SC1015 mini project

Jupyter Notebook

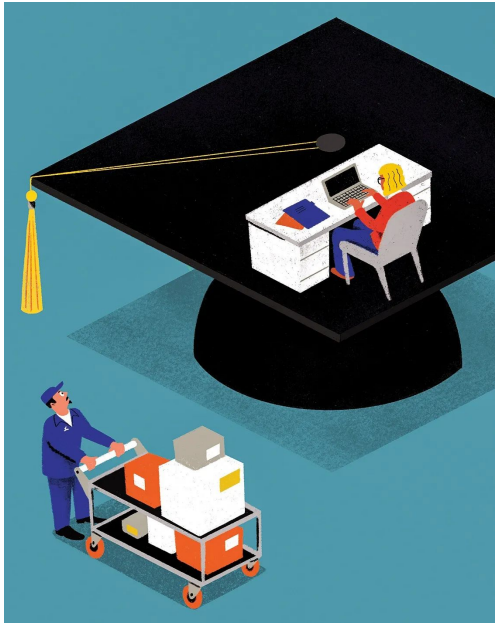
Optimisation of the environment to improve students' grades

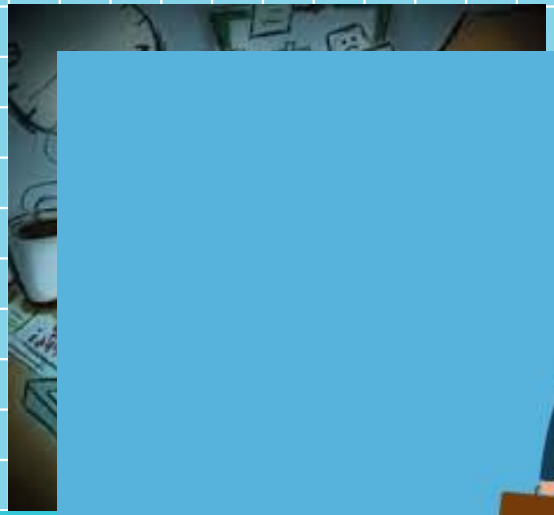
Presented By: THOR JIA YING
HAW JIN YU
ETHAN WONG

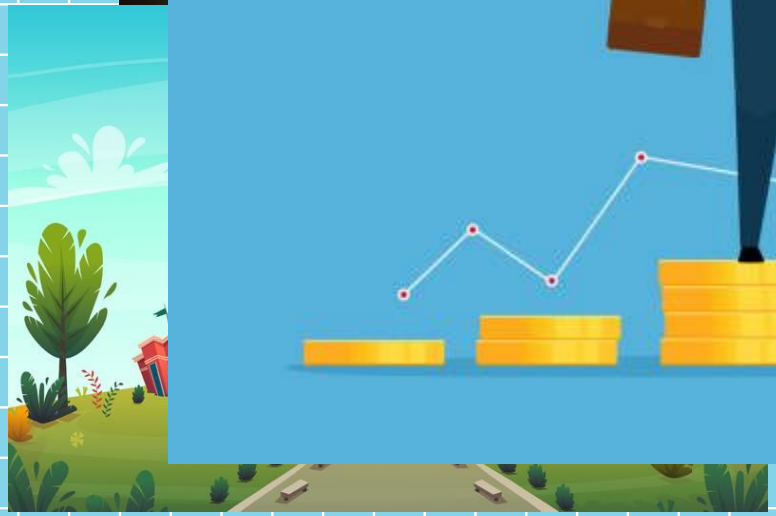
Table of contents

Table of contents

Introduction







Problem Formulation & Motivation

Death of boy, 11, who fell 17 floors after failing his exams for the first time ruled a suicide

-Straits time, 2016

Today Online, 2016 -

Death of Pri 5 student a deliberate act of suicide: Coroner's Court

MOE, MSF 'very concerned' about spike in youth suicides; experts say more support and awareness necessary

-CNA 2019

Problem Formulation & Motivation

How can we optimise the environment to improve Students' Grades?

1

How can we sieve out students who are in need of help?

2

What are the factors that are important in determining our grades?

Problem Formulation & Motivation

- Too ingrained within us to change
- Prevention is better than cure
- Too late to deal with after it happens

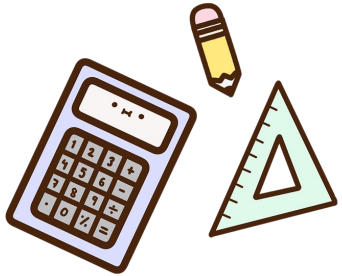
Introduction to dataset

UCI MACHINE LEARNING · UPDATED 7 YEARS AGO

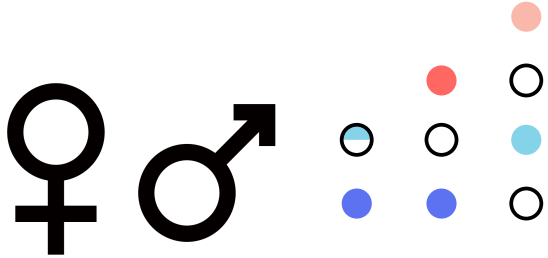
1723 | New Notebook | Download (19 kB)

Student Alcohol Consumption

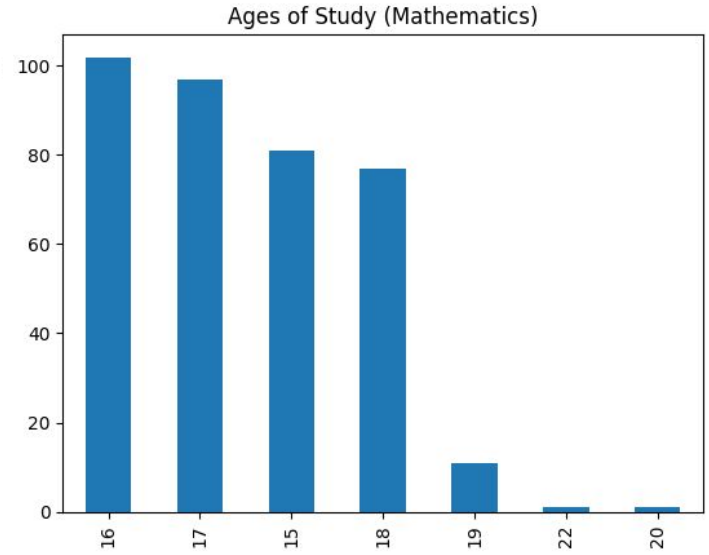
Social, gender and study data from secondary school students



Cleaning the Dataset: Redundant Variables



Cleaning the Dataset: Outlier removal

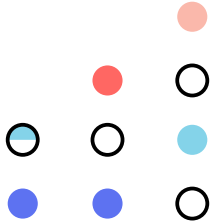


15: 81
16: 102
17: 97
18: 77
19: 11
20: 1
22: 1

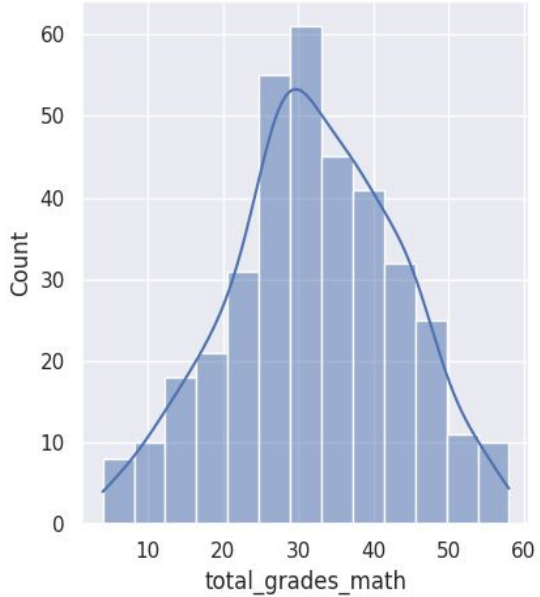
Cleaning the Dataset: Outlier removal

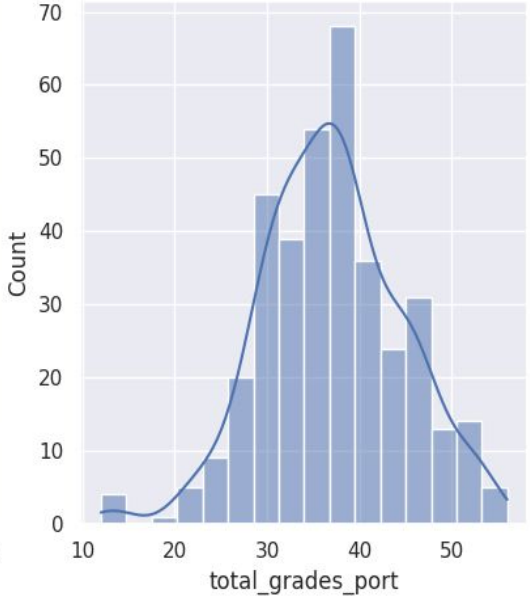
Portugal Secondary School

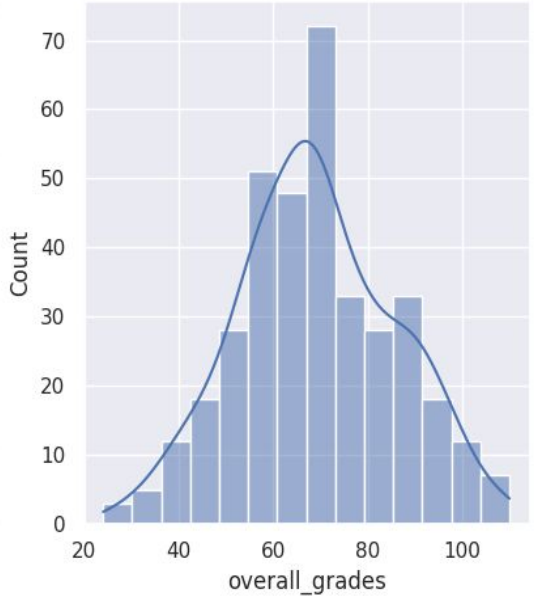
15-18 Years old



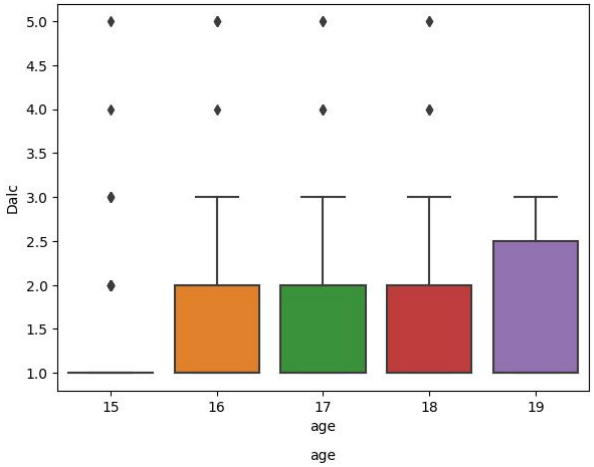
EDA: Grades


Count
total_grades_math

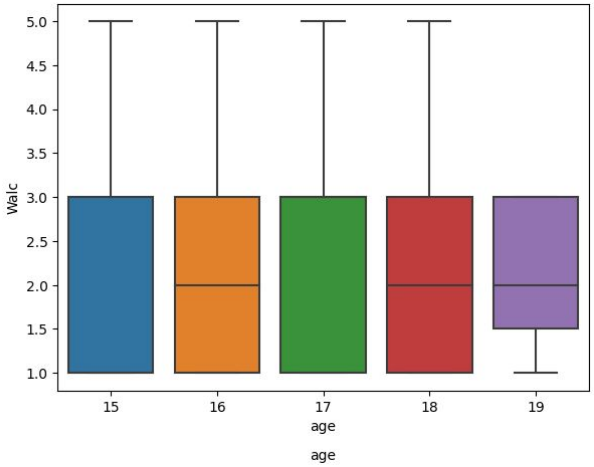

Count
total_grades_port


Count
overall_grades

EDA: Alcohol Consumption



Weekday Alcohol Consumption

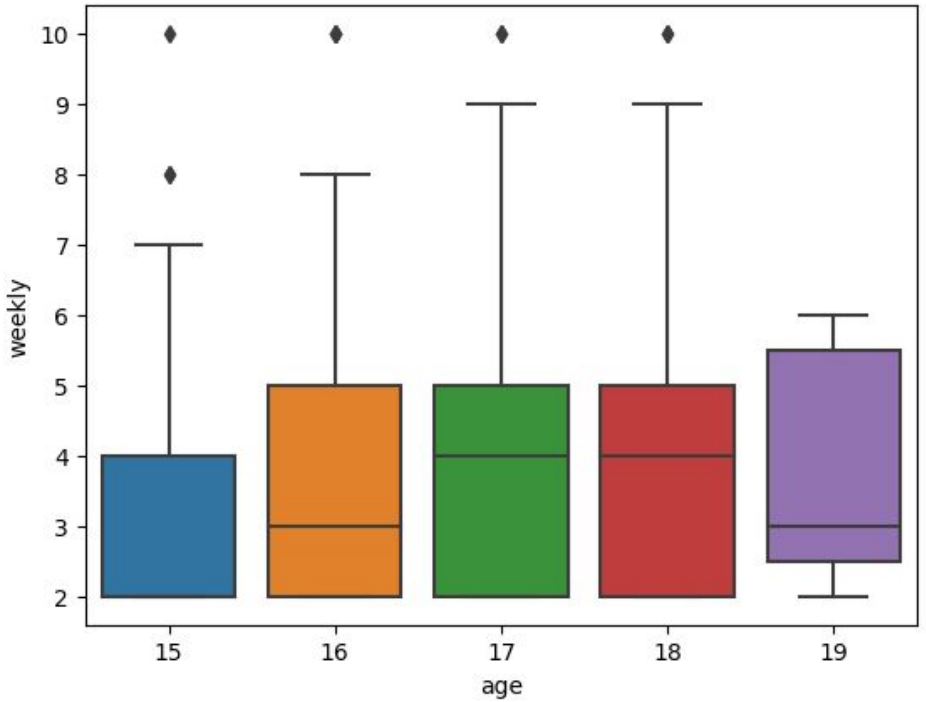


Weekend Alcohol Consumption

1: Very low
2: Low
3: Medium
4: High
5: Very High

EDA: Alcohol Consumption

Weekly Alcohol Consumption
0: Little
1-3: Some
4-6: Medium
>6: High



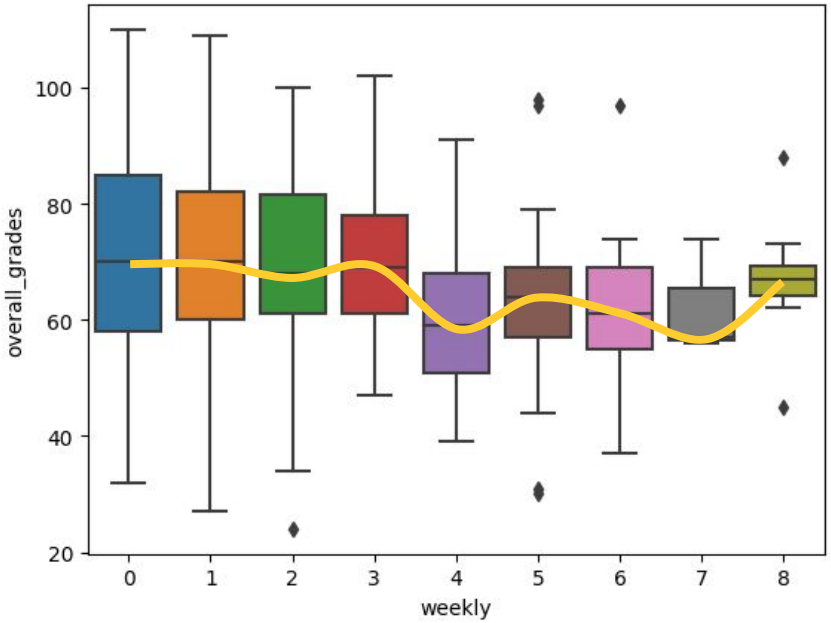
EDA: Alcohol Consumption

<15: Able to drink **under supervision**

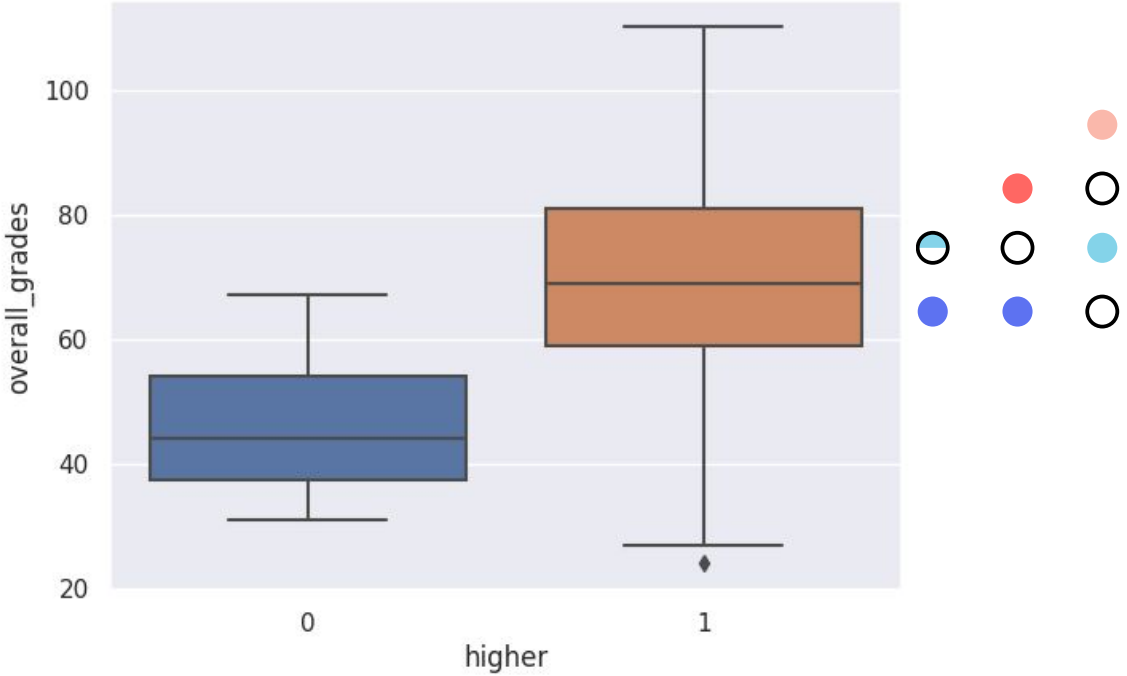
>18: Allowed to **purchase** alcohol

>16: Allowed to drink **wine** and **beer** in **restaurants**

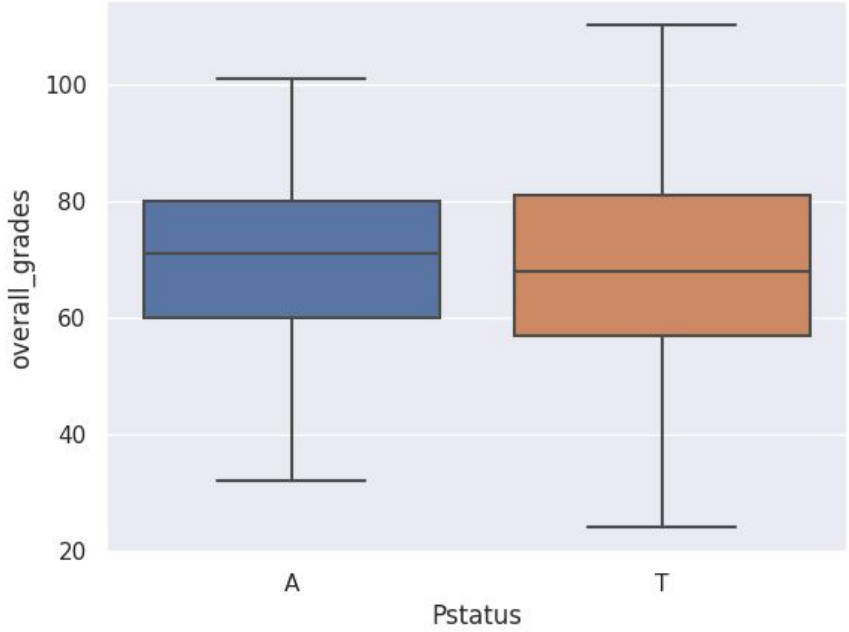
EDA: Alcohol Consumption



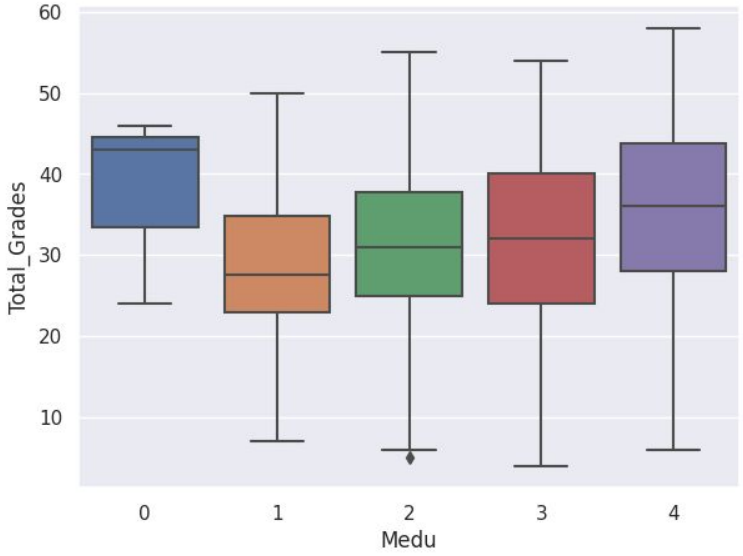
EDA: Desire for higher education

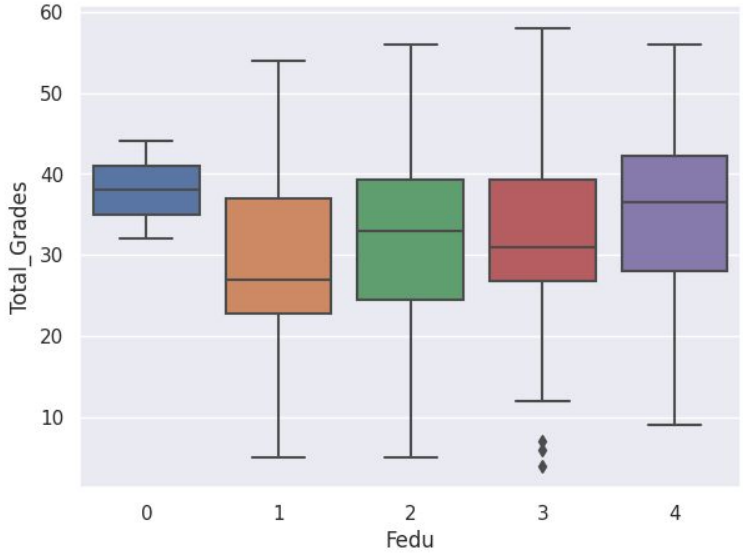


EDA: Parents details

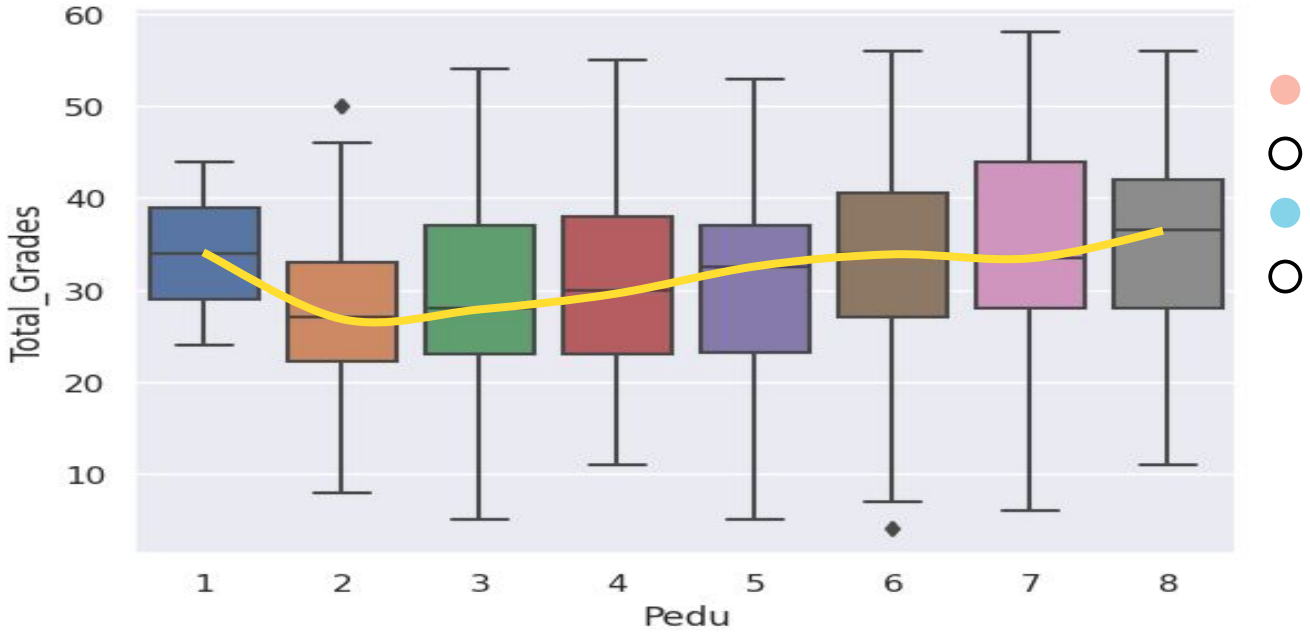


EDA: Parents education

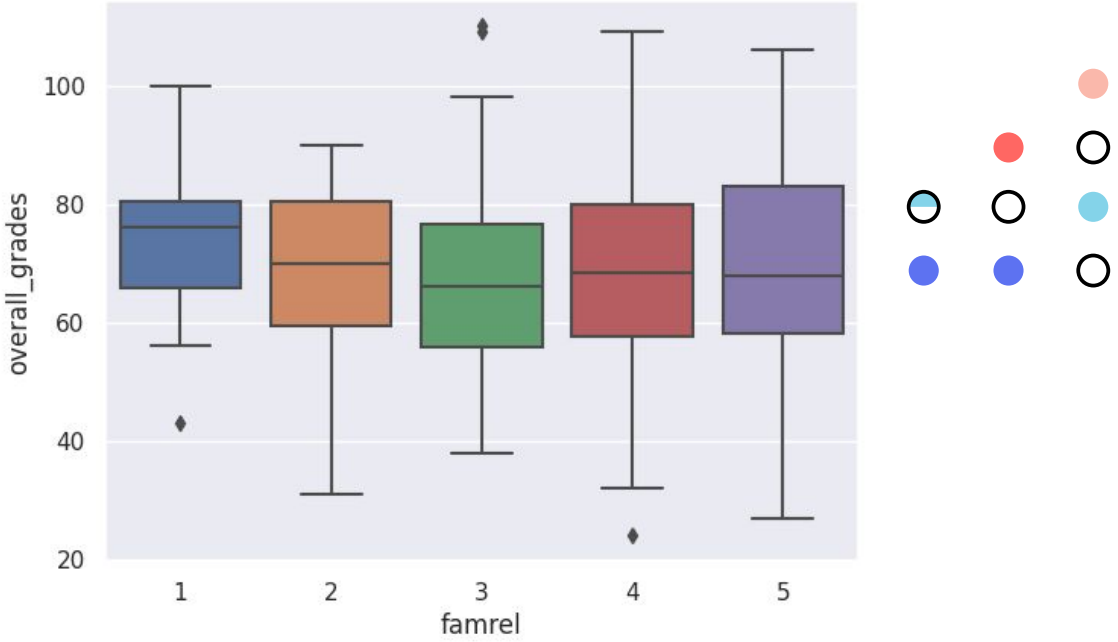




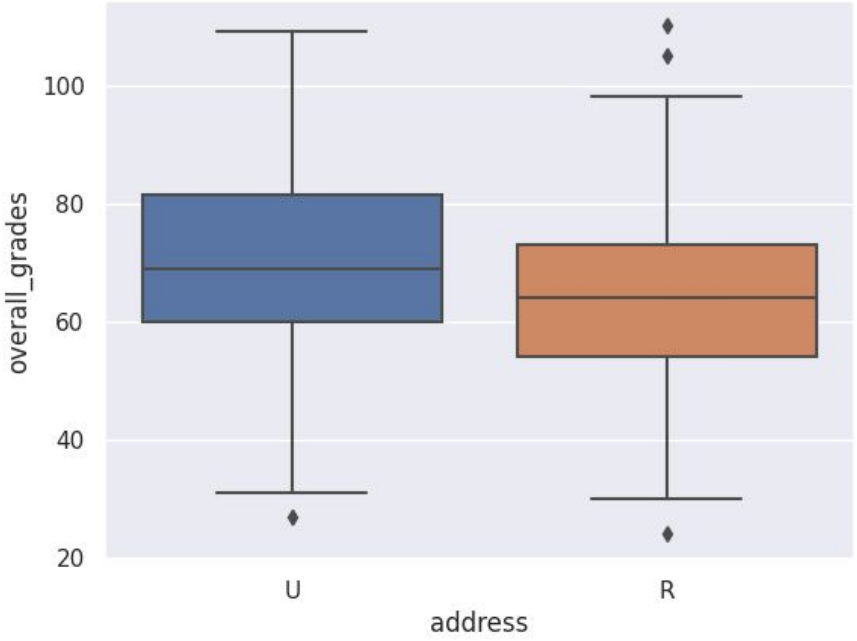
EDA: Parents education



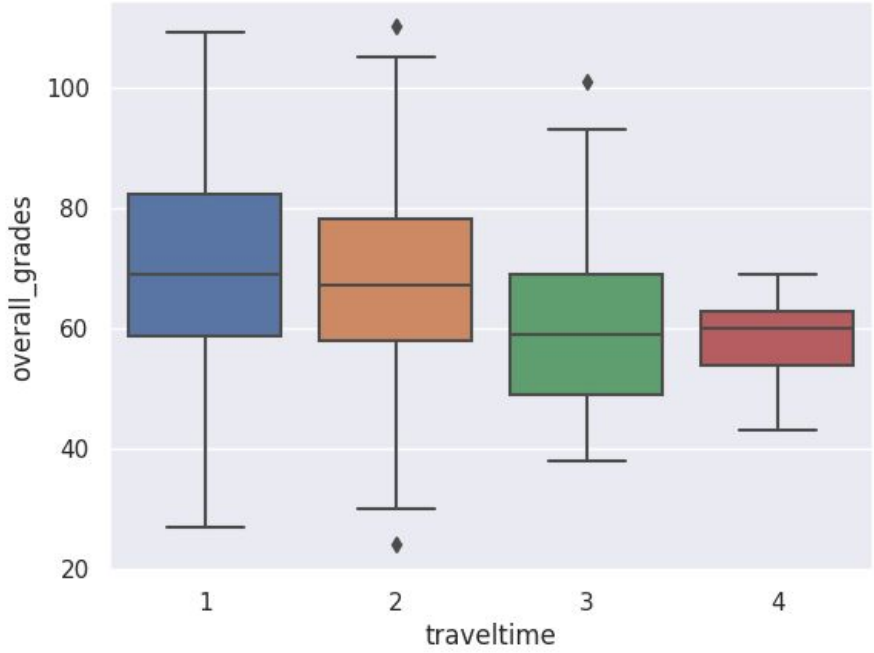
EDA: Family Relationships



EDA: Address



EDA: Travel time



Machine Learning

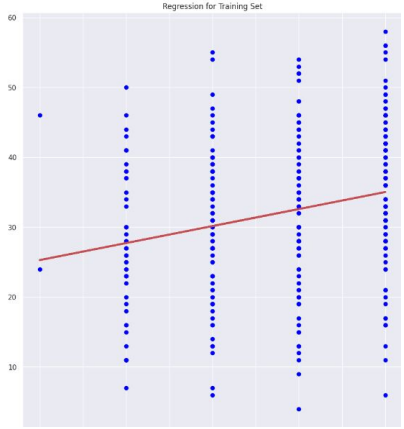
1 Uni-Variate

Linear Regression Model

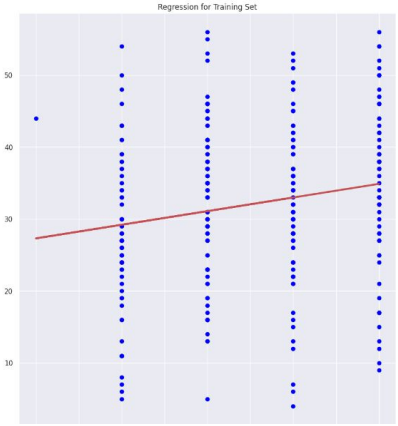
2 Multi-Variate

Linear Regression Model

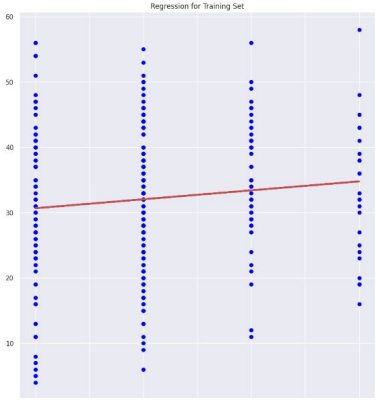
Uni-Variate Regression Model



Medu

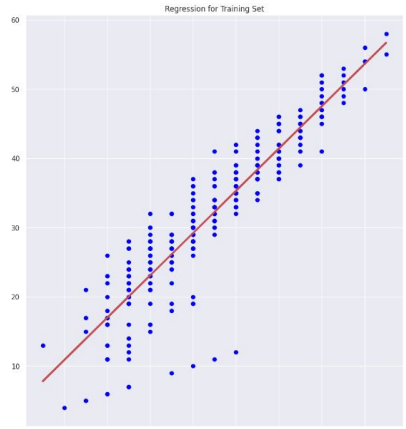


Fedu

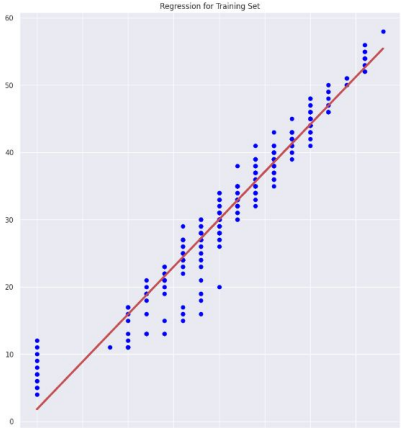


studytime

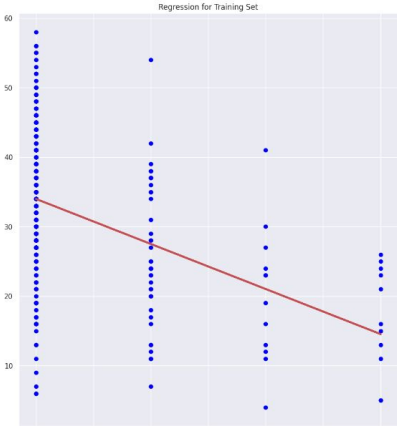
Uni-Variate Regression Model



G1



G2



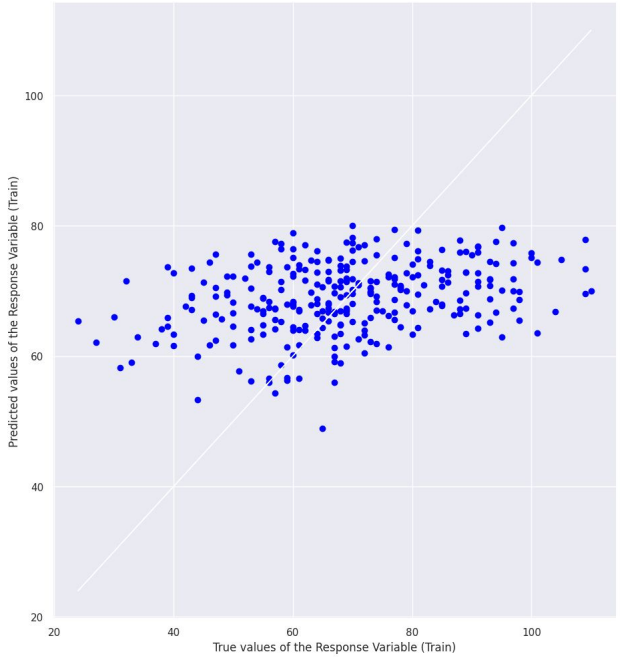
failures

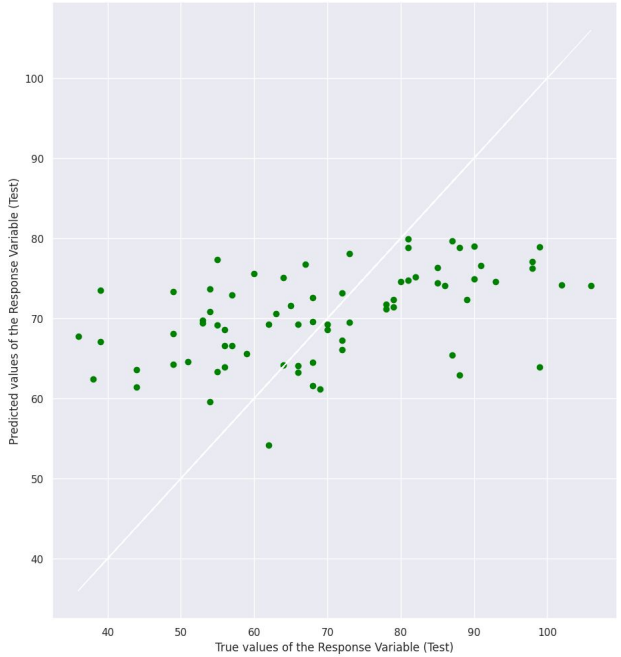
Multivariate Regression Model

	Predictors	Coefficients
0	traveltime	-1.947805
1	activities	0.889190
2	romantic	-2.967059
3	famrel	-0.297625
4	goout	-1.360746
5	total_absences	-0.157849
6	weekly	-0.868633
7	Pedu	1.778763

```
Intercept of Regression
      : b =  [71.36950508]
```

Multivariate Regression Model





Multivariate Regression Model

```
Goodness of Fit of Model          Train Dataset
Explained Variance (R^2)          : 0.12323366964013271
Mean Squared Error (MSE)          : 251.61932840302885
Root Mean Squared Error (RMSE)    : 15.862513306630474

Goodness of Fit of Model          Test Dataset
Explained Variance (R^2)          : 0.15049097506415288
Mean Squared Error (MSE)          : 218.69132979566993
Root Mean Squared Error (RMSE)    : 14.788215909827322
```

New and Beyond!

1 Backward Elimination

OLS (Ordinary Least Square) Regression

2 Feature Importance

Random Forest

OLS Regression

OLS Regression Results

Dep. Variable:	overall_grades	**R-squared:**	0.131
Model:	OLS	**Adj. R-squared:**	0.111
Method:	Least Squares	**F-statistic:**	6.753
Date:	Sat, 22 Apr 2023	**Prob (F-statistic):**	2.94e-08
Time:	06:50:30	**Log-Likelihood:**	-1533.9
No. Observations:	368	**AIC:**	3086.
Df Residuals:	359	**BIC:**	3121.
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	70.8660	5.252	13.493	0.000	60.537	81.195
traveltime	-2.6115	1.223	-2.135	0.033	-5.017	-0.206
activities	1.8803	1.673	1.124	0.262	-1.409	5.170
romantic	-3.3179	1.782	-1.862	0.063	-6.822	0.187
famrel	-0.1153	0.926	-0.125	0.901	-1.936	1.705
goout	-1.8425	0.801	-2.301	0.022	-3.417	-0.268
total_absences	-0.1476	0.077	-1.907	0.057	-0.300	0.005
weekly	-0.6156	0.474	-1.299	0.195	-1.548	0.317
Pedu	1.9909	0.435	4.579	0.000	1.136	2.846

New Model

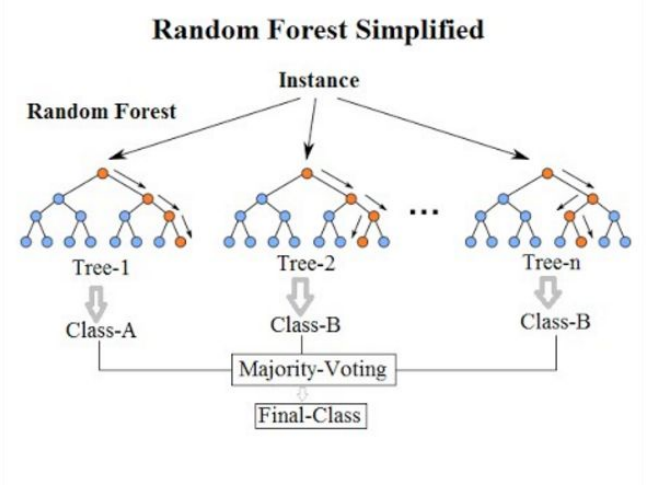
OLS Regression Results

Dep. Variable:	overall_grades	**R-squared:**	0.112
Model:	OLS	**Adj. R-squared:**	0.100
Method:	Least Squares	**F-statistic:**	9.113
Date:	Sat, 22 Apr 2023	**Prob (F-statistic):**	3.54e-08
Time:	06:54:19	**Log-Likelihood:**	-1537.9
No. Observations:	368	**AIC:**	3088.
Df Residuals:	362	**BIC:**	3111.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	70.0269	5.260	13.314	0.000	59.684	80.370
traveltime	-2.8647	1.212	-2.363	0.019	-5.249	-0.480
romantic	-3.5445	1.785	-1.986	0.048	-7.054	-0.035
famrel	0.2678	0.917	0.292	0.770	-1.535	2.071
goout	-2.3768	0.741	-3.208	0.001	-3.834	-0.920
Pedu	1.9896	0.434	4.589	0.000	1.137	2.842

Random forest

Is like a classification tree but done multiple times to get different trees



Feature Importance

	G2	absences	G1	age
Feature Importance	0.840337	0.076088	0.012519	0.009485

goout	studytime	health	failures	Walc
0.008711	0.008008	0.007836	0.006891	0.004808

Intelligent Decision

How can we optimise the environment to improve Students' Grades?

1

How can we sieve out students who are in need of help?

2

What are the factors that are important in determining our grades?

Sub-problem 1

How can we sieve out students who are in need of help?

Sub-problem 2

What are the factors that are important in determining our grades?

Factors

traveltime

Time taken to travel to school

romantic

Whether the student is in a romantic relationship

Pedu

Parents education

famrel

Family Relations

Disclaimer

Since correlation != causation,

We cannot be 100% sure if these solution can improve grades

But there is definitely a good chance that it can impact grades based on our statistic analysis

traveltime

traveltime -2.8647

Provide dorms or integrate home-based learning into curriculum

Romantic

romantic -3.5445

Have lessons on healthy relationships

Pedu

Pedu 1.9896

Have online materials for parents so they can help their child with schoolwork.

famrel

famrel 0.2678

Provide family counseling services

Conclusion

With our solutions to our subproblems

- We will be able to identify students that need help
- We are able to address the social factors that impact grades

Conclusion

Optimisation of the environment to improve students' grades can be achieved!