STUDENT ACADEMIC PERFORMANCE PREDICTION

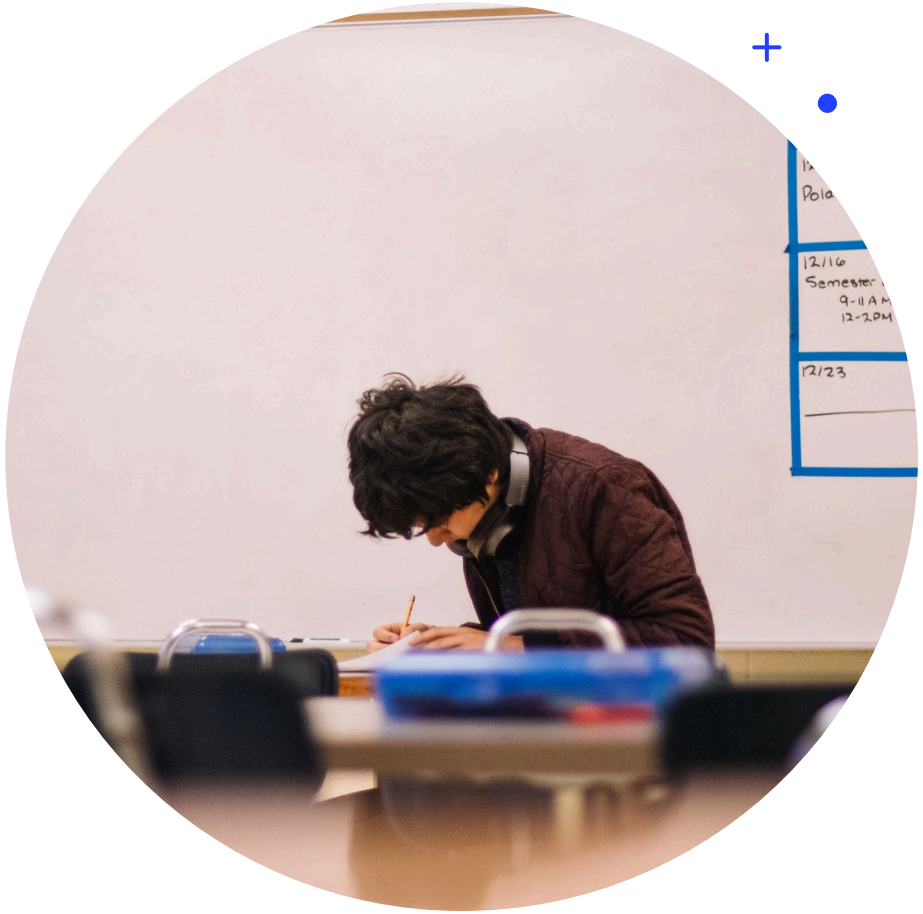
FINDING THE SECRET TO ACADEMIC PERFORMANCE

Gamze Turan

Overview

- My goal is to examine the **effects of several socio-economic factors** on the **grades of secondary school students**.

- These analyses will allow to predict student performance based upon a variety of features.



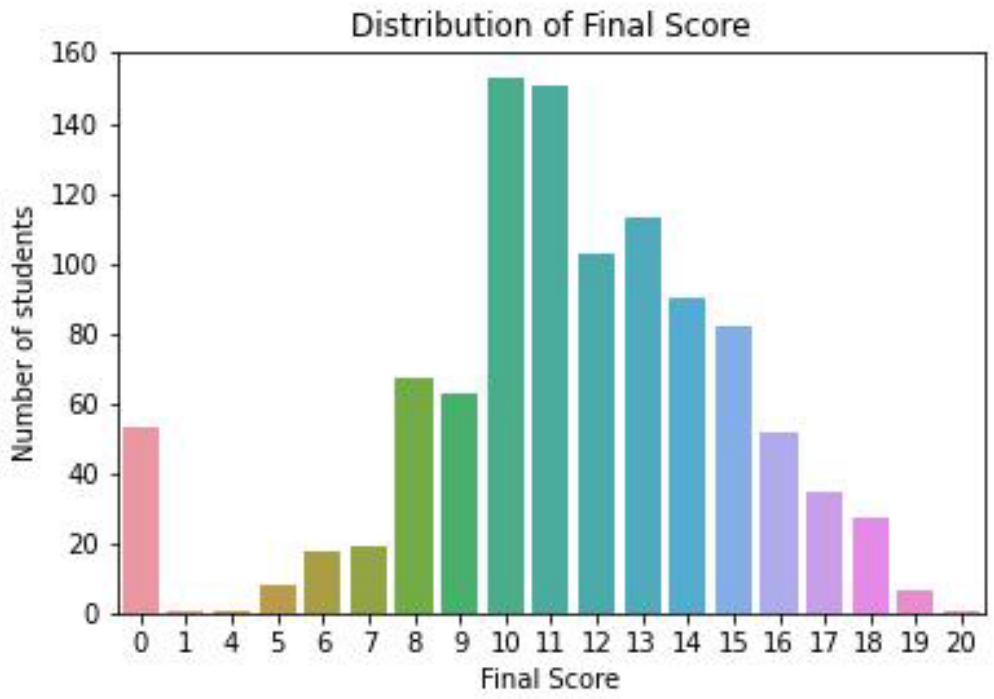
Business Understanding

I will try to build a **personal of a model student** who has the highest chance of academic success.

The result will be used by :

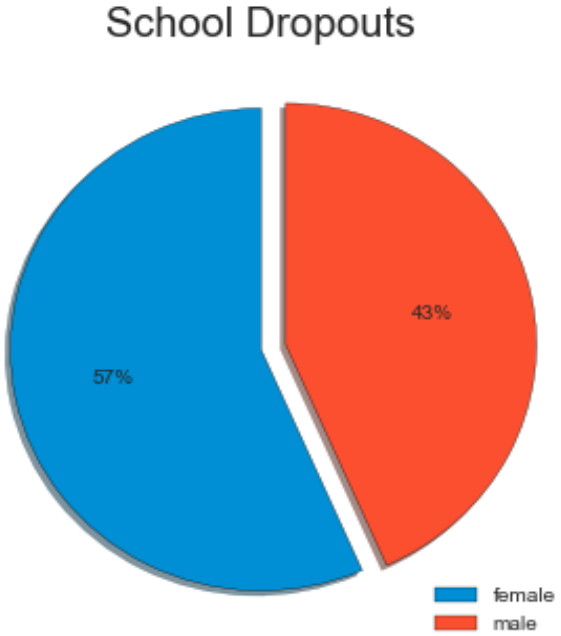
- school districts
- academic counsellors
- parents help guide
- Etc.

Data Understanding



The target is final score, lies between 0 and 20. The average score is 11.34.

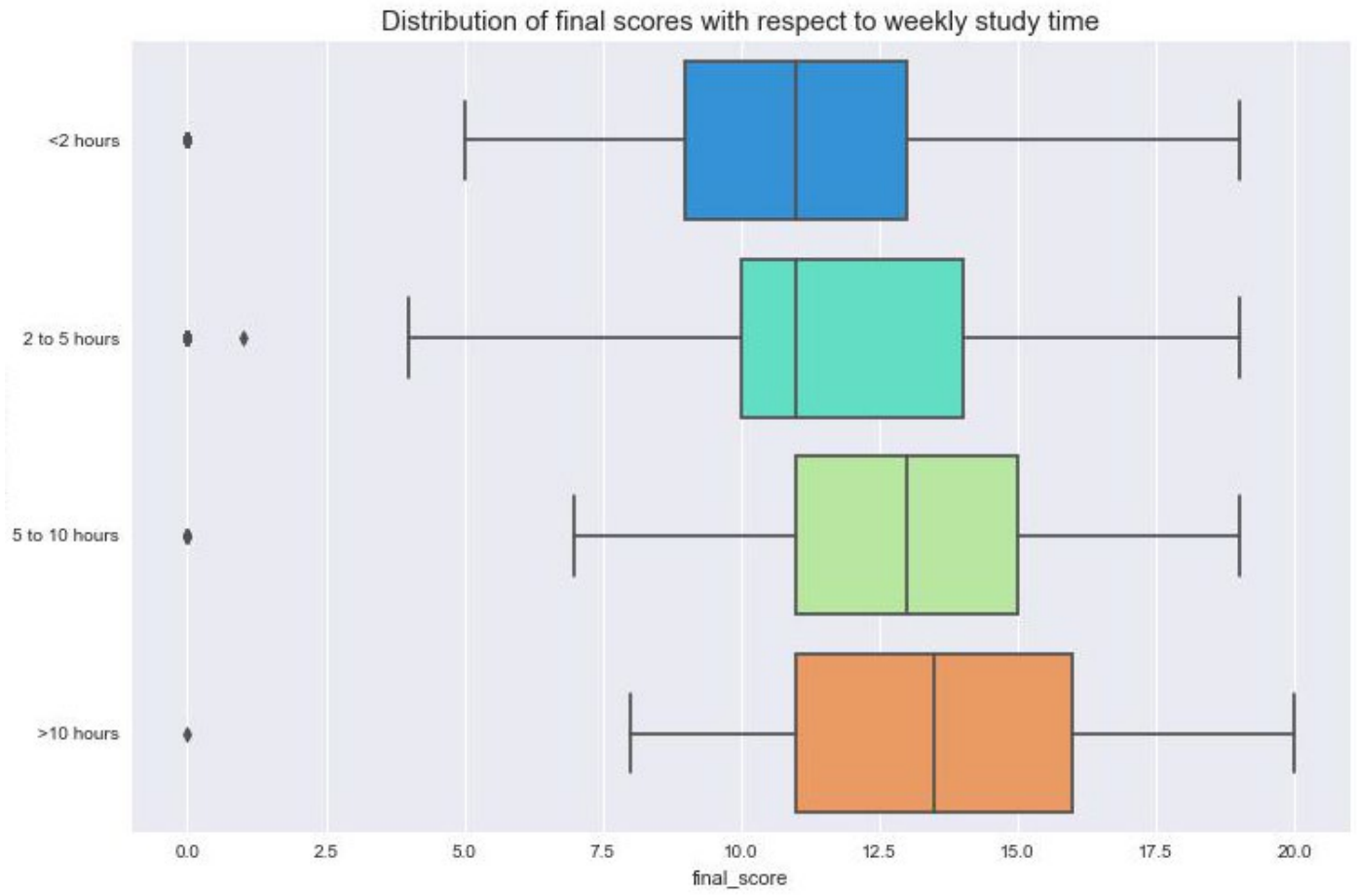
There are a lot of students that score 0 in the final score. These represent students who dropped out of school.



Overall females do better than males in school, but more females drop out of high-school.

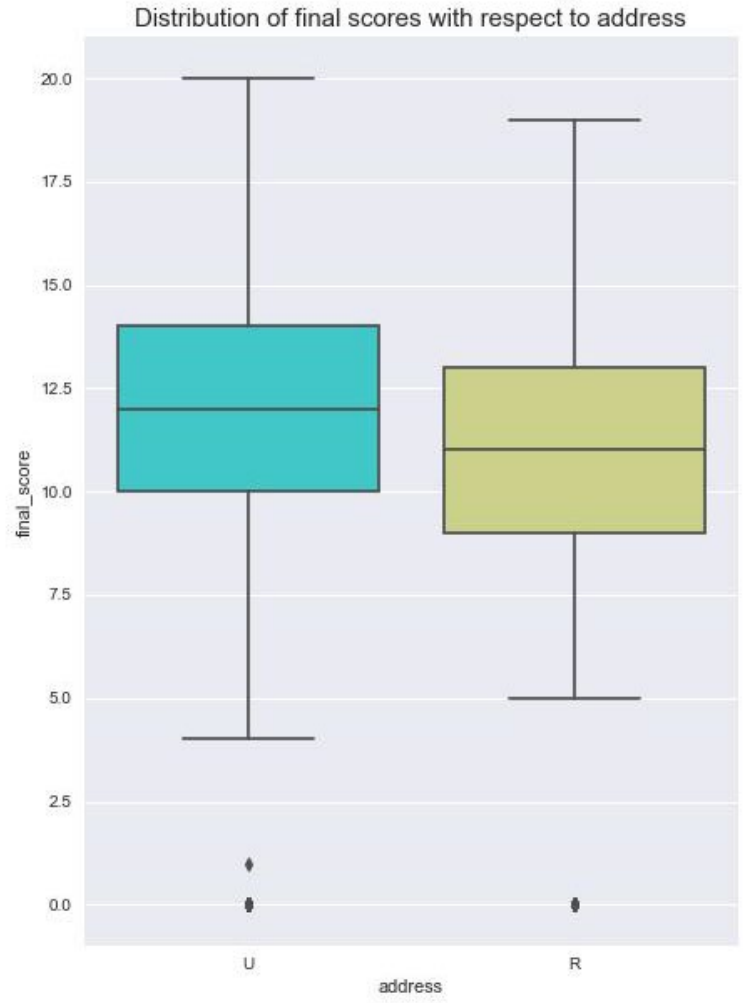
As many as 36% of high-school dropouts can be attributed to teen pregnancy!

Data analyzing: Study Time



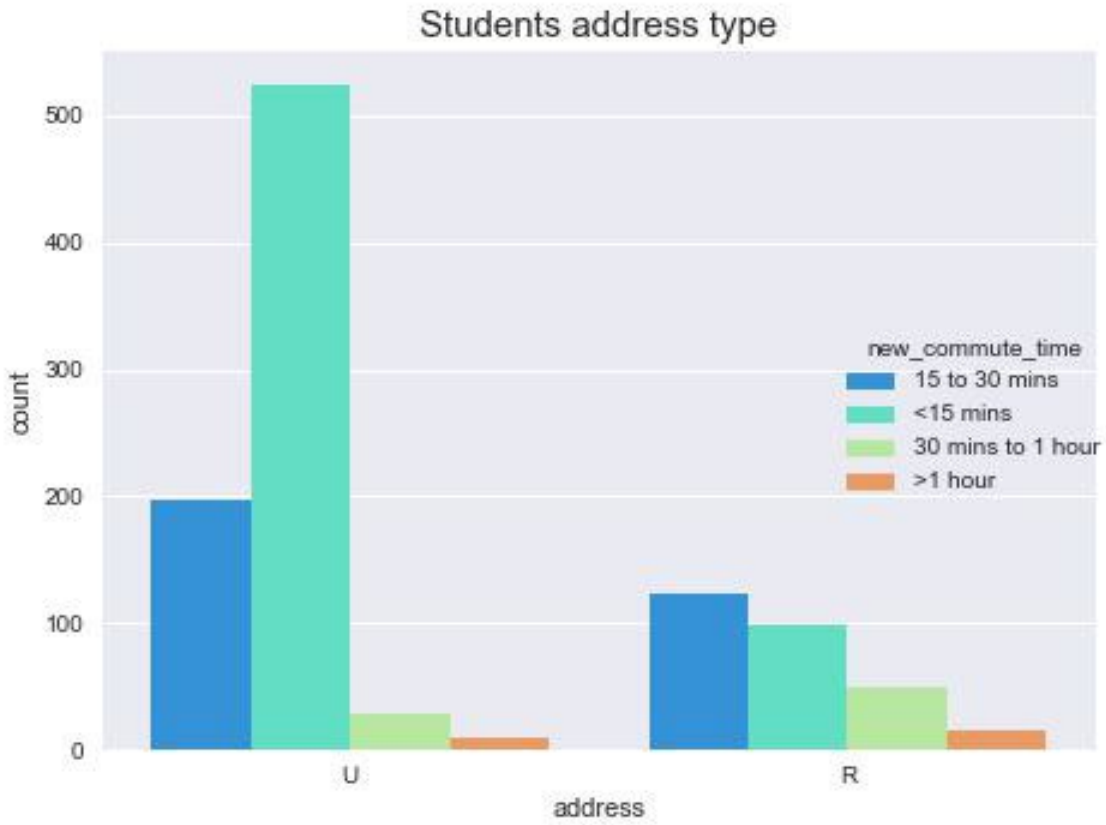
Increasing weekly study time improves final results.

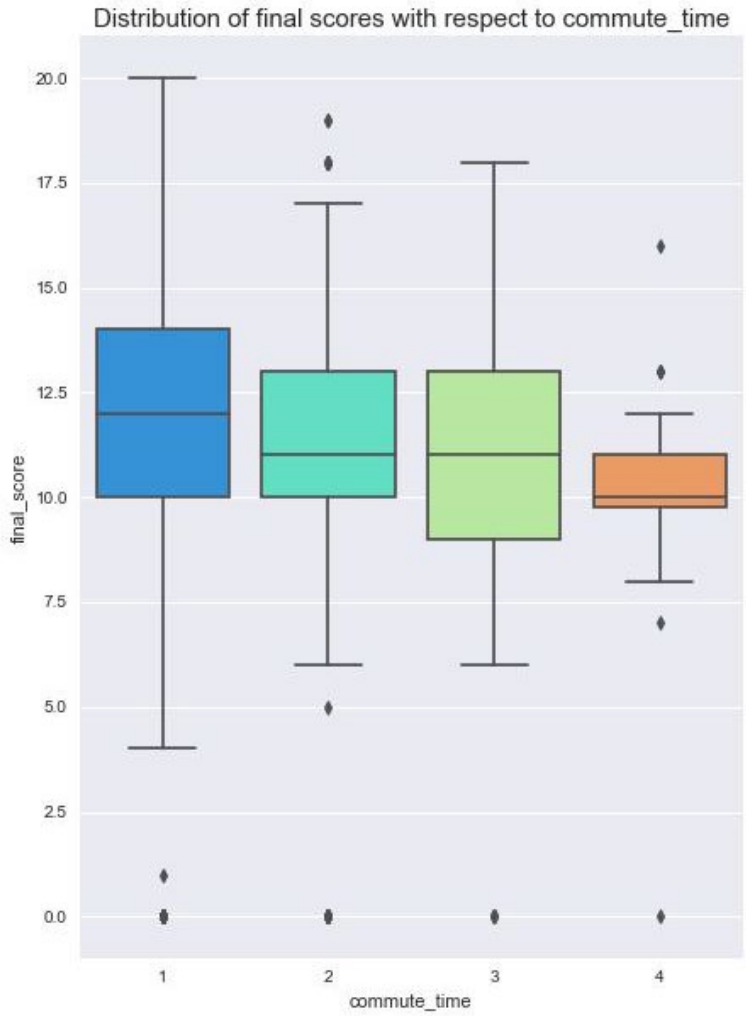
Data analyzing: Urban vs Rural



Urban Students tend to do better. Let's analyze some plausible reasons.

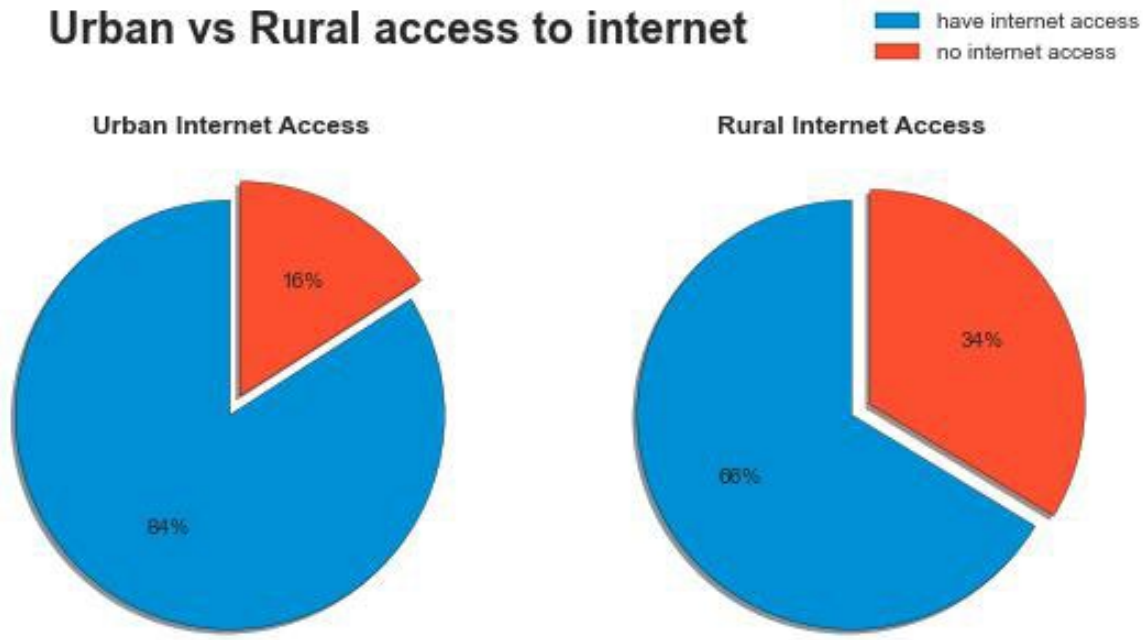
Data analyzing: Commute Time

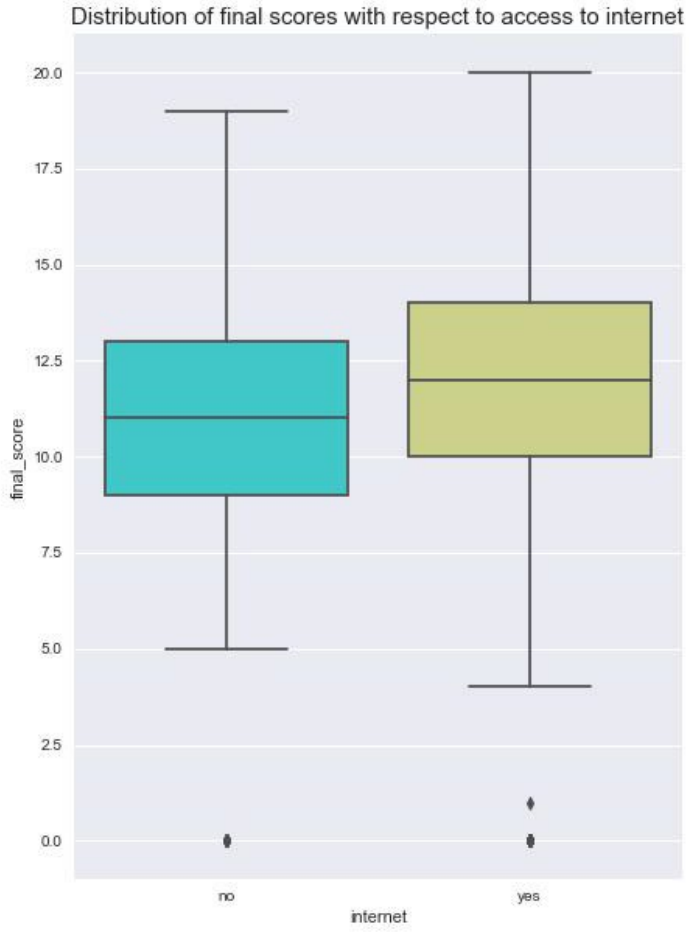




Commute time has an inverse relationship with final score and rural students spend more time commuting to school.

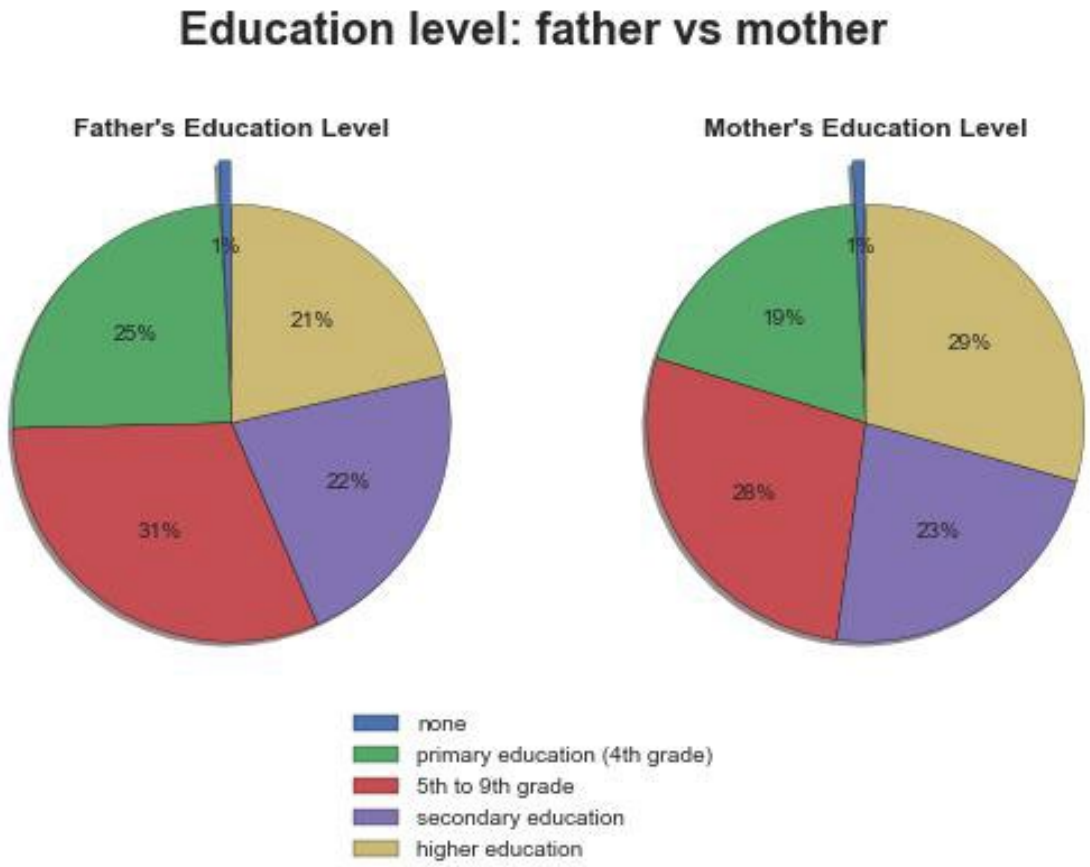
Data analyzing: Internet Access





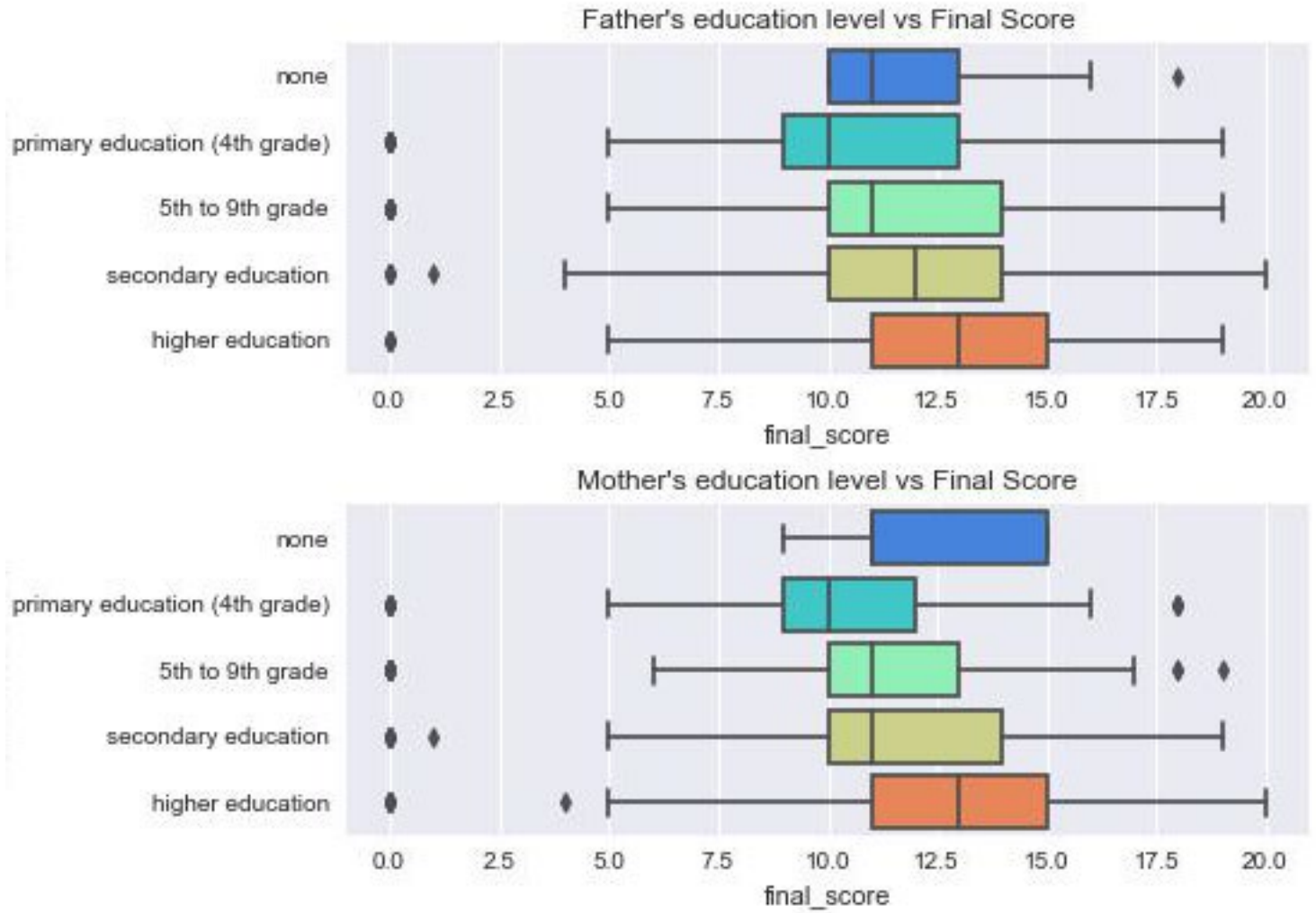
Urban Students have better access to internet and students with access to internet perform better.

Data analyzing: Parents Education



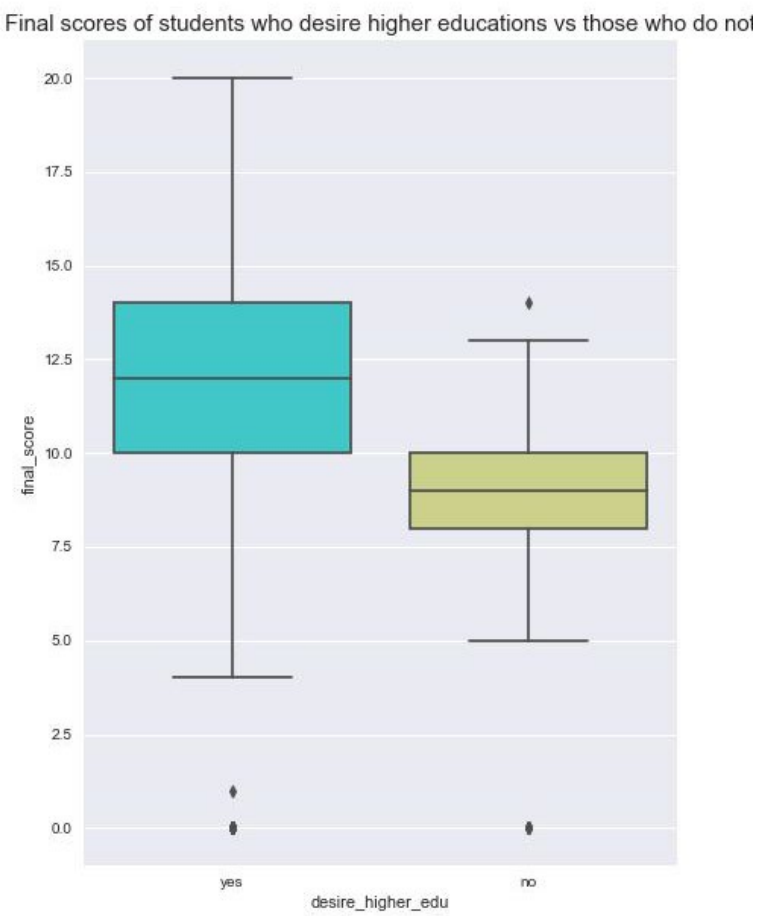
Usually, mothers are better educated than fathers. None represent the cases where no information is available(orphans).

Data analyzing: Parents Education

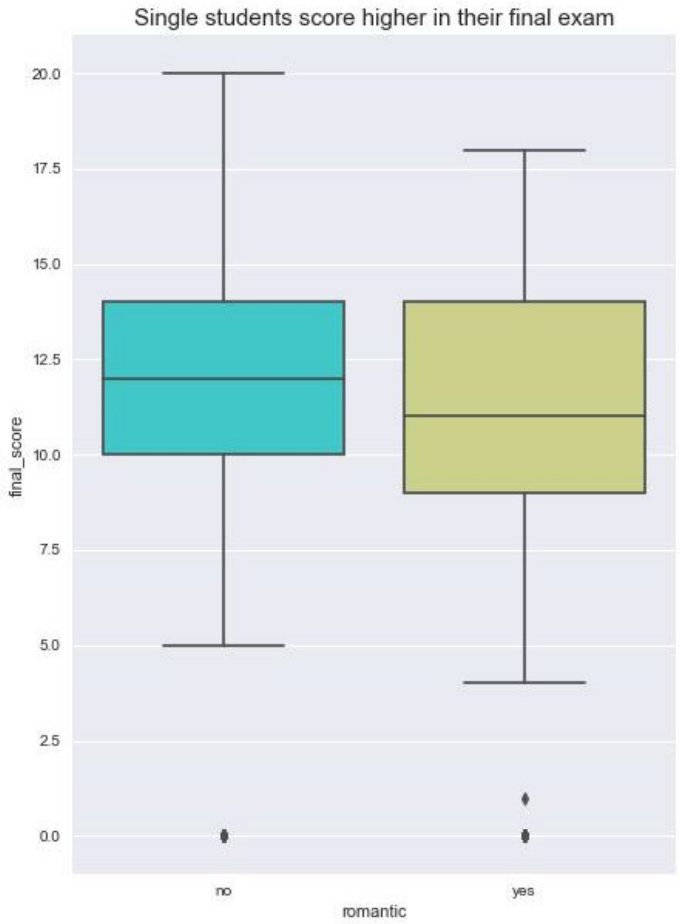


As parent's education level improves so does a child's academic performance.

Data analyzing: Higher Edu and Relationship Status

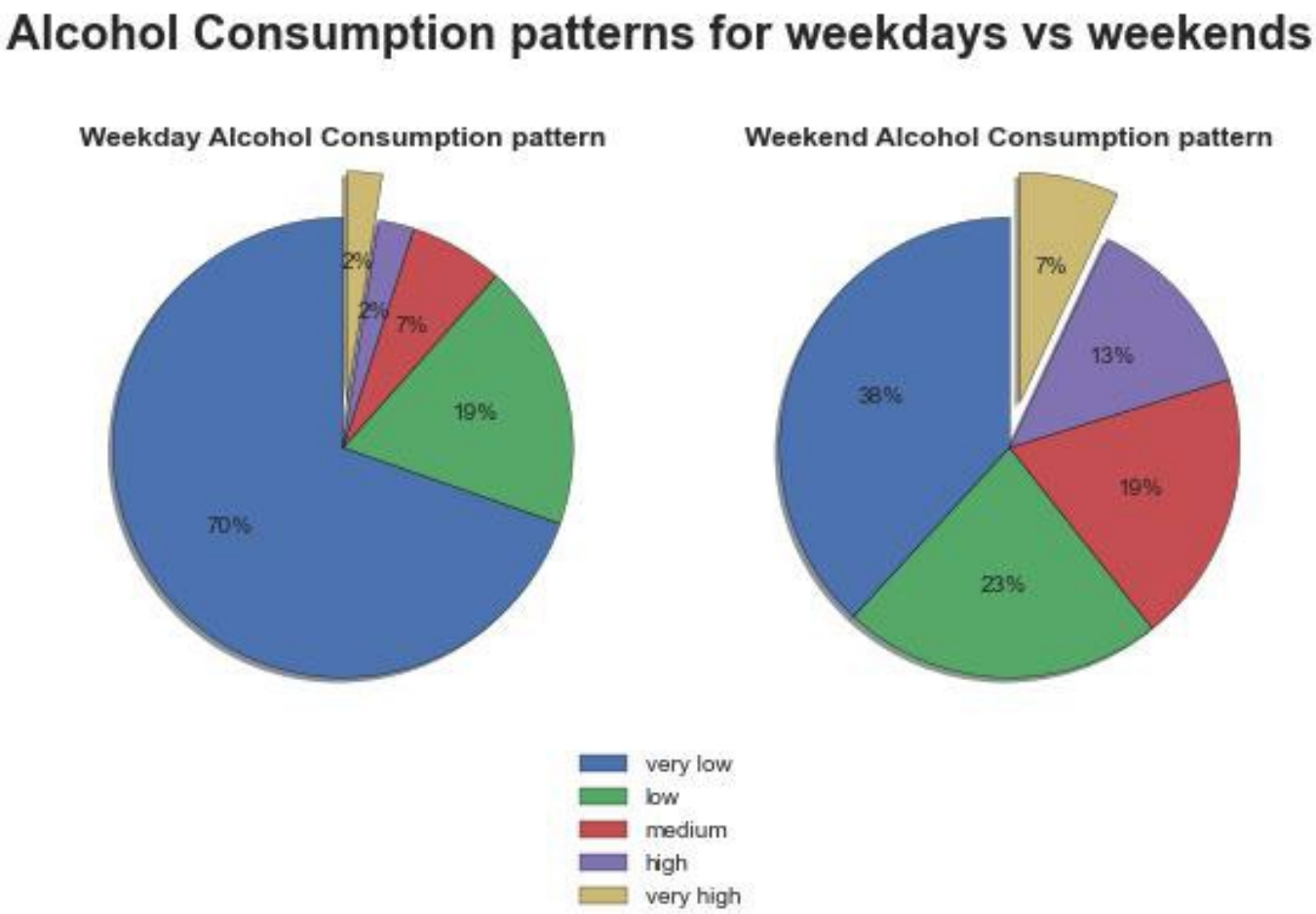


Students who wish to go for higher education perform much better than students who don't.



Students who are not in a relationship perform slightly better than students who are.

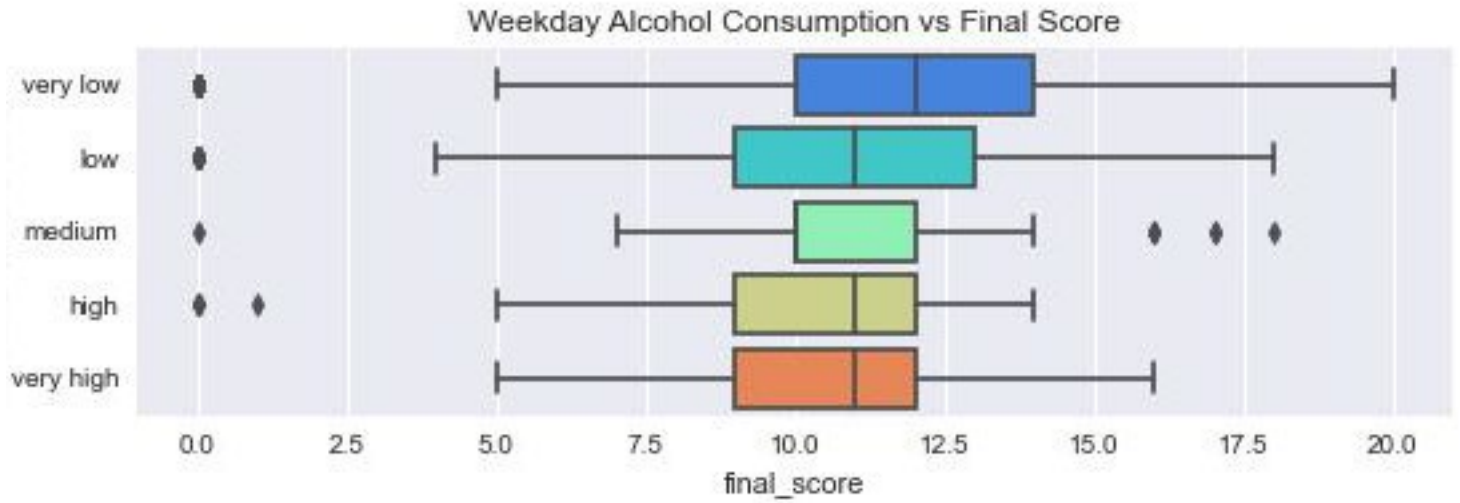
Data analyzing: Alcohol Consumption

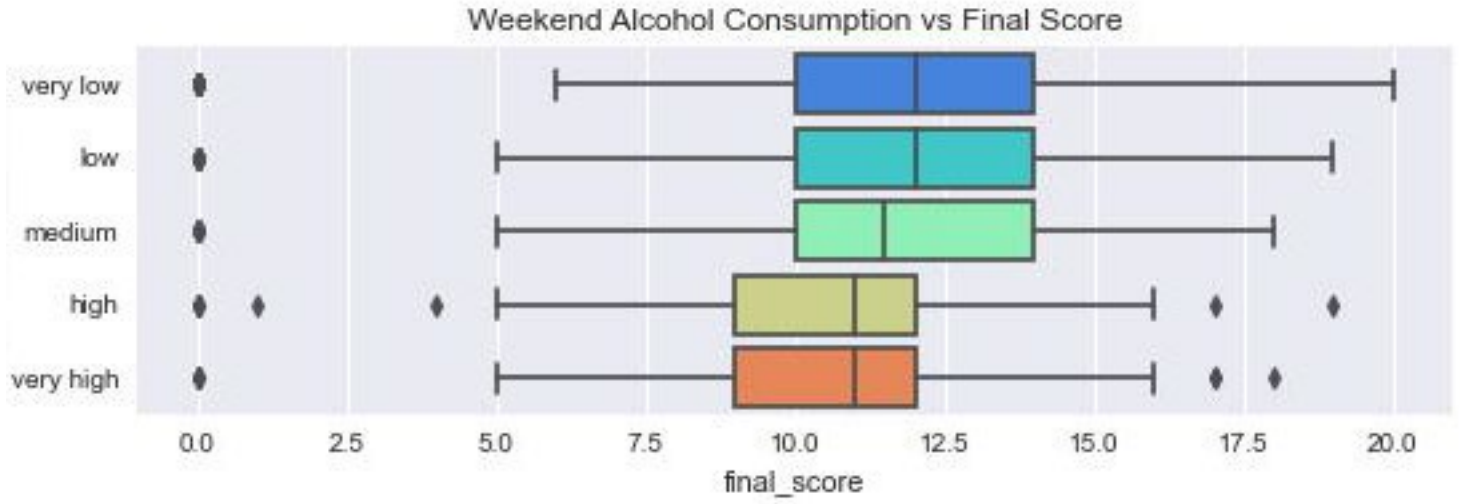


It's good to see that 70% of the students drink very low alcohol on weekdays.

As expected to alcohol usage increases over the weekend

Data analyzing: Alcohol Consumption





Increasing Alcohol consumption has detrimental effects on student performance.

Effect is more profound in the case of weekday drinking.

Modeling:

Objective : Predict Student Final Score
Output would be a predicted final score from 0 to 20

Data Wrangling:
Split the data into training, validation and testing subset

Feature Selection:
Use Random Forest to get a subset of “important” features to use for final modeling.
We receive 16 features as follows:

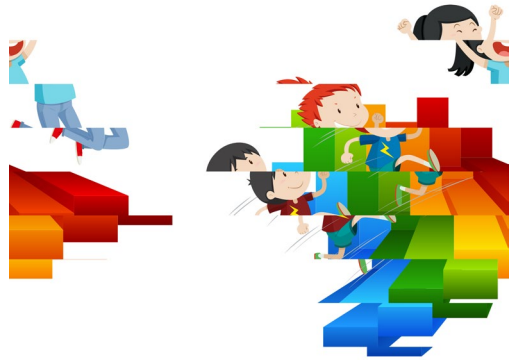
Study Time	Mother Education	Father Education	failures	Free Time	Go out	Weekday alcohol usage	Weekend alcohol usage
health	absences	School support	Paid Classes	Desire higher edu	Period1 score	Period2 score	age

Modeling:

Model Name	RMSE	MAE	R-Squared
Baseline Regressor	2.86	2.37	-0.033
Linear Regression	0.85	0.68	0.908
KNN	0.94	0.71	0.888
Decision Tree	0.83	0.68	0.912
Random Forest	**0.81**	**0.65**	**0.917**
AdaBoost	0.86	0.69	0.906
XGBoost	0.82	0.65	0.915

I modeled the regression task using multiple models. The best accuracy values that I received were for the Random Forest model.

Result:



1. Increasing weekly study time perform better in final results.

2. Increased commute time has a negative impact on academic performance

3. Student’s who has better internet access, perform better.

4. Parent's education status seems to be more relevant to student's success.

5.Students who wish to go for higher education perform better in school.

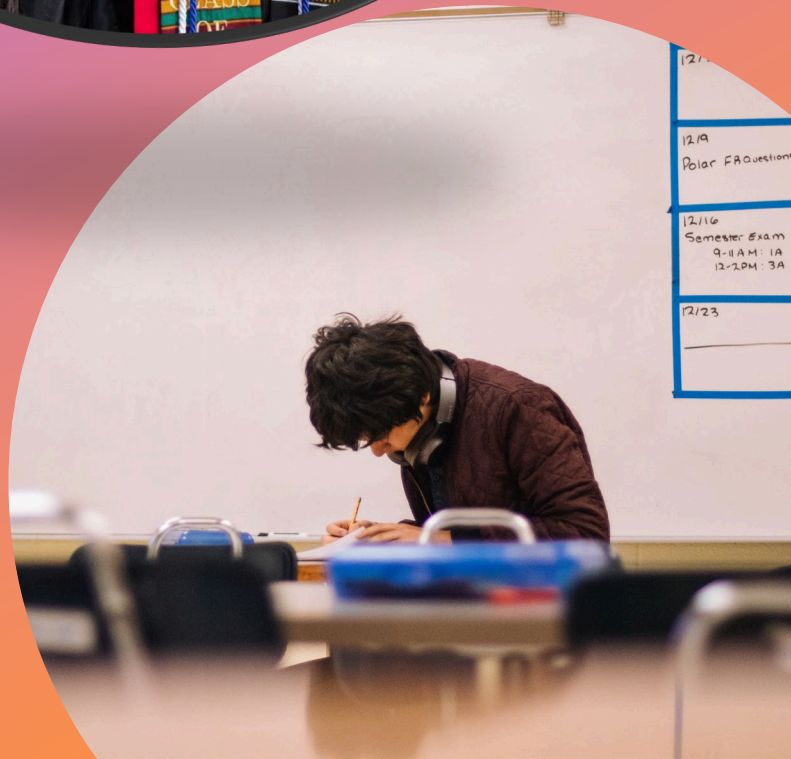
6. Alcohol consumption also has a negative impact on academic performance.

CONCLUSION

Student's academic success is usually thought to be dependent on student's dedication and hard work.

This analyses pointed out that while those parameters are important, there are other background factors that impact student's final score.

These ideas were than supported by the machine learning modeling stage which showed that the same features were important for predicting final score.



Future Work

School Dropouts

Study school dropouts in detail and analyze factors that lead to students leaving school.

Recommendation and Guidelines

Generate guidelines based on current analysis and use them to design experiments with school districts to see what helps students the most.

Deploy model on webapp

Deploy our ML model on a web application which can be integrated with the system used by school districts already.

References

- P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture Business TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7. Available at: [Web Link].
- UCI Machine Learning Repository: Student Performance Data Set
- Teen Pregnancy as cause of School Dropout (link)
- Pradhan, R. & Sinha, Niharika. (2017). Impact of commuting distance and school timing on sleep of school students. Sleep and Biological Rhythms. 15. 10.1007/s41105-017-0091-0. (link)
- Reading Scores by Computer Use and Internet Access at Home (2015), National Center for Education Statistics (link)
- DeVries JM, Rathmann K, Gebhardt M. How Does Social Behavior Relate to Both Grades and Achievement Scores? Front Psychol. 2018 Jun 4;9:857. doi: 10.3389/fpsyg.2018.00857. (link)
- Esser MB, Clayton H, Demissie Z, Kanny D, Brewer RD. Current and Binge Drinking Among High School Students — United States, 1991–2015. MMWR Morb Mortal Wkly Rep 2017;66:474–478. (link)

THANK YOU

Email: ginaturan15@gmail.com

Github: @gt1719

Linkedin: https://www.linkedin.com/in/gamze-turan-958a8012b/