

AI-Driven Insights into Health Awareness among College Students Using Python-Based Data Analytics

Prof.. Sucheta N. Murumkar¹ and Prof. Shraddha D. Patil²

Asst. Prof., Computer, CIMDR, Sangli, India¹

Asst. Prof., Computer, SMGKM, Sangli, India²

Abstract: Health awareness among college students plays a crucial role in promoting long-term well-being and academic performance. The present study investigates health awareness among 211 college students using Python-based data analytics. Data were collected through a structured questionnaire focusing on dietary habits, physical activity, sleep behavior, breakfast consumption, fast food intake, and self-rated health status. The dataset was systematically cleaned, standardized, and numerically encoded to ensure consistency and accuracy prior to analysis. Descriptive statistics, correlation analysis, and one-way ANOVA were applied to examine relationships among various health-related variables. Python libraries such as NumPy, Pandas, and Matplotlib were utilized for efficient data processing, analysis, and visualization. The results indicate that regular physical activity, adequate sleep, and consistent breakfast consumption are positively associated with higher levels of health awareness among students. In contrast, unhealthy dietary practices and irregular meal patterns are linked to lower perceived wellness scores. The study also demonstrates the usefulness of AI-assisted analytical techniques in identifying meaningful patterns within health data. Overall, the findings emphasize the importance of promoting balanced lifestyle habits and highlight the potential of data-driven approaches in supporting effective health awareness and wellness planning in higher education institutions.

Keywords: Artificial Intelligence, Health Awareness, Data Analytics, College Students, Lifestyle Behaviour.

I. INTRODUCTION

Health awareness among young adults plays a crucial role in shaping long-term physical and mental well-being. College students often experience significant lifestyle transitions due to academic pressure, irregular schedules, changing dietary habits, and reduced supervision. These changes may lead to unhealthy behaviours such as meal skipping, insufficient sleep, and increased consumption of fast food.

Traditional surveys often give limited information because the data is processed manually, which is time-consuming and prone to errors. By using data analytics, large amounts of data can be handled quickly and accurately, leading to more reliable conclusions. Python-based tools make it easier to clean data, perform statistical analysis, and identify patterns, helping researchers understand results more clearly and make data-driven decisions.

This research integrates data science techniques with public health analysis to assess health awareness among college students and identify key lifestyle determinants affecting their overall wellness.

II. REVIEW OF LITERATURE

1. Alharthi et al. (2024) conducted a cross-sectional study to assess the knowledge, attitudes, and awareness of health science students regarding the effects of sleep quality on brain function. The study included 361 students from various health science colleges and used a validated self-administered questionnaire. Descriptive statistics along with



independent t-tests and Chi-square tests were applied for data analysis. The findings revealed that although students showed good awareness of the role of sleep in memory, mood, and brain function, unhealthy sleep practices such as excessive screen use before bedtime and lack of physical activity were highly prevalent. Significant gender differences were observed, with female students demonstrating higher awareness of sleep's impact on memory and anxiety-related insomnia. The study emphasized the need for structured sleep education and awareness programs among health science students.

2. Ashraah et al. (2013) conducted a study to assess the level of health awareness among university students in Jordan. The study included 860 undergraduate students from different faculties of the Hashemite University. A structured health awareness scale covering nutrition, personal health, physical activity, and body care was used. Statistical tools such as mean, standard deviation, t-test, and ANOVA were applied. The results revealed that the overall level of health awareness among students was low, with significant differences based on gender, academic specialization, academic level, and GPA. However, the study focused only on traditional health awareness and did not consider the role of modern technology or data-driven approaches in improving student health awareness.

3. Wang and Li (2025) conducted a study to examine the influence of sleep quality, dietary habits, and physical activity on physical health among college students using Structural Equation Modeling (SEM). The study was carried out on a sample of 794 college students in China. The findings revealed that sleep quality had a significant direct and indirect effect on physical fitness, with emotional states acting as a mediating variable. Physical activity showed a strong direct impact on physical fitness, whereas dietary habits did not show a significant relationship with either emotional states or physical fitness. The study highlights the importance of sleep and exercise in maintaining students' physical health.

4. Sheng Jiazhi et al (2025) investigated how different levels of physical activity influence depressive symptoms among university students using standardized depression assessment scales. The study adopted a quantitative survey method and applied statistical techniques such as correlation and regression analysis. The findings indicated a significant negative relationship between physical activity and depression, showing that students who engage in regular physical activity experience lower levels of depressive symptoms. The study emphasizes the importance of physical activity in maintaining students' mental and physical well-being.

5. Wagani and Gaur (2024) conducted a study to assess the effectiveness of awareness media campaigns in spreading health-related information among the population. Using a survey-based methodology, the study analyzed the impact of different media platforms including television, social media, and digital advertisements on public health awareness. The results indicated that awareness media campaigns significantly contribute to improving health knowledge, influencing attitudes, and promoting positive health behaviors. The study highlighted the importance of well-planned media campaigns in strengthening public health awareness.

The present study builds upon this foundation by employing Python-based data analytics and AI-driven statistical techniques to systematically examine health awareness and related behavioral patterns among students.

III. SAMPLE SIZE AND DATASET DESCRIPTION

Out of the collected questionnaires, 211 complete and valid responses were used for analysis to ensure data accuracy and reliability.. Data were obtained through a Google Form questionnaire. The dataset includes the following key variables:

- Number of meals consumed per day
- Breakfast skipping frequency
- Physical activity and exercise habits
- Exercise regularity
- Self-rated health score
- Fast food consumption frequency
- Bedtime patterns



IV. OBJECTIVES OF THE STUDY

- To assess the overall health awareness of college students.
- To identify key lifestyle factors influencing student health.
- To analyze the distribution of dietary, exercise, and sleep-related behaviors.
- To examine the relationship between lifestyle habits and self-rated health.
- To provide data-driven insights for institutional wellness initiatives.

V. RESEARCH HYPOTHESES

Overall Lifestyle Factors and Health Rating

Null Hypothesis (H_{01}):

Lifestyle factors such as physical activity, bedtime, meal frequency, and fast food consumption do not significantly influence health rating.

Alternative Hypothesis (H_{11}):

Lifestyle factors such as physical activity, bedtime, meal frequency, and skipping breakfast significantly influence health rating.

VI. RESEARCH METHODOLOGY

This research adopts a quantitative, descriptive, cross-sectional design. Primary data were collected using a structured Google Form questionnaire consisting of close-ended questions related to health and lifestyle behaviours. Google Forms facilitated standardized data collection, ease of access, and rapid aggregation of responses.

The collected data were exported into CSV format and processed using Python. The pandas library was used for data cleaning, column standardization, handling missing values, and converting categorical responses into numerical formats. Regular expressions were applied to extract numeric values from text-based responses such as meals per day and health ratings. Inconsistent entries, blank responses, and "NaN" values were standardized into uniform categories.

Value count analysis was performed for each major variable, including meals per day, breakfast skipping, physical activity, exercise regularity, health rating, fast food consumption, and bedtime patterns. These processed numerical distributions formed the basis for tabular analysis and interpretation. Statistical computations and normalization were carried out using NumPy

Descriptive Analysis and Value Count Results

The value count analysis revealed important lifestyle trends among students:

Most students reported consuming two or three meals per day, while a smaller proportion consumed only one meal daily.

Breakfast skipping was common, with a majority indicating that they skip breakfast sometimes, highlighting irregular dietary habits.

Physical activity patterns varied, with a significant number of students engaging in daily or weekly exercise, while others reported no activity.

Exercise regularity analysis showed that a majority of students perform some form of physical activity, though a notable proportion do not.

Health ratings were predominantly concentrated around ratings 3 and 4, suggesting moderate to good self-perceived health.

Fast food consumption was frequent among students, with many reporting take occasionally or multiple times per week.

Bedtime analysis indicated that most students sleep between 10 PM and 12 AM, while a smaller group reported late-night sleeping habits.



These distributions provide a clear picture of student lifestyle behaviors and form the empirical basis for further statistical interpretation.

VII. CORRELATION AND INFERENCE ANALYSIS

Correlation analysis was conducted to examine relationships among lifestyle variables. The results indicate a strong positive association between regular exercise, adequate sleep, and higher self-rated health. Students maintaining consistent daily routines demonstrated higher Health Awareness Index scores. Conversely, frequent breakfast skipping and unhealthy eating patterns showed a negative relationship with health awareness. These findings are consistent with existing health and nutrition research.

VIII. INFERENCE STATISTICAL ANALYSIS

Inferential statistical techniques were applied to examine relationships between lifestyle-related variables and students' self-rated health using Python-based data analysis. Statistical significance was evaluated at a 5% significance level ($\alpha = 0.05$).

Correlation Analysis: Meals per Day and Health Rating

Pearson correlation analysis was conducted to examine the relationship between the number of meals consumed per day and self-rated health.

Correlation coefficient (r): 0.060

P-value: 0.38

The Pearson correlation analysis revealed a negligible positive association between meal frequency and health rating ($r = 0.0605$). As the p-value exceeded the 0.05 significance level, the null hypothesis is not rejected. This result suggests that meal frequency does not have a statistically significant association with students' perceived health status.

One-Way ANOVA: Meals per Day and Health Rating

A one-way Analysis of Variance (ANOVA) was conducted to determine whether mean health ratings differ among students based on the number of meals consumed per day (one, two, or three meals).

F-statistic: 0.4398

P-value: 0.6448

Since the p-value is substantially greater than the significance level of 0.05, the null hypothesis is not rejected. This indicates that there is no statistically significant difference in mean health ratings across the three meal-frequency groups.

Statistical Justification:

The very low F-statistic indicates that the variability in health ratings between the meal-frequency groups is much smaller than the variability observed within each group. The high p-value confirms that the observed differences in mean health ratings are likely due to random variation rather than an effect of meal frequency.

Chi-Square Test Results and Statistical Justification

Chi-square tests of independence were conducted to examine the associations between selected lifestyle factors and students' health rating. The results are summarized and interpreted as follows.

The Chi-square value for physical activity is $\chi^2 = 23.19$ with a p-value of 0.00074, which is well below the significance threshold of 0.05. This indicates a statistically significant association between physical activity levels and health rating. The observed frequency distribution differs substantially from the expected distribution under the assumption of independence, confirming that students who engage in regular physical activity are more likely to report better health ratings. This result highlights physical activity as a key determinant of health awareness among students.

For Bedtime and Health Rating, Since the p-value is less than the significance level of 0.05, the null hypothesis is rejected. This indicates that bedtime and health rating are statistically associated.



The chi-square test results demonstrate a statistically significant association between bedtime patterns and self-rated health status. This suggests that variations in sleep timing are related to differences in perceived health among students. Students maintaining regular or earlier bedtimes tend to report better health ratings compared to those with irregular or late sleep schedules. Fast Food Consumption and Health Rating

The Chi-square test for fast food consumption yielded a χ^2 value of 3.879 with a p-value of 0.69, which is substantially higher than the significance level. This indicates no statistically significant association between fast food consumption frequency and health rating. The very high p-value implies that the observed differences in health ratings across fast food consumption categories are largely due to random variation rather than a meaningful relationship.

For breakfast skipping behavior, the Chi-square statistic is $\chi^2 = 23.8407$ with a p-value of 0.00055. As the p-value is less than 0.05, the null hypothesis is not rejected. This result indicates that breakfast skipping does not show a statistically significant association with health rating in this study. While breakfast habits are often linked to health outcomes, the present analysis suggests that breakfast skipping alone does not significantly differentiate students' perceived health status.

Interpretation

The descriptive and inferential analyses provide meaningful insights into the health awareness and lifestyle behaviors of college students. The distribution of responses indicates that most students consume two or three meals per day, yet breakfast skipping remains a common practice, particularly among those consuming three meals daily. This pattern suggests that while overall meal intake may be adequate, meal timing and regularity are inconsistent.

Correlation analysis revealed a negligible and statistically insignificant relationship between meal frequency and self-rated health, indicating that the number of meals consumed per day alone does not strongly influence students' perceived health status. Similarly, one-way ANOVA results confirmed that differences in mean health ratings across meal-frequency groups were not statistically significant.

In contrast, Chi-square test results demonstrated statistically significant associations between physical activity, bedtime patterns, and breakfast skipping behaviour with health ratings. These findings indicate that behavioural and routine-based lifestyle factors have a stronger influence on perceived health than simple dietary frequency measures. Overall, the interpretation highlights that health awareness among students is shaped more by quality and regularity of lifestyle behaviours rather than by isolated habits.

IX. FINDINGS

Based on the analysis, the following key findings emerged:

A majority of students consume two or three meals per day, while a smaller proportion consume only one meal daily.

Breakfast skipping is prevalent among students, with "sometimes skipping breakfast" being the most common behavior.

Physical activity shows a statistically significant association with health rating, indicating that students who engage in regular exercise report better perceived health.

Bedtime patterns are significantly related to health rating, suggesting that consistent and earlier sleep routines contribute positively to student wellness.

Meal frequency alone does not significantly influence self-rated health, as confirmed by both correlation and ANOVA results.

Fast food consumption does not show a statistically significant association with health rating, indicating that its isolated effect may be less pronounced in this sample.

Overall, lifestyle regularity—particularly exercise and sleep—emerges as a stronger determinant of health awareness than dietary frequency.



Discussion

The findings of the present study align with existing research emphasizing the importance of physical activity and sleep in maintaining overall health and well-being among college students. The significant association between physical activity and health rating supports earlier studies that highlight exercise as a key predictor of both physical and mental health outcomes among young adults. Similarly, the observed relationship between bedtime patterns and health ratings reinforces prior evidence suggesting that consistent sleep schedules are essential for optimal cognitive and physical functioning.

Interestingly, the lack of a significant relationship between meal frequency and health rating suggests that merely increasing the number of meals may not lead to improved health perceptions. This finding aligns with recent research indicating that nutritional quality, meal timing, and overall lifestyle balance play a more critical role than meal count alone.

Although breakfast skipping showed significant association in the Chi-square analysis, its independent effect on health rating appears complex. This suggests that breakfast habits may interact with other lifestyle variables such as sleep and physical activity, rather than acting as a standalone determinant of health awareness.

The use of Python-based analytics enabled efficient data cleaning, visualization, and statistical testing, demonstrating the practical value of AI-driven tools in public health research. Such approaches allow educational institutions to move beyond descriptive surveys and adopt evidence-based strategies for student wellness planning.

AI Justification

Artificial Intelligence was employed in this study to efficiently manage, clean, analyze, and interpret health awareness data collected from college students. AI-driven Python tools were used to automate data preprocessing tasks such as handling missing values, standardizing variables, and generating visualizations, thereby reducing manual effort and minimizing human error. The application of AI-supported analytical techniques enabled systematic exploratory analysis and statistical testing to identify meaningful patterns and relationships among lifestyle factors and health perceptions. This approach improved the accuracy, scalability, and reliability of the analysis and supports data-driven decision-making, demonstrating the relevance of artificial intelligence in modern health and educational research.

X. CONCLUSION

This study demonstrates the effectiveness of AI-driven and Python-based data analytics in assessing health awareness among college students. By transforming raw survey responses into structured numerical data, the research provides a systematic and scientific framework for evaluating student lifestyle behaviours.

The findings reveal that regular physical activity and consistent sleep patterns are the most influential factors associated with higher self-rated health, while meal frequency alone does not significantly affect perceived wellness. These results highlight the importance of focusing on behavioural consistency and routine formation rather than isolated health practices.

The integration of data science techniques enhances the accuracy, reliability, and interpretability of health-related research, enabling institutions to design targeted wellness interventions. The study underscores the potential of AI-assisted analytics to support data-driven decision-making in higher education health programs.

Future research may incorporate longitudinal data, larger sample sizes, and machine learning models to predict health outcomes and personalize wellness recommendations. Overall, this research contributes to the growing body of evidence supporting the use of artificial intelligence and data analytics in promoting student health awareness and well-being.

REFERENCES

- [1]. Rana Alharthi¹, Rawan Alamri², WAFAALQARNI³, Muhammad AnwarKhan⁴, Raju Suresh Kumar , Health Sciences Students' Knowledge, Attitudes, and Awareness of the Effects of Sleep Quality on the Brain: A Cross-sectional Study ,Journal of Clinical and Diagnostic Research. 2024 Jun, Vol-18(6): CC05-CC12



- [2]. Mamdouh M. Ashraah1, Ahmad M. Mahasneh, Aeshah A. Al-Sawalmeh\ & Atieh I. Abusheikh : Health Awareness among University Students in Jordan , doi:10.5539/res.v5n5p197 URL: <http://dx.doi.org/10.5539/res.v5n5p197>
- [3]. NingWang and Tao Li (2025), Sleep, Diet, and Exercise as Predictors of Physical Health: A Structural Equation Modeling Study in College Students , DOI: 10.1177/21582440251386694 , journals.sagepub.com/home/sgo
- [4]. Sheng Jiazhi1, Chen Caixia2*, Gong Lamei1 and Zhou Jian1 ,, The relationship between physical activity and depression, anxiety, and stress in college students: a mediating effect of diet , PUBLISHED 07 October 2025, DOI 10.3389/fnut.2025.1611906
- [5]. Rekha Wagani* and Pranita Gaur , Research Article Role of self-awareness in the promotion of health and wellbeing of college students , ORCID: <https://orcid.org/0000-0001-8326-843X>
- [6]. W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2017.
- [7]. Python Software Foundation, Python Language Reference, 2023. [Online]. Available: <https://www.python.org> (Accessed: Dec. 2025)
- [8]. SciPy Community, SciPy Statistical Functions Documentation, 2023. [Online]. Available: <https://docs.scipy.org/doc/scipy/> (Accessed: Dec. 2025)
- [9]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
- [10]. World Health Organization, Healthy Diet and Physical Activity, Geneva, Switzerland, 2023. [Online]. Available: <https://www.who.int> (Accessed: Dec. 2025)
- [11]. J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.
- [12]. A. Field, Discovering Statistics Using IBM SPSS Statistics, 5th ed. London, U.K.: Sage Publications, 2018.

