

Student Performance Project

This project focuses on building a complete **Data Engineering and Data Warehousing system** designed to manage and analyze student performance data efficiently.

Over a four-month academic period, data was collected across five core subjects, including grades, attendance, engagement, and socio-economic factors.

The project establishes a robust **data pipeline** that integrates raw data ingestion, transformation, and loading into a centralized **Data Warehouse**, providing a reliable foundation for analytics and decision-making.

A secondary component involves implementing a simple **predictive model** that utilizes the warehouse data to estimate student performance indicators, such as pass/fail status and GPA, demonstrating the analytical potential of the engineered system.

Core Objectives

- Establish a comprehensive data collection and warehousing system for student academic and behavioral data.
- Implement efficient ETL processes to ensure data quality and readiness .
- **Perform in-depth data analysis** to uncover key patterns and factors influencing student performance.
- Develop predictive machine learning models to identify students at risk of failure and estimate GPA.

Key Technical Components

- **Data Warehouse:** Centralized storage for structured, cleaned, and integrated data.
- **ETL Pipelines:** Automated workflows for Extract–Transform–Load processes.
- **Data Modeling:** Star schema design for efficient querying and reporting.
- **Data Quality & Validation:** Ensuring accuracy, consistency, and completeness of all stored data.
- **Analytics Layer:** Dashboards and reports providing insights into student trends and performance.
- **Predictive Layer (Optional):** A simple ML model to illustrate how the warehouse data can support predictive analytics.

Team Structure and Roles

1

Team Leader

Ahmed Tayel Ibrahim

2

Data Collection

Ahmed Tayel

Mehrael Sawiris

3

ERD

Mehrael Sawiris

4

Preprocess Data

Ahmed Tayel

5

Create Database

Nora Mohamed Elsaheed

6

Sql Queries

Hamza yasser Mohamed

7

Data Warehouse

Mohammed Eslam Mohammed

Abdulrahman Eid

Hamza yasser Mohamed

8

Data Modeling & ML Engineer

Nora Mohamed

Mehrael Sawiris

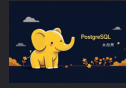
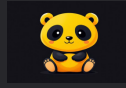
9

Visualization & Reporting Analyst

Abdulrahman Eid

Mohammed Eslam Mohammed

Tools and Technologies



- Data Manipulation & Scripting: **Python** with **Pandas** (for cleaning/ETL) and **Matplotlib/Seaborn** (for static visualizations).
- Database: **PostgreSQL** for persistent, structured storage and efficient querying.
- Machine Learning: **Scikit-learn** to implement Logistic Regression, Random Forest, or Decision Tree models for prediction.
- Dashboard : **Plotly Dash** and **Power BI** for interactive deployment.

Project Milestones and Timeline

Milestone 1: Data Collection and Preprocessing

Target Date: September 30, 2024

Deliverables:

- Cleaned dataset.
- Data collection and preprocessing scripts.

Milestone 2: SQL Integration & Querying & Datawarehouse

Target Date: October 15, 2024

Deliverables:

- Functional SQL database.
- Collection of optimized SQL queries.
- Datawarehouse

Milestone 3: Visualization & Reporting & ML model

Target Date: October 25, 2024

Deliverables:

- Interactive data visualization dashboard.
- Detailed analytical report.
- machine learning model to predict student success/failure based on historical scores

Milestone 4: Final Documentation and Presentation

Target Date: October 30, 2024

Objective: Compile all project documentation and present the findings and solution.

Deliverables:

- Complete project documentation package.
- Final project presentation.

Key Performance Indicators (KPIs) - Data Quality & Efficiency

Data Integrity & Preprocessing Efficiency

Target **99.5%** data integrity with less than **0.5%** data loss/error in cleansed output.

SQL Query Accuracy & Reliability

Achieve a minimum of **98%** accuracy in all SQL query results.

Dashboard Load Time & User Experience

Ensure dashboard loads completely within **5 seconds** for all users.