STUDENTS' PERFORMANCE ANALYSIS

Onyango Jack

4 January 2024

Understanding students' learning curves is a crucial aspect that significantly impacts the educational sector's performance. This project leverages data obtained from Kaggle (link: `https://www.kaggle.com/datasets/mahmoudshogaa/student-alcholoh-data`), focusing specifically on the Mathematics course.

Objective

This analysis targets educators, researchers, and policymakers, aiming to enhance their methods and facilitate students' better understanding of course materials.

About the Data Set

The dataset comprises 33 fields rich in information relevant to understanding various student's perfamance aspects and patterns.
Key fields include:

- **School:** Student's school ('GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
- **Sex:** Student's gender ('F' - female or 'M' - male)
- **Age:** Student's age (numeric: from 15 to 22)
- **Address:** Student's home address type ('U' - urban or 'R' - rural)
- **Famsize:** Family size ('LE3' - less than or equal to 3 or 'GT3' - greater than 3)

- **Pstatus:** Parent's cohabitation status ('T' - living together or 'A' - apart)
- ... (and more)

Problem Statement

In the education sector, understanding students' habits and behaviors is crucial for performance improvement. This project seeks to unravel various aspects influencing students' performance patterns and strategies for enhancing their comprehension.

Research Questions

1. Which gender and age group tends to perform well, and does the school and location influence it?
2. What are the effects of support on students' performance? Does the number of classes missed correlate with the support they receive and their average performance?
3. Does parental status, educational level, and parents' jobs influence students' performance?
4. What is the average alcohol consumption for each gender, its distribution across different ages, and potential correlations with family relationships or support?
5. What is the relationship between support, paid classes, the number of failed classes, and average scores?
6. Do romantic relationships and internet access affect students' study time?

Procedure:

I conducted a comprehensive analysis in MS SQL, covering tasks such as data cleaning, transformation, and visualization tables. Key highlights include:

1. Data Cleaning:

- Renamed columns for better understanding.
- Altered data types for specific columns.
- Checked for null values before using columns.
- Transformed data in various columns (e.g sex, goout, address).

2. Data Transformation:

- Created new columns for total score (`Marks`) and grading (`Grade`).
- Used an interval scale of 12 for grading based on the total possible score a student can attain.

3. Visualization Tables:

- Developed various tables that will help me with visualizations on Power BI and come up with reports

5. Additional Analysis:

- Explored parental jobs, education levels, and their possible combinations.
- Analyzed the relationship between school support, paid classes, failed classes, and average scores.
- Explored social factors such as going out, romance, desire for higher education, and internet access.