TASK-1(REPORT)

Level 1 : Performing EDA to guess unkown features

Level 2 : processing all missing values with appropriate strategies

Level 3 : ASK ATLEAST 5 INSIGHTFUL QUESTIONS ON DATASET

LEVEL 4: APPLYING CLASSIFICATION TECHNIQUES AND ASSESS THEIR PERFORMANCE

LEVEL 5: VISUALIZING MODEL DECISIONS USING SHAP

LEVEL-1

Experiments :

First, evaluated correlation matrix and noted various important positive or negative columns for each features. For each important columns plotted various plots with corresponding feature with which correlation was significant.

IWhile plotting, tried multiple combinations like bar plot of feature_1 with g3 was not interpretable plot so went on with scatter. Like that did all other plots as well. Performed a crucial hist plots of each unkown features count which gave idea of feature_1 being continous, feature_2 and feature_3 being scale (of upto 5).

Feature_1 has positive correlation with failures.Feature_1 shows discrete values (>15), it can be any numerical feature.Higher feature_1 has more likely to have internet access.

Better family relationship has lesser scale of Feature_2.Higher Feature_2 tends to have higher grades.Higher feature_2 tends to have less number of absences.As feature_2 increases upto certain limit grades increases after that it decreases.

For a particular scale of feature_3 ratio of number of people in romantic upon non romantic decreases hinting people with higher feature_3 tends to have nearly same population of those who are in relationship or not.Students with higher Feature_3 are more prone to goout.

Final conclusion : Feature-1 : Weekly screen-time

Feature-2 : Stress levels

Feature-3 : Social activity level

LEVEL-2

Experiments : For famsize: All those students whose parents are apart higher chances of them having family size less than 3 so LE3 and GT3 for parents together

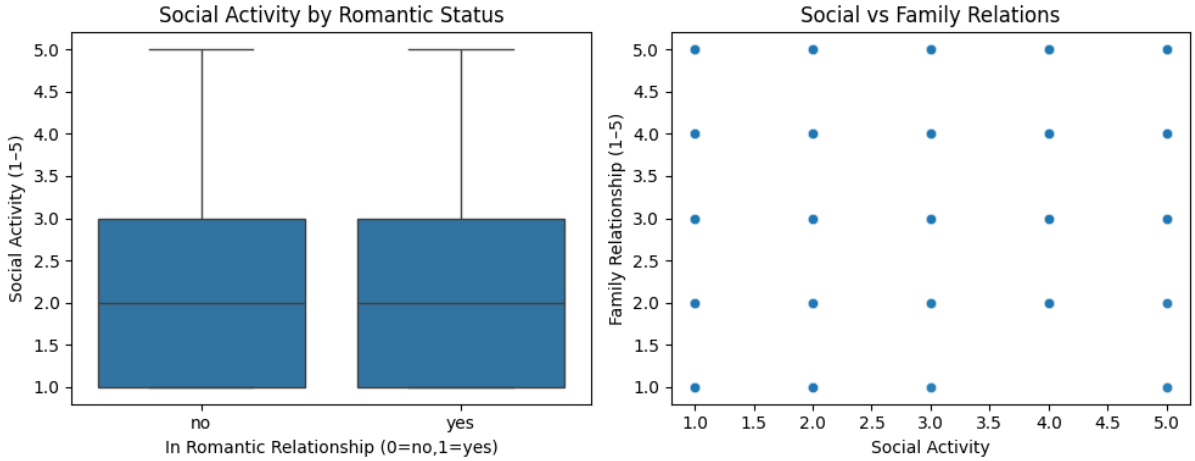
For traveltime, Fedu,higher , freetime,Feature_1 and Feature_2 : Did strategy=”most_frequent” as these were scales , no common pattern.

For Feature_1 : Did strategy=”median” as it was continous data.

LEVEL-3

Q.1 Does social activity differ by romantic status and family relationships?

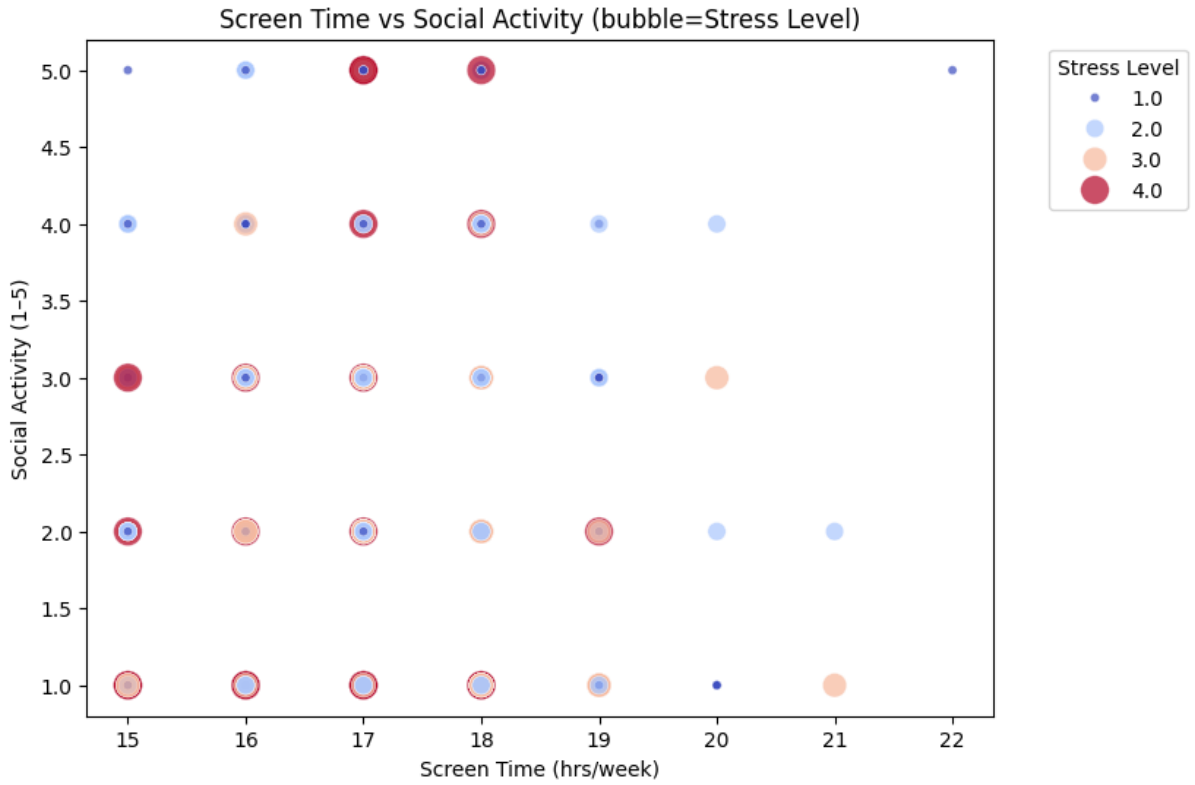
Ans- : Pretty balanced , Social activity isn't changing after being in Romantic relationship.Dataset might be considering activities of couples as social activity (As it is exactly same).Family relationship is also not a hindrance for social activity.



Q.2 What's the interconnection between screen time, social activity, and stress?

Ans : Generally, Students with higher level of social activity tends to have lower stress level.

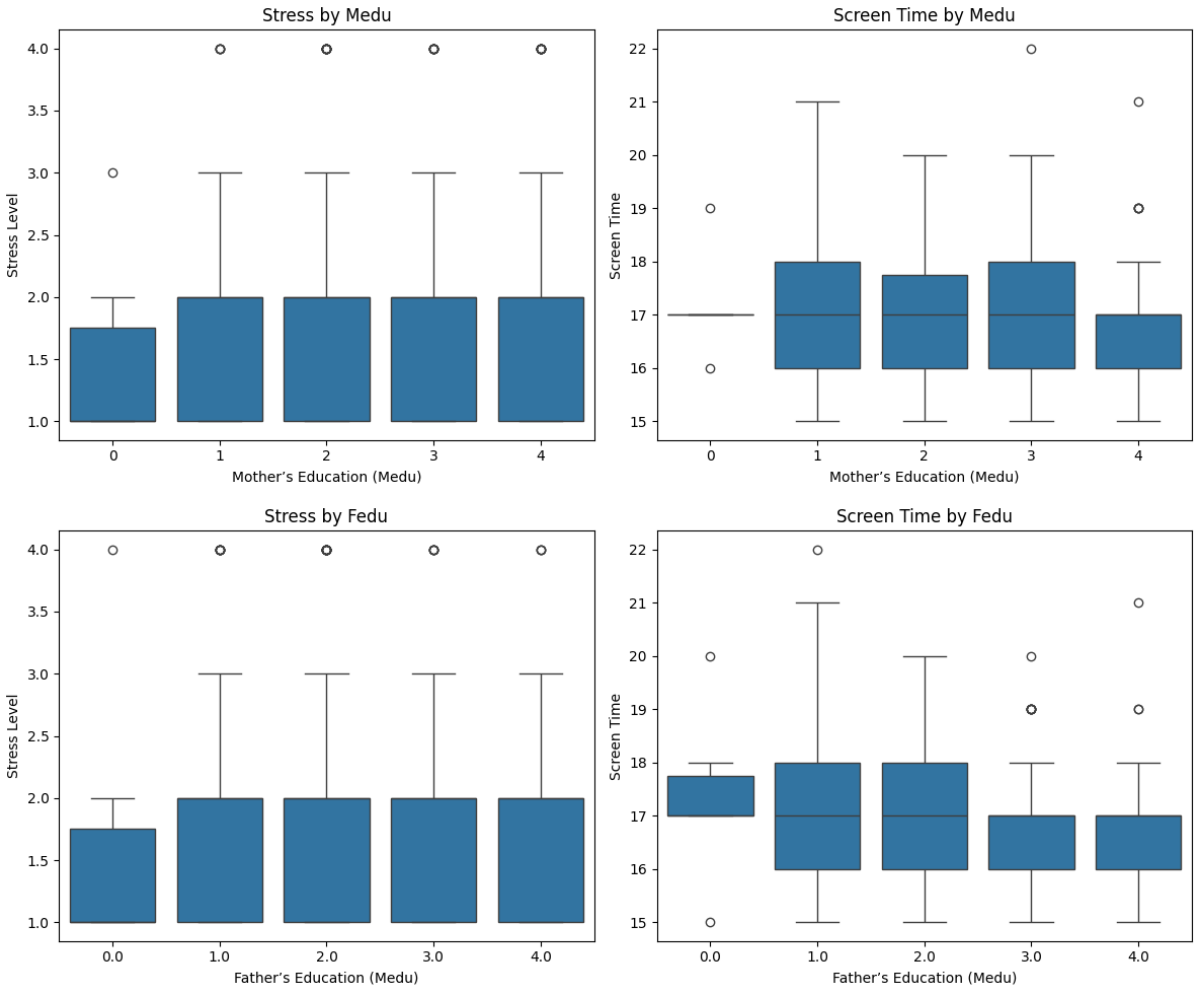
Students with higher screen time tends to have lesser scale of social activity. No such guarantee of stress levels assured by screen time.



Q.3 Do students with low parental education report higher stress and screen time?

Ans : No such trends of stress level found from it.Infact, they all showed similar stress levels from(1-4). However, with very high education level of father or mother or both screen time is the least(generally). While, greater for lowest education level of parents(either or both).

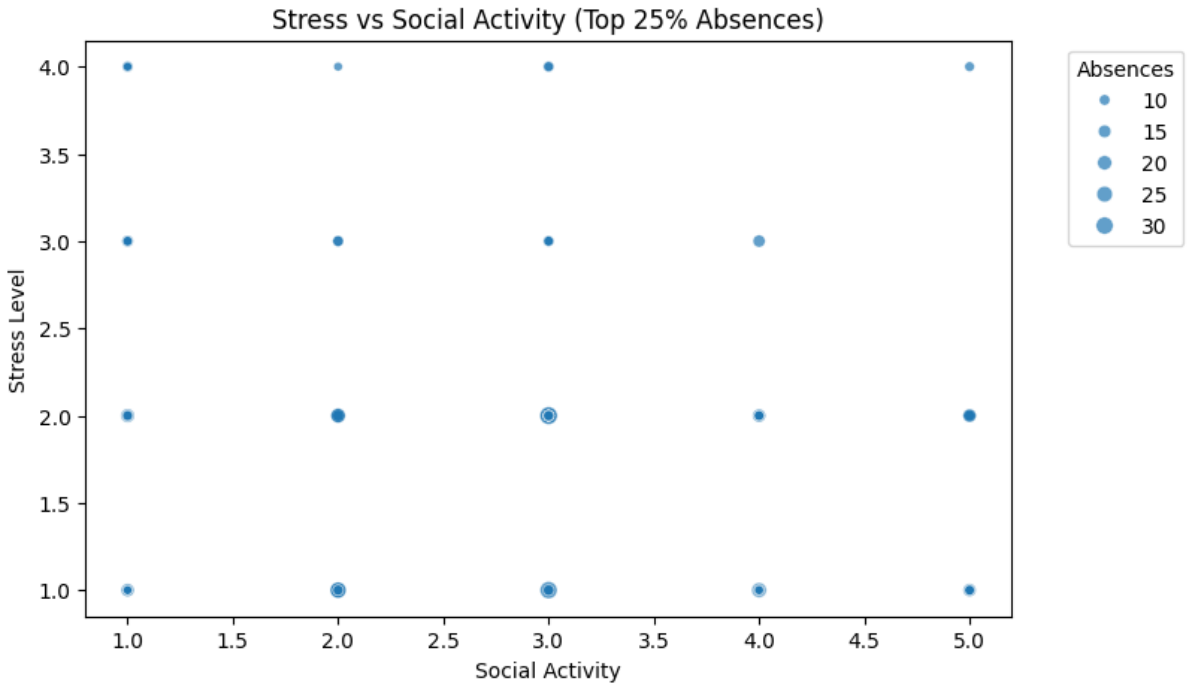
Why it matters: If lower-educated parents' kids are more stressed or glued to screens, targeted academic or counseling support programs could be prioritized.



Q4 Among students with high absence rates, how does social activity vs stress trade off?

Ans: High absences have lesser(or moderate) social activity as well.

Why it matters : To know where are they investing their time and their mental status.



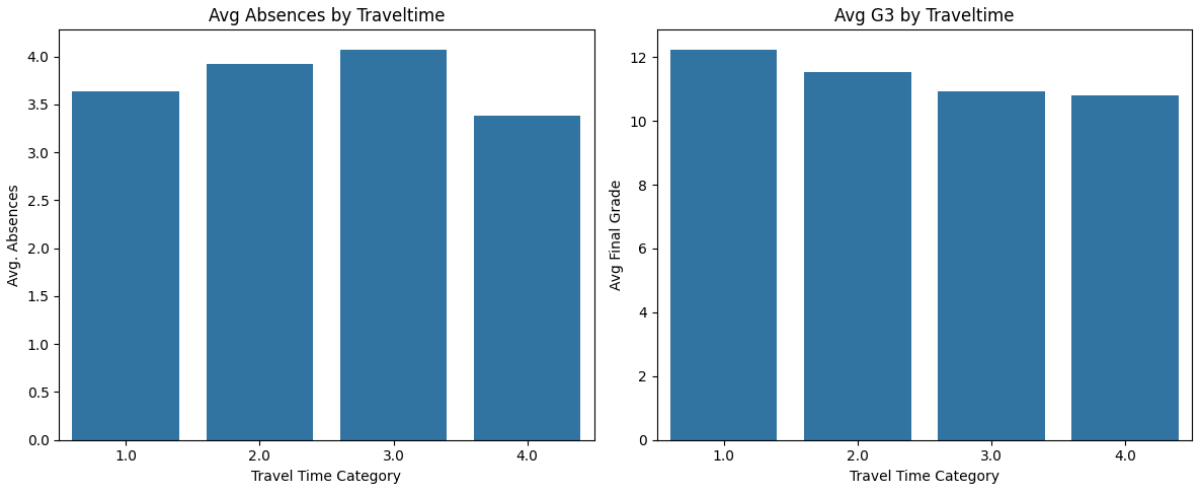
Q5 Does travel time to school (traveltime) affect absences and final grades?

Ans : It turns out that students with lesser travel time are scoring better than those with higher time travel.

On an average, students who fall in higher time travel category are not more absences suggesting time travel category is not the best factor for reason of their absences.

Why it matters: Long commutes can erode attendance and performance,campus transport initiatives could

help.

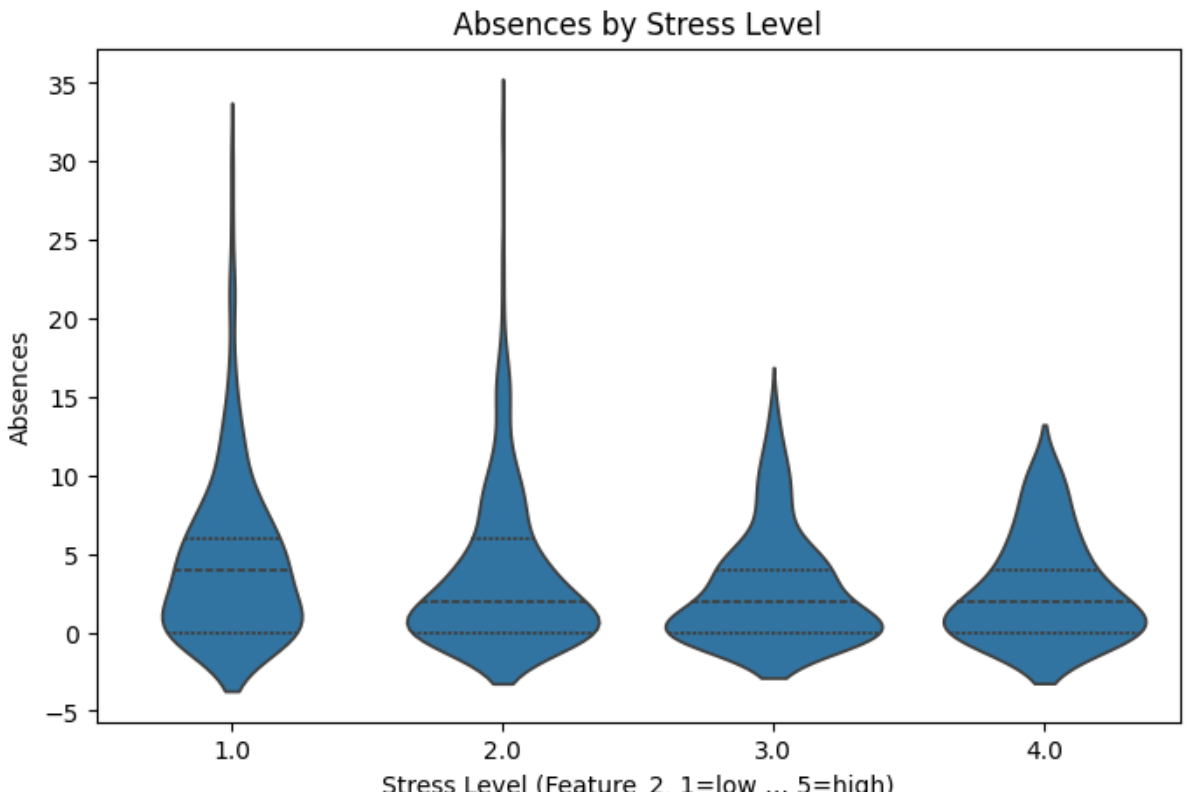


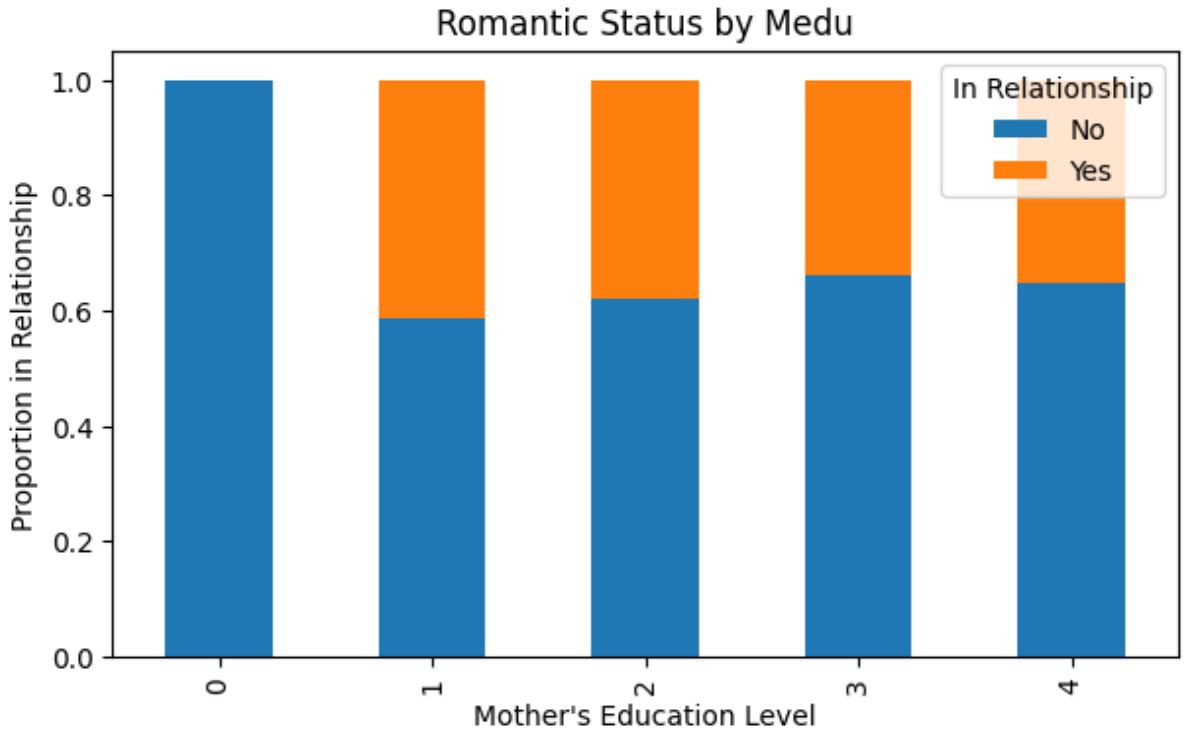
Q6 How do stress levels relate to absences of student?

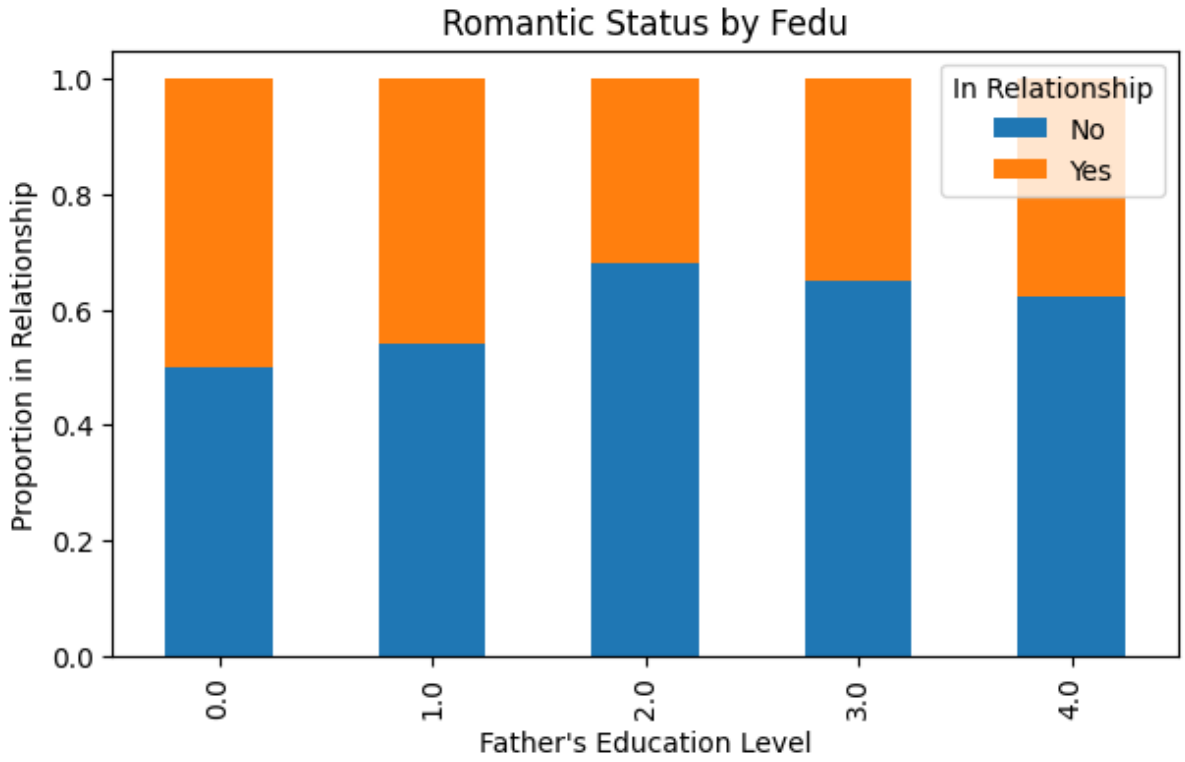
Ans : Higher stress generally corresponds to more absences—the median moves upward from stress 1 to stress 4.the widest spread occurs at moderate stress (level 2). That suggests a subgroup of moderately stressed students who are driving up absenteeism, while the very highest-stress group is more uniformly absent at moderate levels.

Q7 Does kids of parents with higher education more prone to relationship(romantic)?

Ans: No that's not exactly the case however, students with Medu at 0 level are not at all in romantic relationship.Students with lower Fedu are at higher proportion in romantic relationship.


Absences by Stress Level
Absences
35
30
25
20
15
10
5
0
−5
1.0
2.0
3.0
4.0
Stress Level (Feature 2, 1=low ... 5=high)


Romantic Status by Medu
Proportion in Relationship
1.0
0.8
0.6
0.4
0.2
0.0
In Relationship
No
Yes
0
1
2
3
4
Mother's Education Level



LEVEL-4

Logistic regression :

Class	Precision	Recall	F1-score	Support
0	0.71	0.63	0.67	123
1	0.47	0.56	0.51	72
Accuracy			**0.61**	195
Macro avg	0.59	0.59	0.59	195
Weighted avg	0.62	0.61	0.61	195

Confusion matrix:

	Predicted: 0	**Predicted: 1**
Actual: 0	78	45
Actual: 1	32	40

Random Forest :

Class	**Precision**	**Recall**	**F1-score**	**Support**
0	0.65	0.90	0.75	123

Class	Precision	Recall	F1-score	Support
1	0.48	0.15	0.23	72
Accuracy			**0.63**	195
Macro avg	0.56	0.53	0.49	195
Weighted avg	0.58	0.63	0.56	195

SVM :

	Predicted: 0	Predicted: 1
Actual: 0	78	45
Actual: 1	32	40

Random Feature 10 best features:

Feature Name	Importance Score
num_median__G3	0.0674
num_median__G2	0.0638
num_median__G1	0.0634
num_median__absences	0.0546
num_median__Feature_1	0.0536
num_mode__goout	0.0510
num_mode__freetime	0.0454
num_mode__famrel	0.0445
num_mode__health	0.0432
num_mode__Fedu	0.0381

Model prediction for one student :

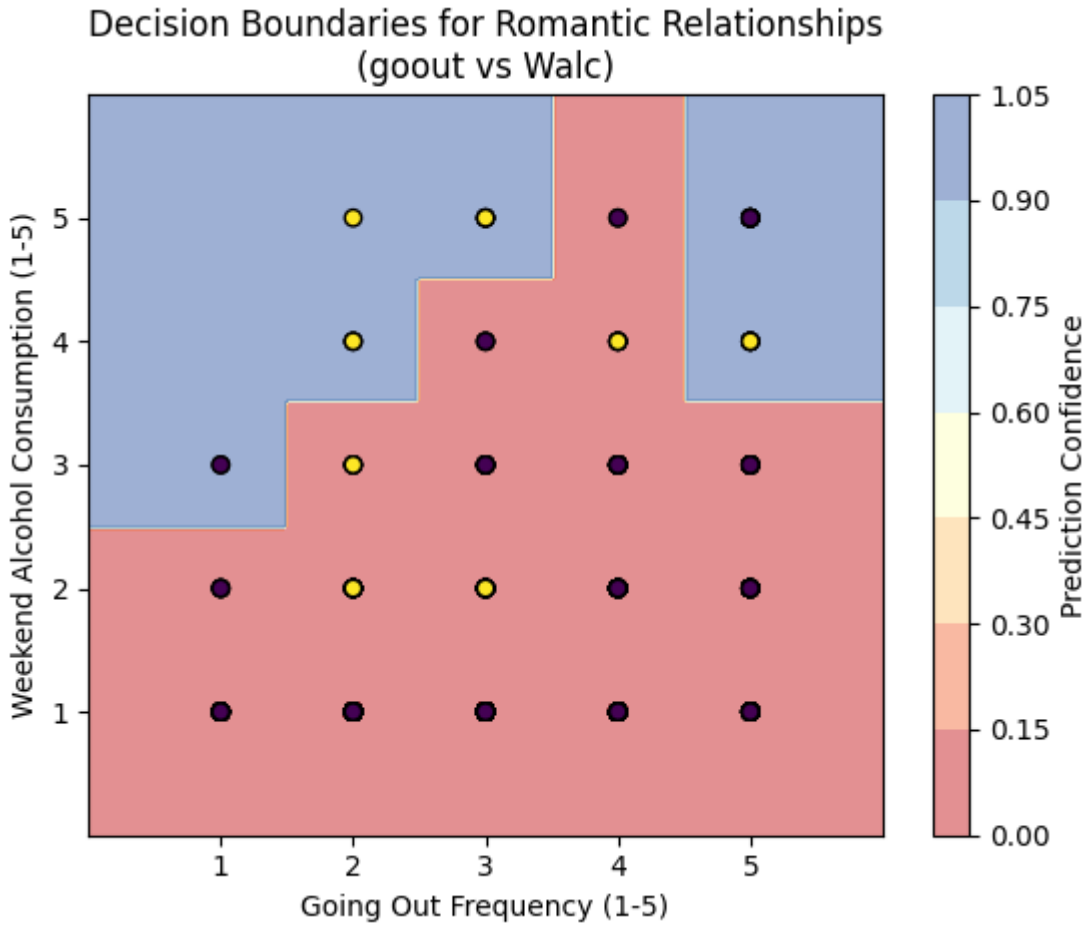
Model	Predicted Status	Probability
Logistic Regression	Single	22.79%
Random Forest	Single	21.00%
SVM	Single	35.97%

LEVEL-5

Individuals with **high alcohol consumption on weekends and moderate going-out frequency** (upper-middle of the graph) are **more likely** to be in a romantic relationship (blue region).

On the other hand, those with **low weekend drinking and varying social activity** are more often predicted **not to be** in a romantic relationship.

The model shows a **non-linear decision boundary**, suggesting some complex interaction between these two features.



Global feature importance :

Most interactions have values close to **zero**, meaning the **model does not rely heavily on interactions** between features — it makes mostly additive decisions.

There are few points far from zero, indicating **some feature pairs may interact significantly**, possibly affecting the romantic prediction in either direction.