Gizem Ezgi AYGÜN-22470

AlcoholConsumptionAndNumberOfAbsenceVsAcademicPerformance-DSA210 PROJECT

PROJECT OVERVIEW

This given dataset researches the impact of alcohol intake and number of absence on students’ grades and academic achievement at two Portuegese secondary school. The dataset includes credits related to student grades, school-related factors, social behaviour and demographics gathered through questionnaires and reports.

The datasets provided, corresponding to 2 different subjects at school. One is Maths, the other one is Portuguese. My goal is to explore how alcohol consumption and number of absence are related to students academic performance. I will explore and analyze this by using effective tools like; statistical tools and data visiualization.

My plan is how alcohol intake, number of absence and level of academic performance is related with each other. I will correlate the results by using hypothesis tests and learn the difference. Im really wondering the results because maybe the results make me quit partying at the weekends.

OBJECTIVES

1. Understanding Performance Influencers:

I want to explore the how alcohol intake and number of absence are related to students grades and their achievement.

1. Identify Key Factors

Determine which variables have more impact on alcohol consumption frequency.

1. Data Driven Insights

Impact of Alcohol Consumption on Grades

Students with higher alcohol consumption, particularly frequent weekend drinking, tend to have lower final grades (G3). This suggests that alcohol use may negatively affect academic performance, potentially due to decreased study time or cognitive impairment.

Correlation Between Prior Grades and Final Performance

Strong correlations exist between G1, G2, and G3, indicating that early academic performance is a strong predictor of final grades. Students struggling in earlier periods may require targeted interventions to improve outcomes.

Influence of Social and Demographic Factors

Socioeconomic status, parental education, and family structure play a role in student achievement. Students from supportive academic environments tend to perform better, while those with fewer resources may be more vulnerable to the negative effects of alcohol.

Behavioral and Lifestyle Patterns

Factors such as absences, extracurricular activities, and study time also influence performance. Students with frequent absences or lower study hours generally show poorer academic outcomes, reinforcing the importance of consistent attendance and effective study habits.

Predictive Modeling for Early Intervention

By analyzing patterns in student performance, educators and policymakers can use predictive models to identify at-risk students early and implement strategies to improve academic success.

1. Apply Real Life Data Science Skills

Using my knowledge in data science, statictics, data visualization and Python in a real life example in this way; improving my skills

Motivation

1. Personal Growth

Partying and studying data science are my two main things in my life. However clearly they have negative correlation. Maybe after seeing results that alcohol intake and academic performance are corralated negatively with each other i can study more by partying.

1. Scientific Approach

It is clearly obvious that alcohol intake and number of absence and grades are reversely corralated with each other however i want to see the scientific results by applying right tools by myself.

1. Practical Case

This is a chance me to improve my computational, statistical and data driven skills.

DATASETS

**Attributes for both Maths and Portuguese Courses**

|  |  |
| --- | --- |
| school | student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira) |
| sex | student's sex (binary: 'F' - female or 'M' - male) |
| age | student's age (numeric: from 15 to 22) |
| address | student's home address type (binary: 'U' - urban or 'R' - rural) |
| famsize | family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3) |
| Pstatus | parent's cohabitation status (binary: 'T' - living together or 'A' - apart) |
| Medu | mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education) |
| Fedu | father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education) |
| Mjob | mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other') |
| Fjob | father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other') |
| reason | reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other') |
| guardian | student's guardian (nominal: 'mother', 'father' or 'other') |
| traveltime | home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) |
| studytime | weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) |
| failures | number of past class failures (numeric: n if 1<=n<3, else 4) |
| schoolsup | extra educational support (binary: yes or no) |
| famsup | family educational support (binary: yes or no) |
| paid | extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) |
| activities | extra-curricular activities (binary: yes or no) |
| nursery | attended nursery school (binary: yes or no) |
| higher | wants to take higher education (binary: yes or no) |
| internet | Internet access at home (binary: yes or no) |
| romantic | with a romantic relationship (binary: yes or no) |
| famrel | quality of family relationships (numeric: from 1 - very bad to 5 - excellent) |
| freetime | free time after school (numeric: from 1 - very low to 5 - very high) |
| goout | going out with friends (numeric: from 1 - very low to 5 - very high) |
| Dalc | workday alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| Walc | weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| health | current health status (numeric: from 1 - very bad to 5 - very good) |
| absences | number of school absences (numeric: from 0 to 93) |

**These grades are related with the course subject, Math or Portuguese:**

| **Grade** | **Description** |
| --- | --- |
| G1 | first period grade (numeric: from 0 to 20) |
| G2 | second period grade (numeric: from 0 to 20) |
| G3 | final grade (numeric: from 0 to 20, output target) |

TOOLS

* **Python**: For data cleaning and statistical analysis
* **Pandas**: To manipulate and preprocess data
* **Matplotlib and Seaborn**: For creating visualizations (scatter plots, heatmaps, time series)
* **SciPy**: For hypothesis testing and regression analysis

ANALYSIS PLAN

**Analysis Plan**

1. **Data Exploration and Cleaning**
   * Load the dataset , identifying inconsistencies and its missing values.
   * Handle inconsistent data through removal if it is necessary.
   * Convert categorical variables into numerical representations where required with the help of binary variables
2. **Descriptive Statistics and Visualization**
   * Compute summary statistics (mean, median, standard deviation) for key variables.
   * Visualize relationships between alcohol consumption, number of absence and academic performance using histograms, box plots, and scatter plots etc.
   * Compute correlations between student grades (G1, G2, G3) and other features using heatmaps etc.
3. **Feature Selection and Engineering**
   * Identify the most influential factors affecting student grades through correlation analysis and feature importance ranking among alcohol consumption and number of absence
   * Create new features if beneficial (e.g., combining weekday and weekend alcohol consumption)
4. **Model Selection and Implementation**
   * Apply classification and regression models (e.g., logistic regression, clustering, decision trees, random forests) to predict student performance based on alcohol consumption and other attributes like number of absence
   * Train models using a portion of the data and evaluate their performance using metrics like accuracy, RMSE, or F1-score.
5. **Insights and Interpretation**
   * Interpret model results to understand the extent to which alcohol consumption and number of absence influence grades.
   * Compare findings with existing literature or studies on the topic.
   * Discuss potential implications for educators and policymakers.
6. **Conclusion and Recommendations**
   * Summarize key findings and their significance.
   * Suggest strategies for mitigating the negative effects of alcohol on student performance, maybe.
   * Highlight areas for further research or data collection improvements.

DATA SOURCE

\*From Kaggle on the internet

REFERENCES

Whenamancodes. (Year). Alcohol Effects on Study [Data set]. Kaggle. <https://www.kaggle.com/datasets/whenamancodes/alcohol-effects-on-study>