Main research question and its relevance-

My research investigates the effects of cannabis use on the number of days college students skip class per month, aiming to understand whether cannabis consumption is associated with decreased academic engagement. Given the rising rates of cannabis use among young adults, this study seeks to explore its potential impact on students' academic performance, which may inform university policies and interventions.

Hypothesis-

The hypothesis is that there is a difference in the mean number of classes skipped in the past month between college students who have used marijuana and those who have not.

Descriptive statistics-

The following table shows the descriptive statistics for the variables examined.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
AGE3	5.238	5.000	1.559	3.000	11.00
eduskpcom	0.7129	0.000	2.290	0.000	30.00
micatpy	0.5302	0.000	0.9366	0.000	3.000
substanceuse	2.105	2.000	1.268	1.000	4.000
health	2.035	2.000	0.9084	1.000	5.000
IRFAMIN3	4.604	5.000	2.301	1.000	7.000
new_aldaypmo	1.856	0.000	3.523	0.000	30.00
new_mrdaypmo	0.4886	0.000	2.553	0.000	30.00

The following table shows the frequency distributions for the categorical variables as well as variable descriptions.

Frequency distributions for categorical variables with observations 1-1752				
Variable	**Categories**	**Frequency**	**Relative**	**Cumulative**
AGE3: age category	3=Respondent is 16 or 17 years old	28	1.60%	1.60%
	4=Respondent is between 18 and 20 years old	714	40.75%	42.35%
	5=Respondent is between 21 and 23 years old	496	28.31%	70.66%
	6=Respondent is 24 or 25 years old	182	10.39%	81.05%
	7=Respondent is between 26 and 29 years old	124	7.08%	88.13%
	8=Respondent is between 30 and 34 years old	95	5.42%	93.55%
	9=Respondent is between 35 and 49 years old	97	5.54%	99.09%
	10=Respondent is between 50 and 64 years old	12	0.68%	99.77%
	11=Respondent is 65 years old or older	4	0.23%	100.00%

micatpy: severity of mental illness in the past year **Note: 28 missing observations (1.60%)**	0=No Past Year MI	1218	70.65%	70.65%
	1=Past Year Mild Mental Illness	231	13.40%	84.05%
	2=Past Year Moderate Mental Illness	142	8.24%	92.29%
	3=Past Year Serious Mental Illness	133	7.71%	100.00%
Substanceuse: type of substance use in the past year	1=Sober	834	47.60%	47.60%
	2=Used marijuana and alcohol	392	22.37%	69.98%
	3=Used only marijuana	34	1.94%	71.92%
	4=Used only alcohol	492	28.08%	100.00%
health: overall health rating	1=Excellent	556	31.74%	31.74%
	2=Very good	698	39.84%	71.58%
	3=Good	397	22.66%	94.24%
	4=Fair	83	4.74%	98.97%
	5=Poor	18	1.03%	100.00%
IRFAMIN3: total family income	1=Less than $10,000 (Including Loss)	257	14.67%	14.67%
	2=$10,000-$19,999	226	12.90%	27.57%
	3=$20,000-$29,999	152	8.68%	36.24%
	4=$30,000-$39,999	129	7.36%	43.61%
	5=$40,000-$49,999	153	8.73%	52.34%
	6=$50,000-$74,999	225	12.84%	65.18%
	7=$75,000 or more	610	34.82%	100.00%
new_aldaypmo: number of days used alcohol in the past month				
New_mrdaypmo: number of days used marijuana in the past month				
eduskpcom: number of school days skipped per month				

Regression Model-

Predicted equation: ^eduskpcom = 1.25 - 0.140*AGE3 + 0.395*micatpy - 0.068*IRFAMIN3 + 0.097*health + 0.037*new_aldaypmo + 0.070*new_mrdaypmo

Interpretation of all of the model's slope coefficients:

- For a two year increase in age, the number of school days skipped decreases by 0.140 days, holding other Xs constant.
- For a one level increase of mental illness severity, the number of school days skipped increases by 0.395 days, holding other Xs constant.
- For a $10,000 increase in family income, the number of school days skipped decreases by 0.068 days, holding other Xs constant.
- For a one level decrease of health, the number of school days skipped increases by 0.097 days, holding other Xs constant.
- For a one day increase of alcohol use, the number of school days skipped increases by 0.037 days, holding other Xs constant.

- For a one day increase of marijuana use, my main predictor variable, the number of school days skipped increases by 0.070 days, holding other Xs constant.

The variables with a statistically significant slope coefficient are age, intensity of mental illness, family income, number of days used alcohol, and number of days used marijuana. The multiple regression model has an adjusted R-squared of 0.054. The adjusted R-squared increased from a simple regression, suggesting that the additional predictor variables did improve explanatory power. However, 0.054 is also significantly small, which might show a weak correlation between variables.

Results-

As shown in the table, the variables with a statistically significant slope coefficient are age, intensity of mental illness, family income, number of days used alcohol, and number of days used marijuana. The multiple regression model has an adjusted R-squared of 0.054. The adjusted R-squared increased from a simple regression, suggesting that the additional predictor variables did improve explanatory power. However, 0.054 is also significantly small, which might show a weak correlation between variables.

```
Model 8: OLS, using observations 1-1752 (n = 1724)
Missing or incomplete observations dropped: 28
Dependent variable: eduskpcom

                coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const          1.25023      0.242091      5.164    2.69e-07 ***
  AGE3          -0.139691     0.0355365    -3.931    8.80e-05 ***
  micatpy        0.394615     0.0606294     6.509    9.93e-11 ***
  IRFAMIN3      -0.0684695    0.0235280    -2.910    0.0037   ***
  health         0.0973979    0.0627935     1.551    0.1211
  new_aldaypmo   0.0365210    0.0157403     2.320    0.0204   **
  new_mrdaypmo   0.0704574    0.0216594     3.253    0.0012   ***

Mean dependent var   0.710557   S.D. dependent var   2.301753
Sum squared resid    8609.459   S.E. of regression   2.239251
R-squared            0.056866   Adjusted R-squared   0.053571
F(6, 1717)           17.25447   P-value(F)           1.84e-19
Log-likelihood      -3832.531   Akaike criterion     7679.062
Schwarz criterion    7717.228   Hannan-Quinn         7693.181

Excluding the constant, p-value was highest for variable 16 (health)
```

This research has shown that an increase in marijuana use in college students is likely to affect student's attendance in higher education. Therefore, schools should include more interventions and support services to address substance use, not just alcohol use. If marijuana use in college students continues to rise in the following years, our research indicates that more students will have lowered attendance. However, because this dataset provided by the 2022 National Survey on Drug Use and Health uses self reported information, there are limitations that could influence these findings. For this reason, it is advised that each school conducts further research to determine how marijuana use affects their students' population.