

DataGood x School Alc Consumption

Datathon Spring 2024



The Team







Herman



Michael



Will



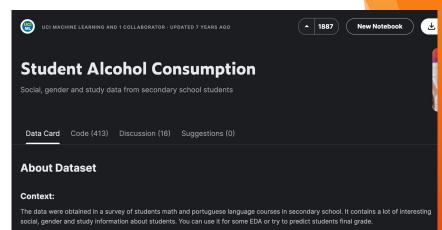
Project Area: Health & Education

Dataset: Student Alcohol Consumption, Social, gender and study data from secondary school students in Portugal Secondary Schools

Hypothesis

We predict that **Alcohol Consumption**, both daily & weekend, **will negatively affect** a student's academic performance.

https://www.kaggle.com/datasets/uciml/student-alcohol-consumption







Factors to Consider in the Study

- Data collected 8 years ago
- Secondary Schools in Portugal
- Habits could be different than the United States

Project Goals

Using the Student Alcohol Consumption Dataset, we strove to:

Identify correlations Understand between alcohol and confounding factors grades Recommendation to Risk Analysis & school for improving **Factors for Risk** education



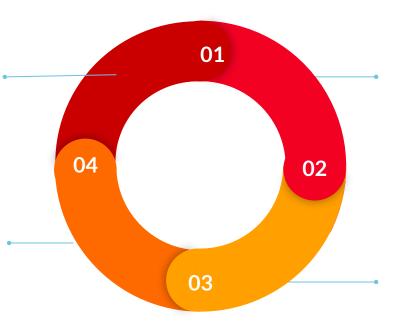
Methodology

Data Collection

• Finding the dataset through online research

Model Creation

- Creating machine learning Sklearn models for linear regression & predictions
- Develop recommendation for schools



Data Cleaning & Processing

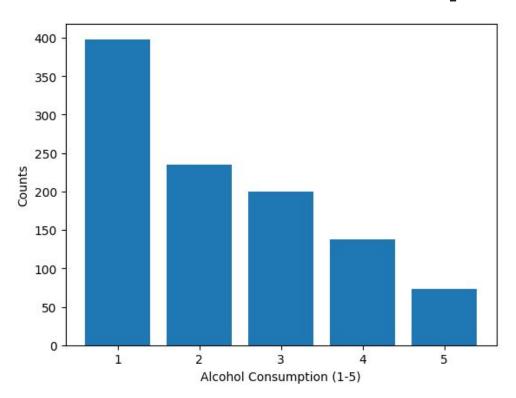
 The data was loaded into shared notebooks to fix or remove incorrect, corrupted, improperly formatted, duplicate, or incomplete data within the dataset.

Exploratory Data Analysis

Data visualizations were made in DeepNote notebooks using Python libraries.

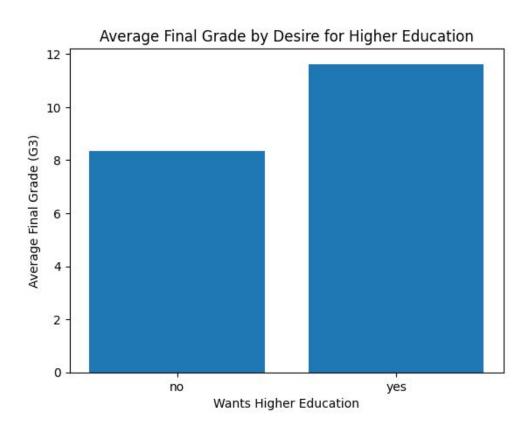


How Common Were Different Levels of Alcohol Consumption?



Did a Desire of Higher Education Impact Final Grades?



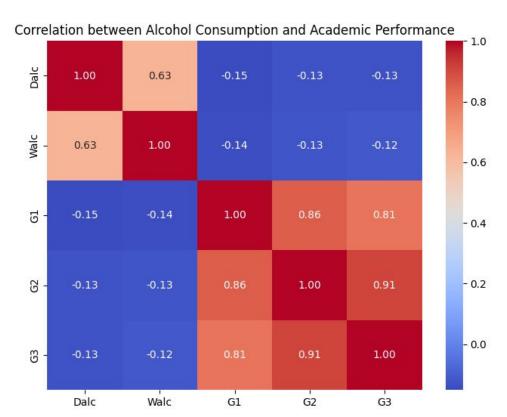




2. Alcohol Affecting Grades?

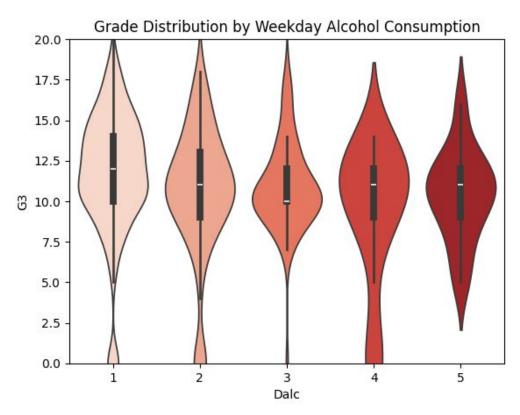


Interesting Findings



- There is a very slight negative correlation between alcohol consumption and on grade performances
- Due to the small correlation, this could imply there are other factors that affect academic performance, such as number of past class failures, biological sex, age, and more.

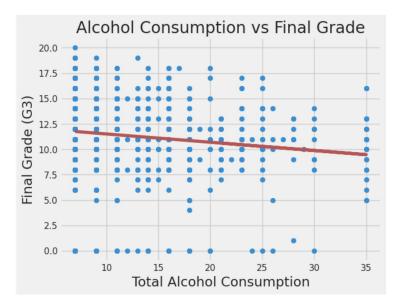




- Looking at this violin plot, as weekday alcohol consumption increases, the median grade decreases and the grade distribution becomes more variable
- However, the presence of high grades even in higher alcohol consumption categories indicates that the relationship is not deterministic, and other factors may play an important role.



Negative Correlation



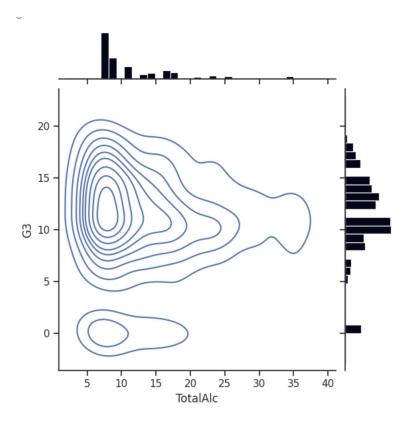
```
df_alc_combined = df.copy()
df_alc_combined['TotalAlc'] =
df['Dalc'] * 5 + df['Walc'] * 2
df_alc_combined
```

```
# solve for a and b
def best_fit(X, Y):
   xbar = sum(X)/len(X)
   ybar = sum(Y)/len(Y)
   n = len(X) # or len(Y)
   numer = sum([xi*yi for xi,yi in zip(X, Y)]) - n * xbar * ybar
   denum = sum([xi**2 for xi in X]) - n * xbar**2
   b = numer / denum
   a = ybar - b * xbar
   print('best fit line:\ny = \{:.2f\} + \{:.2f\}x'.format(a, b))
  return a, b
average_grades = df.groupby('higher')['G3'].mean()
X = df alc combined['TotalAlc']
y = df_alc_combined['G3']
# solution
a, b = best_fit(X, y)
#best fit line:
\#v = 0.80 + 0.92x
# plot points and fit line
plt.title('Alcohol Consumption vs Final Grade')
plt.xlabel('Total Alcohol Consumption')
plt.ylabel('Final Grade (G3)')
plt.scatter(X, y)
yfit = [a + b * xi for xi in X]
plt.plot(X, yfit, color='r')
plt.show()
```





Negative Correlation



Joint Density Plot

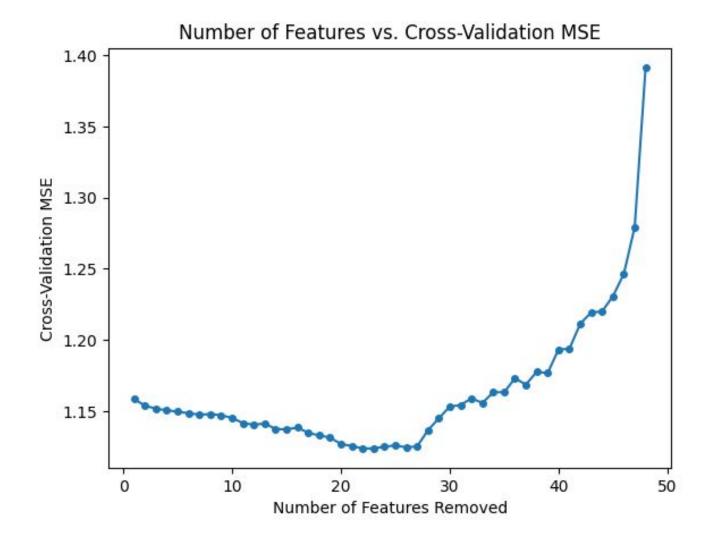
- Higher density of low alcohol drinkers, generally performing average
- Higher alcohol consumption → tends to lower final grades



3.

Predicting Alcohol Consumption

```
walc X = X[walc coefs]
   walc X2 = sm.add constant(walc X)
   walc_model = sm.OLS(walc, walc_X2)
   walc_model = walc_model.fit()
   print(walc_model.summary(slim = True))
 ✓ 0.0s
                            OLS Regression Results
                                 Walc
Dep. Variable:
                                        R-squared:
                                                                          0.357
Model:
                                  0LS
                                        Adi. R-squared:
                                                                          0.343
No. Observations:
                                        F-statistic:
                                 1044
                                                                          25.77
Covariance Type:
                            nonrobust
                                        Prob (F-statistic):
                                                                       1.58e-82
                     coef
                             std err
                                                     P>|t|
                                                                 [0.025
                                                                             0.9751
const
                  -0.1698
                               0.626
                                         -0.271
                                                     0.786
                                                                -1.398
                                                                              1.058
                                                     0.000
                                                                             -0.457
                  -0.5947
                               0.070
                                         -8.460
                                                                -0.733
sex F
                                                                              0.131
                   0.0748
                               0.029
                                          2.597
                                                     0.010
                                                                 0.018
age
famsize GT3
                               0.072
                                                     0.010
                                                                             -0.045
                  -0.1867
                                         -2.579
                                                                -0.329
paid
                               0.081
                                                                 0.079
                                                                              0.398
                   0.2388
                                          2.941
                                                     0.003
                               0.083
                                         -2.789
                                                     0.005
                                                                -0.396
                                                                             -0.069
nursery
                  -0.2325
famrel
                  -0.2266
                                         -6.403
                                                     0.000
                                                                             -0.157
                               0.035
                                                                -0.296
                   0.4079
                                         14.191
                                                                              0.464
goout
                               0.029
                                                     0.000
                                                                 0.351
```

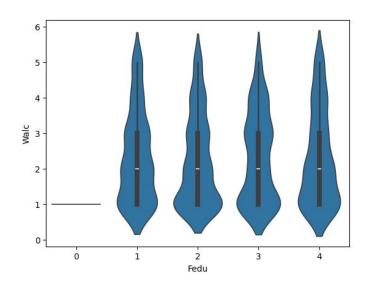


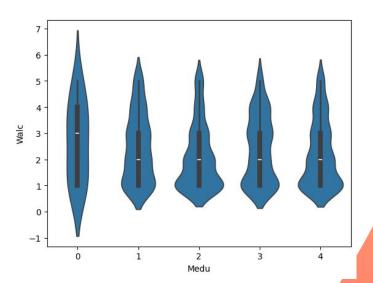




Interesting Findings

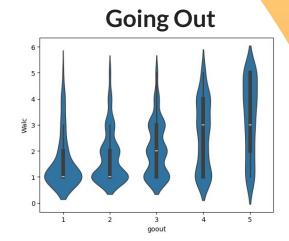
 While Father's Education was positively correlated with alcohol consumption, Mother's Education was negatively correlated



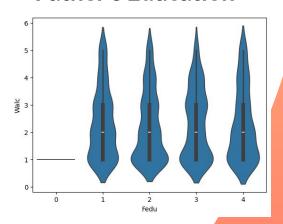


Strongest Predictors of Alcohol Consumption

Biological Sex 5.0 4.5 4.0 3.5 2.5 2.0 1.5 sex F



Father's Education





Comparisons to Existing Research



- **Biological Sex**: Males have been found to consume more alcohol in all kinds of countries (low, medium, high income)
- This aligns with our findings from this dataset that males tend to have higher rates of alcohol consumption

https://www.sciencedirect.com/science/article/pii/S0376871619302790

Comparisons to Existing Research

- Grade
 Distributions: more
 drinks → lower
 grades
- Suggested by the data



Comparisons to Existing Research (cont.)



- Parental Education: A study on adolescents in Sweden found that adolescents whose parents have higher education tend to drink more frequently, but in less quantity than those with less educated parents
- This aligns with our finding that father's education is positively correlated with alcohol consumption, but conflicts with our finding that mother's education is negatively correlated with alcohol consumption

SSM Popul Health, 2018 Dec; 6: 91–97.	PMCID: PMC614374
Published online 2018 Sep 5. doi: 10.1016/j.ssmph.2018.09.001	PMID: 3023805
Deported advection differentially predicts young adults' faceyoney on	d avantity of alashal
Parental education differentially predicts young adults' frequency and quantity of alcohol use in a longitudinal Swedish sample	
use in a longitudinal swedish sample	
<u>Laura Wells* and Viveca Östberg</u>	
► Author information ► Article notes ► Copyright and License information PMC Disclaime	ſ
Abstract	Go to:
Background	
Alcohol consumption contributes to health inequalities, but few studies have exa differentiated alcohol use develops across the life course. In this study, we exam childhood socioeconomic position (parental education) relates to two often-conf patterns: drinking frequency and quantity per occasion. Using a life course persy whether parental drinking patterns or young adults' own educational attainment associations.	ine how one aspect of lated young adult drinking pective, we also explore
Methods	
This study used longitudinal data from the nationally representative Swedish Le (LNU). Young adults' (aged 20 – 28 , n = 803) drinking patterns and educational through the LNU 2010 and official registers. A decade earlier, parents self-repordrinking patterns in the LNU 2000 and Partner-LNU 2000 .	attainment were determine
Results	
Logistic regression models showed that high parental education predicted young	adult frequent drinking.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6143747/

Comparisons to Existing Research (cont.)



- **Going Out**: A study on teen parties found that adolescents who are involved in social activities with peers like parties are more likely to start drinking
- This aligns with our finding from this dataset that going out is positively correlated with alcohol consumption

https://www.ncbi.nlm.nih.gov/pmc/articles/P MC4512649



4. Conclusion





^{*}But many other factors involved



5.Tech Stack



Tech Stack

Tools: VSCode, LiveShare



Languages: Python

Libraries: Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Statsmodels



Takeaways

o Don't drink too much, kids!



Thanks!

Any questions?

