

Task 1: The CampusPulse Initiative

Software Versions :

Python 3.13.2

Numpy 2.3.2

Libraries imported :

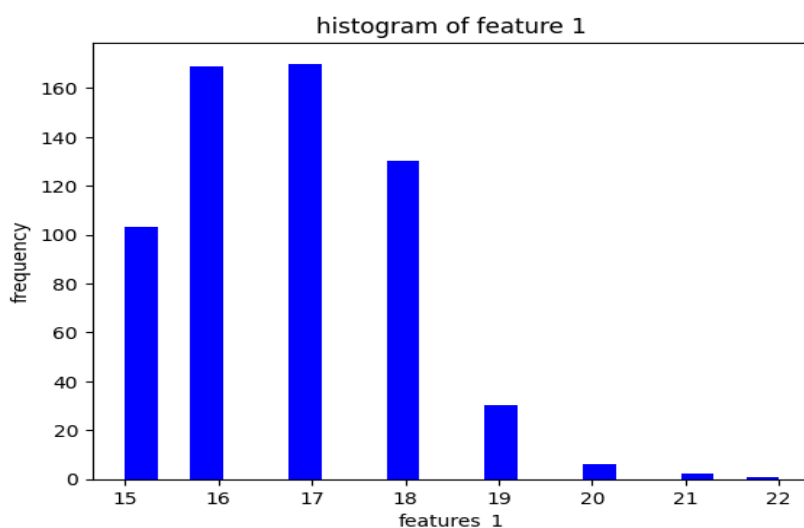
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
```

Level 1: Variable Identification Protocol

EDA techniques :

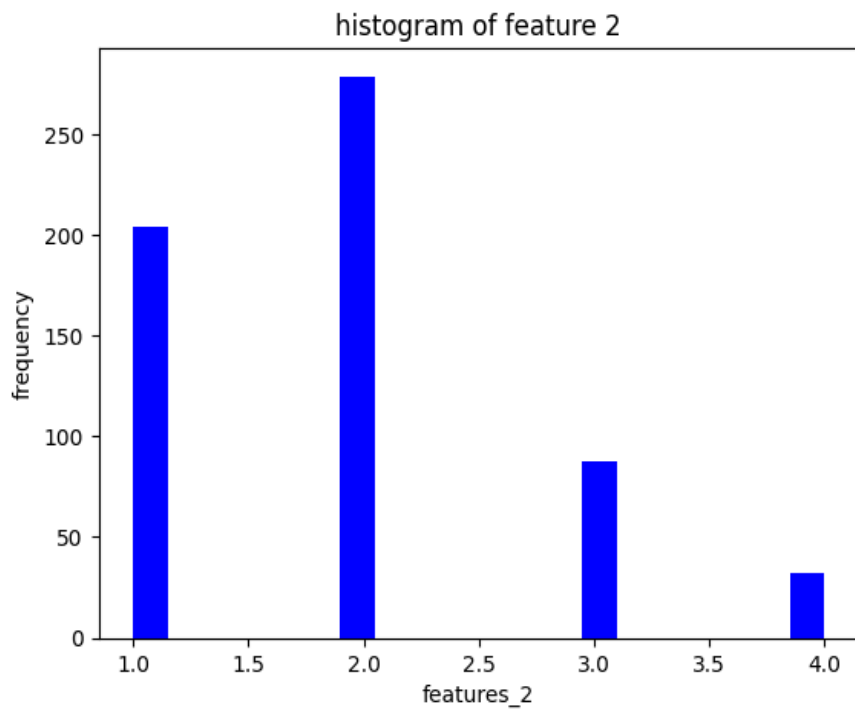
First plotted histogram for each Feature_1 , Feature_2 and Feature_3.

For Feature_1 :



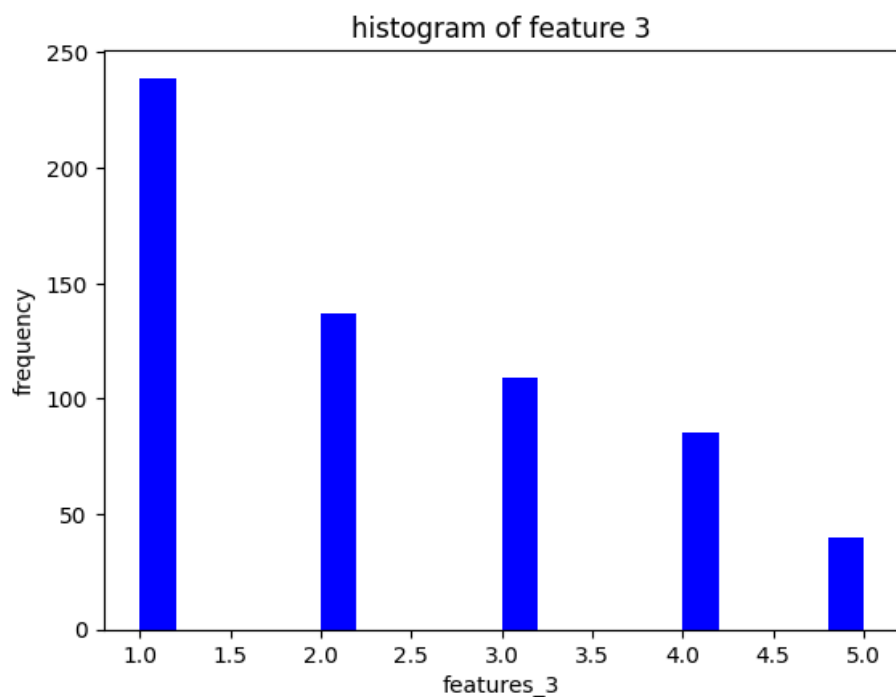
1. Range between 15 to 22
2. Continuous type of Data .
3. Have a mean of 16.75
4. Most likely to represent age.

For Feature_2 :



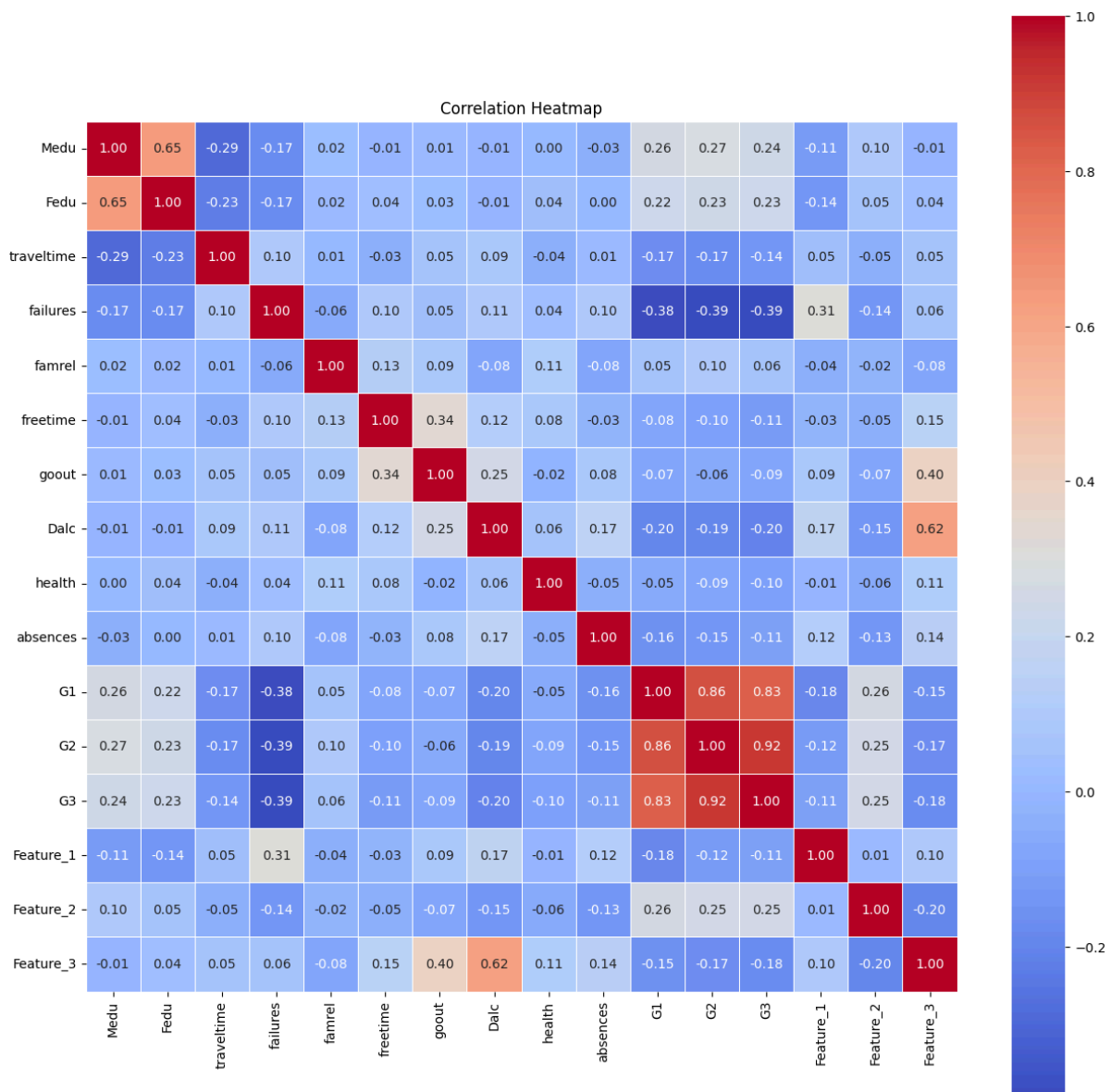
1. Range between 1 to 4.
2. Have a discrete type of
3. May be a scale of something having max value of 4 and Min value 1.
4. Median of this feature is 2.

For Feature_3 :



1. Range between 1 to 5.
2. It also have discrete type of data.
3. May be represent level of something.

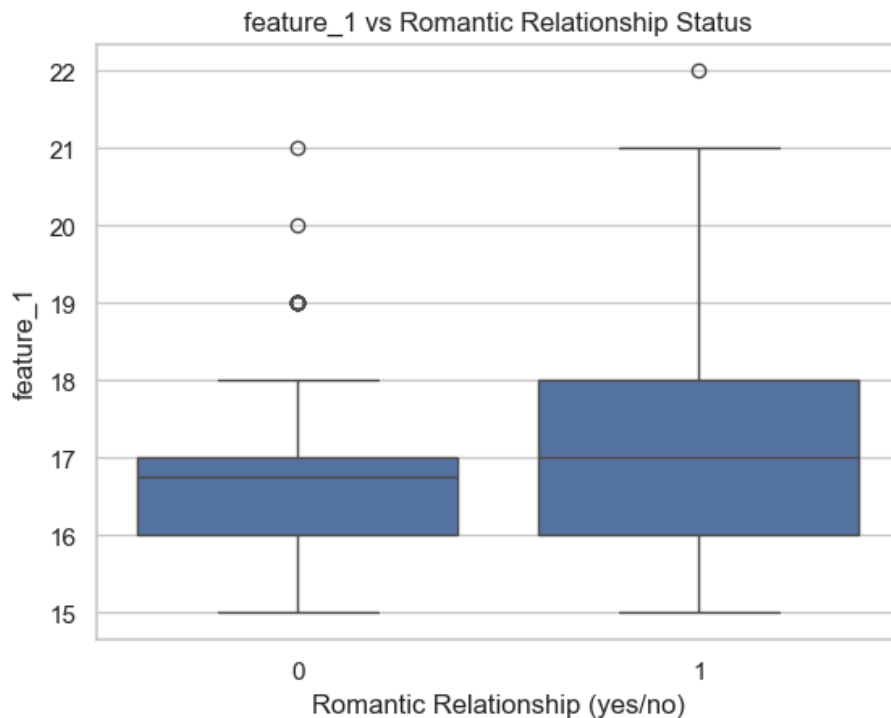
Correlation Heatmap :



1. As we can see, Feature_1 moderately correlates with grades (G1 , G2 , G3). suggesting academic relation of this feature.
2. Feature_2 has no correlation with grades (G1 , G2 , G3). means this feature is not directly related to academics .
3. Feature_2 has mild inverse correlation with "absences", and a small positive bump with "goout" . means Feature_2 is related to behavior or something psychological .
4. Feature_3 has weak positive correlation with "gout" , "health" and "Dalc" .

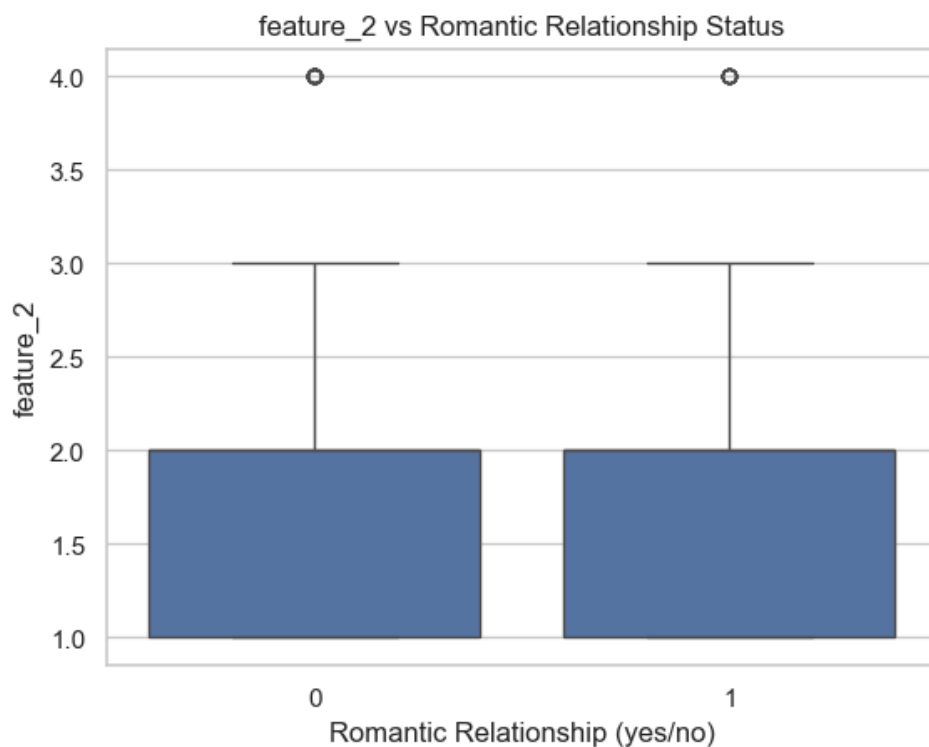
Box plot :

For Feature_1 :



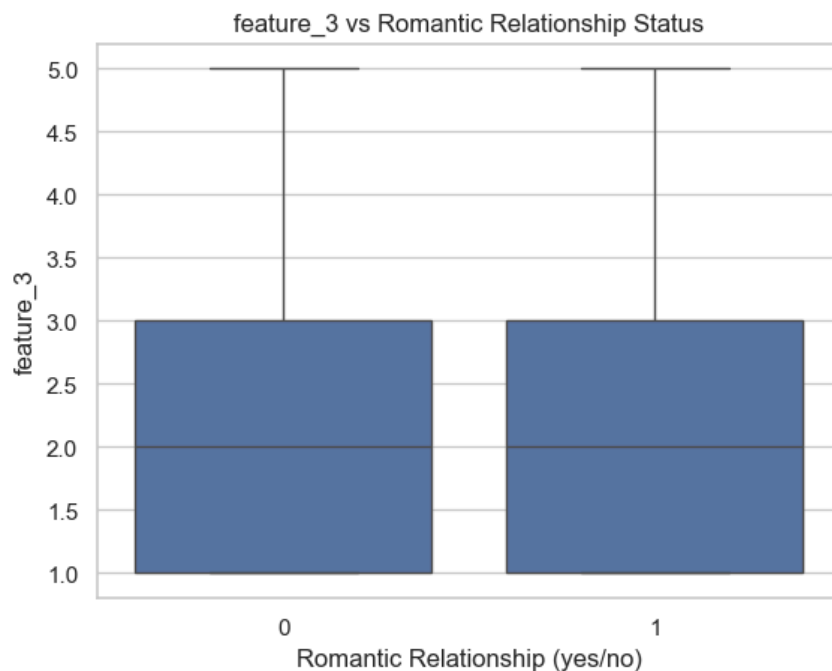
1. The interquartile range is slightly wider for students in a relationship .
2. Students not in a relationship ("no") show few outliers above 19-21
3. Those in a relationship have a high outlier close to 22.

For Feature_2 :



1. The spread is similar for Both groups.
2. This suggests that Feature is not a Strong differentiator between students who are in a romantic relationship

For Feature_3 :



- 1.The median of feature_3 is similar for both groups.
- 2.The minimum and maximum values are also same.
3. The distribution is similar and almost identical for Both groups.

Conclusion :

1. Feature_1 may represent age because the age factor is missing from the dataset and the 15 to 22 range is highly indicating that it is age.
2. Feature_2 may represent screen time hours because of a positive correlation with absences .
3. Feature_3 may represent a romantic satisfaction score because students who are in a romantic relationship score significantly higher in this feature.

Level 2: Data Integrity Audit

First calculating the missing values from the dataset.
Here's the list of missing values.

famsize	50
Fedu	73
traveltime	73
higher	76
freetime	45
absences	69
G2	35
Feature_1	38
Feature_2	46
Feature_3	39

Dividing list into according to their data types :

Continuous :	Discrete :	Categorical :
` Feature_1`	`Fedu`	`famsize`
` G2`	`traveltime`	`higher`
	`freetime`	
	`Feature_2`	
	`Feature_3`	
	`absences`	

Filling method :

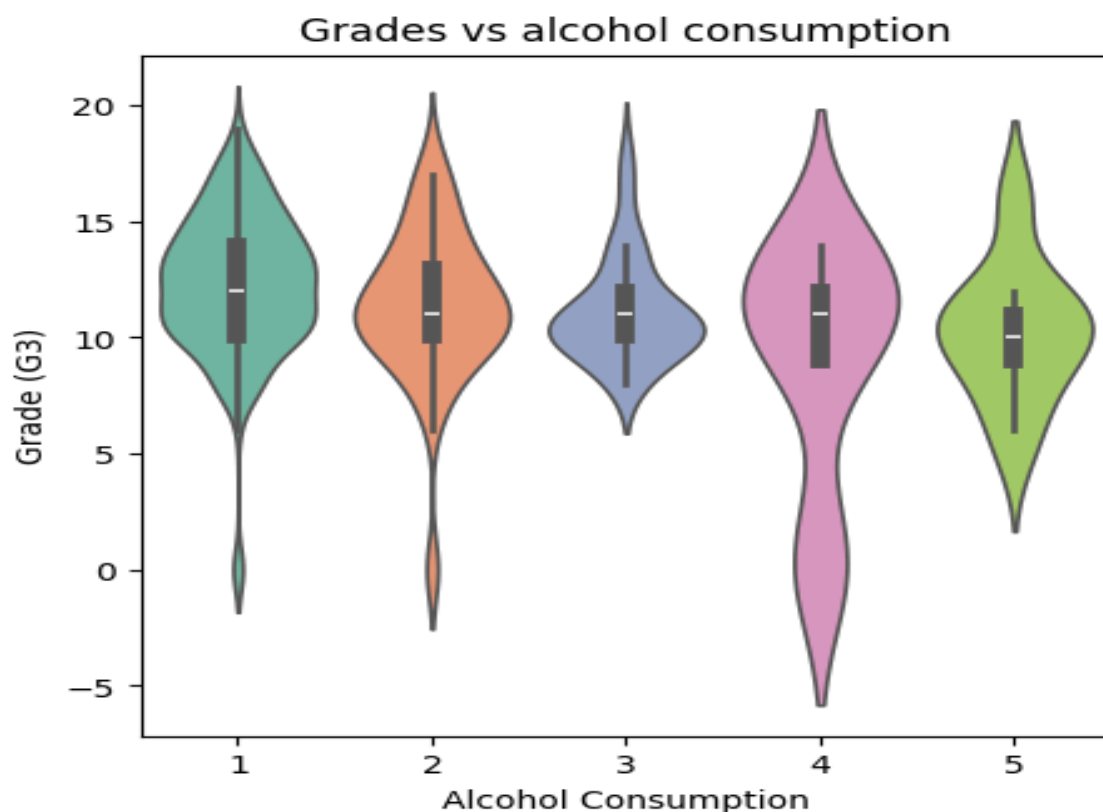
1. For continuous data features like `Feature_1` (age) and `G3` filling with mean will be best .

2. 'Fedu' , 'traveltime', 'absences', 'freetime', 'Feature_2', 'Feature_3' all have discrete type data Numeric but have maximum and minimum values so filling with median will make it more stable.
 3. 'famsize' represents no. of family members having categorical type of data so filling it with Most common value (mode).
 4. 'higher' has discrete data values 'Yes' or 'No' so filling with mode will be best.
-

Level 3: Exploratory Insight Report

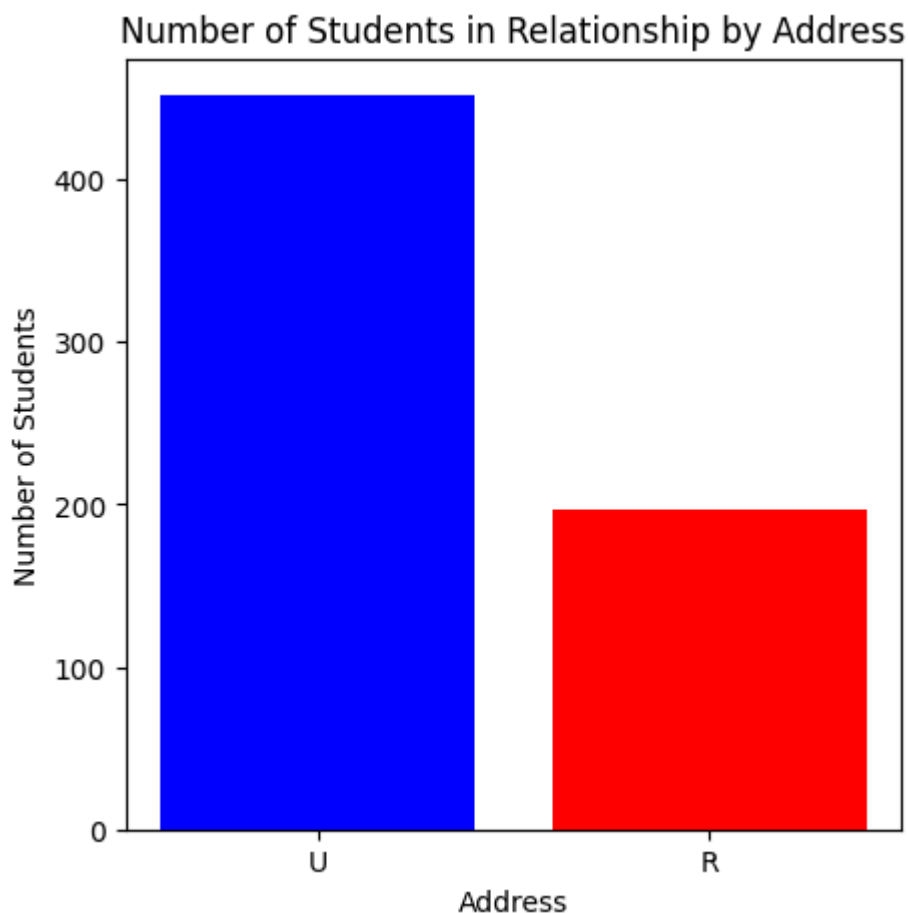
Insightful questions and their plots :

1.Does drinking alcohol affect how well students do in academics ?



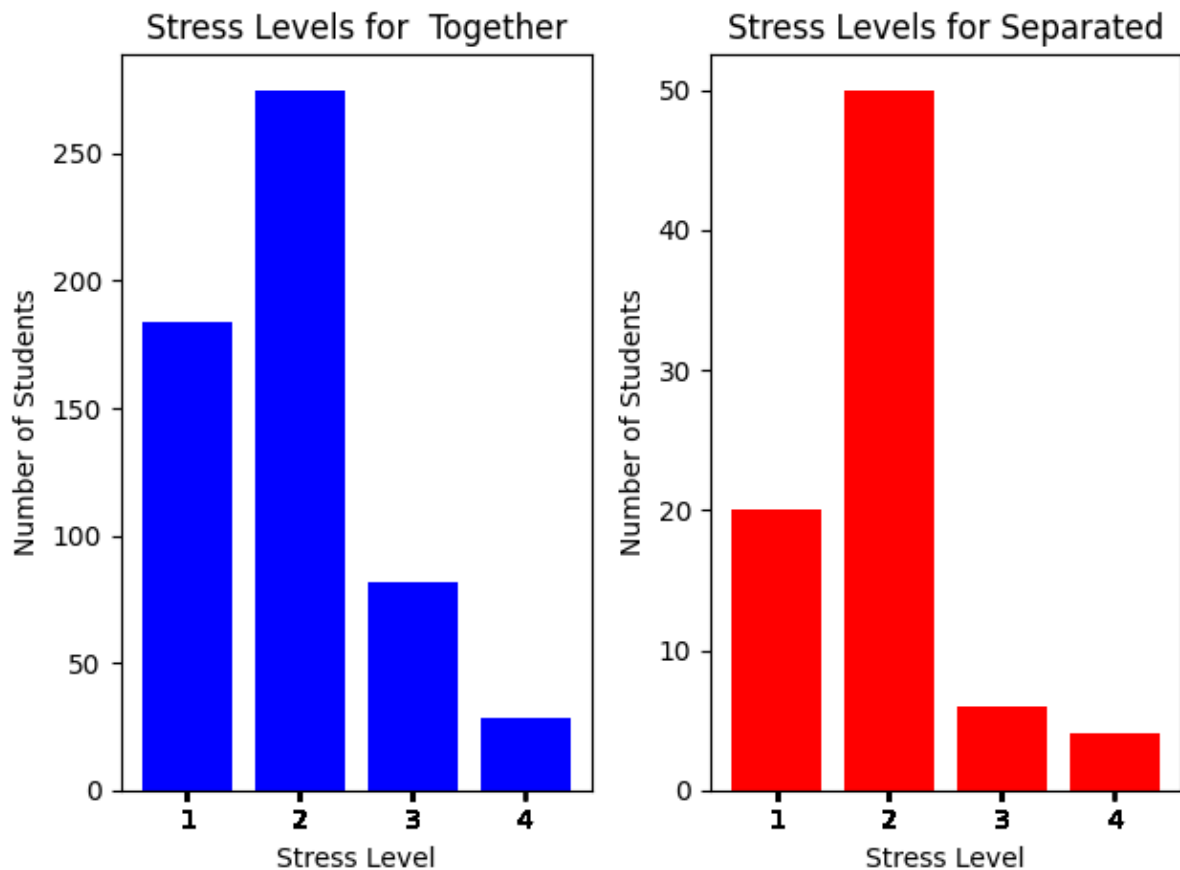
- a. Grades are decreasing with increased Alcohol Consumption.
 - b. Students with low alcohol consumption (1 and 2) have generally higher median grades, with most grades above 10.
 - c. As alcohol consumption increases (levels 3 to 5), the median grade drops.
 - d. At alcohol levels 4 and 5, some students perform very poorly.
 - e. This plot indicates a negative relationship between alcohol consumption and student grades.
-

2. Are students from urban areas more likely to be in romantic relationships than those from rural areas?



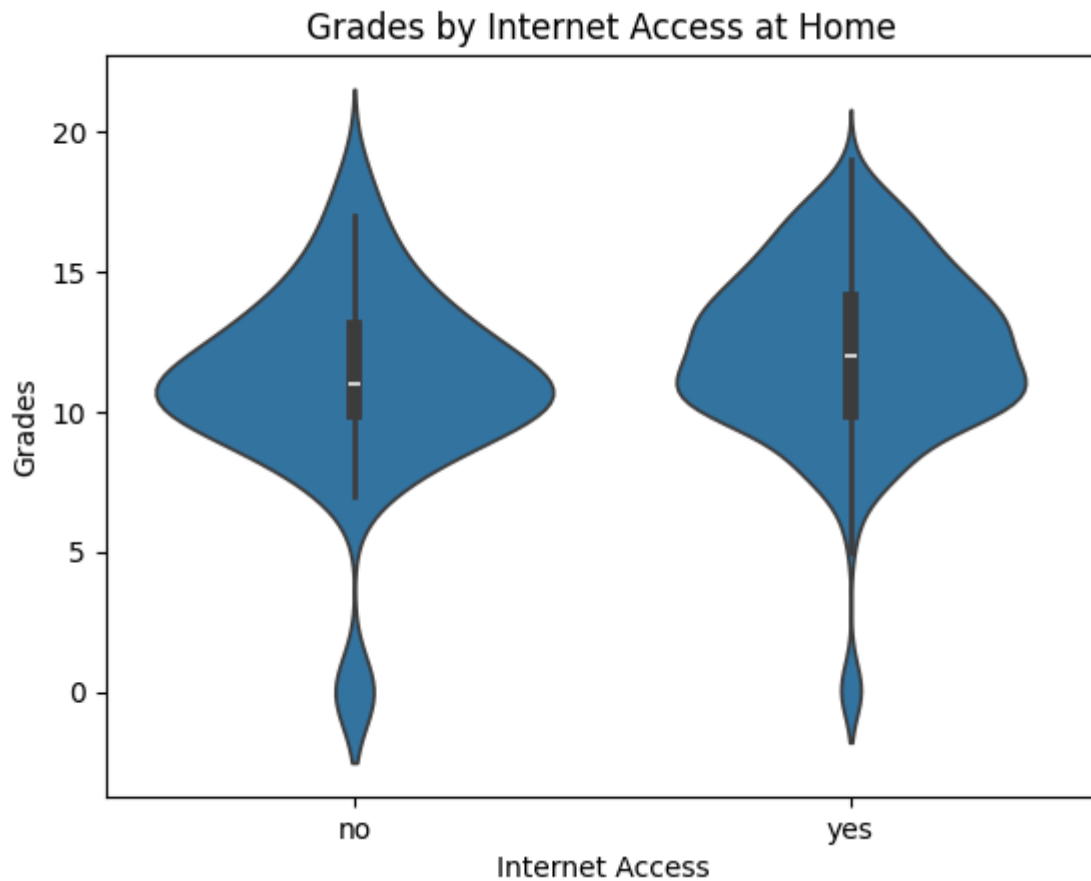
- a. Urban students (U) are more likely to be in a relationship than rural students (R).
 - b. Indicating address is the dominant feature for our model.
-

3. How does parental cohabitation status relate to stress levels among students?



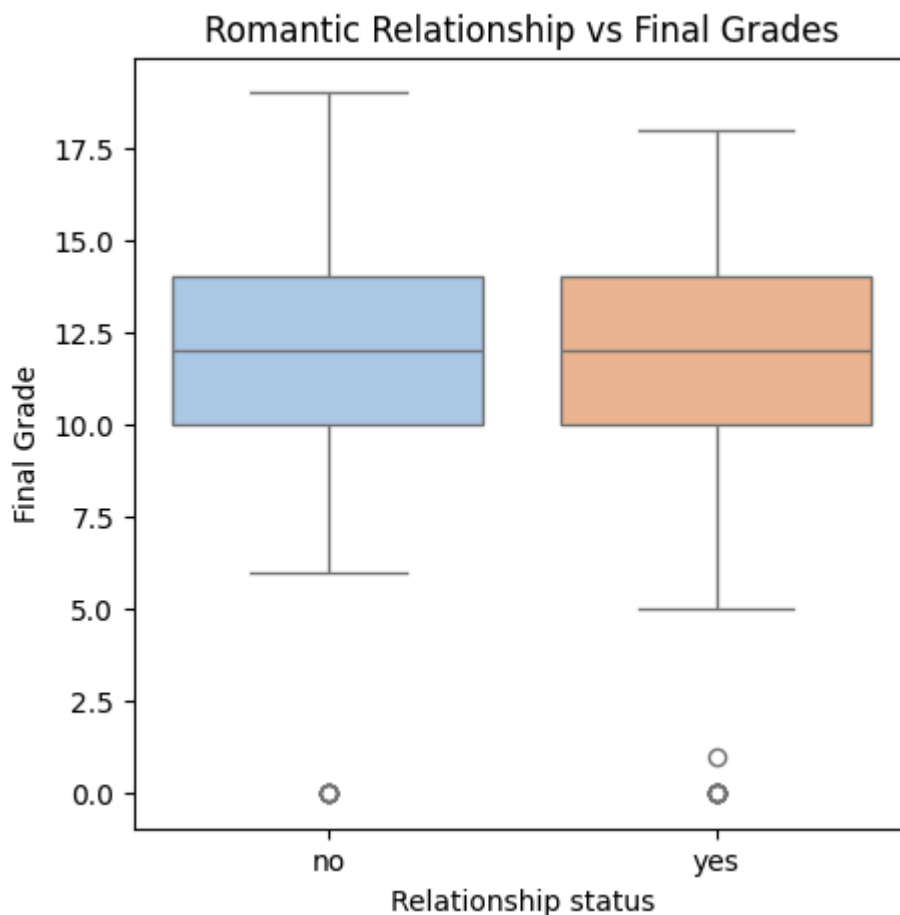
- a. In both the groups stress level 2 is almost common.
 - b. Students whose parents are together have lower stress levels, with most in levels 1 and 2.
 - c. Students whose parents are separated report higher stress levels (more level 3 and 4).
-

4. Does internet access at home affect academic performance?



- a. Both groups have similar median grades around 11–12.
 - b. Students with internet access have a slightly higher concentration of top grades .
 - c. Students without internet access have a more compressed distribution.
 - d. suggesting that home internet may provide an advantage in academic performance.
-

5. Does being in a romantic relationship impact a student's final grades?



- a. Both groups have very similar median grades around 12.
 - b. There are few outliers with very low grades in both groups.
 - c. There is no significant difference in final grades between students who are in a relationship and those who are not.
 - d. suggesting romantic relationship status has no impact on academic performance.
-

Level 4: Relationship Prediction Model

Applying classification Techniques :

1. Logistic regression model :

- Accuracy of Model -> **61.02564102564103**
- Weightage of features in prediction —>

1. Mjob_services	0.629879
2. Mjob_health	0.611198
3. guardian_other	0.496696
4. guardian_father	0.448005
5. Fjob_teacher	0.423354
6. paid_no	0.417814
7. Fjob_at_home	0.389758
8. Mjob_teacher	0.385497
9. Pstatus_T	0.381797
10. sex_M	0.374552
11. schoolsup_yes	0.353612
12. school_GP	0.337946
13. guardian_mother	0.324961
14. internet_no	0.312916
15. address_R	0.274742
16. Feature_2	0.258733
17. famsize_LE3	0.250911
18. nursery_yes	0.250628
19. famsup_yes	0.249003
20. activities_no	0.248259
21. Fjob_other	0.209362
22. higher_yes	0.206285
23. reason_reputation	0.191569
24. G2	0.188256
25. Fedu	0.184180
26. Feature_1	0.170334
27. freetime	0.162842
28. Mjob_other	0.150217
29. paid_yes	0.141544
30. reason_course	0.118692

2. Random forest model :

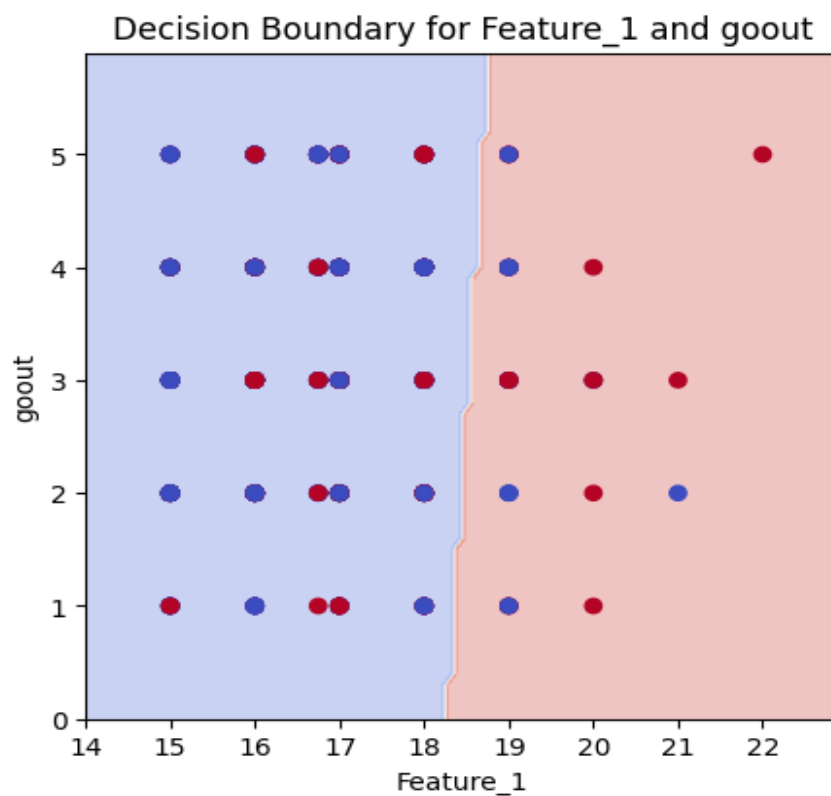
- Accuracy of model -> **56.92307692307692**
- Weightage of features in prediction :

1. G2	0.055364
2. absences	0.050527
3. G1	0.050419
4. G3	0.047867
5. Feature_1	0.047083
6. Fedu	0.037296
7. health	0.034100
8. Feature_3	0.033341
9. freetime	0.032859
10. Medu	0.032707
11. Feature_2	0.032581
12. famrel	0.032503
13. goout	0.032126
14. Dalc	0.024404
15. traveltime	0.023774
16. activities_yes	0.016477
17. Mjob_services	0.014997
18. sex_M	0.014534
19. sex_F	0.013439
20. guardian_other	0.013419
21. reason_course	0.013410
22. famsup_no	0.013246
23. reason_home	0.013092
24. failures	0.013052
25. famsup_yes	0.012604
26. Mjob_other	0.012426
27. address_R	0.012157
28. nursery_no	0.012058
29. school_GP	0.012032
30. activities_no	0.011986

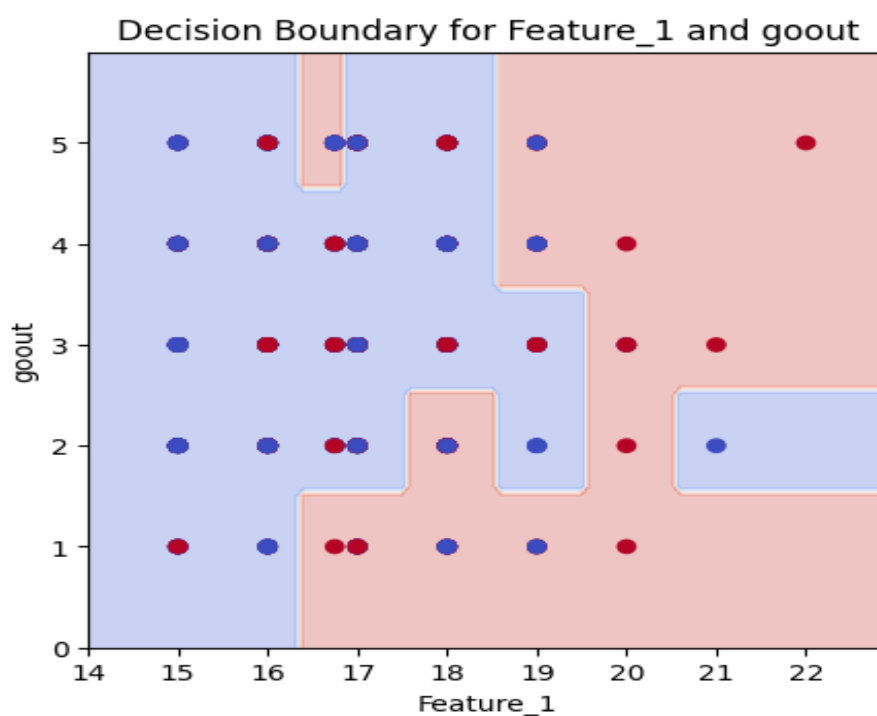
Level 5: Model Reasoning & Interpretation

Decision boundary plot for 'Feature_1' and 'goout'

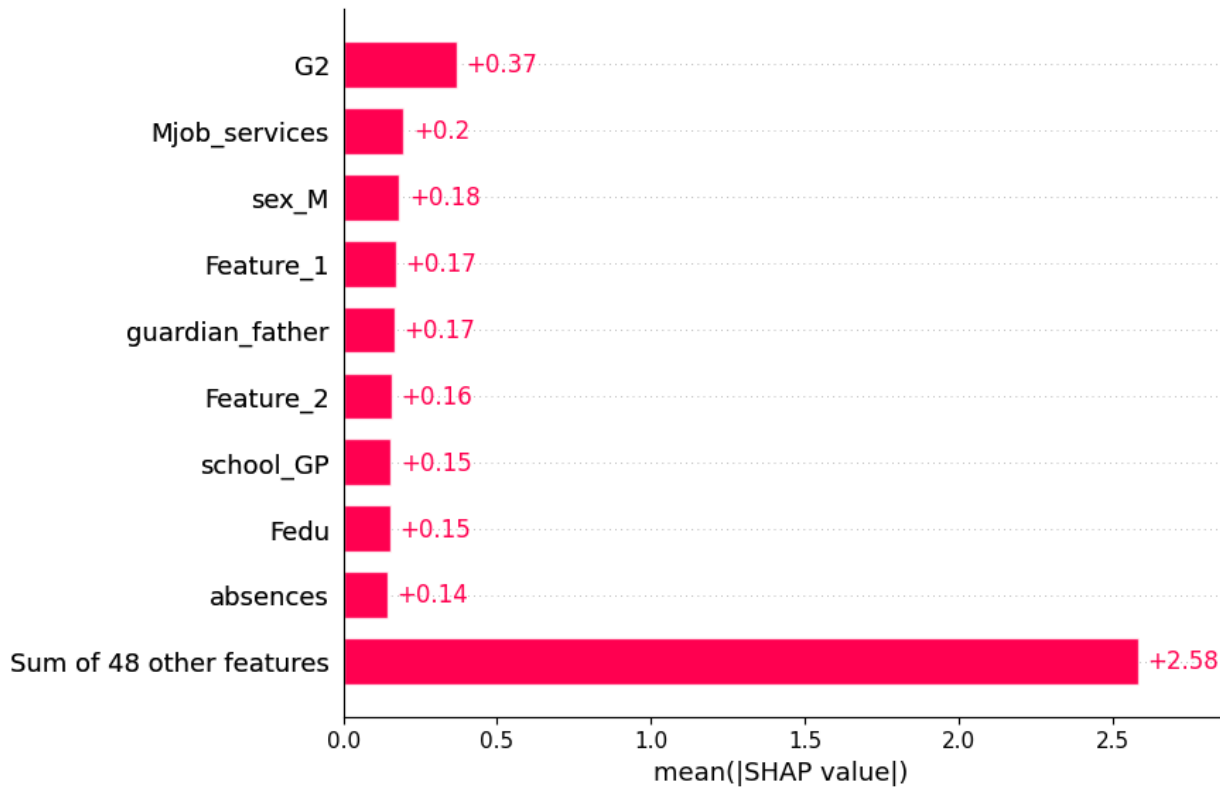
1. For Logistic regression model :



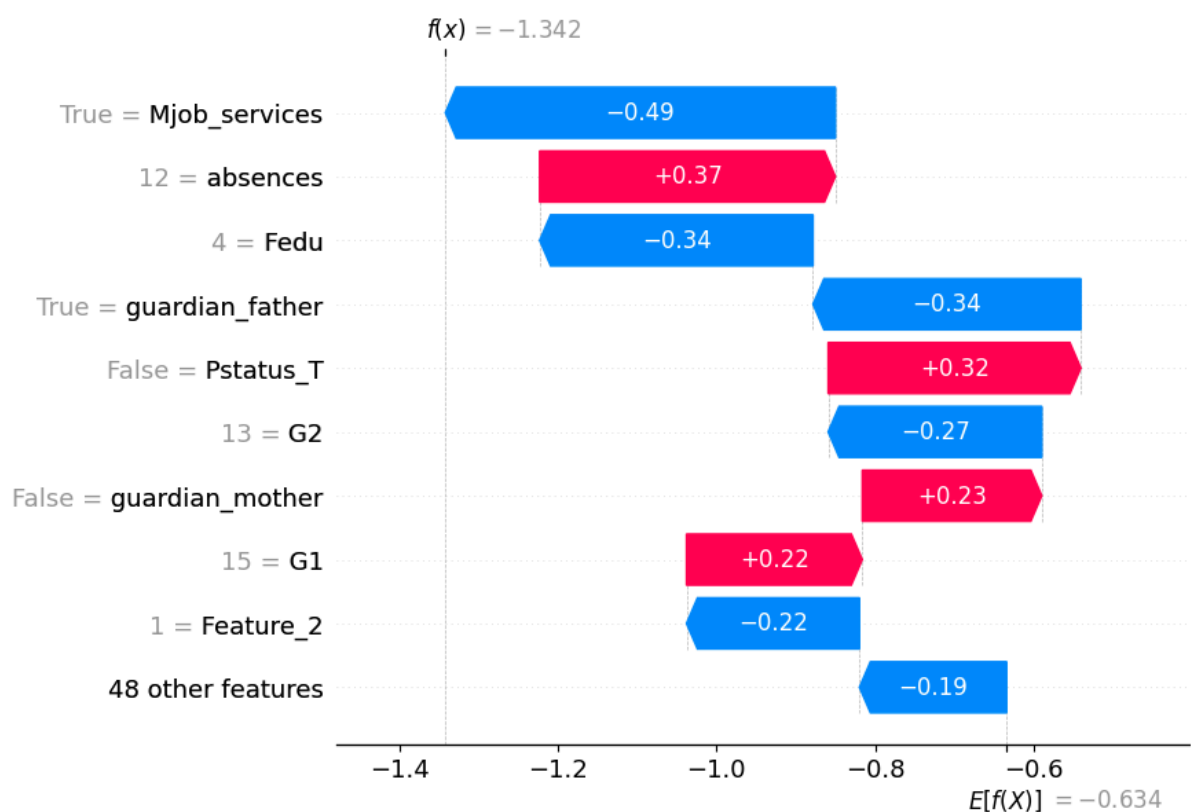
2. For Random forest model :



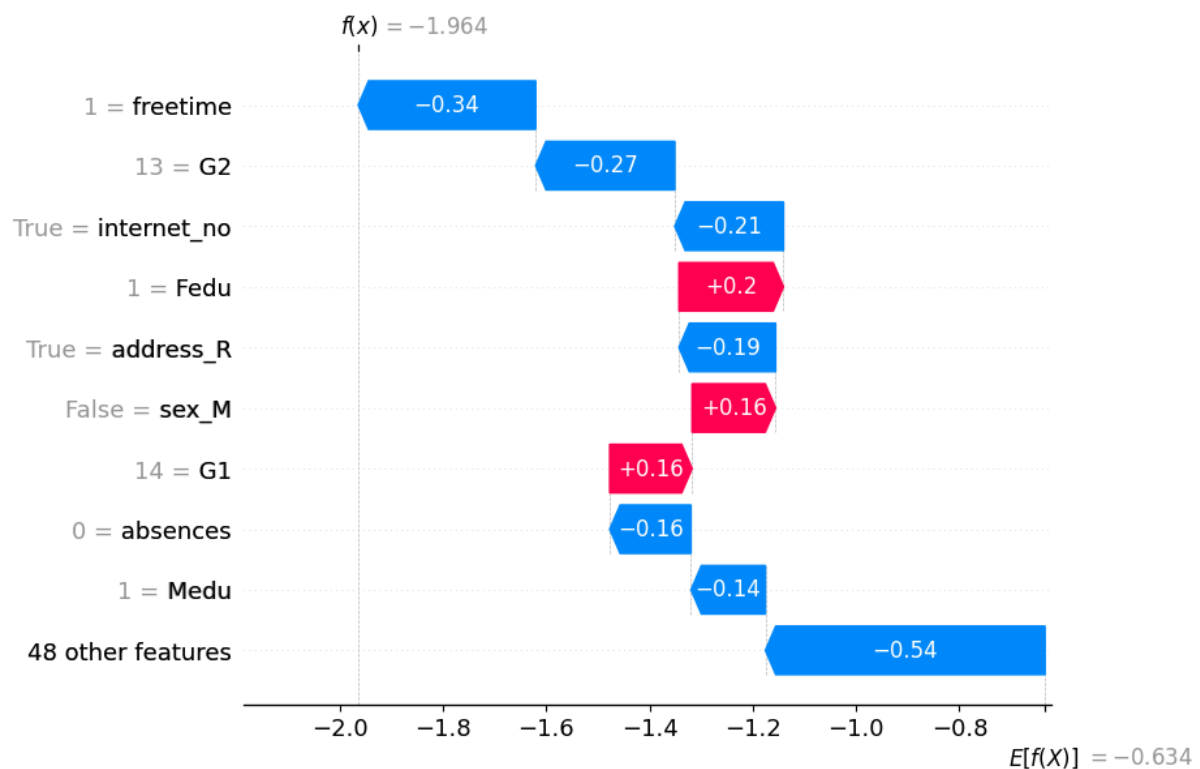
Global feature importance using SHAP :



1.SHAP Explanation for a student predicted YES :



2. SHAP Explanation for a student predicted NO :



Conclusion:

- A biggest contribution in the relationship prediction model is from academic grades 'G2'.
- Features like sex_M, guardian_father, and Fedu show that demographic and familial condition also has predictive power.
- For students predicted yes the biggest contributor is Mjob_services.
- Biggest positive driver is 'absences'.

Task 2 : Enter LangGraph

Gemini model = **"models/gemini-2.0-flash"**

Gemini API key

=**"AIzaSyAfAqf22-jWISyx5IDCDkM6_CmW9pP3Sq8"**

Weather API key

=**"7b6ee732596e9a0d8259a9fd7cd1c5b7"**

Visualisation of level 1 :

