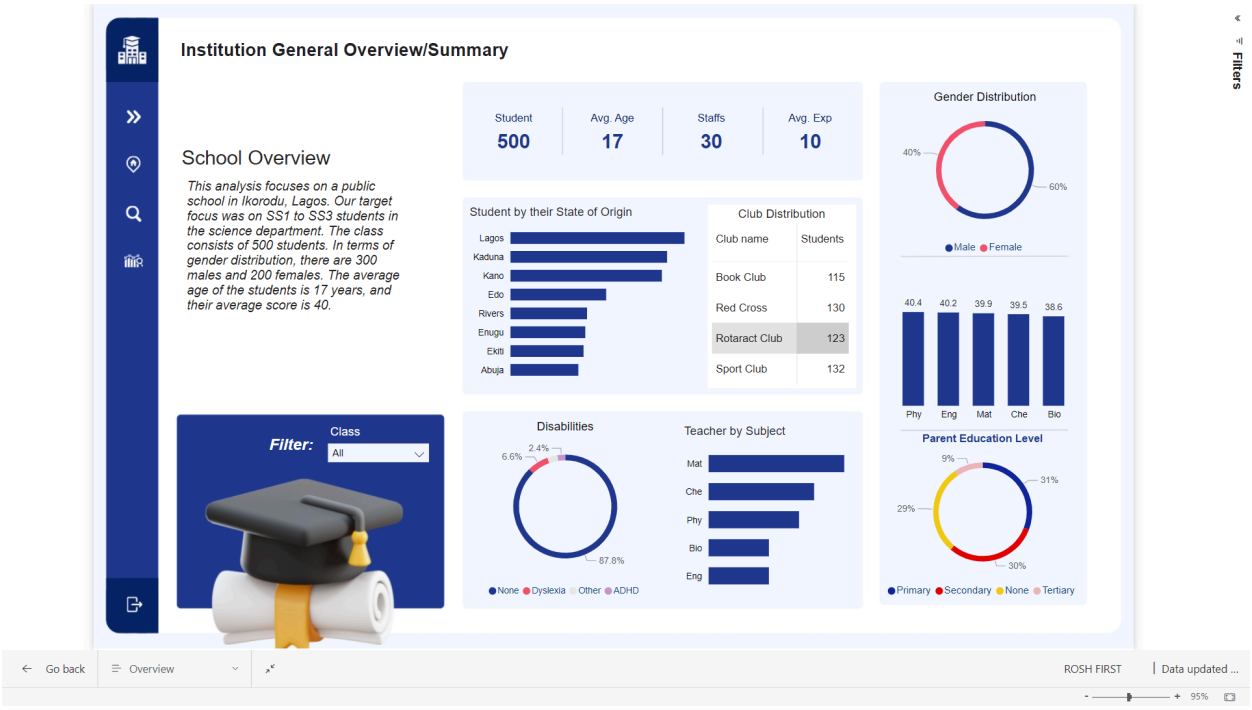
4. Reporting and Visualization

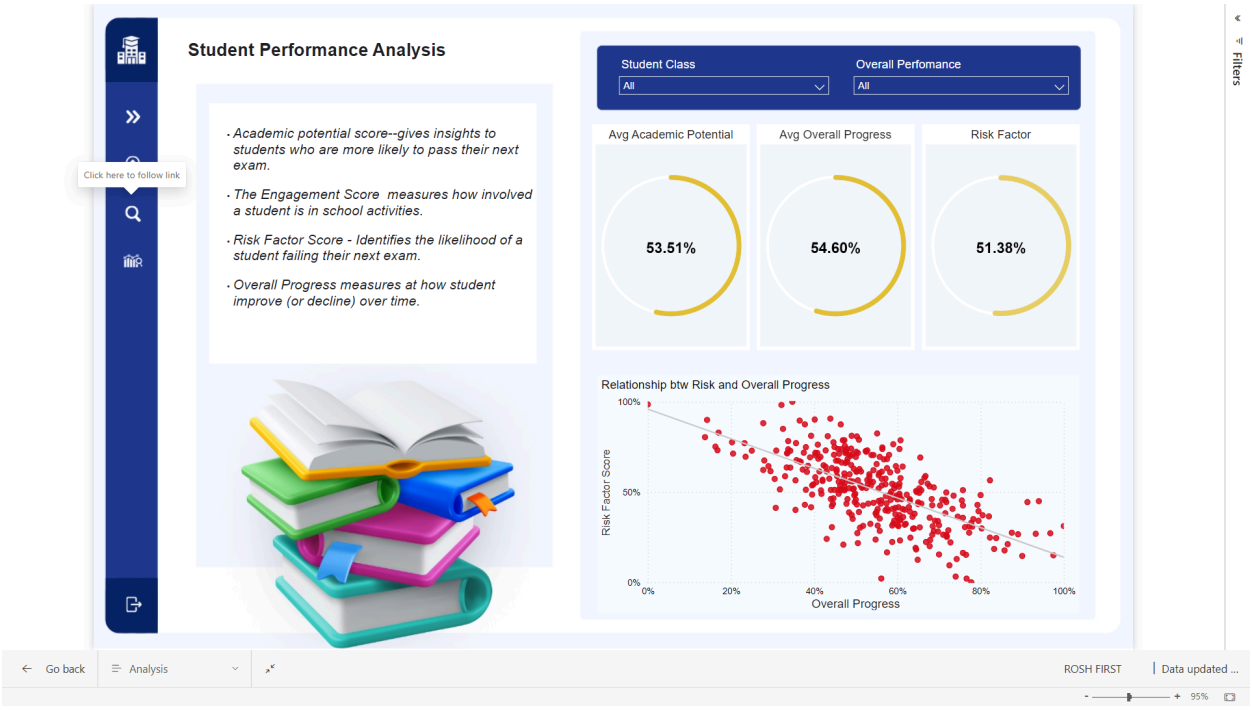
Interactive Dashboard:

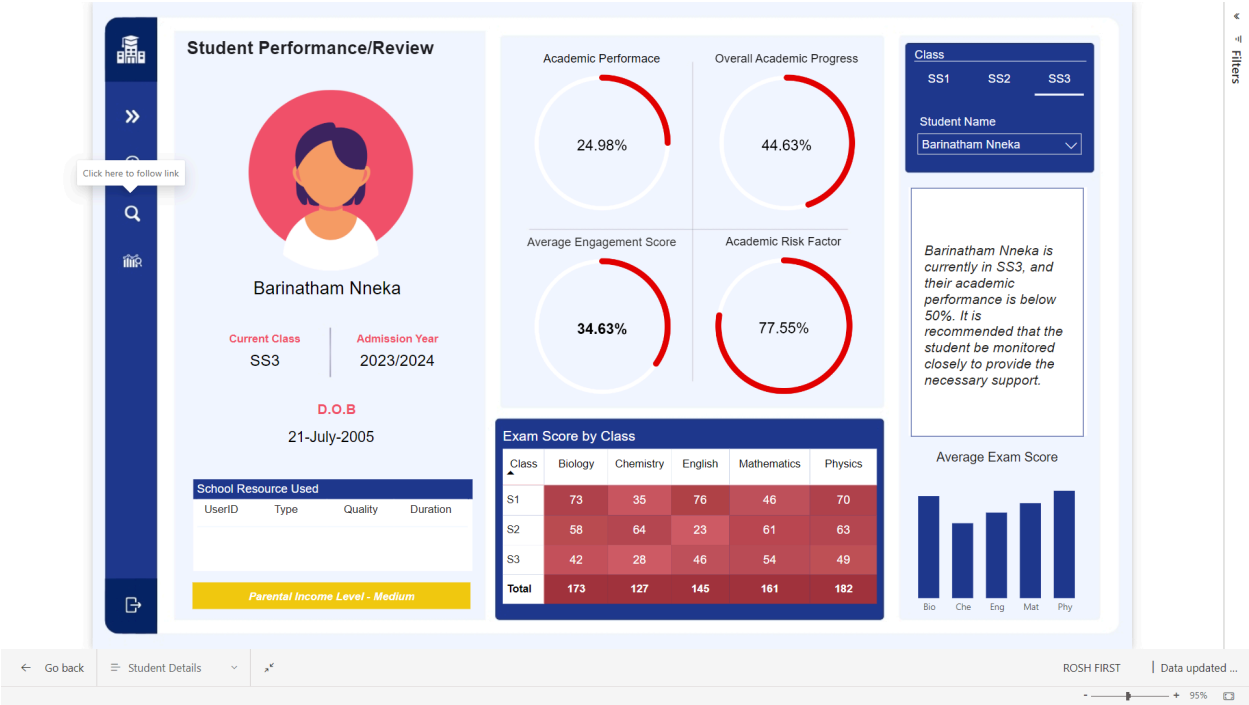
- **Features:**
  - Individual and collective performance trends.
  - At-risk student identification.
  - Teacher impact assessment and resource utilization metrics.
- **Audience:** School administrators, teachers, parents.

Automated Reporting:

- Weekly/monthly performance reports with actionable insights.
- Alerts for underperforming students and interventions for improvement.


Institution General Overview/Summary
School Overview
This analysis focuses on a public school in Ikorodu, Lagos. Our target focus was on SS1 to SS3 students in the science department. The class consists of 500 students. In terms of gender distribution, there are 300 males and 200 females. The average age of the students is 17 years, and their average score is 40.
Student 500
Avg. Age 17
Staffs 30
Avg. Exp 10
Student by their State of Origin
Lagos
Kaduna
Kano
Edo
Rivers
Enugu
Ekiti
Abuja
Club Distribution
Club name
Students
Book Club 115
Red Cross 130
Rotaract Club 123
Sport Club 132
Gender Distribution
40%
60%
Male Female
40.4 40.2 39.9 39.5 38.6
Phy Eng Mat Che Bio
Parent Education Level
9%
31%
29%
30%
Primary Secondary None Tertiary
Filter:
Class
All
Disabilities
2.4%
6.6%
87.8%
None Dyslexia Other ADHD
Teacher by Subject
Mat
Che
Phy
Bio
Eng
Filters
Go back
Overview
ROSH FIRST
Data updated ...
95%


Student Performance Analysis
Click here to follow link
· Academic potential score--gives insights to students who are more likely to pass their next exam.
· The Engagement Score measures how involved a student is in school activities.
· Risk Factor Score - Identifies the likelihood of a student failing their next exam.
· Overall Progress measures at how student improve (or decline) over time.
Student Class
All
Overall Perfomance
All
Avg Academic Potential
53.51%
Avg Overall Progress
54.60%
Risk Factor
51.38%
Relationship btw Risk and Overall Progress
Risk Factor Score
100%
50%
0%
0% 20% 40% 60% 80% 100%
Overall Progress
Filters
Go back
Analysis
ROSH FIRST
Data updated ...
95%



Click HERE to interact with the dashboard.

5. Key Recommendations

1. **Early Interventions:**
   Use predictive insights to provide additional support for students identified as high risk (tutoring, counseling, etc.).
2. **Teacher Development:**
   Provide continuous professional development for teachers based on their performance data and student outcomes.
3. **Resource Allocation:**
   Prioritize schools and students with limited access to learning resources and ensure equitable distribution.
4. **Parental Involvement:**
   Increase communication with parents on their child's progress and offer guidance on how to support learning at home.

6. Conclusion

This solution leverages data analytics to improve education outcomes by identifying key performance factors, predicting student success, and offering actionable insights. With a comprehensive data infrastructure and predictive model, schools can implement timely interventions to significantly enhance student performance in exams like JAMB and WASSCE.