Analyzing the Impact of Alcohol Consumption on Academic Performance Using AI-Based Tools

1. Introduction

This project explores the relationship between alcohol consumption and academic performance among high school students. With increasing concerns regarding youth behavior and the quality of education, this study aims to uncover meaningful patterns using data from a real-world dataset. By applying AI-based tools, particularly GPT models, we analyzed and interpreted the complex data to extract relevant insights.

2. Dataset Description

The dataset used for this project was sourced from Kaggle. It contains 649 records with 31 features covering various aspects of student life, including demographics, parental background, academic grades, and alcohol consumption. Key variables include daily and weekend alcohol use (Dalc and Walc), academic grades (G1, G2, G3), and other influential factors such as study time, parental education levels, and student absences.

3. Technical Stack

For data analysis, we used Google Colab as our development environment. Python was our primary programming language, supported by libraries such as pandas, numpy, matplotlib, seaborn, and sklearn. We also used GPT-based AI tools to interpret and summarize the results. Visualizations were created to illustrate key relationships and trends in the data.

4. Exploratory Data Analysis (EDA)

Our exploratory data analysis revealed several significant patterns. Box plots demonstrated that students who consumed more alcohol tended to score lower on all grade metrics (G1, G2, and G3). Bar charts showed that students with parents who had higher education levels generally performed better academically. There was a clear positive relationship between study time and student grades, while absences showed a negative correlation with performance. Lastly, age appeared to influence alcohol consumption, with older students reporting higher daily and weekend usage.

5. Extended Data Interpretation

Although we did not implement machine learning models in this project, we conducted a thorough exploratory analysis of the dataset. This allowed us to interpret how various social and behavioral factors are associated with student performance. The patterns found offer a strong foundation for future predictive modeling and educational intervention planning.

6. GPT-Based Insight Generation

We utilized GPT models to generate narrative explanations of our findings. For example, GPT helped reframe technical insights into accessible language, such as stating that students who drink daily are 30% more likely to underperform. Additionally, the model highlighted how parental education might mitigate some of the negative effects of alcohol use. These insights were especially useful in presenting the results to a broader audience.

7. Visualization Summary

Throughout the project, we used visualizations to illustrate our findings. Box plots were used to show the relationship between alcohol consumption and grades. Correlation heatmaps allowed us to see how various variables were related. Bar charts were used to demonstrate the influence of study time and parental education on grades, while scatter plots revealed the negative trend between absences and academic performance.

8. Challenges & Limitations

This project faced a few limitations. First, the dataset relied on self-reported data, which may introduce biases or inaccuracies. Second, our survey responses were limited to a local group of participants, which might not fully represent broader populations. Finally, while we observed correlations between alcohol use and academic performance, we cannot conclusively determine causation from this analysis.

9. Team Contributions

Each team member contributed in a specific way:

- Leonie searched for and identified the datasets that were considered during the survey process and also contributed to the insights analysis.
- Gentiana wrote the detailed report and collaborated with Linda to gather around 30 participants for the initial survey used to select the topic.
- Linda also worked on finding participants for the survey and also on interpreting the final insights.
- Gabriel designed and wrote the survey questions and he created the final presentation.

10. Conclusion

Our findings suggest that alcohol consumption, particularly on weekdays, is associated with lower academic performance. Meanwhile, factors such as increased study time and higher parental education positively impact student success. By leveraging AI tools like GPT, we were able to better understand and communicate these patterns, contributing to meaningful dialogue around student health and education.

11. References

- Kaggle Dataset: https://www.kaggle.com/datasets/gabrielluizone/high-school-alcoholism-and-academic-performance
- Google Colab Notebook:
https://colab.research.google.com/drive/1pclpyWptryqBZfFNfx2YKgYegZJ2jA9
- Python Libraries: pandas, sklearn, matplotlib, seaborn