Visual Analysis of Student Performance

Visual Analytics Course Project - Fall 2025

Leonardo Ricca & Federico Turrini

Project Goal

Problem: Traditional ‘Black Box’ Grade Reports.
Hide the root causes of failure

Solution: Interactive Exploratory Environment.
Uncovers hidden patterns and relationships

Goal: Develop a Visual Analytics System.
To support educational decision-making

Key Capability: Understand Multidimensional Risks.
Empower experts with integrated Social + Academic insights

Potential Users

1. School Counselors:

- Need to identify at-risk students before they fail.
- Use the tool to differentiate between behavioral issues and academic struggles.

2. Educational Policymakers:

- Need to validate hypotheses (e.g., 'Does affect grades more than study time?').
- Design broader intervention strategies based on data clusters.

The Dataset

Source:
UCI Machine Learning Repository (Student Performance Data Set)

Volume & Subject:
Volume: 395 Students | Subject: Mathematics (Secondary School)

Dimensions:
33 Attributes per student

Type:
Mixed data (Numerical grades, Categorical demographics, Ordinal surveys)

Data Structure (4 Pillars)

1. Demographics:

Age, Sex, Address (Urban/Rural), Family Size

2. Social Context:

Parent's Job, Relationships, Alcohol consumption (Walc/Dalc)

3. School Habits:

Study time, Travel time, Absences, Failures

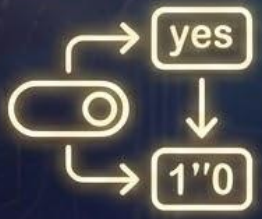
4. Performance (Targets):

G1 (1st Period), G2 (2nd Period), G3 (Final Grade)

Data Preprocessing

Parsing:

Custom CSV parsing implemented in D3.js.

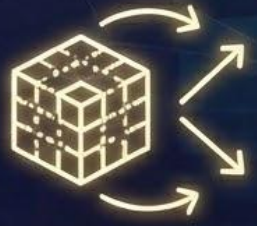


Encoding:

Converting categorical strings (e.g., 'yes'/'no') to binary integers.

Normalization:

Scaling numerical features to ensure ensure fair weighting.



Dim. Reduction:

PCA calculation to condense 33 features into 2 principal components.

Visualization Strategy

Approach: Coordinated Multiple Views (CMV)

1. Macro View:
PCA Projection
(Global clustering)

2. Correlation View:
Scatter Plot
(Grade progression)

3. Micro View:
Parallel Coordinates
(Individual profiles)

4. Context View:
Bar/Histograms
(Distribution analysis)

1. Parallel Coordinates

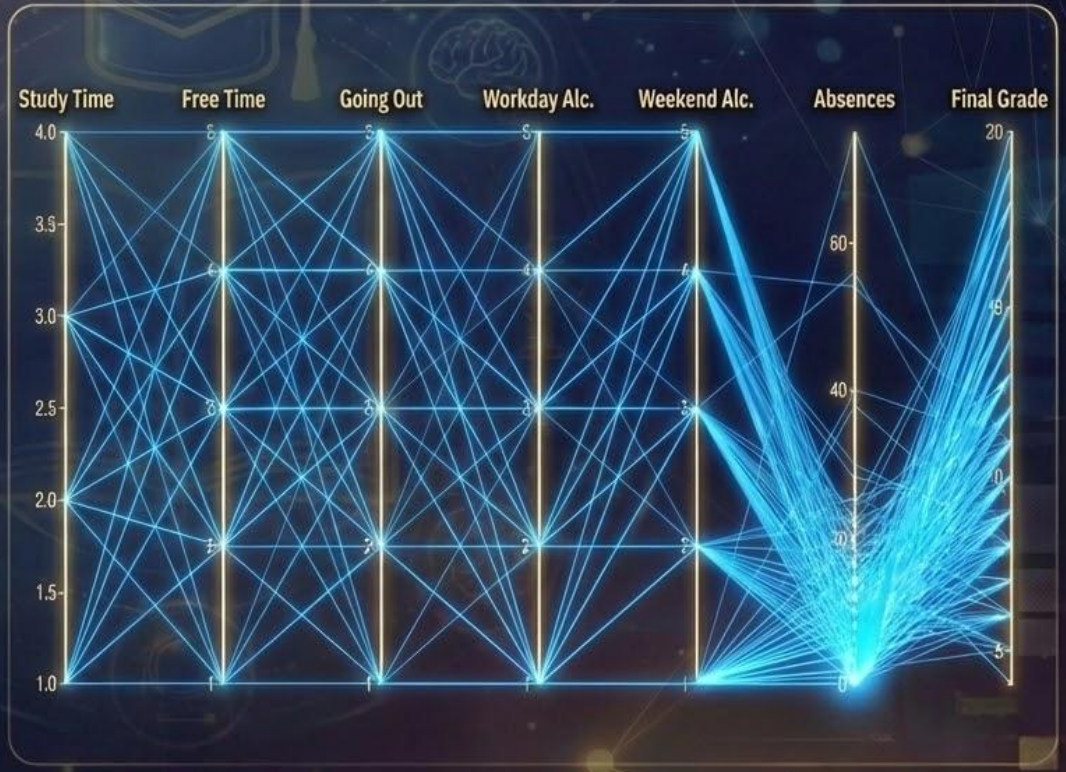
Key Features & Insights

Purpose:
Multidimensional Profiling

Function: Each student is a line crossing multiple axes (Grades, Alcohol, Study Time)

Why?: Only way to visualize N dimensions simultaneously

Insight: Detects specific profiles (e.g., 'High Alcohol' + 'High Going Out Time' -> 'Low G3')



2. Scatter Plot

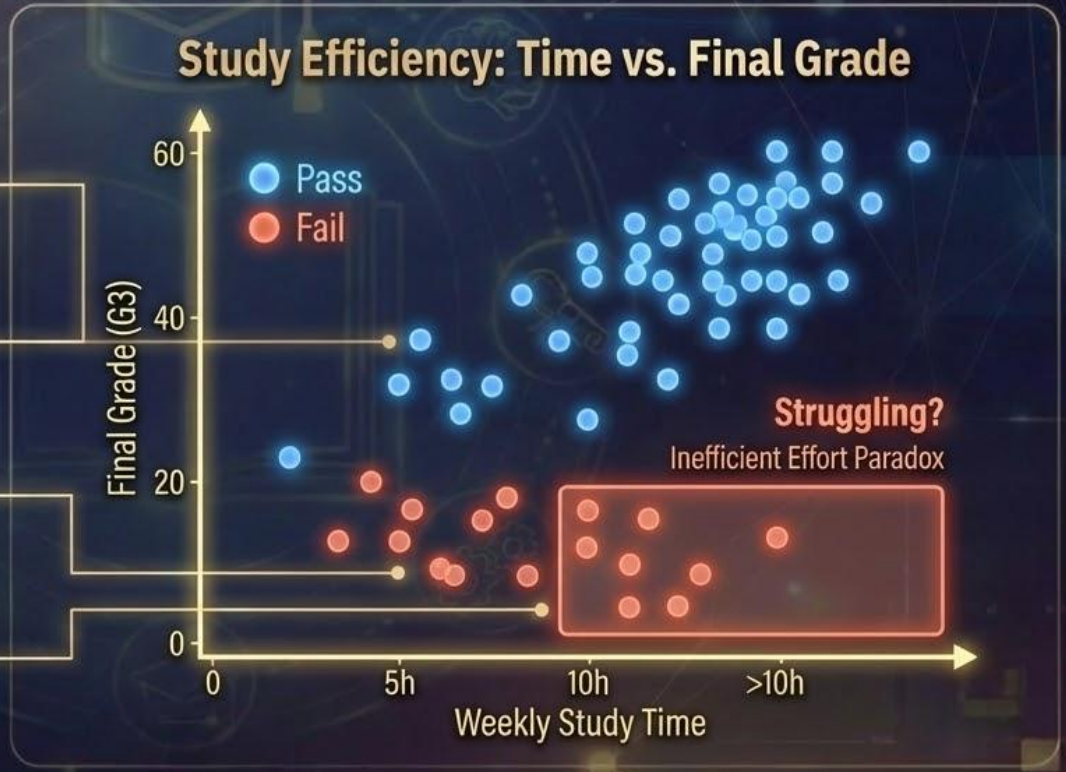
Key Features & Insights

Purpose:
Performance Correlation
(Study Time vs. Final Grade)

Why?: Tests assumption:
More study ≠ Better grades

Insight: Reveals 'Inefficient Effort' paradox (High study time, low grades)

Value: Identifies students needing intervention in study methods



3. PCA Projection

Key Features & Insights

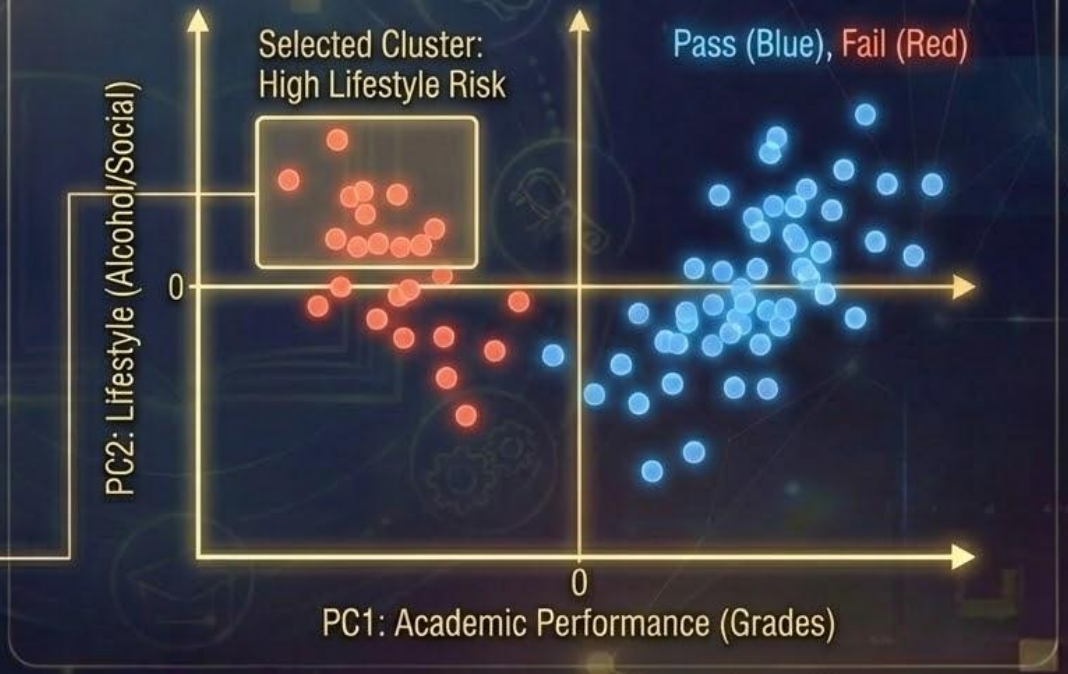
Purpose:
Cluster Analysis

Technique:
Principal Component Analysis (2D Projection)

Why?: Reveals natural groupings based on similarity across ALL 33 variables

Interaction: 2D brushing allows selecting a 'cluster' to see their profile in other views

PCA Projection: Student Clusters



4. Context Charts

Key Features & Insights

Purpose: Global Filtering & Distribution Analysis

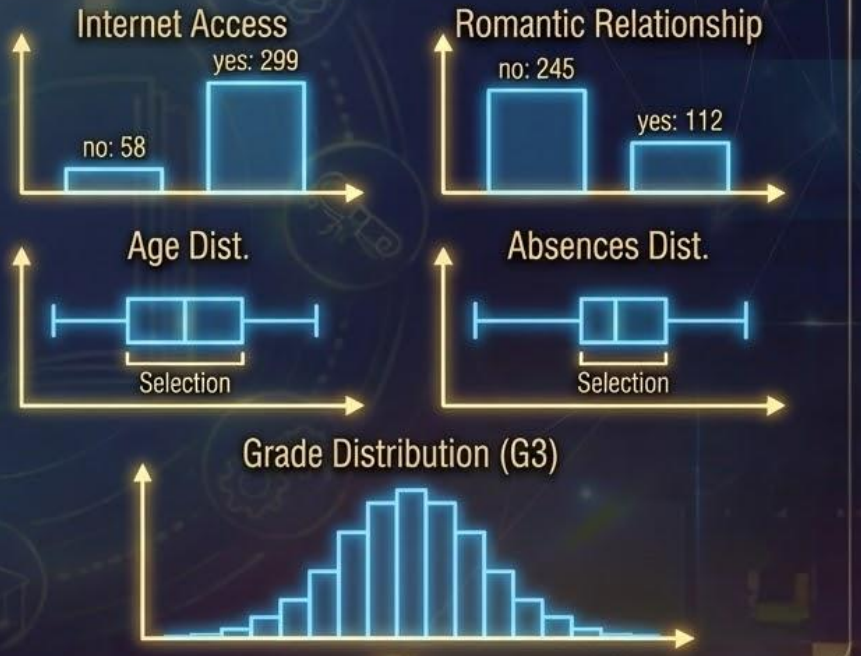
Bar Charts: Binary filters for 'Internet Access' and 'Romantic Relationships'

Box Plots: Statistical summaries (Median/Quartiles) for 'Age' and 'Absences'

Histogram: Visualizes the distribution of Final Grades (G3) to assess the 'shape' of performance

Interaction: Acts as the controller; selecting bars or ranges here updates all other views

Context Charts: Data Distribution & Filtering



System Architecture

Frontend & Styling: HTML5, SASS (SCSS) for modular styling.

Visualization Engine: D3.js (Data-Driven Documents).

Logic: JavaScript (ES6+).

Specific Libraries: ‘pca-js’ used for real-time dimensionality reduction in the browser.

Case Studies & Insights

The ‘Inefficient Effort’ Paradox:

- Students studying >10h but failing (<10/20). Parallel plot reveals they are not drinkers/absentees. Problem is study method, not motivation.

The ‘Gifted Underachiever’ Risk:

- Students studying <2h with high grades (>15/20). High correlation with ‘Going Out’. Risk of future failure when raw intellect is insufficient.

Diminishing Returns of Study Volume:

- Scatter plot reveals a saturation point. Studying >10h yields marginal gains over 5-10h, likely due to burnout.

Conclusion

Summary: The system moves beyond static reporting to active exploration.

Value: Integrates statistical complexity (PCA) with interpretable details (Parallel Coords).

Impact: Allows detection of subtle student risks, such as the ‘Inefficient Effort’ paradox.