

DataGood **x** School Alc Consumption

Datathon Spring 2024



The Team



Manan



Herman



Michael



Will



Background

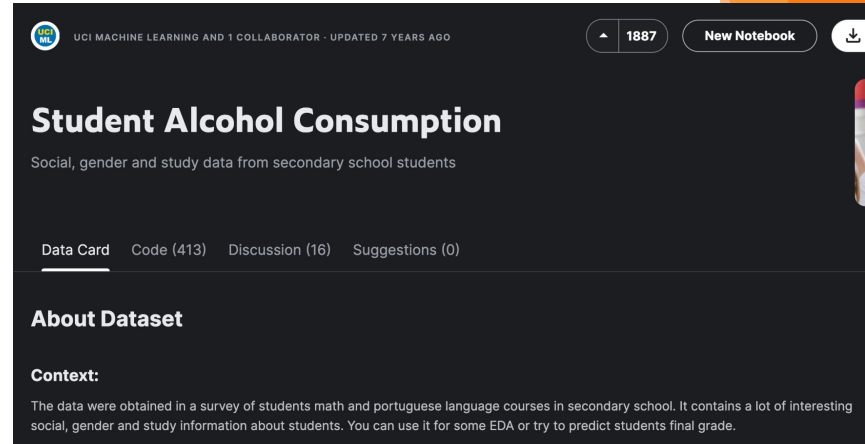
Project Area: Health & Education

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>



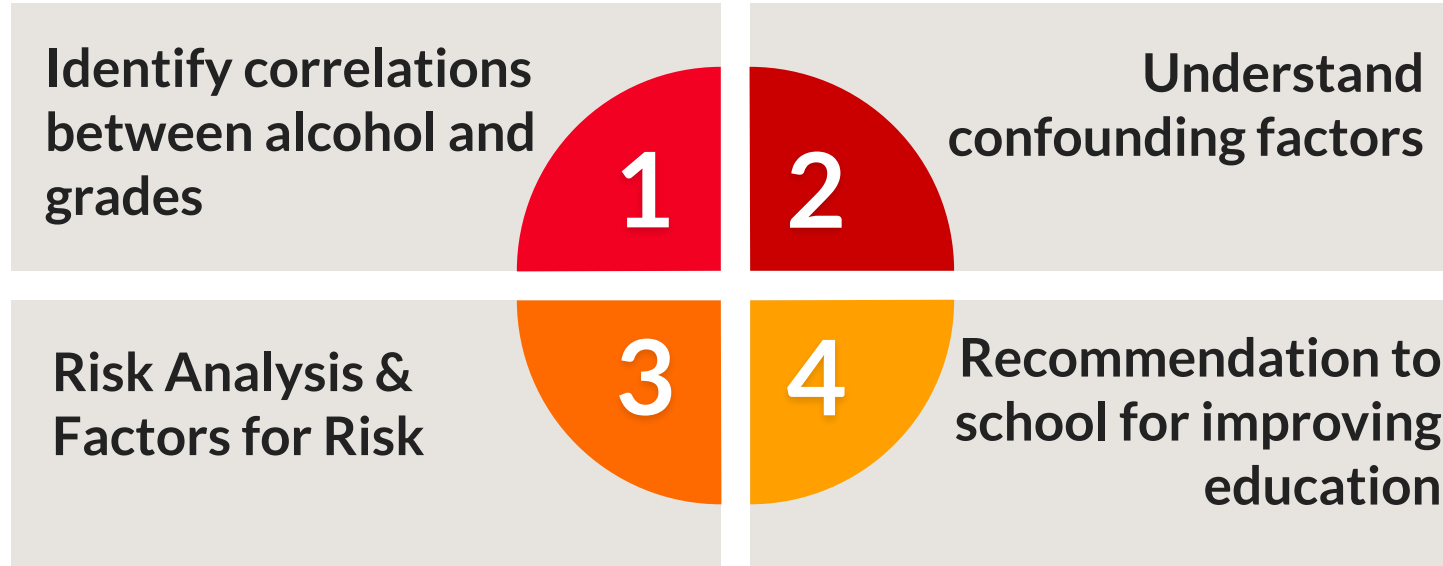


Factors to Consider in the Study

- 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:



Methodology

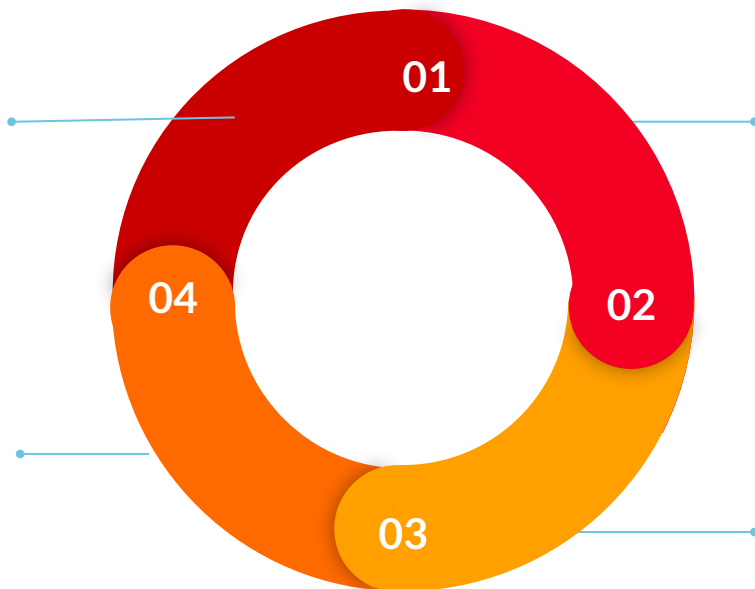


Data Collection

- Finding the dataset through online research

Model Creation

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



Data Cleaning & Processing

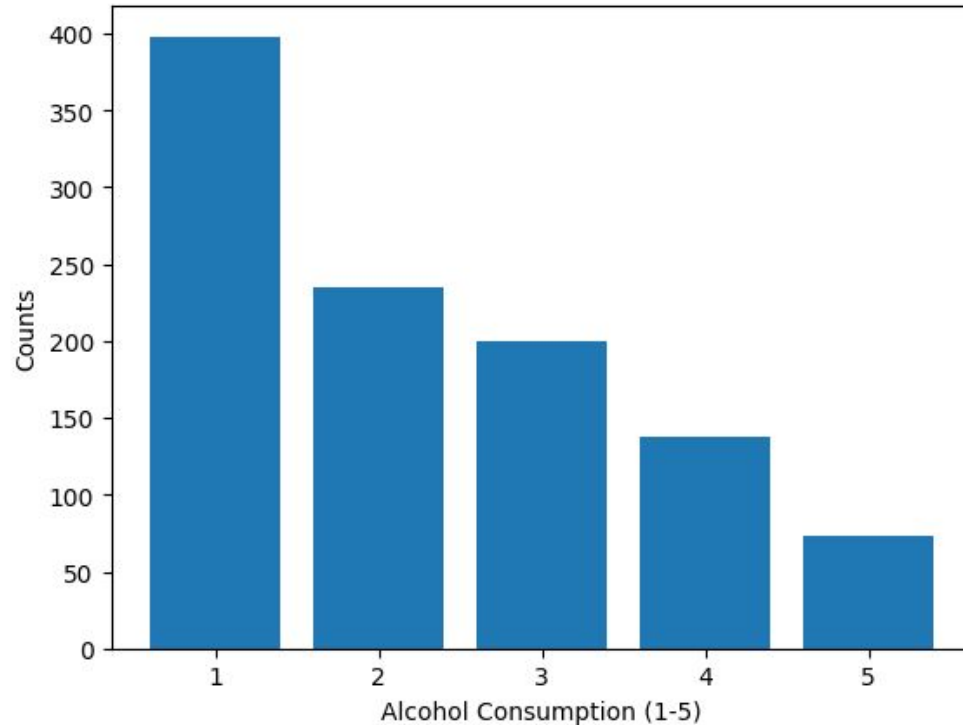
- The data was loaded into shared notebooks to fix or remove incorrect, corrupted, improperly formatted, duplicate, or incomplete data within the dataset.

Exploratory Data Analysis

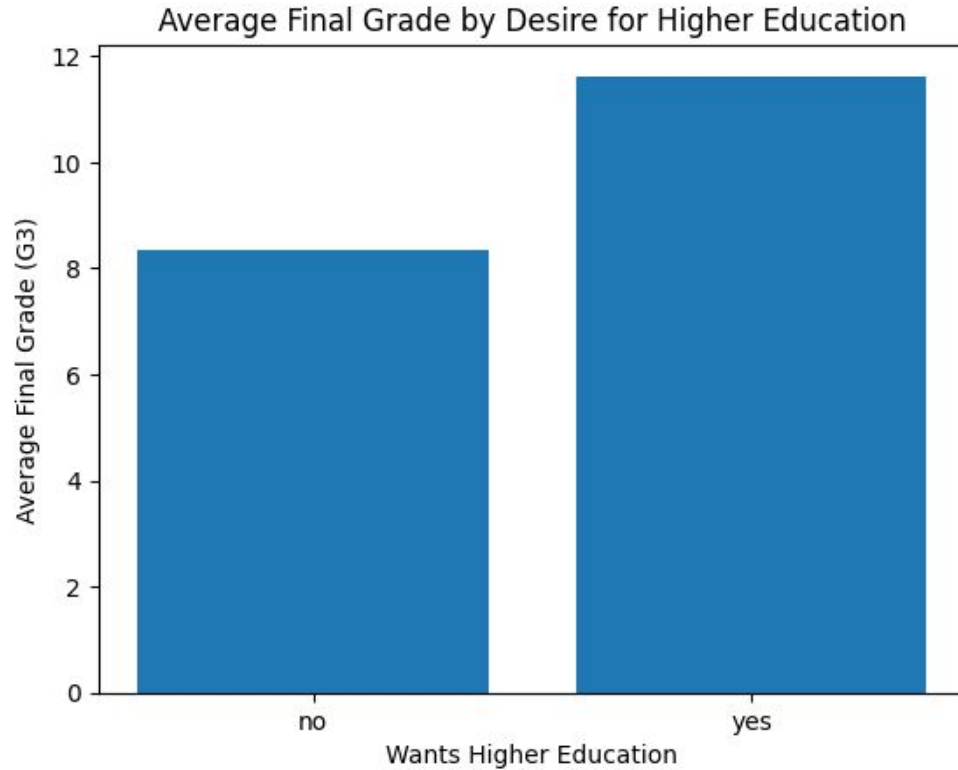
- Data visualizations were made in DeepNote notebooks using Python libraries.

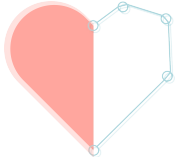


How Common Were Different Levels of Alcohol Consumption?



Did a Desire of Higher Education Impact Final Grades?





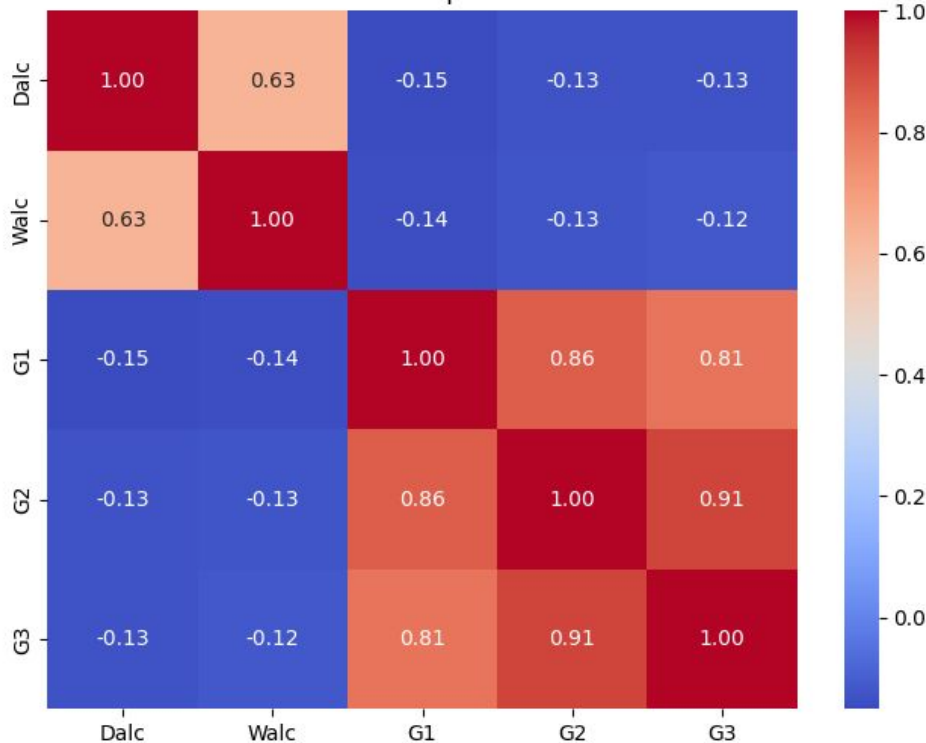
2.

Alcohol Affecting Grades?



Interesting Findings

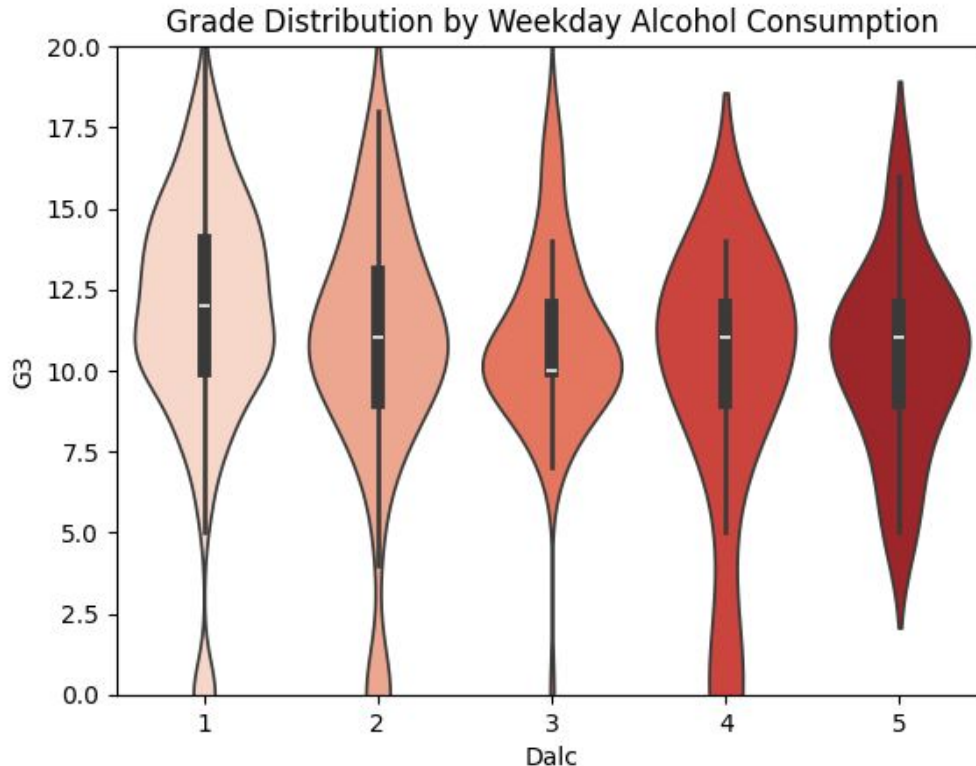
Correlation between Alcohol Consumption and Academic Performance



- 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.

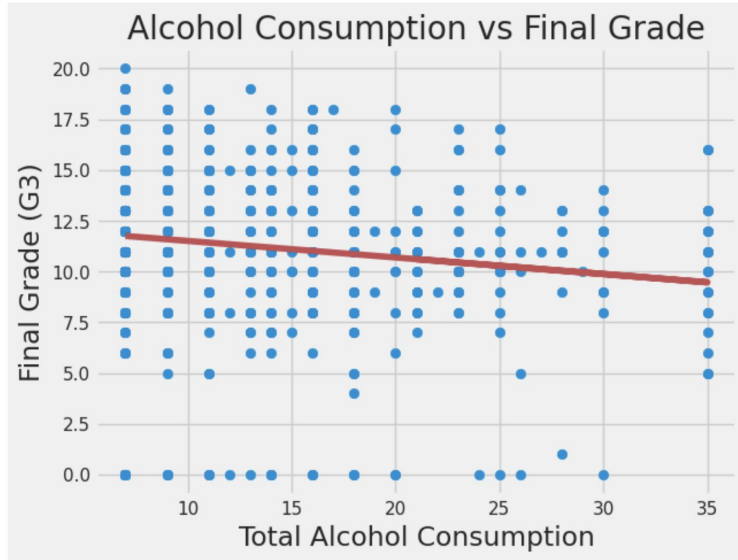


Interesting Findings



- 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
    denom = sum([xi**2 for xi in X]) - n * xbar**2

    b = numer / denom
    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']
```

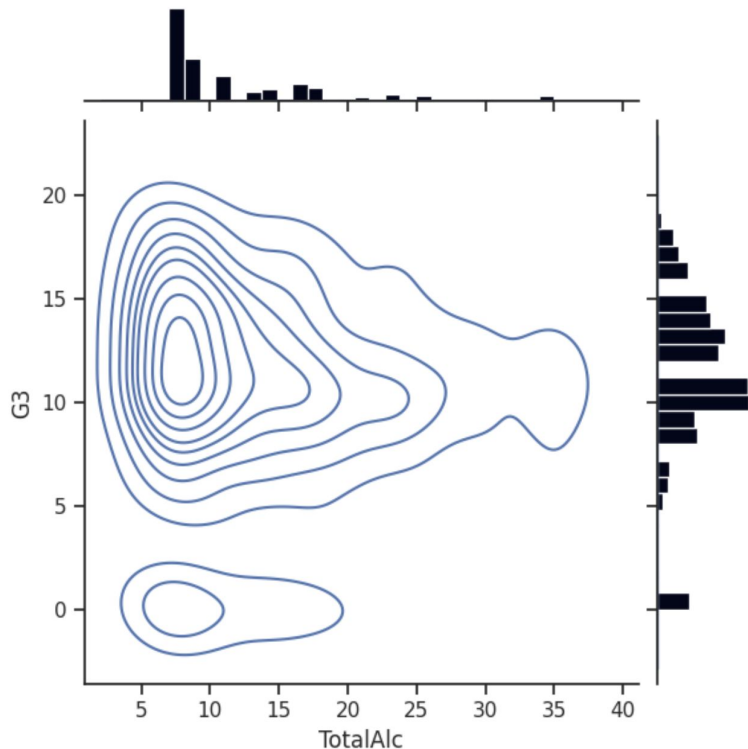
```
# solution
a, b = best_fit(X, y)
#best fit line:
#y = 0.80 + 0.92x
```

```
# 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()
```



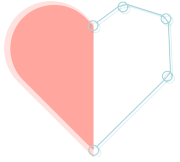


Negative Correlation



Joint Density Plot

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



3.

Predicting Alcohol Consumption



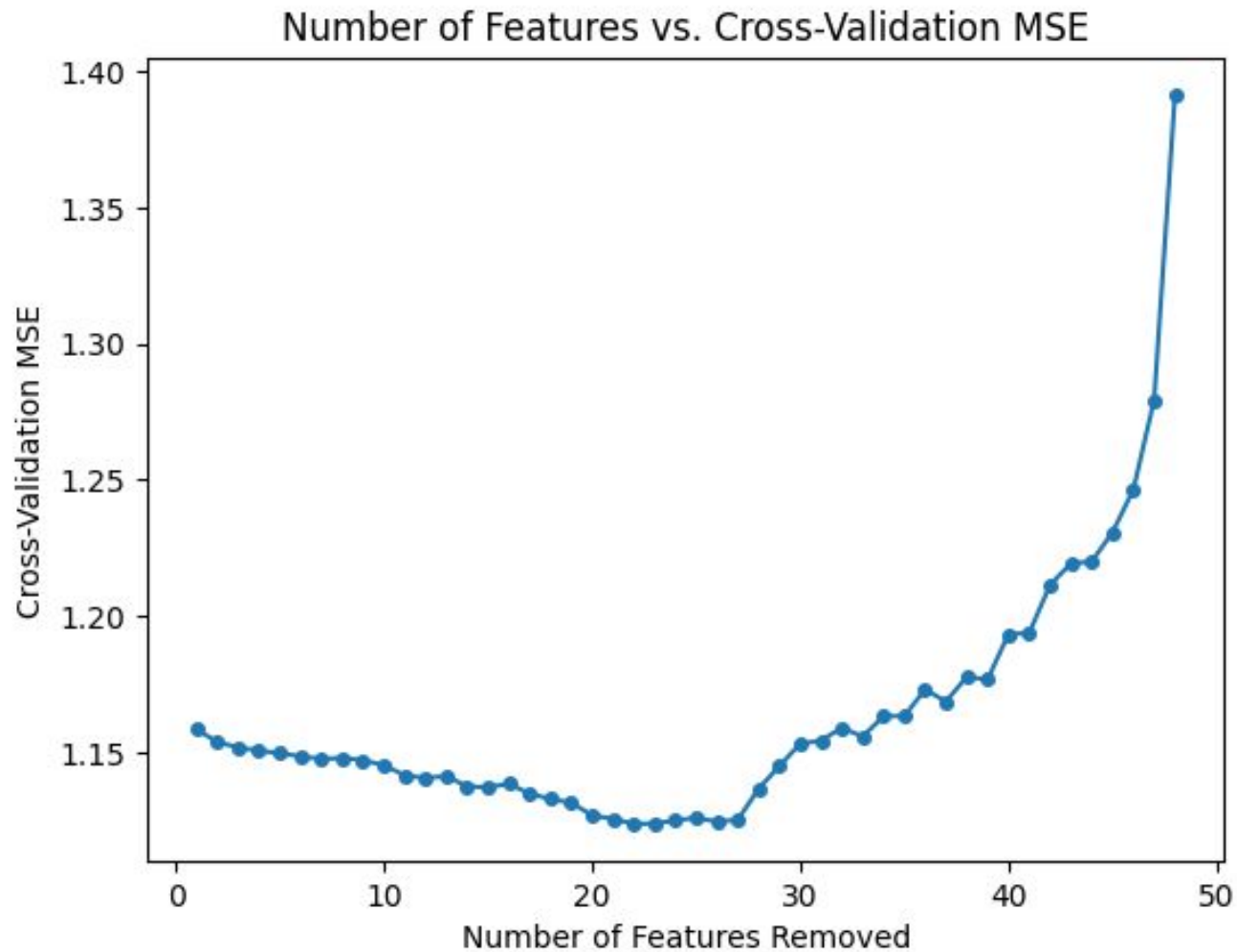
```
walc_X = X[walc_coefs]
walc_X2 = sm.add_constant(walc_X)
walc_model = sm.OLS(walc, walc_X2)
walc_model = walc_model.fit()
print(walc_model.summary(slim = True))
```

✓ 0.0s

OLS Regression Results

```
=====
Dep. Variable:          walc    R-squared:                0.357
Model:                  OLS     Adj. R-squared:           0.343
No. Observations:      1044    F-statistic:            25.77
Covariance Type:        nonrobust Prob (F-statistic):      1.58e-82
=====
```

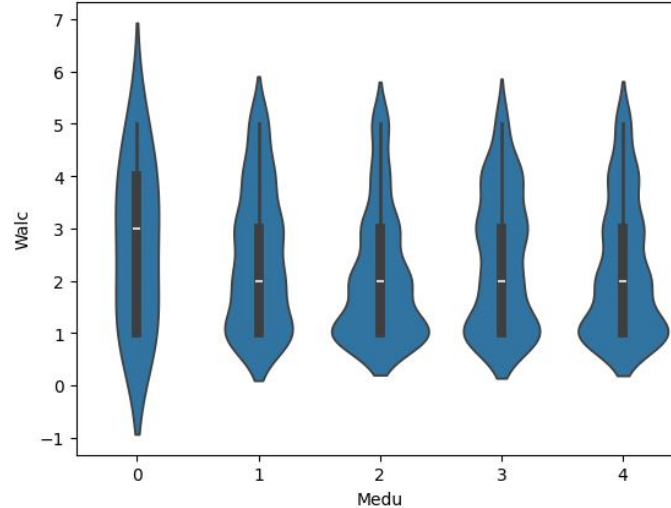
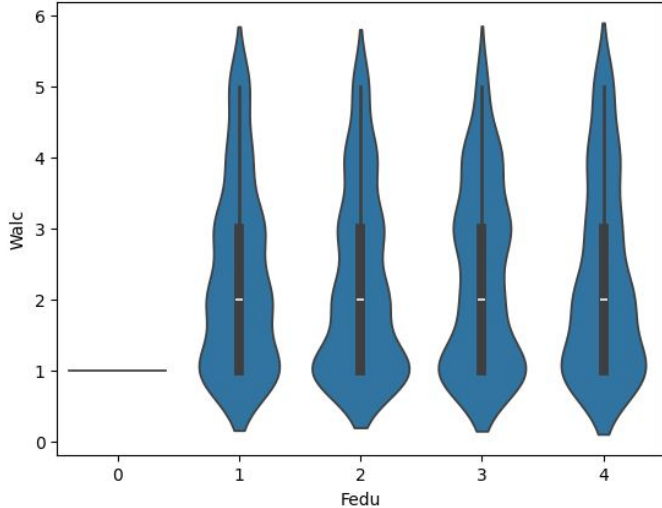
	coef	std err	t	P> t	[0.025	0.975]
const	-0.1698	0.626	-0.271	0.786	-1.398	1.058
sex_F	-0.5947	0.070	-8.460	0.000	-0.733	-0.457
age	0.0748	0.029	2.597	0.010	0.018	0.131
famsize_GT3	-0.1867	0.072	-2.579	0.010	-0.329	-0.045
paid	0.2388	0.081	2.941	0.003	0.079	0.398
nursery	-0.2325	0.083	-2.789	0.005	-0.396	-0.069
famrel	-0.2266	0.035	-6.403	0.000	-0.296	-0.157
goout	0.4079	0.029	14.191	0.000	0.351	0.464



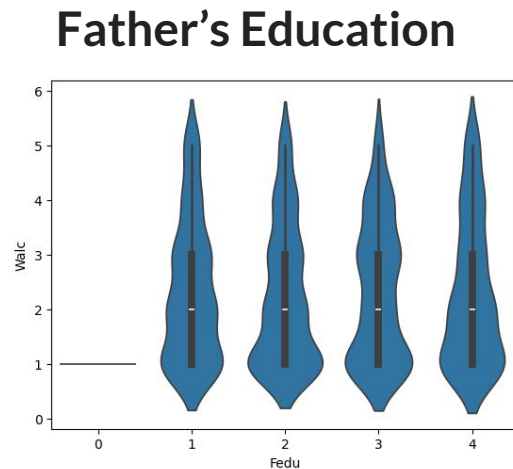
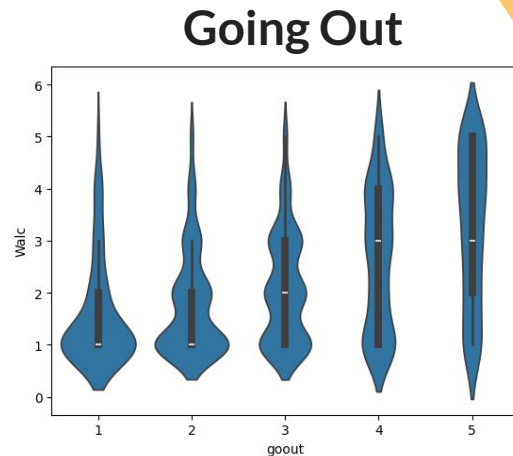
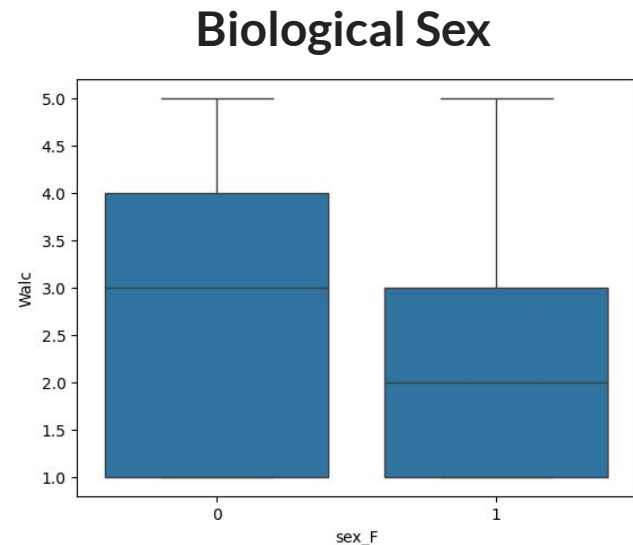


Interesting Findings

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



Strongest Predictors of Alcohol Consumption



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



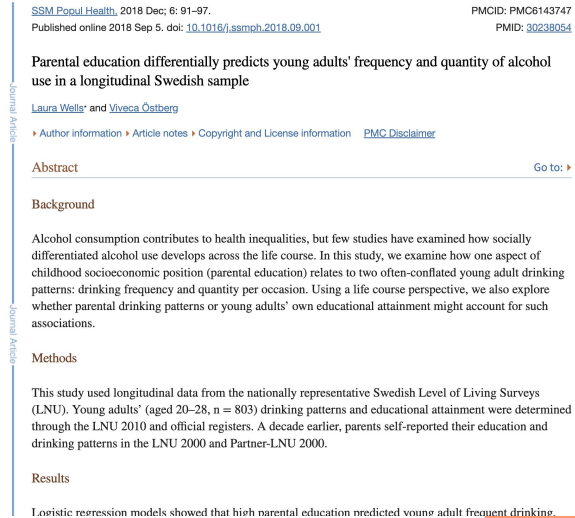
- **Grade Distributions:** more drinks → lower grades
- Suggested by the data



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



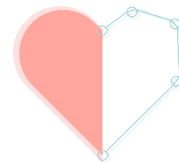
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6143747/>

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

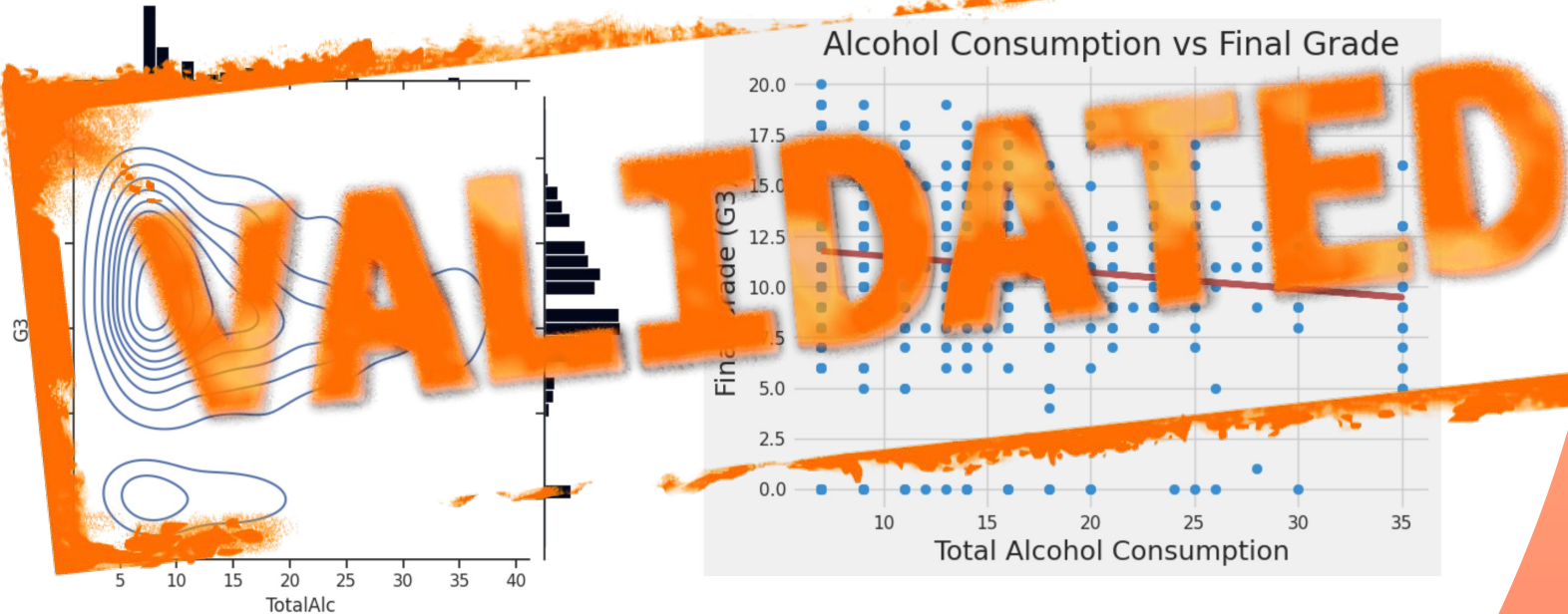
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512649>



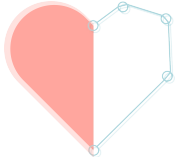
4. Conclusion



YES! Alcohol Negatively Affects Academic Performance*



*But many other factors involved



5. Tech Stack



Tech Stack

Tools: VSCode, LiveShare



Languages: Python

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



Takeaways

- Don't drink too much, kids!



Thanks!



Any questions?

