PROJECT REPORT

BY-

Abhiraj Singh

INCLUDES-
TASK1
TASK2

TASK -1

Goal

This project investigates relationship by analyzing a variety of demographic and academic factors. The main objective is to develop robust machine learning models capable of predicting relationship .

Dataset
The dataset used in this project is available here
"https://drive.google.com/drive/folders/1RBkKBDC2HOH0_paOWMy yWdvQV82kX0SZ"
Feature 1
EXPERIMENT 1: HISTOGRAM

The histogram indicate the continuous value of the data in the range from 15-23.Morever more or less Data is consistent in the range 15-17 and decrease with greater values which may coincide with the age Of Students as students with more age are less than students with less age in colleges.
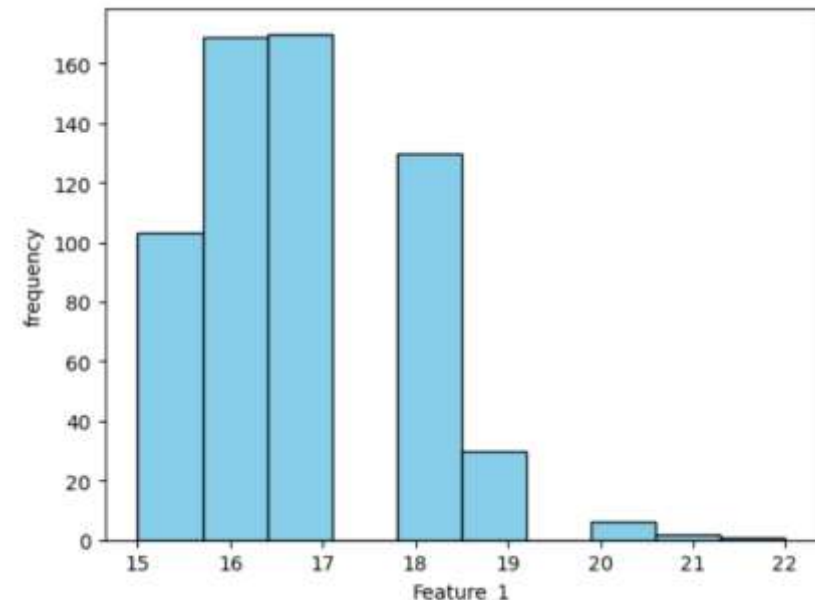
Mean: 16.74795417348609
Mode: 17.0
Median: 17.0
Max: 22.0
Min: 15.0

Since the mean, median, and mode are all close to 17, the distribution is fairly symmetric, but with a slight right skew

EXPERIMENT 2:CORRELATION

The side correlation graph with feature 1 indicate that feature 1 Does not have strong correlation with other demographic and Academic factors such as grades,goout etc.since age also does not have strong correlation with these factors . Clearly, it has mild coorelation with failures which age even has as more age might mean more chance of failures too..
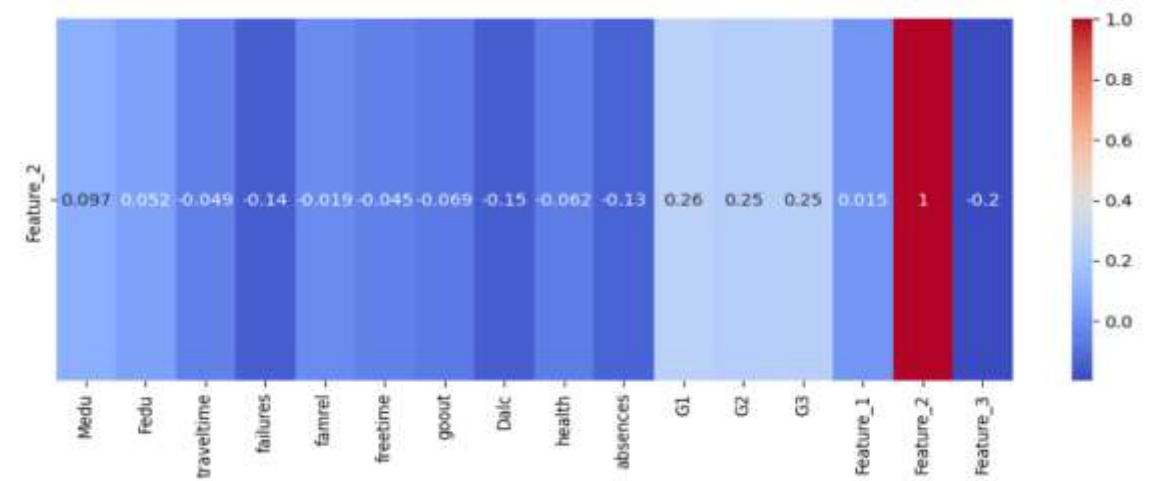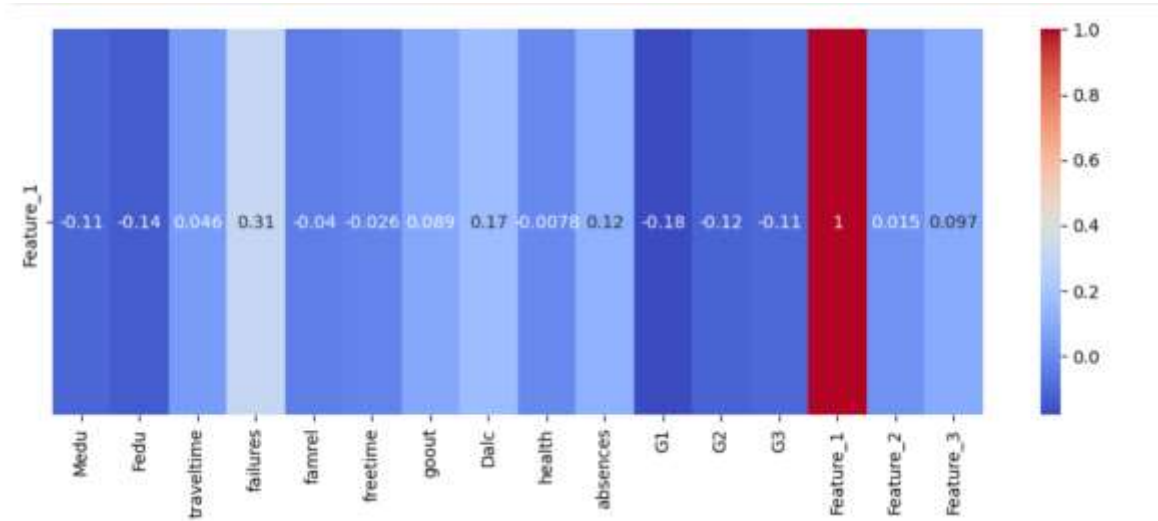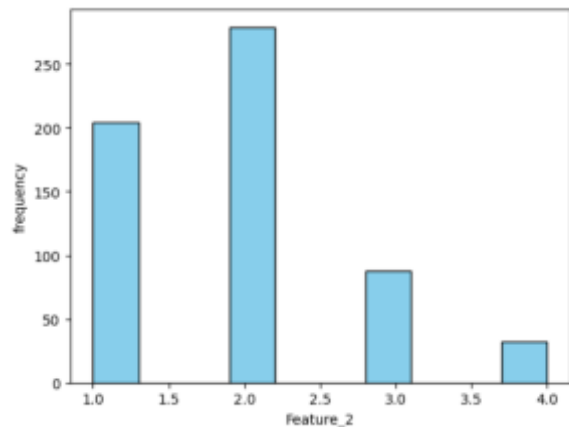Therefore,
FEATURE_1=AGE OF STUDENTS.

FEATURE_2

EXPERIMENT 1:CORRELATION

EXPERIMENT 2: Feature_2 has a strong correlation with Grades G1,G2,G3.so, it could be something related to academic Such as study time.It cannot be year of student as It has a weak coorelation with age of student.
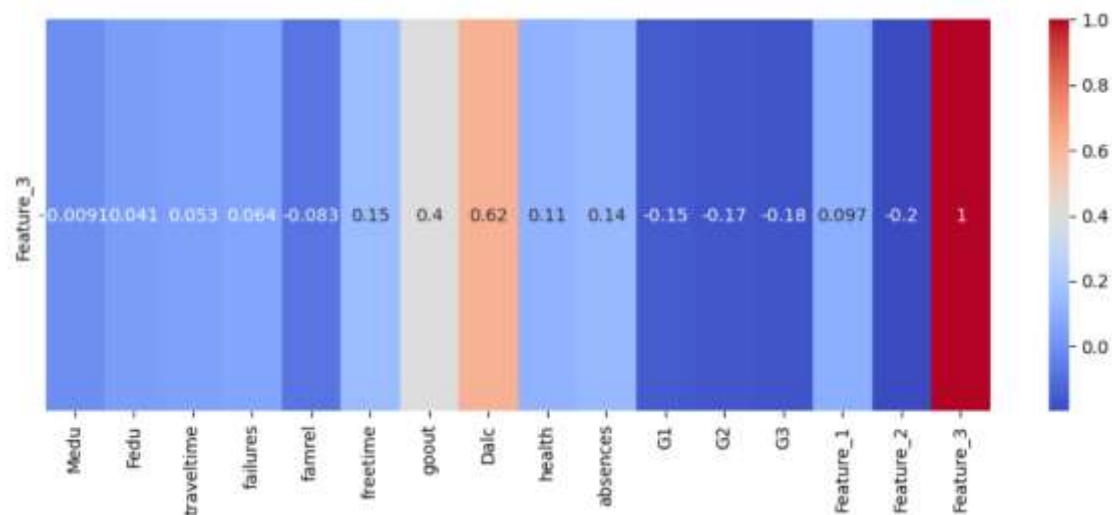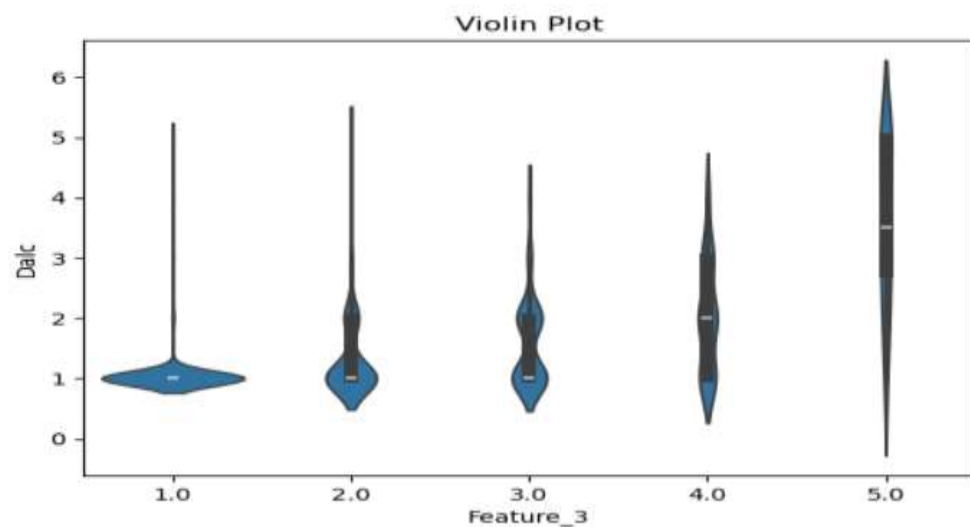
Frequency first increase then decrease, which more or less like the no. of hour
A student spent on study
So,**FEATURE_2=STUDY TIME**

FEATURE_3

It has a strong cooreltion with Dalc and gout.
Therefore, it could be parameter
Related to weekly consumption of alcohol.

Cleaning Dataset

In below graph, the yellow lines indicate the null values, so from here we find that famsize,fedu,traveltime,higher,absecnces,g1,g2,g3,feature 1,feature 2, feature 3 are having missing data
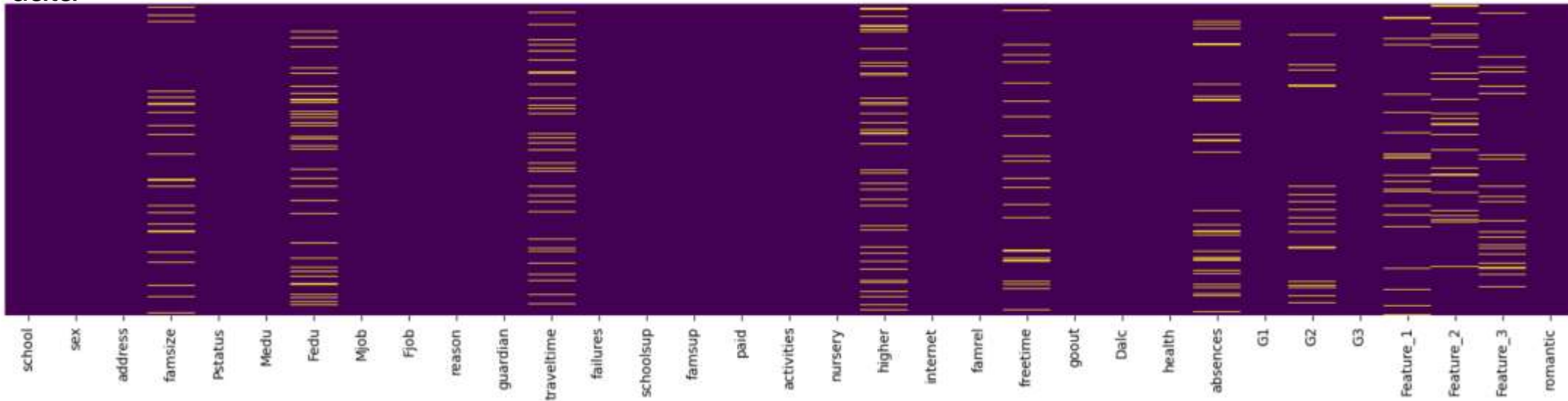
Most of the feature do not have strong correlation except G2 and G3, and Feature_3 and Dalc.so we will divide features into categorical features and numerical feaures.Famsize and higher

Are categorical feaures so, we use mode for their null val,while

Other are numerical ,we use median to fill
EX like here feature_3 is not symmetrical.so
We are not using mean as data is scewed.
Can't use mean to replace null values.
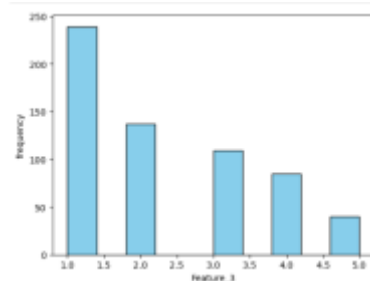
Data Analysis

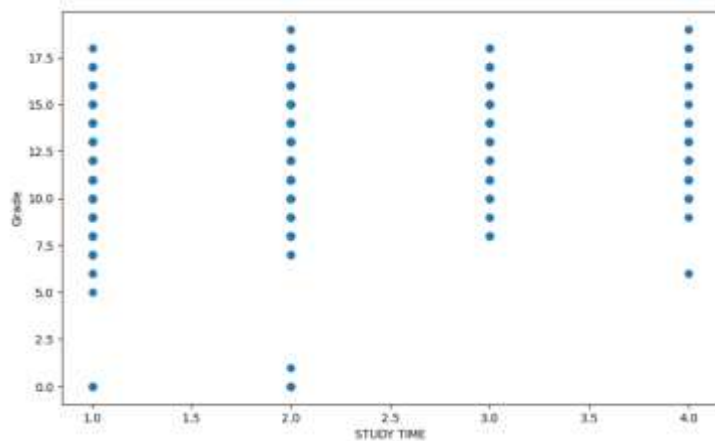## 1.Daily alcohol consumption with gender-
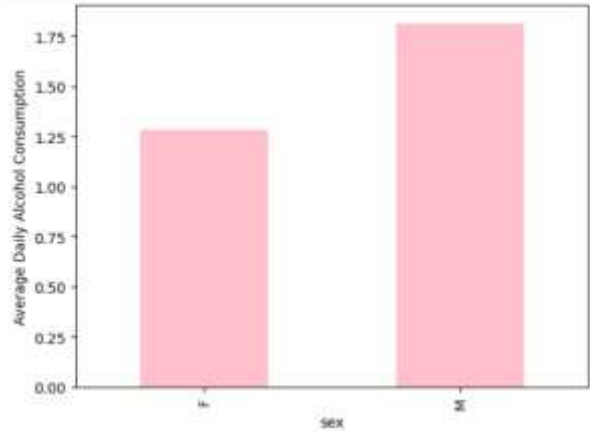
This indicate that average daily alcohol consumption in males is greater than that of females. So males have higher alcohol consumption.

2.How alcohol consumption vary with family relationship?

It is clear from bar graph that much better the family relation is ,lesser is the alcohol consumption.
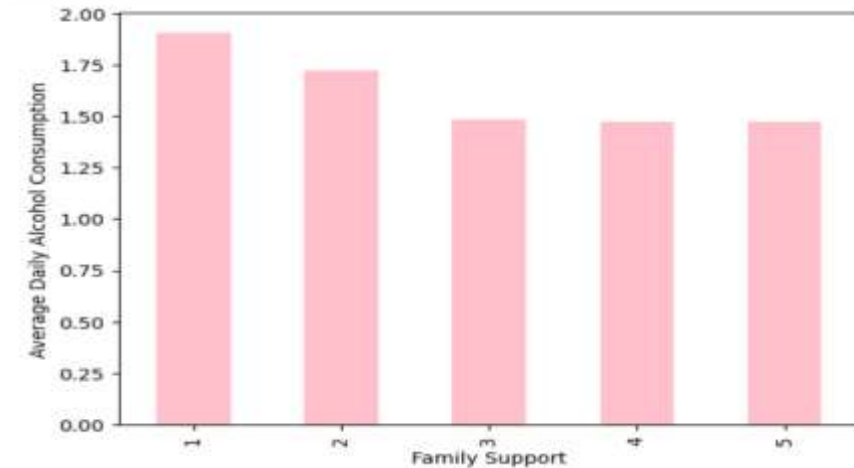
3.How study time affect Grades?

The scatter plot indicae that higher study time mainly have more good grades than lower study time.

4.How Family support affect pursuit of higher education?

It could be seen from pie chart that with no family support,
The pursuit of higher education decreases.

Pursuit of Higher Education (No Family Support)

WANT TO PURSUE

DO NOT WANT TO PURSUE

Pursuit of Higher Education (Family Support)

WANT TO PURSUE

DO NOT WANT TO PURSUE

5.How romanticism vary with age?

It can be seen that proportion of students in relationship
Increase with age . Clearly , indicate that chance of
Student be in relationship increase with age.
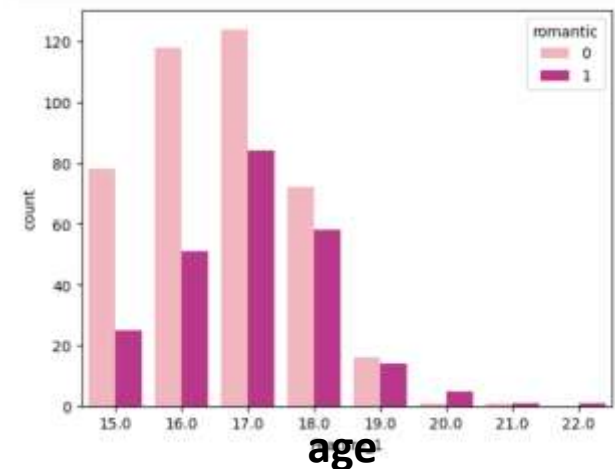
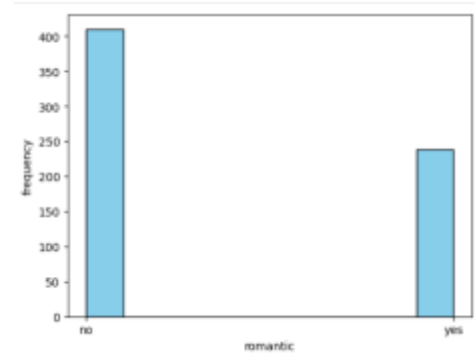**Features affecting Relationship**
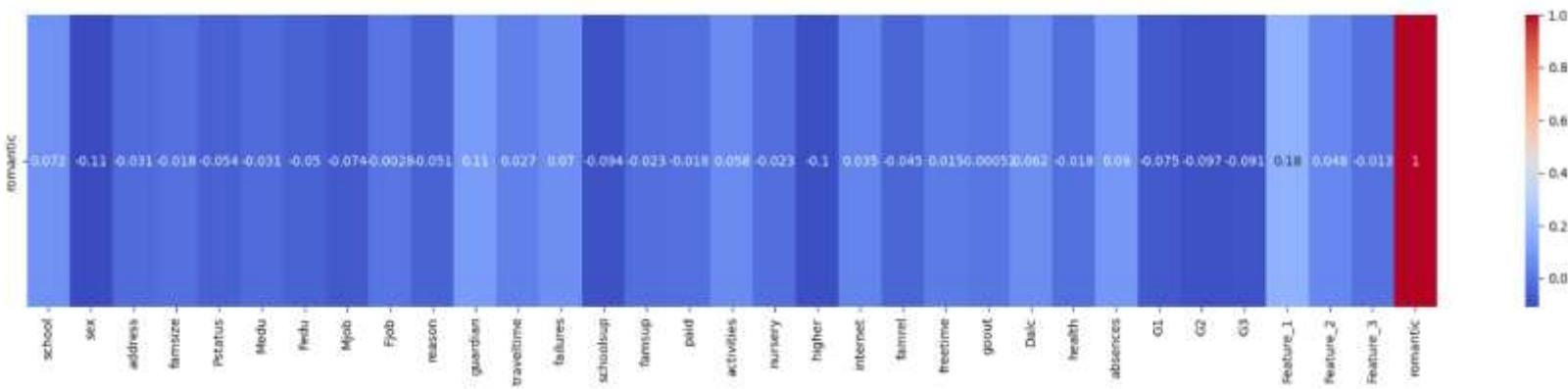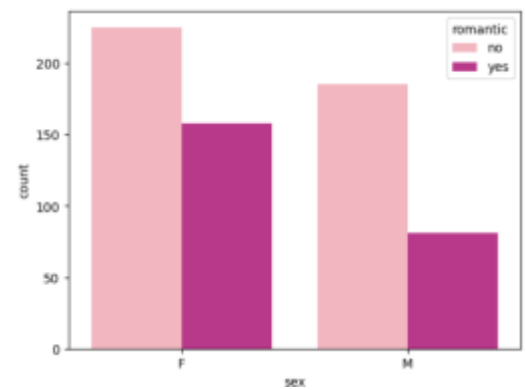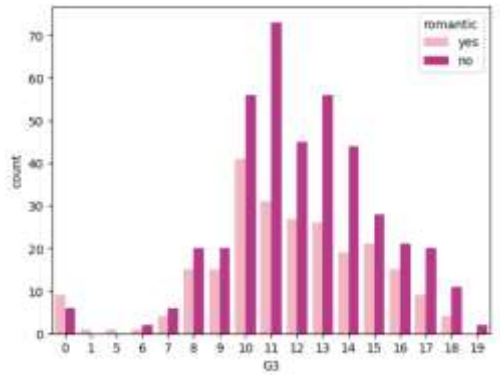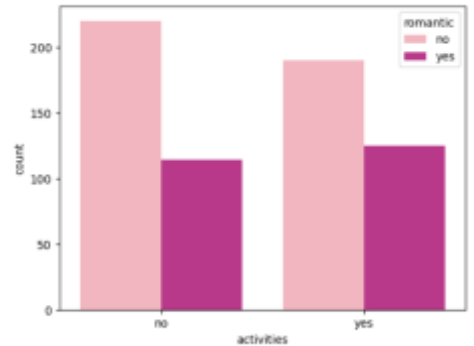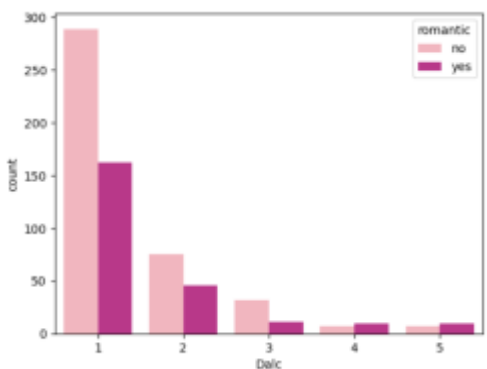
Count Of yes /No is Unbalanced-
Yes-239
No-410 indicating that no of students not in relationship>in relationship

How romanticism vary with other factors?

Relationship do not have strong correlation with other features

Models-

It is a classification problem, therefore we are using three different models-Logistic regression,Decision Tree,Random Forest.

MODEL	Accuracy	F1 score(for class 0)	F1 score (for class 1)
Decision Tree	0.62	0.71	0.46
Random Forest	0.55	0.66	0.34
Logestic Regression	0.59	0.71	0.33

Class Imbalancing could easily be seen.All models have average accuracy with huge class imbalance with less f1 ratio for minority class.
Key reasons are-
1.No of people not in relationship are nearly double of people in relationship.
2.No strong coorelation has been found between features and Relationship.

Then we try feature engineering and class weight=balanced ,result is still bad. F1 score (for class 1) drop to 0.44 and accuracy too decrease to 0.54 .

Global Feature Importance-

For decision tree model,We use shap to plot his bar graph. This indicate that Feature_1 which is age has highest impact on predicting relationship , and so on .Similarly,we do for random forest model

Decision boundary-

1.For decision tree, here the model will predict if any student fall in region yellow will be in relationship, otherwise if it is purple it will predict in not in relationship. here dots indicate what it actually represent. Now ,see how it works-

1.If age ≤ 19.5:
  1. If study_time ≤ 2.5: → Class 0 (purple)
  2. If study_time > 2.5 and age ≤ 16.5: → Class 1 (yellow)
  3. If study_time > 3.5 and age ≤ 17.5: → Class 1 (yellow)
  4. Otherwise: → Class 0 (purple)
2.If age > 19.5:
  1. If study_time ≤ 2.5 or study_time > 3.5: → Class 1 (yellow)
  2. If 2.5 < study_time ≤ 3.5: → Class 0 (purple)

Decision Tree Decision Boundary
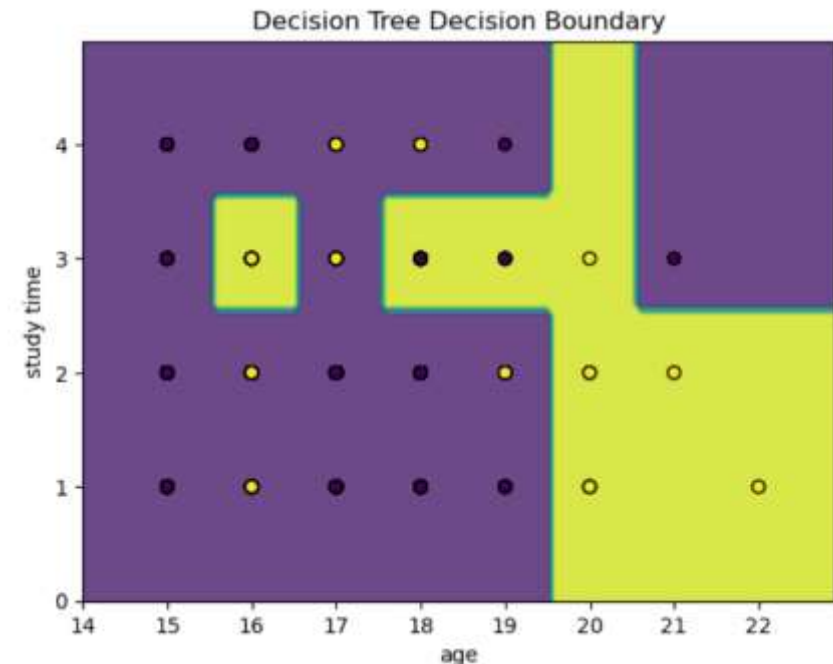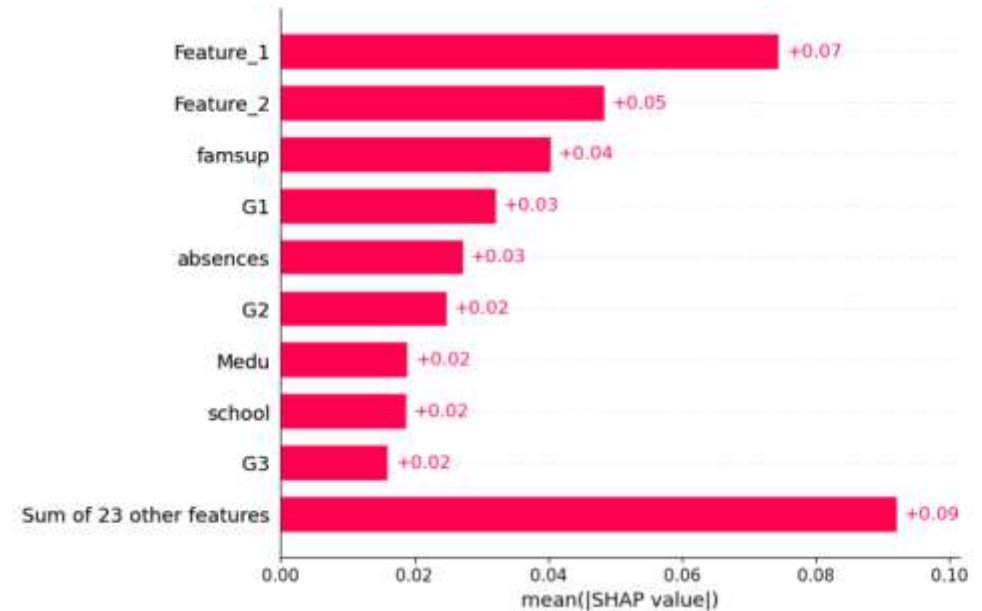
TASK -2

STEPS TO MAKE CHATBOT

We are using google gemini model for this task

Then, we are defining some fxns

1.Calculator-use to calculation using bodmas rule

2.ADD

3.MULTIPLY

4.DIVIDE.

5.SUBTRACT

We add prompts to them to instruct the language model to use them accurately

```python
def calculator(expression:str)->float:
    """Evaluate math expressions using BODMAS rules."""
    return eval(expression)
```

Then we add these fxn to tools and bind the tools to LLM

GRAPH BUILDING-

THE GRAPH START FROM THE START NODE, AND THEN EDGE HAS BEEN ADDED FROM THE STARTNODE TO ASSISTANT WHICH IS THE CHATBOT. A CONDITIONAL EDGE HAS BEEN ADDED WHICH CHATBOT COULD USE DEPENDING ON INPUT OF USER.

Image describing Graph –

USER : " Multiply 8 by 2 and add 5 to ans obtain earlier.Tell me weather in newdelhi"
Analysis-
ALL this user input will go to assistant,now the message read multiply which could be covered by tools, therefore assistant Goes to tools and use the tool multiply and then tool Add is called .Then it back to assistant as weather fxn Right now is not under tools.therefore,can't do weather task.

Weather and Fashion Models-
1.For weather we are using openweathermapAI wrapper
2.For fashion, we are using DuckDuckgosearchTool,which is a web search tool. It search on web things related to fashion Trends.

Now, It could calculate and also tell the weather and fashion trends.
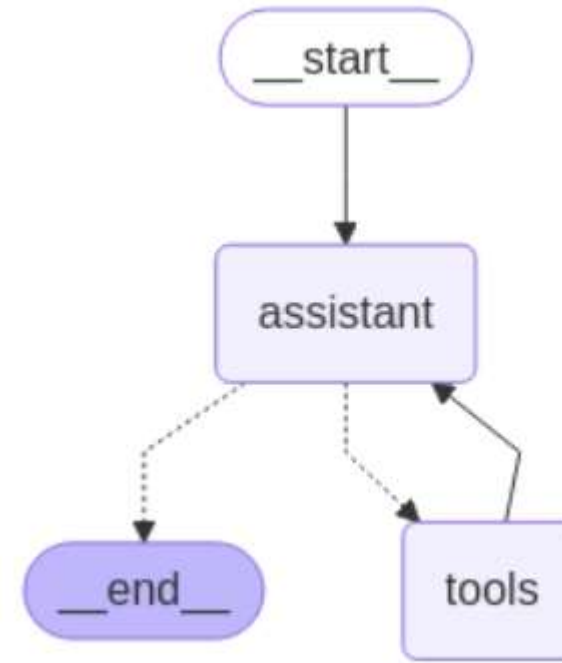
```
================================ Ai Message ==================================

In Delhi, the current weather is haze with a temperature of 37.05°C and a feels like temperature of 41.73°C. The wind speed is 4.12 m/s.
The calculation of 4/5*2 is 1.6.
```

USER:"What is fashion trend in Tokyo?"

```
================================ Ai Message ===================================


Based on the latest observations, fashion trends in Tokyo include:

*    **Genderless Fashion:** Styles that blur the lines between menswear and womenswear are being adopted by more young people.
*    **Techwear & Functional Streetwear:** Urban utility-inspired looks with straps, buckles, and layers are popular.
*    **Summer 2024 Trends:** Wide-leg pants, kitten heels, short sleeve blouses, loose dresses, and basket bags were prominent.
*    **Spring 2025 Trends:** A bold fusion of current styles and grunge-inspired flair, including edgy model-off-duty looks and layered sportswear with st
atement accessories.
```

Memory:

To store the memory of the previous conversations, we are using checkpointers and memory stores.For same configuration,It store the memory.

MULTI AGENTS

We are using various agents such as

1.Research agent-It use duckduckgosearch tool to do various researches.We add prompt to tell it to do only search task

2.Math agent-It use the various tools like calculator etc, defined earlier and do only math related task

3.Weather agent-It use weather_run tool to do this specific task.

Example-

USER:"Who is director of IIT Guwahati?"

```
============================== Ai Message =================================
Name: research_agent


Prof. Devendra Jalihal is the Director of IIT Guwahati.
```

Routing using supervisor

Component	Functions
Supervisor	Task routing between agents based on query type
Agents	- Research Agent: Handles research tasks - Math Agent: Solves math problems - Weather Agent: Provides weather info
Prompt	Explicit instructions for task delegation
Memory	checkpointer=memory preserves conversation history

MULTI-AGENT CONNECTIONS