EDA Report – High School Alcoholism and Academic Performance

# A brief description of the dataset and summary of attributes

The dataset, High School Alcoholism and Academic Performance was sourced from Kaggle. The data was collected through a survey conducted among high school students in Portugal and includes biographical and behavioral data of students relating to alcohol consumption and their academic population such as school, age, family dynamics, structure, school attendance, grades, etc. These features may prove to be valuable in understanding the impact of alcohol consumption on academics as well as identifying potential causes of teenage alcohol consumption.

## SIZE, AND Quality

The dataset contains 31 features (columns) and 649 data points (rows). Feature data types include integers as well as objects (strings). Some features appear to have missing values which will be handled appropriately for proper data analysis. Many of these features contain non-numeric data which will be encoded to numeric values. Refer to Appendix A, Screenshots 1, and 2 for more details.

## List of Features

Following is the list of features in the dataset and descriptions with valid values.

|  |  |
| --- | --- |
| **Feature** | **Description** |
| School | Student's school (Gabriel Pereira, Mousinho da Silveira) |
| Gender | Student's sex (Female, Male) |
| Age | Student age (numeric: 15 to 22) |
| Housing\_Type | Type of student's residential address (Urban, Rural) |
| Family\_Size | Family size (Upto 3, Above 3) |
| Parental\_Status | Parents' cohabitation status (Separated, Living Together) |
| Mother\_Education | Mother's education level (0 - none, 1 - Elementary School 1, 2 - Elementary School 2, 3 - High School or 4 - Higher Education) |
| Father\_Education | Father's education level (0 - none, 1 - Elementary School 1, 2 - Elementary School 2, 3 - High School or 4 - Higher Education) |
| Mother\_Work | Mother's job (nominal: teacher, health, services, at\_home or Other) |
| Father\_Work | Father's job (nominal: teacher, health, services, at\_home or Other) |
| Reason\_School\_Choice | Reason for choosing this school (nominal: home, reputation, course or other) |
| Legal\_Responsibility | Student's guardian (nominal: mother, father or other) |
| Commute\_Time | Travel time from home to school (time intervals: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour or 4 - >1 hour ) |
| Weekly\_Study\_Time | Weekly study time (time intervals: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours or 4 - >10 hours) |
| Extra\_Educational\_Support | Extra educational support (binary: yes or no) |
| Parental\_Educational\_Support | Family educational support (binary: yes or no) |
| Private\_Tutoring | Private classes on subjects related to the course (binary: yes or no) |
| Extracurricular\_Activities | Performs extracurricular activities (binary: yes or no) |
| Attended\_Daycare | Attended daycare (binary: yes or no) |
| Desire\_Graduate\_Education | Desire to pursue a degree (binary: yes or no) |
| Has\_Internet | Internet access at home (binary: yes or no) |
| Is\_Dating | Are you in a romantic relationship (binary: yes or no) |
| Good\_Family\_Relationship | Quality of family relationships (categorical: from 1 - very bad to 5 - excellent) |
| Free\_Time\_After\_School | Free time after school (categorical: from 1 - very low to 5 - very high) |
| Time\_with\_Friends | Time with friends (categorical: from 1 - very low to 5 - very high) |
| Alchohol\_Weekdays | Alcohol consumption on the work day (categorical: from 1 - very low to 5 - very high) |
| Alchohol\_Weekends | Alcohol consumption on the weekend (categorical: from 1 - very low to 5 - very high) |
| Health\_Status | Current health status (categorical: from 1 - very bad to 5 - very good) |
| School\_Absence | Number of school absences (numeric: from 0 to 93) |
| Grade\_1st\_Semester | First-semester grade (numeric: from 0 to 20) |
| Grade\_2nd\_Semester | Second-semester grade (numeric: from 0 to 20) |

# Quality of data

While the dataset hardly contains any missing fields/values, the sample size is small and restricted to students of only two schools, Gabriel Pereira, and Mousinho da Silveira.

## REQUEST FOR ADDITIONAL DATA

quality of this data set and a request for additional data if needed

# Appendix A: SCREENSHOTS

## 1) First 5 rows of data

A screenshot of a computer screen

Description automatically generated

A computer screen shot of a computer program

Description automatically generated

## 2) FEATURE INFORMATION

A screenshot of a computer program

Description automatically generated

# REFERENCES

P. Cortez e A. Silva. Usando a Mineração de Dados para Prever o Desempenho do Aluno do Ensino Médio. Em A. Brito e J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, abril de 2008, EUROSIS, ISBN 978-9077381-39-7.